Economic Theory **19**, 105–144 (2002)

Economic Theory

© Springer-Verlag 2002

Minimal belief change, Pareto-optimality and logical consequence[*]

Oliver Schulte

Department of Philosophy and School of Computing Science, Simon Fraser University,
Burnaby, B.C., V5A 1S6, CANADA (e-mail: oschulte@cs.sfu.ca)

Received: August 20, 2000; revised version: March 19, 2001

Summary. A rational agent changes her beliefs in response to new information; a widely held idea is that such belief changes should be minimal. This paper is an overview of the theory of minimal belief revision. I employ a decision-theoretic framework to compare various principles for minimal belief revision. The main topics covered include the AGM postulates for belief revision, belief contraction, Grove's representation theorem, axioms for conditionals, and the connections between minimal belief change and questions in formal logic. I characterize under what conditions belief revision functions are consistent with the Levi Identity, and under what conditions belief contraction functions are consistent with the Harper Identity.

Keywords and Phrases: Belief revision, Mathematical logic, Conditionals, Iterated belief change.

JEL Classification Numbers: A12, C70, D83.

1 Belief revision

Belief change pervades human life. As we interact with our environment, we continually update our beliefs about it. In such situations, we encounter the problem of how should an agent change his beliefs in light of new information. In the last three decades or so, logicians, philosophers and computer scientists have developed answers to this question under the heading of "belief revision theory". Two main features characterize belief revision theory: the formal model

[*] I am grateful to Mamoru Kaneko for detailed comments on drafts of this paper, and to Hans Rott for helping me draw on his work on belief revision. Work on this paper was supported by a grant from the Social Sciences and Humanities Research Council of Canada.

of belief, and the guiding principle for belief change, namely minimizing the extent of change.

1.1 Sets of beliefs

A set of assertions is the basic formal structure that represents the epistemic state of an agent in belief revision theory. The intended interpretation is that they are the assertions that the agent believes or accepts. Section 2 discusses the relationship between belief as accepted propositions and probabilistic representations of belief. Belief revision theorists assume that the beliefs of a rational agent satisfy the principles of logical reasoning. Thus in belief revision theory, logic determines the structure of belief sets, and the concepts and techniques of mathematical logic are as characteristic of belief revision theory as probability theory is of statistics.

Logically related assertions form a highly abstract and general mathematical structure. As a result, the formal apparatus and theorems of belief revision theory lend themselves to applications besides modelling belief changes. For example, we may identify sets of assertions with records in a database, or with laws in a legal code. On the first interpretation, belief revision theory becomes a theory of revising and updating databases given new inputs, and on the second, an account of revising legal codes given new laws or changes to existing ones. The most influential interpretation, however, and the one that I shall pursue in this paper, is that belief revision theory models the belief changes of logical reasoners. The guiding principle of this theory is that belief changes should be *minimal*.

1.2 Minimal belief change

The core project of belief revision theory is to answer the question: What is a minimal belief change? In this paper, I describe several common approaches to this question. My own is based on decision-theoretic principles, especially the Pareto principle (applied to choice among objects with multi-dimensional attributes). I compare the results of this approach with other important analyses of minimal belief change, notably the Alchourrón-Gärdenfors-Makinson (AGM) axioms.

An interesting aspect of belief revision is its close connection with conditionals – statements of the form “if p, then q”. I describe a formal method for establishing correspondences between belief revision postulates and conditional axioms known as the Ramsey Test. It turns out that the conditional axioms corresponding to Pareto-minimal belief revision are part of a widely used conditional logic.

Finally, in keeping with the theme of this special issue on Logic, I point out the role that various fundamental facts of Mathematical Logic play in belief revision theory as we go along.

Unless otherwise noted, proofs of formal results are in Section 12.

2 Theories, logical consequence and belief

I begin with the representation of an agent's current beliefs as a *theory*. By the term "theory" logicians usually mean a set of assertions that is closed under logical consequence, in the sense that if a proposition p follows from the statements in a theory, then p is also part of the theory. There are two main ways to represent theories, syntactically or semantically. For a syntactic representation, we assume that some formal language L has been fixed, and take a theory to be a deductively closed set of sentences or formulas from L (more below). On a semantic approach, we take theories to be propositions, where propositions are suitable abstract objects such as sets of possible states of affairs, possible worlds, or sets of points in an "outcome space". Probability theory employs a form of the semantic approach: One begins with an abstract set of points, which in statistical applications are often thought of as "outcomes". Sets of points are referred to as "events". Logical operations such as conjunction ("and") and disjunction ("or") correspond to set-theoretic operations (intersection and union) on the powerset of the outcome space. In decision theory and game theory, it is more usual to speak of "possible states of the world" rather than "outcomes". In philosophically motivated developments such as modal logic (which includes among others the logics of belief, knowledge, possibility, and tense) the standard concept is that of a "possible world", and sets of possible worlds are referred to as "propositions" rather than "events". From the point of view of formal logic, it does not matter how one labels the basic set of points and sets of these points; I will employ the logician's usage of "propositions".[1]

For the bulk of this paper, I represent theories syntactically to facilitate comparison with the large part of the literature on belief revision that takes a syntactic approach. Another advantage of the syntactic approach, from the point of view of computer science, is that a syntactic formulation is indispensable if we want to apply belief revision theory to computational agents. However, it should be noted that all of the developments to follow are valid in a purely semantic, propositional setting as well.

A syntactic representation of belief begins with a set of *formulas,* which I denote by L. I denote typical formulas by lower case Roman letters such as p, q etc. Formulas are strings of symbols, somewhat comparable to the way in which English sentences are strings of words. We think of formulas being interpretable, such that a formula expresses a proposition, or state of affairs, again comparable to the way in which an English sentence expresses a state of affairs. Although in this paper I do not go into the theory of interpreting formulas – a part of logic called *semantics* – the reader's intuitions will be helped by thinking of

[1] Incidentally, John Maynard Keynes came out strongly against the "event" terminology: "With the term 'event', which has taken hitherto so important a place in the phraseology of the subject, I shall dispense altogether. Writers on Probability have generally dealt with what they term the 'happening' of 'events'. ... But these expressions are now used in a way which is vague and [un]ambiguous; and it will be more than a verbal improvement to discuss the truth and the probability of *propositions* instead of the occurrence and the probability of *events*." (Keynes, 1921, Ch.1.4), emphasis Keynes's.

formulas as expressing a proposition and as being either true or false. Kaneko's introductory paper gives a precise development of the notion of a formula.

For example, suppose that we want a formal language for describing a very simple situation: there are three objects and one table. Our language has three propositional letters a, b, c. To provide some intuition, we interpret a to mean "the first object is on the table", b to mean "the second object is on the table", and c as "the third object is on the table". This is the kind of example discussed by Katsuno and Mendelzon (1991), Chou and Winslett (1994), and Ginsberg and Smith (1987). I will use the scenario throughout the paper to illustrate definitions.

This language allows us to say very little. For example, we cannot express the proposition "all three objects are on the table". What we need is a way of combining the basic statements given by the propositional letters. So we enrich our language with *operators* and *connectives*. An operator yields a new formula from another; a binary connective yields a new formula from two others. I begin with three connectives that in a propositional setting correspond to the three basic set-theoretic operations of complementation, intersection and union. All told, we have the following definition of a *language*.

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 $(p \rightarrow q)$.

The implication connective $\rightarrow$ is intended to correspond to *material implication*, the kind of "if-then" that is used in mathematical statements. In the presence of the negation operator, disjunction ("or") can replace material implication and vice versa, because $p \rightarrow q$ is equivalent to "*not* p or q" (the only way that "if p, then q" is false in mathematical statements is for p to be true and q to be false). For our purposes, it is technically convenient to use material implication rather than disjunction.

To make formal expressions more readable, I will follow standard practice and omit parentheses around formulas when the intended reading is clear (e.g., write $p \wedge q$ instead of $(p \wedge q)$.)

Logical reasoners derive conclusions from a given set of premises. Formally this corresponds to a *consequence operation* Cn: $2^L \rightarrow 2^L$, where $\mathrm{Cn}(\Gamma)$ gives the set of formulas derivable from the formulas in the set Γ. A set of formulas Γ *entails* another set of formulas Γ', written $\Gamma \vdash \Gamma'$, iff $\mathrm{Cn}(\Gamma) \supseteq \Gamma'$. A set of formulas Γ entails a formula p, written $\Gamma \vdash p$, iff $p \in \mathrm{Cn}(\Gamma)$. Thus $\Gamma \vdash \Gamma'$ iff every formula $p \in \Gamma'$ is a consequence of Γ, that is, iff $\Gamma \vdash p$ for all $p \in \Gamma'$.

A *theory* is a deductively closed set of formulas. That is, a set of formulas $T \subseteq L$ is a theory iff $\mathrm{Cn}(T) = T$. I denote the set of all theories by $\mathbf{T}$. As another piece of terminology, a *theorem* is a formula p that holds without any assumptions; formally p is a theorem iff $\emptyset \vdash p$. It is customary to abbreviate $\emptyset \vdash p$ as $\vdash p$.

I assume that Cn satisfies a number of properties, for all sets of formulas Γ, Γ'. The first three do not depend on the structure of our formal language.

Inclusion	$\Gamma \subseteq \mathrm{Cn}(\Gamma)$.
Montonicity	$\mathrm{Cn}(\Gamma) \subseteq \mathrm{Cn}(\Gamma')$ whenever $\Gamma \subseteq \Gamma'$.
Iteration	$\mathrm{Cn}(\mathrm{Cn}(\Gamma)) = \mathrm{Cn}(\Gamma)$.

These properties are known as the *Tarskian properties* in honour of the great logician Alfred Tarski. Inclusion expresses the idea that any statement p follows from itself. The motivation for Monotonicity is that the more premises we are given, the more conclusions we can derive. Iteration says that if a conclusion p follows from conclusions of an original set of premises, then p follows from the original premises. Clearly these properties hold for mathematical reasoning. Recently there has been a surge of interest in nonmonotonic logics in which the addition of information may render previous inferences invalid (Brewka, Dix and Konolige, 1997) (see Section 10.1).

I further assume that the entailment relation $\vdash$ (and hence the consequence operation Cn) is related to the propositional connectives as follows.

Modus Ponens	If $\Gamma \vdash p, (p \rightarrow q)$, then $\Gamma \vdash q$.
Implication	If $\Gamma \vdash q$, then $\Gamma \vdash (p \rightarrow q)$.
Deduction	$\Gamma \cup \{p\} \vdash q$ iff $\Gamma \vdash (p \rightarrow q)$.
Conjunction	$\Gamma \vdash (p \wedge q)$ iff both $\Gamma \vdash p$ and $\Gamma \vdash q$.
Consistency	Suppose that $\Gamma \nvdash p$. Then $\Gamma \cup \{\neg p\} \nvdash p$.
Inconsistency	$\{p \wedge \neg p\} \vdash L$.
Double Negation	$\Gamma \vdash p$ iff $\Gamma \vdash \neg\neg p$.

These conditions bring the behaviour of the propositional connectives in line with their intended interpretation. As Kaneko's introductory paper shows, the provability relation $\vdash$ of classical logic as well as those of modal logics satisfies these requirements.

Belief revision theorists usually assume that the consequence relation Cn is compact. A consequence relation Cn is compact iff for all formulas p and sets of formulas Γ, we have that $p \in \mathrm{Cn}(\Gamma)$ only if $p \in \mathrm{Cn}(\Gamma')$ for some *finite* subset Γ' of Γ. There is a clear analogy with topological compactness, though not more than an analogy since we are not working in a topological space. Compactness corresponds to the idea that proofs are finite sequences of finite objects (lines in the proof). Thus any given proof can make use of at most finitely many premises. So if we have a proof of p from premises Γ, there must be a finite subset of Γ that suffices for a proof of p, namely the set of all assertions mentioned in the proof of p. Logics that allow infinitely long statements or infinitely long proofs are generally not compact. When the set of possible worlds (or the "event space") is infinite, semantic entailment relations are typically not compact either. None of the results in this paper require compactness.

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. It will help to understand this model of belief based on logic if we relate it to formal models of belief used in economics. One connection is that we may interpret the agent's belief set as his *knowledge*, often modelled with *information partitions* (Osborne and Rubinstein, 1994, Ch.5). The set of assertions that an agent knows – that are entailed by his current information – form a deductively closed set. Another interpretation (consistent with the first) is that an agent's belief set represents the assertions to which the agent assigns a personal probability of 1 (cf. Gärdenfors, 1988, Ch.5). We can then see belief revision theory as an extension of the Bayesian theory of belief change through updating, an extension that addresses the case in which the agent receives new information to which he had assigned probability 0. This combination of belief revision and Bayesian updating suggests some fruitful analogies between probabilistic conditionalization and belief revision (see Section 8). For discussion of the concept of belief employed in belief revision theory, including comparisons with probabilistic conceptions of belief, see for example Gärdenfors (1988), Levi (1980, 1983), Harper (1975, 1976), van Fraassen (1976), Spohn (1987).

I conclude this section with two simple lemmas regarding consequence relations. Though we will not need them until later, I state them here to emphasize that they are a consequence only of our assumptions about consequence relations.

Unless otherwise noted, proofs are in Section 12.

Lemma 1 *Let T_1, T_2 be two theories. Then $T_1 \cap T_2$ is a theory.*

We will often have occasion to consider the logical consequences of adding a formula p to a theory, that is Cn$(T \cup \{p\})$. In belief revision theory, this operation is called *expansion*. Introducing a special symbol for expansion will simplify the notation in what follows.

Definition 1 *For a set of formulas Γ and a formula p define $\Gamma + p = Cn(\Gamma \cup \{p\})$.*

Note that in this notation, the Deduction Principle is expressed as $\Gamma \vdash p \rightarrow q$ iff $\Gamma + p \vdash q$.

Another useful fact is that, given our assumptions about the consequence relation Cn, expansion distributes over the intersection of two theories.

Lemma 2 *Let T_1, T_2 be two theories. For any formula p, $(T_1 \cap T_2) + p = (T_1 + p) \cap (T_2 + p)$.*

3 Additions and retractions

I now begin the analysis of what a minimal theory change is. An obvious approach to this question would be to define a metric ρ between theories, such that $\rho(T_0, T_1)$ is a real number that measures the "distance" between two theories. If we had such a metric ρ at our disposal, we could define a minimal change from a current theory T_0 to be another "closest" theory T_1, that is, a theory T_1

such that there is no "closer" theory T_2. In symbols, a theory T_1 is ρ-closest to T_0 if $\rho(T_0, T_1) \leq \rho(T_0, T_2)$ for all theories T_2. However, so far no generally satisfactory metric between theories has been designed. An interesting approach would be to assume that an agent has some *subjective* metric among theories, but different equally rational agents may have different metrics. Then universal, intersubjective principles of minimal belief revision would be those that hold for any agent who revises his theories minimally according to his subjective metric. This approach would be analogous to a subjective approach to probabilistic belief change, in which agents revise their subjective priors by Bayesian updating. In Section 8 I describe an account of minimal theory change based on subjective orderings that is similar in spirit.

Another approach is to aim for less than a metric provides. Note that a metric ρ between theories defines a *total order* $\leq_T^\rho$ among possible new theories given a current theory T: define $T_1 \leq_T^\rho T_2$ iff $\rho(T, T_1) \leq \rho(T, T_2)$, where $\leq$ denotes the standard ordering of the real numbers. My approach is to consider *partial orders* $\prec_T$ where we read $T_1 \prec_T T_2$ as "T_1 is a smaller change from T than T_2 is". We can think of a given partial order $\prec$ as defining a set of minimal theory changes, namely the minimal elements in that order. So the minimal changes of a theory T according to the $\prec_T$-criterion would form the set $\{T' :$ for all T^*, either $T' \preceq_T T^*$ or T^* and T' are not comparable with respect to $\preceq_T\}$.

In these terms, the project of the first part of this paper is this: Define naturally motivated partial orders, and then characterize their minimal elements in terms of a belief revision operation $*$, such that $*$ produces a minimal element if and only if $*$ satisfies certain axioms.

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.

Definition 2 *Let T, T' be two theories.*

1. T' retracts *the formula p from T $\iff$ $T \vdash p$ and $T' \nvdash p$.*
2. T' adds *the formula p to T $\iff$ $T \nvdash p$ and $T' \vdash p$.*

For example, consider again our simple setting with three objects and a table and the language with propositional letters a, b, c for asserting that a given object is on the table. Let $T = \mathrm{Cn}(\{a, b\})$ (the first two objects are on the table). Then $T^- = \mathrm{Cn}(\{a\})$ retracts b from T, and $T^+ = \mathrm{Cn}(\{a, b, c\})$ adds c to T.

Next, I define two partial orders among theory changes by applying the principle of *dominance*. The first partial order defines a notion of a theory T_1 "retracting more" from a previous theory T than another 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 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.

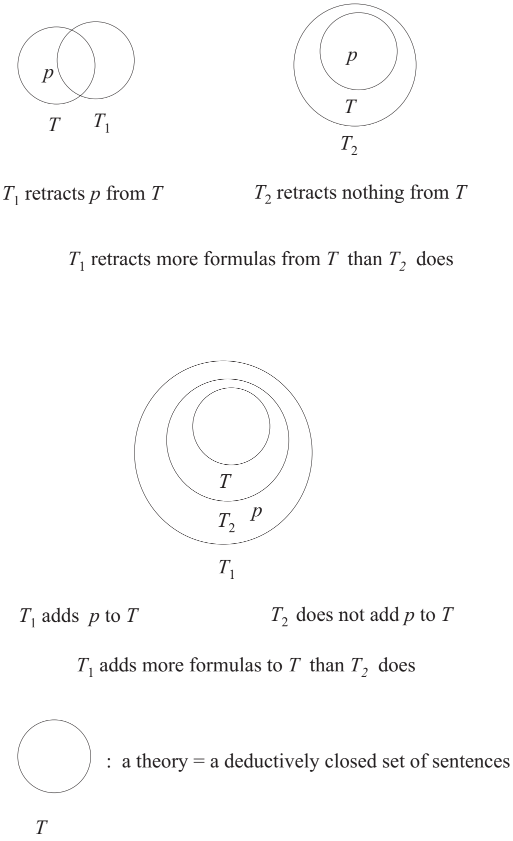
p
T
T1
p
T
T2
T1 retracts p from T
T2 retracts nothing from T
T1 retracts more formulas from T than T2 does
T
T2
p
T1
T1 adds p to T
T2 does not add p to T
T1 adds more formulas to T than T2 does
: a theory = a deductively closed set of sentences
T

Figure 1. Dominance in additions and retractions

Definition 3 *Let T, T_1, T_2 be three theories.*

1. T_1 retracts more *formulas from T than T_2 does $\Longleftrightarrow$*

(a) for all formulas p, if T_2 retracts p from T, then T_1 retracts p from T, and

(b) for some formula p, T_1 retracts p from T but T_2 does not retract p from T.

2. T_1 adds more *formulas to T than T_2 does $\Longleftrightarrow$*

(a) for all formulas p, if T_2 adds p to T, then T_1 adds p to T, and

(b) for some formula p, T_1 adds p to T but T_2 does not add p to T.

Thus T_1 retracts more formulas from T than T_2 iff $T - T_2 \subset T - T_1$, and T_1 adds more formulas to T than T_2 iff $T_2 - T \subset T_1 - T$, where $\subset$ stands for proper set inclusion (recall that $T = \mathrm{Cn}(T)$ for any theory T). Figure 1 illustrates these definitions.

In the example above, with $T = \mathrm{Cn}(\{a, b\})$, we have that $T^- = \mathrm{Cn}(\{a\})$ retracts more formulas from T than $T^+ = \mathrm{Cn}(\{a, b, c\})$ does, and that T^+ adds more formulas to T than T^- does.

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.

For an example, suppose again that $T = \mathrm{Cn}(\{a, b\})$, meaning that the agent believes that the first two objects are on the table and is uncertain about the third. Suppose that the agent learns that both the first and on the second are on the table, that $\neg(a \wedge b)$ holds. Consider two possible revisions, first $T_1 = \mathrm{Cn}(\{a, \neg b\})$ and second $T_2 = \mathrm{Cn}(\{\neg(a \wedge b)\})$. In the sense of Definition 3, we have that T_1 adds more to T than T_2 does. For instance, T_1 adds the formula $\neg b$ to T, whereas T_2 does not (upon learning that either the first or the second object are off the table, the agent comes to hold a new belief about the second object). Also, T_2 retracts more from T than T_1 does, in the sense of Definition 3. For instance, T_2 retracts a from T whereas T_1 does not.

This example illustrates the general tension between avoiding additions and avoiding retractions. The results below characterize the extent of this tension; 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. I will come back to this idea in Section 8. 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 basic principle of *Pareto-optimality*. For minimal theory change, we can render it as follows.[2]

Definition 4 *Let T, T_1, T_2 be three theories. T_1 is a* greater change *from T than T_2 is* $\Longleftrightarrow$

[2] To obtain the appropriate definition for the propositional setting, replace the word “formulas” by “propositions” in the following definition.

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 that 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 ($A \triangle B = [A - B] \cup [B - A]$).[3]

For an example, suppose that $T = \mathrm{Cn}(\{a\})$ and that the agent learns b. Let $T_1 = \mathrm{Cn}(\{a, b, c\})$ and let $T_2 = \mathrm{Cn}(\{a, b\})$. Then T_1 adds more formulas to T than T_2 does (for example, c) and T_1 retracts all the formulas from T that T_2 retracts (which is none). Hence by Clause 2 of Definition 4, it follows that T_1 is a greater change from T than T_2 is. To illustrate Clause 1, suppose that the agent learns a. Then $T_3 = \mathrm{Cn}(\{b\})$ retracts more formulas from T than T does, and adds all formulas that T adds (which is none). So T_3 is a greater change from T than T is.

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 $\prec_T$-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" (cf. Lewi, 1988, p. 52, Condition (1); Lewis, 1981, p. 313). 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 $\preceq_T$-ordering among the theories that entail p.

Definition 5 *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 $\Longleftrightarrow$*

1. *$T_1 \vdash p$, and*
2. *for all theories T_2 such that $T_2 \vdash p$, T_1 is not a greater change from T than T_2 is.*

Now we are ready to give necessary and sufficient conditions for a theory revision to be a Pareto-minimal change. It is not difficult to see that the following three conditions are necessary. Let us write $T * p$ for the revision of theory T given new information p. First, it is our basic constraint that the revision $T * p$ must entail p. Second, since the least change of a theory T is T itself, we don't change the current theory at all if it already entails the new information p; in symbols, $T * p = T$. Third, the revision $T * p$ must follow from the result of simply adding the new information to the old theory; formally, it must be the

[3] Rott considers this set-difference criterion for comparing theory changes (Rott, 2000). Norman Foo made the observation that Pareto-optimality with respect to additions and retractions is equivalent to the set-difference criterion.

case that $T + p \vdash T * p$. For suppose that a revision $T * p$ does not satisfy this condition. Then $T * p$ entails a sentence q that is not entailed by $T + p$, and hence not by T. Consider the theory T' that entails a sentence r just in case both $T * p$ and $T + p + \neg q$ entail r. Clearly T' adds less to T than $T * p$ does because T' is weaker than $T * p$; in particular, T' does not add q to T whereas $T * p$ does. Furthermore, T' retracts from T exactly those sentences that $T * p$ retracts from T. For let r be a sentence entailed by T but not by T'. Then $T + p + \neg q$ entails r and so by the definition of T', it must be the case that $T * p$ does not entail r. This argument shows that $T * p$ is a greater change from T than T' is. Hence $T * p$ is not a Pareto-minimal change unless $T + p$ entails $T * p$. The next theorem shows that the three conditions listed are sufficient as well, that is, any theory revision that satisfies them is Pareto-minimal. Thus we have the following characterization of Pareto-minimal theory change that incorporates a given piece of new information (see Figure 2 in Section 5).

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 $\Longleftrightarrow$*

1. *$T * p \vdash p$, and*
2. *$T + p \vdash T * p$, and*
3. *if $T \vdash p$, then $T * p = T$.*

The proof is in Schulte (1999).[4] 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 ($T + p$) to entailing nothing but the evidence and its consequences $\mathrm{Cn}(\{p\})$.

Pareto-minimality appears to be a basic necessary requirement for any minimal theory change: it is hard to imagine a context in which a theory change $T * p$ violates Pareto-minimality and can yet be considered minimal. But in a given context, we might well require more than Pareto-minimality to accept a theory change as minimal. For example, Pareto-minimality is consistent with an agent retracting just about all her beliefs when her current theory T is consistent with the new information p but does not entail it: in that case it follows from Theorem 1 that $T * p = \mathrm{Cn}(\{p\})$ is a Pareto-minimal theory change. The reason why $T * p = \mathrm{Cn}(\{p\})$ is Pareto-minimal is that by retracting beliefs, the agent avoids adding new ones; we can think of such an agent as viewing additions to be a higher "cost" than retractions. Katsuno and Mendelson have drawn a well-known distinction between two different kinds of contexts in which, they argue, different weightings of additions and retractions are appropriate (Katsuno and Mendelzon, 1991). If belief revision represents the process of receiving increasing information about a *static* world, they recommend avoiding retractions.

[4] Theorem 1 entails that when the new information p is inconsistent with the current theory T, any theory T' entailing p is a Pareto-minimal revision. Rott observed this part of Theorem 1 independently (Rott, 2000, Observation 1).

But if the agent's environment is dynamically changing, they would treat additions and retractions on a par, as Pareto-minimality does. (I return to Katsuno and Mendelson's "update" theory in Section 7.)

In the remainder of the paper, I describe a number of proposals for further constraints on theory change beyond Pareto-minimality, most of which place higher weight on avoiding retractions than on avoiding additions.

5 Retraction-minimal theory change and the AGM postulates

The previous section determined the implications of the Pareto principle for making the trade-off between additions and retractions in minimal theory change. *Lexicographic* ordering is another standard principle for making such trade-offs without invoking weightings of cost dimensions. There is support among belief revision theorists for the view that retracting beliefs is to be especially avoided.[5] What happens in theory revision if we assign maximum priority to avoiding retractions? Formally, this means that we choose *retraction-minimal* theory revisions, which are defined as follows. A revision $T * p$ of a theory T on new information p is *retraction-minimal* if (1) $T * p$ entails p, and (2) no other theory T' entailing p is such that $T * p$ retracts more from T than T' does (in the sense of Definition 3). Now it is easy to see that no retraction-minimal revision $T * p$ retracts any belief from T. For if $T * p$ retracts some belief q, then it retracts more than the revision $(T * p) + q$. In the case in which the new information p is consistent with the current theory T, the only revision that is both Pareto-minimal and retraction-minimal is $T + p$, which may be a satisfactory result. But when the current theory T entails $\neg p$, there is trouble: since no retraction-minimal revision $T * p$ retracts anything from T, we have that $T * p$ entails $\neg p$. But since $T * p$ entails the new information p as well, $T * p$ is inconsistent. In other words, when the new information contradicts the current theory, the only retraction-minimal revision is the inconsistent theory.

One possible remedy is to look for a *consistent* retraction-minimal theory revision. But it turns out that when T is inconsistent with the new information p, every consistent retraction-minimal revision $T * p$ is a *complete* theory, in the sense that for every formula q, either $T * p$ entails q or $\neg q$. The proof of this result is in (Alchourrón and Makinson, 1982, Observation 3.2). Intuitively, the reason is this: If $T \vdash \neg p$, then by Deduction and Inconsistency, for every formula q, it is the case that $T \vdash p \rightarrow q$ and $T \vdash p \rightarrow \neg q$. To make T consistent with p, one of these implications must be removed. But retraction-minimality forces us to keep at least one of them, say $T * p \vdash p \rightarrow q$. Thus by Modus Ponens, $T * p \vdash q$. Since q is an arbitrary formula, any retraction-minimal revision $T * p$ is complete.

[5] "The next postulate for expansions can be justified by the 'economic' side of rationality. The key idea is that, when we change our beliefs, we want to retain as much as possible of our old beliefs – information is in general not gratuitous, and unnecessary losses of information are therefore to be avoided. This heuristic criterion is called the criterion of *information economy*" (Gärdenfors, 1988, p. 49); emphasis is Gärdenfors'.

For example, take $T = \mathrm{Cn}(\{a, b\})$. Then $T \vdash \neg a \rightarrow c$ and $T \vdash \neg a \rightarrow \neg c$, because $T \vdash \neg p \rightarrow q$ whenever $T \vdash p$ (by Deduction and Inconsistency). Now if the agent learns $\neg a$, minimizing retractions does not allow him to give up both $\neg a \rightarrow c$ and $\neg a \rightarrow \neg c$, or else he could retract less by keeping one of the implications among his beliefs. If the agent does not retract $\neg a \rightarrow c$, then retraction-minimality implies that his revised theory is $T' = \mathrm{Cn}(\{\neg a, b, c\})$, and if the agent keeps $\neg a \rightarrow \neg c$, retraction-minimality implies that his revised theory is $T' = \mathrm{Cn}(\{\neg a, b, \neg c\})$. The same is true for the other basic sentences a and b, and so the agent has complete beliefs about all three objects in our example scenario.

More details on retraction-minimal revisions (known as "maxichoice revisions" in the belief revision literature) are in (Schulte, 1999), (Gärdenfors, 1988, Ch. 4), (Alchourrrón and Makinson 1982).

Another approach to minimal belief change, indeed the most common one, is to investigate various belief revision axioms directly without relating them to decision-theoretic principles. A standard set of axioms has emerged from these investigations known as the AGM axioms (after their originators, Alchourrón, Gärdenfors, and Makinson). In my notation, the AGM axioms for theory revision are the following, for a theory T and sentences p, q (Gärdenfors, 1988, Ch. 3.3).

K*1 $T * p$ is a theory.
K*2 $T * p \vdash p$.
K*3 $T + p \vdash T * p$.
K*4 If $T + p$ is consistent, then $T * p \vdash T + p$.
K*5 $T * p$ is inconsistent just in case p is inconsistent.
K*6 If p and q are logically equivalent, then $T * p = T * q$.
K*7 $(T * p) + q \vdash T * (p \wedge q)$.
K*8 If $(T * p) + q$ is consistent, then $T * (p \wedge q) \vdash (T * p) + q$.

I will discuss the justification for these axioms in Sections 7 and 8; see also Gärdenfors (1988, Ch.3). For comparison with Pareto-minimal and retraction-minimal theory change, the relevant axioms are K*1–K*4. From the decision-theoretic standpoint that I have adopted so far, these axioms are a mixture of Pareto-minimality and retraction-minimality. K*3 expresses the characteristic property of Pareto-minimal revisions. K*4 expresses the characteristic property of retraction-minimal revisions, but its application is restricted to the case in which the current theory is consistent with the new information, because as we have seen retraction-minimality leads us to inconsistent or at least complete beliefs when the current theory is inconsistent with the new information. Because K*4 requires an agent to preserve all her beliefs when the new information is consistent with her current beliefs, the postulate is often called the *preservation principle*.

Figure 2 compares the three theories of minimal theory change outlined so far.

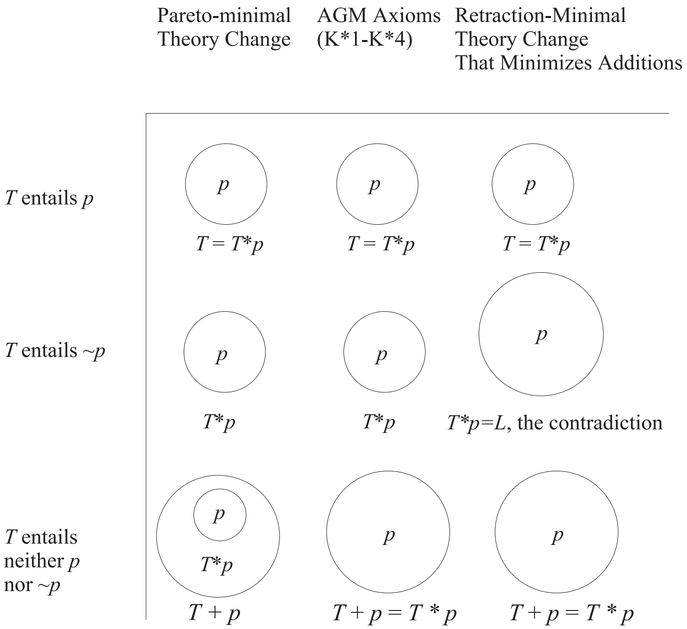
Pareto-minimal Theory Change
AGM Axioms (K*1-K*4)
Retraction-Minimal Theory Change That Minimizes Additions
T entails p
p
T = T*p
T entails ~p
T*p
T*p=L, the contradiction
T entails neither p nor ~p
T + p
T + p = T * p

Figure 2. The behaviour of Pareto-minimal, retraction-minimal and AGM belief revision functions

6 Belief revision functions and belief contraction functions

A major part of the theory of minimal belief change is the analysis of belief contraction. Roughly speaking, contracting beliefs T with respect to a formula p means withdrawing beliefs from T so that T does not entail p. The idea is to develop the analysis of belief contraction independently of the analysis of belief revision. We can then relate the two accounts in ways described below. If we find that two independently motivated theories support each other, this provides an argument in favour of each.

We first need the notions of belief revision and belief contraction functions. So far we have considered a single theory change of a theory T in light of some specific new information p; a belief revision function specifies a new revised theory for every new piece of information p. Extending the previous notation, I write $T * p$ for the result of applying a revision function $*$ for T to a formula p.

Definition 6 *A belief revision function for a theory T is a function $* : L \to \mathbf{T}$ such that for all formulas p, $T * p \vdash p$.*

Definition 6 makes it part of the notion of a belief revision function that the result of applying the function is a theory entailing the new information. In other words, the result of applying the function satisfies the AGM axioms K*1 and K*2. This definition simplifies the theorems below. However, in a more general setting we may wish to investigate theory changes that do not satisfy K*1 and K*2 (see Schulte, 1999, Sect. 7; Levi, 1996).

As we will see presently, in the context of belief contraction there is special motivation to pay attention to belief revision functions that avoid the inconsistent belief set. If the new information p is a theorem – that is, if $\vdash \neg p$ – then any revision on p leads to an inconsistent theory. But we may require that in all other cases, the revision produces a consistent theory. This leads to the following definition.

Definition 7 *A belief revision function* $*$ *for a theory* T *is* consistent *if for all formulas* p*,* $T * p = L \iff \vdash \neg p$.

In other words, a belief revision function is consistent if it satisfies the AGM postulate K*5. The consistency requirement is more complex than it may seem at first. Conceptually, the goal of achieving consistent beliefs is distinct from the goal of minimizing belief change. This is especially clear in the case in which the current theory T is inconsistent. Since T is inconsistent, it entails any new information. And since the smallest change is clearly no change (cf. Sect. 4), the minimal revision of an inconsistent theory T that entails the new information is to stay with the inconsistent theory. A technical point is that requiring the minimal revision of the inconsistent theory to be consistent causes difficulties in relating belief revision postulates to conditional logics (Arló-Costa, 1990) (see Sect. 10.2).

In general, we will avoid confusion if we distinguish clearly between belief revision principles that serve the aim of minimizing the extent of belief change, and belief revision principles that express other constraints, such as principles of rational belief. Without any further conditions, the smallest belief change is no belief change. In Sections 4 and 5 we considered the consequences of the constraint that a belief revision has to entail the new information. We may also examine the consequences of another constraint, namely the consistency requirement K*5. In all these cases, we can apply the principle of Pareto-minimality to find principles of minimal belief change for a given set of constraints. Decision-theoretically, a set of constraints on the possible revisions limits the range of available alternatives. Since an option is Pareto-optimal iff it is not Pareto-dominated by another *available* option, limiting the range of available alternatives can and typically does change the set of Pareto-optimal outcomes – in our setting, the set of Pareto-minimal theories. (Schulte gives a general definition of Pareto-minimality in terms of arbitrary constraints on acceptable revisions (Schulte, 1999, Sect.7).)

Next, I introduce a requirement on belief contraction functions that corresponds to the consistency requirement K*5. First define a belief contraction function as follows.

Definition 8 *A belief contraction function* $\dot{-}$ *for a theory* T *is a function* $\dot{-} : L \rightarrow \mathbf{T}$ *such that for all formulas* p*,* $T \vdash T \dot{-} p$.

In the terms of Section 3, a belief contraction only retracts, but does not add. Usually belief revision theorists require that a belief contraction on a formula p yields a theory that is consistent with the negation of p. Clearly this is possible if and only if p is not a theorem. Thus we have the following definition.

Definition 9 *A belief contraction function* $\dot{-}$ *for a theory* T *is* consistent *if for all formulas* p, $T \dot{-} p \vdash p \iff \vdash p$.

For example, we might have $\mathrm{Cn}(\{a, \neg b\}) \dot{-} \neg b = \mathrm{Cn}(\{a\})$ for a consistent belief contraction function $\dot{-}$. In words, this belief contraction withdraws the belief to contract on, and nothing else.

Let's compare Definition 8 with Gärdenfors' postulates for belief contraction, which in my notation are the following.

$K^{-}1$ $T \dot{-} p$ is a theory.
$K^{-}2$ $T \vdash T \dot{-} p$.
$K^{-}3$ If $T \nvdash p$, then $T \dot{-} p = T$.
$K^{-}4$ If $\nvdash p$, then $T \dot{-} p \nvdash p$.
$K^{-}5$ If $T \vdash p$, then $(T \dot{-} p) + p \vdash T$.
$K^{-}6$ If p and q are logically equivalent, then $T \dot{-} p = T \dot{-} q$.
$K^{-}7$ $T \dot{-} (p \wedge q) \vdash T \dot{-} p \cap T \dot{-} q$.
$K^{-}8$ If $T \dot{-} (p \wedge q) \nvdash p$, then $T \dot{-} p \vdash T \dot{-} (p \wedge q)$.

In my usage, Gärdenfors' postulates $K^{-}1$ and $K^{-}2$ define a belief contraction function, and adding $K^{-}4$ yields a consistent belief revision function. Sometimes in the literature the term "belief contraction" is used to refer to what I call consistent belief contraction, and sometimes it is used to refer to a function that satisfies all of Gärdenfors' postulates.

6.1 Mappings between belief revision and belief contraction functions: the Levi and Harper Identities

One of the early ideas about belief revision was Levi's proposal that a theory change from T on new information p ought to proceed in two stages (Levi, 1980, Ch. 3). First, we may contract the theory T on the assertion $\neg p$. If the contraction is consistent, this yields a theory T' that does not entail $\neg p$, and hence is consistent with p. Then we add the new information p to T', which is guaranteed to yield a consistent theory. Thus using consistent belief contractions to define belief revision via Levi's two-step process guarantees that revisions satisfy the consistency postulate K*5. This is the sense in which it is a consistency requirement to stipulate that a belief contraction on p should not entail p.

Another way to think about Levi's proposal is that it provides a recipe for constructing a belief revision function from a belief contraction function. That is, Levi's proposal maps contraction functions to belief revision functions (see Fig. 3). Gärdenfors refers to this mapping as the "Levi Identity"; his formal definition is as follows.

Definition 10 (The Levi Identity) *Let* $\dot{-}$ *be a belief contraction function for a theory* T. *The belief revision function* $*$ *associated with* $\dot{-}$ *is defined by* $T * p = T \dot{-} \neg p + p$.

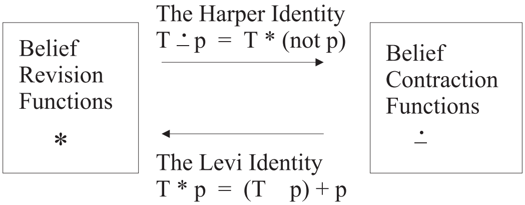

Figure 3. The Levi and Harper identity

I write levi($\dot{-}$) to denote the belief revision function associated with $\dot{-}$. To illustrate the Levi Identity, consider again the belief contraction $\text{Cn}(\{a, \neg b\}) \dot{-} \neg b = \text{Cn}(\{a\})$. The associated belief revision is $\text{Cn}(\{a, \neg b\}) * b = (\, \text{Cn}(\{a, \neg b\}) \dot{-} \neg b) + b$, which is $\text{Cn}(\{a\}) + b = \text{Cn}(\{a, b\})$. In words, we can think of the revision as first withdrawing nothing but the negation of the new information b, and then adding the new belief b.

It is easy to see that the Levi Identity yields a belief revision function for a given belief contraction function, and that it yields a consistent belief revision function for a consistent belief contraction function.

Lemma 3 *Let $\dot{-}$ be a belief contraction function for a theory T, and let $*$ be the function associated with $\dot{-}$. Then*

1. *$*$ is a belief revision function for T, and*
2. *if $\dot{-}$ is consistent, then $*$ is consistent. (In other words, if $\dot{-}$ satisfies $K^{-}4$, then $*$ satisfies K*5).*

How should we define a belief contraction function given a revision function $*$ for a theory T? Harper made the following proposal (translated into our syntactic framework) (Harper 1975). Consider the revision $T * \neg p$. If $T * \neg p$ is a minimal revision of T on $\neg p$, then the difference between $T * \neg p$ and T is minimal, and so $T * \neg p$ has as much in common with T as is possible given the requirement of accommodating $\neg p$. Thus the overlap $T \cap T * \neg p$ ought to be as large as it can be while conforming with $\neg p$. This means that $T \cap T * \neg p$ is a plausible candidate for a minimal retraction of T that makes room for $\neg p$, that is, a contraction of T on p. Hence the following definition.

Definition 11 (The Harper Identity) *Let $*$ be a belief revision function for T. The contraction function associated with $*$ is defined by $T \dot{-} p = T \cap T * \neg p$.*

As the Levi Identity yields a contraction function given a revision function, the Harper Identity defines a revision function from a contraction function; see Figure 3. I write harper($*$) to denote the belief contraction function associated with $*$.

To illustrate the Harper identity, consider again the revision $\text{Cn}(\{a, \neg b\}) * b = \text{Cn}(\{a, b\})$ from before. (The agent initially believes that the first object is on the table and the second is not. On learning that the second is on the table, her new beliefs are that the first two objects are on the table.) Assume that the revision function treats b and $\neg\neg b$ identically, such that $\text{Cn}(\{a, \neg b\}) * \neg\neg b = \text{Cn}(\{a, b\})$. The associated belief contraction is $\mathrm{Cn}(\{a, \neg b\}) \dot{-} \neg b = \mathrm{Cn}(\{a, \neg b\}) \cap (\mathrm{Cn}(\{a, \neg b\}) * \neg\neg b)$, which is $\mathrm{Cn}(\{a, \neg b\}) \cap \mathrm{Cn}(\{a, \neg b\}) * b = \mathrm{Cn}(\{a, \neg b\}) \cap \mathrm{Cn}(\{a, b\})$. It is possible to show that $\mathrm{Cn}(\{a, \neg b\}) \cap \mathrm{Cn}(\{a, b\}) = \mathrm{Cn}(\{a\})$. All told, we have that $\mathrm{Cn}(\{a, \neg b\}) \dot{-} \neg b = \mathrm{Cn}(\{a\})$.

Note that in this case, the Levi and Harper Identities invert each other. If we start with the revision $\mathrm{Cn}(\{a, \neg b\}) * b = \mathrm{Cn}(\{a, b\})$, the Harper Identity yields the contraction $\mathrm{Cn}(\{a, \neg b\}) \dot{-} \neg b = \mathrm{Cn}(\{a\})$. And as we saw above, applying the Levi Identity to the contraction $\mathrm{Cn}(\{a, \neg b\}) \dot{-} \neg b = \mathrm{Cn}(\{a\})$ yields the revision $\mathrm{Cn}(\{a, \neg b\}) * b = \mathrm{Cn}(\{a, b\})$.

It is easy to see that the Harper Identity yields a belief contraction function for a given belief revision function, and that it yields a consistent belief contraction function for a consistent belief revision function.

Lemma 4 *Let $*$ be a belief revision function for a theory T, and let $\dot{-}$ be the function associated with $*$. Then*

1. *$\dot{-}$ is a belief contraction function for T, and*
2. *if $*$ is consistent, then $\dot{-}$ is consistent. (In other words, if $*$ satisfies K*5, then $\dot{-}$ satisfies $K^{-}4$).*

6.2 The content of the Levi and Harper Identities

The Levi Identity stipulates a constraint on revision functions for minimal belief change by connecting them to belief contraction. What is the content of this constraint? That is, what properties must belief revision functions satisfy if they follow the Levi Identity? The answer is that the Levi Identity picks out those revision functions that satisfy K*3 – the requirement that the expansion $T + p$ must be at least as strong as the revision $T * p$. Thus the Levi Identity and Pareto-minimality turn out to characterize a very similar class of belief revision functions; the only difference is that Pareto-minimality does not allow any change when the evidence is entailed by the agent's current theory, whereas the Levi Identity does.

Let us say that a function $*$ *satisfies the Levi Identity*, or is *generated by the Levi Identity*, if there is a belief contraction function $\dot{-}$ such that $*$ is the function associated with $\dot{-}$ (i.e., $* = \mathrm{levi}(\dot{-})$). It is easy to see that if a belief revision function satisfies the Levi Identity, then it also satisfies K*3, the characteristic property of Pareto-minimal revisions.

Lemma 5 *Let $\dot{-}$ be a belief contraction function for a theory T with associated belief revision function $*$. Then for all formulas p, $T + p \vdash T * p$.*

To illustrate the lemma, let us consider an example of a revision that does not satisfy the Levi Identity. For example, let $T = \mathrm{Cn}(\{a\})$ and consider the revision $T * b = \mathrm{Cn}(\{a, b, c\})$ (from the belief that the first object is on the table, and the information that the second is on the table, infer that the third one is as well). Consider any contraction $T^{-} = \mathrm{Cn}(\{a\}) \dot{-} \neg b$. By Definition 8, we have that $T \vdash T^-$, so T^- is consistent with b and $\neg b$ as well as c and $\neg c$ (T^- entails nothing about the whereabouts of the second and third object). Thus any revision T' defined by $T * b = T^- + b$ is consistent with c as well as $\neg c$. Therefore the revision $\mathrm{Cn}(\{a\}) * b = \mathrm{Cn}(\{a, b, c\})$ does not satisfy the Levi Identity.

What about the converse of Lemma 5? The converse requires us to show that if a belief revision function $*$ satisfies K*3, then there is some contraction function $\dot{-}$ that generates $*$ via the Levi Identity. The obvious candidate for such a contraction function is the function harper($*$) that the Harper Identity associates with the revision operator. It turns out that indeed, applying the Levi Identity to harper($*$) yields the original belief revision function $*$; in other words, the Levi Identity inverts the Harper Identity, but only with some provisos. The first proviso is that $*$ must satisfy K*3, as Lemma 5 requires. The second is that $*$ must treat doubly negated formulas like unnegated formulas. Thus I say that a belief revision function $*$ for T *respects double negation* if for all formulas p, we have that $T * p = T * \neg\neg p$. Respect for double negation is much weaker than the AGM postulate K*6 which requires that the respective results of revising on logically equivalent formulas be the same. With these conditions in place, the postulate K*3 is a necessary and sufficient condition for the Levi Identity to invert the Harper Identity.

Proposition 1 *Let $*$ be a belief revision function for T that respects double negation. Then the Levi Identity inverts the Harper Identity applied to $*$ $\iff$ for all formulas p, $T + p \vdash T * p$.*

Proposition 1 immediately yields a characterization of the belief revision functions that are consistent with the Levi Identity.

Corollary 1 *A belief revision function $*$ for a theory T that respects double negation can be generated by the Levi Identity $\iff$ for all formulas p, $T + p \vdash T * p$.*

Next, I investigate the content of the Harper Identity. Let us say that a function $\dot{-}$ *satisfies the Harper Identity,* or is *generated by the Harper Identity,* if there is a belief revision function $*$ such that $\dot{-}$ is the function associated with $*$ (i.e., $\dot{-}$ = harper($*$)). It is not hard to prove that the following condition is necessary for a belief contraction function to be generated by the Harper Identity.

Lemma 6 *Suppose that $*$ is a belief revision function for a theory T, and that $\dot{-}$ is the contraction function associated with $*$. Then for all formulas p, $T \dot{-} p + p = T + p$.*

To illustrate the lemma, let us consider an example of a contraction that does not satisfy the Harper Identity. For example, let $T = \mathrm{Cn}(\{a, b\})$, and suppose that $T \dot{-} a = \mathrm{Cn}(\emptyset)$ (to withdraw the belief that the first object is on the table, contract to being uncertain about all three objects). Then $T \dot{-} a + b = \mathrm{Cn}(\{b\})$, which is different from $T + b = \mathrm{Cn}(\{a, b\})$. Hence Lemma 6 entails that $T \dot{-} a$ does not satisfy the Harper Identity.[6]

[6] To verify this fact directly, consider any revision $T * \neg a$ and apply Lemma 2 to $(T \cap T * \neg a) + a$.

In the case in which $T \vdash p$, the condition that $T \dot{-} p + p = T + p$ is essentially equivalent to Gärdenfors' postulate $\mathrm{K}^{-}5$, viz. $T \dot{-} p + p \vdash T$. Since $T + p = T$ if $T \vdash p$, the condition of Lemma 6 entails $\mathrm{K}^{-}5$. And since $T \vdash T \dot{-} p$ for any contraction function $\dot{-}$, it is immediate that $T + p \vdash T \dot{-} p + p$.

In the case in which $T \nvdash p$, Gärdenfors' postulate $\mathrm{K}^{-}3$ stipulates that $T \dot{-} p = T$, which clearly entails the condition of Lemma 6. However, the requirement that $T \dot{-} p + p = T + p$ does not entail $\mathrm{K}^{-}3$ because for example $T \dot{-} p$ might retract a statement q from T provided that it does not retract $p \rightarrow q$.

The postulate $\mathrm{K}^{-}5$ is often referred to as a *recovery postulate* because it asserts that after first contracting on p and then adding p "back in", the agent recovers all of the beliefs in her original theory T. The intuition behind the recovery principle is this. To contract beliefs on p means to "give $\neg p$ a hearing", or to entertain the possibility that p may be false. If the agent gives $\neg p$ a hearing, but then finds that p is correct after all, the agent should restore confidence in any proposition q that he may have believed but called into doubt along with q. The condition of Lemma 6 is a formulation of the recovery principle. As we will see, the recovery postulate characterizes contraction functions that satisfy the Harper Identity.

Before establishing a converse to Lemma 6, I ask under what circumstances the Harper Identity inverts the Levi Identity, as before in the case of the Levi Identity. The recovery postulate turns out to be sufficient as well as necessary, provided that the consequence relation satisfies two more conditions.

First, as with belief revision functions, I say that a belief contraction function for a theory T *respects double negation* if for all formulas p, it is the case that $T \dot{-} \neg\neg p = T \dot{-} p$. Respect for double negation is an instance of Gärdenfors' postulate $\mathrm{K}^{-}6$. Second, a consequence relation Cn *satisfies disjunctive syllogism* if for all sets of formulas Γ it is the case that if $\Gamma \vdash p \rightarrow q$ and $\Gamma \vdash \neg p \rightarrow q$, then $\Gamma \vdash q$. If a consequence relation satisfies disjunctive syllogism, it licenses arguments of the form "If p, then q. And if $\neg p$, then q. Therefore q." Clearly the standard logic of mathematical practice satisfies this principle. With these conditions in place, the recovery principle is a necessary and sufficient condition for the Harper Identity to invert the Levi Identity.

Proposition 2 *Assume that the consequence relation Cn satisfies disjunctive syllogism, and let $\dot{-}$ be a belief contraction function for a theory T that respects double negation. Then the Harper Identity inverts the Levi Identity applied to $\dot{-} \iff$ for all formulas p, $T \dot{-} p + p = T + p$.*

Proposition 2 immediately yields a characterization of the belief revision functions that are consistent with the Harper Identity.

Corollary 2 *If the consequence relation Cn satisfies disjunctive syllogism, a belief contraction function $\dot{-}$ for a theory T that respects double negation can be generated by the Harper Identity $\iff$ for all formulas p, $T \dot{-} p + p = T + p$.*

7 In search of contraction functions

The Levi Identity suggests that we construct belief revision functions from contraction functions. Corollary 1 shows that Pareto-minimal belief revision functions can be so constructed. And conversely, if a contraction function generates a belief revision function, the belief revision function is Pareto-minimal provided that it makes no change to the current theory when the current theory is consistent with the evidence. Thus Pareto-minimality and the Levi Identity together strongly support basing minimal belief change on belief contraction.

So to find further constraints on minimal belief revision, we may look first for further constraints on belief contraction. An obvious idea is to require that a belief contraction function $\dot{-}$ should be such that it satisfies the Harper Identity. By Corollary 2, this amounts to requiring that the belief contraction function $\dot{-}$ should satisfy the recovery principle: $T\dot{-}p + p = T + p$ for all theories T and formulas p. Unfortunately, whereas the recovery principles does constrain belief contraction, combining it with the Levi Identity does not yield any constraints on belief revision. Makinson discusses this issue in detail (Makinson, 1987; see also Gärdenfors 1988, Ch.3.6). Basically, the reason is this. Suppose we stipulate that a revision function $*$ ought to be generated by a contraction function $\dot{-}$ that satisfies the Harper Identity. From Proposition 1, we know that a contraction function that generates $*$ is given by $T\dot{-}p = T \cap T * \neg p$ (if one exists at all). By Proposition 2, we have that $T\dot{-}p$ satisfies the Harper Identity if and only if $T\dot{-}p + p = T + p$. So for $\dot{-}$ to satisfy the Harper Identity, the required constraint on the associated revision function $*$ is that $(T \cap T * \neg p) + p = T + p$. Now by Lemma 2, $(T \cap T * \neg p) + p = (T + p) \cap (T * \neg p + p)$, which is $T + p \cap L = T + p$ since $T * \neg p \vdash \neg p$ and so $T * \neg p$ is inconsistent with p. This requires only that $T * \neg p \vdash \neg p$, which is the case for any revision function. Thus if a revision function $*$ is generated by any contraction function $\dot{-}$, then $*$ is generated by a contraction function that satisfies the Harper Identity. Therefore the Harper Identity does not yield constraints on belief revision.

Another suggestion is to choose retraction-minimal contraction functions. A contraction $T\dot{-}p$ is retraction-minimal just in case there is no other contraction T' of T such that T' retracts less from T than $T\dot{-}p$ does (in the sense of Definition 3). Clearly the only retraction-minimal contraction on a theory T retracts nothing, that is, $T\dot{-}p \vdash T$. Then by the Levi Identity, it follows that $T * p = T\dot{-}\neg p + p = T + p$. Thus if belief contraction minimizes retractions, then belief revision is just belief expansion: adopting the logical consequences of adding the new information to the current beliefs. And we saw that in terms of our decision-theoretic approach to belief revision, the expansion function $+$ is the only Pareto-minimal belief revision function that minimizes retractions. Although there are thus several considerations pointing towards expansion as minimal belief change, there is one big problem: when the new information is inconsistent with the current theory, just adding the new information leads to inconsistent beliefs.

The obvious remedy is to require that contraction functions should be *consistent* and retraction-minimal. Recall from Lemma 6 that if a belief contraction function $\dot{-}$ is consistent, then the belief revision function $*$ associated with $\dot{-}$ produces consistent beliefs (provided that the new information p itself is consistent). A belief contraction $T \dot{-} p$ is consistent and retraction-minimal just in case (1) $T \dot{-} p$ is consistent in the sense of Definition 9, and (2) there is no other consistent retraction T' of T on p such that T' retracts less from T than $T \dot{-} p$ does. However, it turns out that applying the Levi Identity to consistent retraction-minimal contractions leads to a complete theory $T * p$ whenever the current theory T contradicts the new information p. We saw the basic reason for this in Section 5: If $T \vdash \neg p$, then by Implication, $T \vdash p \rightarrow q$ and $T \vdash p \rightarrow \neg q$, for any formula q. If the contraction $T \dot{-} \neg p$ is consistent and retraction-minimal, it will retract one of these implications but not both. Thus either $T \dot{-} \neg p + p \vdash q$ or $T \dot{-} \neg p + p \vdash \neg q$ for any formula q; in other words, the revision $T * p = T \dot{-} \neg p + p$ produces a complete theory.

As a claim about minimal belief change, it intuitively seems false that whenever an agent's current beliefs are inconsistent with new information, all minimal revisions of her beliefs should lead her to have a definite opinion about every possible fact. Thus belief revision theorists do not require contraction functions to be retraction-minimal [“maxichoice” (Gärdenfors, 1988, Ch.4.2)] (see Gärdenfors, 1988, pp.58–59; Levi, 1996, p.22). Much of belief revision literature has the aim of developing constraints on belief contraction that do not lead an agent into complete theories whenever the new information contradicts her current beliefs. There is no space here to review all proposals comprehensively, but I will briefly describe three of the main approaches that are still subject of current research. The first employs so-called belief bases, the second “update” approach essentially abandons the idea that contraction functions should avoid retractions above all additions, and the third project is the standard AGM axiomatization.

Belief Bases. A belief base is just a set of formulas (Nebel, 1989, 1994; Schulte, 1999). Hence a belief base is not necessarily closed under logical consequence. What does a belief base represent? An interesting interpretation of belief bases is that they capture the *fundamental* or *basic* beliefs of an agent. A rational agent will also have other beliefs, but these we may view as “mere consequences” of the basic beliefs. An interesting nonepistemic interpretation of what a belief base represents is to view a database as a belief base by identifying data records with formulas. Clearly the records in a database entail all sorts of facts that are not themselves part of the database (e.g., “there is only one employee who smokes”).

Now we may define Pareto-minimal revision of belief bases much as we did for theories, by considering only changes in the basic beliefs, not in the consequences of basic beliefs. [For more details, see Schulte (1999, Sect. 6).] A Pareto-minimal revision of B on new information p is just $B \cup \{p\}$. This returns us to the problem that revising a belief base on information inconsistent with it yields an inconsistent set of beliefs. As in the case of theories, the obvious remedy is to restrict attention to consistent Pareto-minimal revisions of belief bases. One might think that this will again lead to a complete set of beliefs, as in the case of theories, but as Alchourrón and Makinson observed in an important paper, this undesirable result does not obtain for belief bases (Alchourrón and Makinson, 1982). For the simplest possible example, consider the belief base $B = \{\neg a\}$ (the first object is not on the table). Note that B *entails* the "irrelevant implications" $a \rightarrow b$ and $a \rightarrow \neg b$ (for example), but B does not *contain* these implications. Given new information a, the consistent Pareto-minimal revision of B is just $B * a = \{a\}$. This revision $B * a$ no longer entails the implication $a \rightarrow b$ (for example), but that implication was not a *basic* belief. The only basic belief that $B * a$ retracts from B is $\neg a$, which any consistent revision of B on p must retract. Hence $B * p$ is retraction-minimal with respect to basic beliefs. But clearly $B * p$ is not a complete theory.

Schulte proves that the Levi identity applied to retraction-minimal base contractions characterizes the Pareto-minimal base revisions (Schulte, 1999, Th. 14). More precisely, say that a belief base B' is a consistent contraction of B on p if $B \dot{-} p \subseteq B$ and $B \dot{-} p \nvdash p$. A belief base B' is a consistent retraction-minimal contraction of B on p if (1) B' is a consistent contraction of B on p and (2) no other consistent contraction B^* of B on p retracts less from B than B' does. Then we have the following result: A consistent revision $B * p$ is a Pareto-minimal contraction $\iff$ there is a consistent retraction-minimal contraction $B \dot{-} \neg p$ such that $B * p = B \dot{-} \neg p \cup \{p\}$.

Thus the Levi Identity, applied to retraction-minimal contractions, gives exactly the necessary and sufficient conditions for Pareto-minimal base revisions. And unlike in the case of theories, consistent retraction-minimal contractions do not lead to a definite belief about every possible assertion. A sample of papers on base revision with further references is Nebel (1994, 1989), Hansson (1993, 1998), and Meyer (1999).

Update: abandoning retraction-minimality. Let us return to modelling the agent's beliefs by logically closed theories. In that case the aim of minimizing retractions causes the difficulty that it leads to complete beliefs, at least when formulated as the directive to choose retraction-minimal theories. If we abandon this aim, the difficulty disappears. In general, we may allow any Pareto-minimal trade-off between additions and retractions, with no special weight given to avoiding retractions. In some influential papers, Katsuno and Mendelzon argued that treating additions and retractions on a par is appropriate when the agent receives information about changes in the world ("update") rather than new information about a static world (Katsuno and Mendelzon, 1991). The result is a theory of minimal change that discards the preservation principle K*4, but otherwise is generally compatible with the AGM postulates. Specifically, Katsuno and Mendelzon introduce an update function $\diamond$ such that $T \diamond p$ is a set of sentences. Their postulates for update functions $\diamond$ are as follows.[7]

[7] For ease of comparison, I have adapted Katsuno and Mendelzon's notation to the one used in this paper. Also, they consider finite belief bases rather than deductively closed theories.

U1 $T \diamond p \vdash p$.
U2 If $T \vdash p$, then $T \diamond p = T$.
U3 If T and p are each consistent, then $T \diamond p$ is consistent.
U4 If p and q are logically equivalent, then $T \diamond p = T \diamond q$.
U5 $(T \diamond p) + q \vdash T \diamond (p \wedge q)$.
U6 If $(T \diamond p) \vdash q$ and $(T \diamond q) \vdash p$, then $T \diamond p = T \diamond q$.
U7 If T is complete, then $(T \diamond p) \cup (T \diamond q) \vdash T \diamond (p \wedge q)$.
U8 $(T \cap T') \diamond p = (T \diamond p) \cap (T' \diamond p)$.

The AGM *Axiomatic Approach*. The most wide-spread approach has been to follow the preservation principle so long as it "works", that is, to restricts its application to the case in which the current theory T is consistent with the new evidence p. In that case, the AGM theory chooses the logically strongest Pareto-minimal revision, namely $T + p$. In terms of contraction functions, this amounts to choosing $T \dot{-} p = T$ whenever $T \nvdash p$, which is Gärdenfors' postulate $K^{-}3$. For the case in which the new information is inconsistent with the current beliefs, we may seek a supplementary account of minimal contraction, which need not be reflected in axioms (cf. (Gärdenfors 1988, Ch.4)). This approach has advantages and disadvantages. I will briefly review a few considerations; more discussion, especially pertaining to the preservation principle may be found in Gärdenfors (1988, Ch.7.4), Schulte (1999, Sect. 7), and Levi (1988).

One disadvantage is that from our decision-theoretic perspective, the AGM axioms are perched between retraction-minimality and Pareto-minimality. When the new information is consistent with the current theory, the AGM axioms give lexicographic preference to avoiding retractions rather than avoiding additions. When the new information is inconsistent with the current theory, the AGM axioms allow any Pareto-minimal trade-off whatsoever. It seems that there ought to be a principled reason why the extent of belief change should be assessed so differently in the two situations.

Another disadvantage is that, as it turns out, the preservation principle K*4 makes it difficult to connect axioms for minimal belief change with axioms for reasoning about conditionals (statements of the form "if p, then q"). I will come back to this issue in Section 10.

On the other hand, many theorists find K*4 and its concomitant, $K^{-}3$ intuitively plausible, certainly plausible enough to pursue a theory that incorporates these principles.

Perhaps the most influential reason for accepting the AGM axioms is that they define axiomatically what many researchers view as a natural and plausible model of belief change. This is the content of the important *Grove Representation Theorem* which I describe in the next section.

8 The Grove representation theorem

As its name suggests, the Grove Representation Theorem is analogous to other representation theorems such as Savage's for maximizing expected utility (Savage, 1954). A comparison might be helpful to clarify the import of Grove's theorem. Let us think of maximizing expected utility as a plausible and natural model for how a rational agent chooses. Savage's theorem shows that we can think of an agent as choosing in this way just in case her binary choices between options (which reveal her preference ordering among these options) satisfy a certain set of axioms. Grove's approach also begins with a natural model of theory choice, and then proves that we can think of an agent as revising her theories in this way just in case her choice of a new theory (her belief revision function) satisfies certain axioms (Grove, 1988).

It is easiest to describe Grove's model of belief change in terms of possible worlds. Aside from formal convenience, this approach will also provide an opportunity to illustrate a technique for passing from a syntactic to a semantic setting that is often useful in formal logic.

In the dominant philosophical tradition, part of the notion of a possible world is that every meaningful statement p is either true or false in a given possible world. In standard logical notation, we write $w \vDash p$ if the statement p is true at a world w. If w is a possible world, no contradiction is true in w, and we thus expect that the set of all statements true at w (i.e., $\{p : w \vDash p\}$) forms a consistent set. Thus the set of statements true at w forms a *maximal consistent set*: First, $\{p : w \vDash p\} \nvdash L$, and second, if $q \notin \{p : w \vDash p\}$, then $\{p : w \vDash p\} \cup \{q\} \vdash L$. Now, since the possible worlds ("possible states of the world", "outcome space") are generally just uninterpreted points, nothing prevents us from taking maximal consistent sets of formulas to be possible worlds. Thus for a language L and consequence relation $\vdash$, there is a natural set of possible worlds $W_{\vdash}$, namely $W_{\vdash} = \{T : T$ is a maximal consistent theory$\}$. When w is a maximal consistent set, we clearly have that $w \vDash p \iff p \in w$.

In our example scenario with three objects, a maximal consistent set of formulas contains either a or $\neg a$, and either b or $\neg b$, and either c or $\neg c$. Thus there is a one-to-one mapping between maximal consistent sets of formulas and the $2^3 = 8$ sets $\{a,b,c\}$, $\{a,b,\neg c\}$, $\{a,\neg b,c\}$, $\{a,\neg b,\neg c\}$, $\{\neg a,b,c\}$, $\{\neg a,b,\neg c\}$, $\{\neg a,\neg b,c\}$, $\{\neg a,\neg b,\neg c\}$. For purposes of illustration, let us take these sets as possible worlds. Then we may write, for example, that $\{a,b,c\} \vDash a$.

Note that we may identify any consistent theory T, for example one modelling an agent's beliefs, with a set of possible worlds: the set of possible worlds in which all assertions in T are true. Formally, we may write $[T] = \{w :$ for all $p \in T$, it is the case that $w \vDash p\}$. For example, if the agent's belief is $T = \mathrm{Cn}(\{a,b\})$, then $[T] = \{\{a,b,c\},\{a,b,\neg c\}\}$. Conversely, for a set of possible worlds P, the set of all sentences true at all worlds in P is a theory, which I denote by $\langle P\rangle$. For example, if $P = \{\{a,b,c\},\{a,\neg b,c\}\}$, then $\langle P\rangle = \mathrm{Cn}(\{a,b,c\}) \cap \mathrm{Cn}(\{a,\neg b,c\})$, which is equal to $\mathrm{Cn}(\{a,c\})$.

Now consider an agent who ranks possible worlds according to a prewellorder $\leq$. This means that the ordering $\leq$ is total, reflexive and transitive; furthermore, the equivalence classes defined by $w \sim w' \iff w \leq w'$ and $w' \leq w$, are well-ordered. Thus in every non-empty set of equivalence classes, there is a minimal element. The intuitive interpretation of the $\leq$ ordering is that it

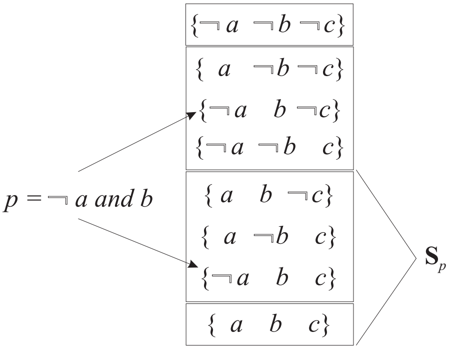

Figure 4. A grove sphere system centered on $[T] = \{\{a, b, c\}\}$

represents something like the plausibility rank that an agent assigns to a possible world. The minimal worlds at the bottom of the ordering are the most plausible worlds, which correspond to the agent's current beliefs. Worlds at higher levels of the ordering represent possibilities that the agent does not believe to obtain but which he is nonetheless able to rank by plausibility.

For an intuitive representation of the pre-wellorder $\leq$, imagine a system of nested "spheres" where each sphere contains an equivalence class of possible worlds as well as all the preceding equivalence classes. Figure 4 illustrates such a system, albeit with "boxes" rather than the customary spheres for the sake of readability. A system of spheres derived from a pre-wellorder satisfies the following conditions.

Definition 12 *Let W be a set of possible worlds. A collection $\mathbf{S}$ of subsets of W (i.e., $\mathbf{S} \subseteq 2^W$) is a Grove sphere system centered on $[T] \iff$*

1. *$\mathbf{S}$ is totally ordered by $\subseteq$ (i.e., if S and S' are in $\mathbf{S}$, then $S \subseteq S'$ or $S' \subseteq S$).*
2. *$[T]$ is the $\subseteq$-minimum of $\mathbf{S}$ (i.e., $[T] \in \mathbf{S}$ and for all $S \in \mathbf{S}$, $[T] \subseteq S$).*
3. *W is in $\mathbf{S}$.*
4. *If p is a formula and there is some sphere $S \in \mathbf{S}$ intersecting $[p]$ (i.e., $S \cap [p] \neq \emptyset$), then there is a smallest sphere in $\mathbf{S}$ intersecting $[p]$.*

To illustrate, suppose that the agent currently believes that all three objects are on the table; take $T = \mathrm{Cn}(\{a, b, c\})$. Let us measure the "distance" *dist* between two possible worlds by the number of propositional letters on which they disagree. Thus for example $\mathit{dist}(\{a, b, c\}, \{\neg a, b, c\}) = 1$, and $\mathit{dist}(\{a, b, c\}, \{\neg a, \neg b, \neg c\}) = 3$ (this is the Hamming distance, a commonly used metric; cf. (Chou and Winslett 1994)). Then we can rank possible worlds according to their distance from the agent's current beliefs; formally, I define $w < w' \iff \mathit{dist}(\{a, b, c\}, w) < \mathit{dist}(\{a, b, c\}, w')$. Figure 4 shows the resulting Grove sphere system.

If a formula p is consistent, then there is a smallest sphere $\mathbf{S}_p$ that intersects $[p]$. Thus for any consistent formula p, the set $C(p) = \mathbf{S}_p \cap [p]$ contains the "most plausible" or "closest" worlds in which p is true. If p is inconsistent, we let $C(p) = \emptyset$. For example, let $p = \neg a \wedge b$. Then $[p] = \{\{\neg a, b, c\}, \{\neg a, b, \neg c\}\}$. Thus the smallest sphere intersecting $[p]$ is $\mathbf{S}_p = \{\{a, b, c\}, \{\neg a, b, c\}, \{a, \neg b, c\}\{a, b, \neg c\}\}$. So $C(p) = \mathbf{S}_p \cap [p] = \{\{\neg a, b, c\}\}$. In our interpretation, given that the agent believes that all three objects are on the table, upon learning that the first is not and the second is on the table, the closest scenario to his current beliefs, according to Hamming distance, is that the first object is off the table, the second is on and the third is off.

Given a Grove sphere system centered on $[T]$, Grove's method for revising a theory T on new information p is simply to let the new theory contain all and only the formulas true at the most plausible worlds consistent with p; formally, we have that $T * p = \langle C(p) \rangle$. In the example above, $T * p = \mathrm{Cn}(\{\neg a, b, c\})$.

The following theorem shows that this model of theory change exactly represents AGM theory revision.

Theorem 2 (Grove, 1988) *Let language L and consequence relation $\vdash$, and hence $W_\vdash$, be given. Suppose that $*$ is a belief revision function for theory T. Then $*$ satisfies the AGM axioms K^*1 through K^*8 $\iff$ there is a Grove sphere system centered on $[T]$ such that for all formulas p, $T * p = \langle C(p) \rangle$.*

The power of Grove's approach stems from enriching the representation of an agent's epistemic state. Modelling an agent's epistemic state by a theory T allows us to say whether an agent believes or disbelieves an assertion, or is undecided about it. An ordering $\leq$ over possible worlds provides more information, since it tells us how the agent ranks possibilities that she believes do not obtain. Grove's representation shows that we can interpret K*7 and K*8 as ensuring that the agent's ranking of alternative revisions, or worlds in which his current beliefs are false, form a pre-wellorder. We may also connect the pre-wellorder $\leq$ over possible worlds with a ranking of formulas in the language (see Grove, 1988, Sect. 3; Gärdenfors, 1988, Ch.4.8); in that case belief revision theorists often interpret the ordering as representing something like "epistemic entrenchment", the degree to which an agent is committed to various assertions.

In what sense is Grove belief revision minimal belief change? (A question considered in some detail by Rott (Rott, 2000, Sect.III.B).) One answer is that a Grove revision minimizes the change in the pre-wellorder $\leq$ associated with a Grove sphere system. Consider the ranking on possible worlds defined by $w \leq_{\mathbf{S}} w' \iff \mathbf{S}_w \subseteq \mathbf{S}_{w'}$. Given a formula p, there are many ways to form a revised Grove system $\mathbf{S}*p$. But as long as (1) the minimum of $\mathbf{S}*p$ entails p, and (2) for any two worlds w, w' consistent with p, it is the case that $w \leq_{\mathbf{S}} w' \iff w \leq_{\mathbf{S}*p} w'$, we will have that the minimum of $\mathbf{S}*p$ is $\mathbf{S}_p \cap [p]$. This observation suggests an analogy with changing degrees of belief by Bayesian conditioning: In Bayesian updating, the ratio $P(w)/P(w')$ remains constant for any two worlds w, w' consistent with the new information (i.e., $P(w|p)/P(w'|p) = P(w)/P(w')$ whenever $w, w' \models p$; recall that $P(p|w) = P(p|w') = 1$ if $w, w' \models p$). Thus we may compare a Grove update to Bayesian updating, in which (ordinal) "degrees of plausibility" play the role of (real-valued) "degrees of belief" (Spohn, 1987).

It is clear that the principle of keeping constant the relative ranking of possible worlds that are consistent with the new information leads to the preservation principle, K*4. For retracting beliefs corresponds to adding worlds. But if any world w is not among the most plausible ones to begin with (i.e., w is not in the minimum sphere), then a Grove update can include w among the most plausible worlds only if the new information p is inconsistent with all worlds in the minimum sphere, which corresponds to the agent's current theory T. In other words, retracting beliefs is possible only if the new information p is inconsistent with the current theory T, just as K*4 says. In light of our previous discussion, this shows that minimizing change of an agent's plausibility ranking $\leq$ and minimizing change in his actual beliefs is not the same thing: When the new information is consistent with the agent's current beliefs, a Grove update induces essentially no change in the plausibility ranking, but may lead the agent to entertain many new beliefs (cf. Schulte, 1999, Sect. 6; Rott, 2000, p.513).

On the view of belief change suggested by the Grove representation theorem, a theory of belief change amounts to a theory of plausibility rankings. In specific applications, it may be possible to provide more constraints on such rankings. For example, Stalnaker has shown how to incorporate plausibility rankings into game-theoretic models of decision-making (Stalnaker, 1996). This allows us to express a principle such as this (sometimes associated with forward induction (Stalnaker, 1996, Sect. 6; Battigalli, 1996, p.179): player 1 believes that player 2 is rational and hence that player 2 will make a certain move m. But of the two possibilities, (1) player 2 deviates from m by mistake, and (2) player 2 deviates from m because she is irrational, player 1 should consider possibility (1) the more plausible one. Thus if player 1 were to revise her beliefs given the information that 2 did not play m, then by the Grove representation theorem player 1 would come to believe that 2 deviated from m by mistake. Moreover, it could be common knowledge among the players that each considers mistakes more plausible than irrationality; in other words, various aspects of their plausibility rankings – and hence, belief revision functions – could be part of common knowledge among the players.

9 Iterated belief change

At the level of general belief revision theory, much of the work that departs from Grove's theorem has been concerned with iterated belief revision: modelling situations in which an agent first revises her beliefs on p, then on some other formula q, etc. After a revision $T * p$ has taken place, what belief revision function $*'$ should guide the next revision? Since we have now identified a belief revision function $*$ for a given theory T with a Grove sphere system centered on $[T]$, the issue becomes updating an entire sphere system, rather than just finding the next set of beliefs. Even though Grove's representation theorem determines the next theory $T * p$ given a sphere system $\mathbf{S}$, it leaves some freedom in what the next sphere system $\mathbf{S} * p$ should be. There are various proposals for rules for updating plausibility rankings, some of them with axiomatic characterizations (several are presented in Kelly's paper (Kelly, 1999), such as Darwiche and Pearle (1997), Nayak (1994), and Boutilier (1996)).

Once we have iterated belief change in view, it is natural to ask what the long-run behaviour of methods for iterated belief change is. This is a familiar question for Bayesians: results that show long-run convergence to correct beliefs via conditioning have long since been part of Bayesian statistics (Savage, 1954, Ch.3.6; Halmos, 1974, Sect. 49, Th.B). Roughly speaking, it can be shown that given a countably additive probability measure μ over a set of possible worlds, and a proposition p, then as the agent revises her beliefs by conditioning μ on the evidence, with μ-probability 1 the agent's updated degrees of belief converge to 1 if p is true and to 0 if p is false (assuming that the total body of evidence is such that it entails either p or not p). In short, the agent is sure that he will eventually arrive at correct beliefs via Bayesian updating. Is there any counterpart to the Bayesian convergence result for iterated minimal belief change?

At first glance, minimal belief change looks bad for arriving at correct beliefs. Consider an infinitely iterated Prisoner's Dilemma. Suppose that player 1 initially believes the assertion p ="if I always cooperate, player 2 will eventually start cooperating too". Given this belief, he might continue to cooperate waiting for player 2 to start cooperating. But p may be false without ever being falsified by the evidence, for example if player 2 always defects while player 1 cooperates. Since the new information q_n = "player 2 has defected for the last n rounds" is logically consistent with p, the proposition that 2 will eventually cooperate, by the preservation principle K*4 player 1 will never retract his belief in p no matter how many rounds player 2 defects. Thus if in fact player 2 defects as long as player 1 cooperates, player 1 will forever hold a false belief to the effect that his opponent will eventually cooperate – not to mention that player 1's payoffs will be miserable! The general difficulty is that mere logical consistency with the evidence is a very weak test for the correctness of an empirical hypothesis. It is possible for a hypothesis to have much evidence against it without being logically proven false by the evidence (any statistical hypothesis test illustrates this), and it is possible for a hypothesis to be false without ever being *proven* false by the evidence. But AGM belief change is stubborn in the sense that it will not retract a belief until the belief is proven false (Kelly, Schulte, and Hendricks, 1995).

However, the issue warrants a second look. Notice that in the infinitely iterated Prisoner's Dilemma, if player 1 had started out believing that player 2 will always defect, then that belief would be eventually falsified by the evidence if it is false, because if it is false there will be some stage at which player 2 cooperates. This observation suggests that minimal belief change might lead to correct beliefs after all provided we *start* with the right sort of initial beliefs, and update them in the right way when they are logically falsified by the evidence. Indeed, it is possible to prove the following learning-theoretic completeness result for AGM belief revision (stated roughly): Suppose that there is any belief revision method at all that is guaranteed to converge to correct beliefs about a given set of questions, or hypotheses. Then there is a belief revision method that is guaranteed to converge to correct beliefs in the same sense *and* satisfies the AGM axioms.

What does this completeness result tell us about AGM belief revision? The negative conclusion is that we cannot rely in general on AGM belief revision to lead us to correct beliefs, the way that the Bayesian convergence theorems suggest that we can rely on Bayesian updating. The positive view is that the AGM axioms provide only weak guidance as to what an agent should believe, and one might welcome a source of principled constraints on what the agent's initial beliefs and future belief revisions ought to be. The learning-theoretic completeness theorem shows that we can derive such constraints from the precept that belief revision methods ought to be as powerful learners as possible, a principle that a rational agent may well endorse. Kelly has shown how learning-theoretic analysis yields similar constraints on iterated belief revision (Kelly, 1988). It turns out that many of the standard proposals for updating plausibility rankings differ in their ability to lead the agent to correct beliefs and predictions, and for many of these updating methods their learning power depends on subtle specifications on their parameters, in ways that Kelly describes precisely.

The formal and computational learning theory involved in analyzing belief revision methods is quite substantive, and further details are beyond the scope of this paper. Kelly et al. discuss various interpretations of iterated belief revision as learning, and give a completeness result for AGM belief revision in a somewhat limited but fairly simple setting (Kelly, Schulte, and Hendricks, 1995). Martin and Osherson prove another completeness theorem that is more general (and more complicated) (Martin and Osherson, 1998). Schulte provides a brief, informal summary of their work (Schulte, 2000). Osherson and Martin's book treats many aspects of AGM revision operators guaranteed to converge to correct beliefs (Martin and Osherson, 1998).

10 Conditionals

One of the major insights of contemporary logic is that various forms of nondeductive, nonprobabilistic reasoning are closely related, especially belief revision, nonmonotonic reasoning and conditional logic. A conditional is a statement of the form "if p, then q". Arguably conditionals are closely connected with causality (Gibbard and Harper, 1981). Conditionals are also important for strategic thinking (Stalnaker, 1996): In game theory, agents must think about what would happen if they were to make certain moves. In this section I consider some of the characteristics of conditionals and their relationship with nonmonotonic reasoning and belief revision.

10.1 Conditionals and defeasibility

Some of the main properties of conditionals come out clearly when we contrast them with material implication. In material implication – in a mathematical theorem – a statement of the form "if p, then q", or $p \to q$ in my formal notation, is false just in case p is true and q is false, and otherwise true. If p is false, then $p \rightarrow q$ is vacuously true, but true nonetheless. Thus the truth or falsehood of a material implication $p \rightarrow q$ is determined if we know whether p and q are true or false. With the other standard logical connectives, it is also the case that the truth values of their compounds (p, q) determine the truth value of the compound; for example, $p \wedge q$ is true iff both p and q are true. Logicians call such connectives *truth-functional*. Conditionals are not truth-functional. That is, if we write $p > q$ for "if p, then q", the truth values of p and q do not necessarily determine the truth value of $p > q$. For example, the conditional "if I were the pope, then 2>4" is false. And the conditional "if Dole had defeated Clinton in the presidential election, he would not be advertising Viagra" is true. However, in each example both antecedent and consequent are false.

Another very important property that distinguishes conditionals from material implication is their *defeasibility*. For material implication, it is the case that whenever $p \rightarrow q$, then $(p \wedge r) \rightarrow q$. But it is not always, or even typically, the case that whenever $p > q$, then $(p \wedge r) > q$. To use Nelson Goodman's example, I can consistently assert that "if I strike the match, it will light" along with "if I strike the match and it is wet, it will not light".

The essence of defeasibility is that knowing more facts undoes conclusions that are warranted in the light of fewer facts. In other words, the set of correct conclusions does not grow monotonically with the set of premises. One approach to capturing defeasibility is therefore to consider consequence relations that do not satisfy the monotonicity property. The field of nonmonotonic reasoning has received extensive development (a good overview is Brewka, Dix, and Konolige, 1997). In the remainder of this paper, I consider the logic of defeasible conditionals with a consequence relation that is monotonic.

As we have seen, the truth values of p and q by themselves do not determine the truth value of $p > q$. What more information is needed to settle the truth or falsehood of $p > q$? Let's consider again our example. Suppose that Nelson is about to strike an apparently dry match, and asserts that "if I strike this match, it will light". If Mary challenges his assertion by noting that if the match were wet, it wouldn't light, Nelson would reasonably reply that he meant that the match would light assuming that things generally stay as they are now. So one might say that conditionals implicitly assume an "other things equal" condition: $p > q$ holds if, whenever p is true and "all other circumstances remain the same", then q is true. This condition is too strict, however, because it is impossible for one fact to change and for others to remain exactly the same. For example, if the match is struck, someone has expended energy, the match is subject to a new force, etc. Arguably, the correct analysis is that $p > q$ holds if, whenever p is true and all other circumstances are as similar as possible, then q is true. The challenge, then, is to analyze the notion of one set of circumstances being "similar" to another, at least with enough precision to give a foundation for a formal logic of conditionals. This is the project of Lewis' and Stalnaker's famous work on conditionals (Lewis, 1981; Stalnaker, 1981). The result of their work is a formal logic of conditionals, which is demonstrably sound and complete with respect to the interpretation of conditionals that I have sketched, and whose mathematical properties have been studied in considerable detail. For example, the computational complexity of Lewis-Stalnaker conditional logics is well understood, as is their relationship to nonmonotonic reasoning and belief revision principles (see for example Arló-Costa, 1995). It turns out that various axiom systems for conditionals are characterized by properties of Lewis sphere systems, which are much like Grove sphere systems. If we interpret the ranking specified by a sphere system as representing similarity, the sphere system defines truth conditions for conditionals along the lines sketched above. Similarly, systems of nonmonotonic logic can be characterized by various kinds of rankings of possible worlds, such as pre-wellorders and metrics. Kraus, Lehmann and Magidor provide a careful examination of various relationships between possible worlds that supply an interpretation for nonmonotonic consequence relations (Kraus, Lehmann, and Magidor, 1990). Rott examines the connection between nonmonotonic reasoning and belief revision in light of constraints on selection functions familiar from rational choice theory (Rott, 1998). To see the general idea, consider a nonmonotonic relation *Inf*: $L \to \mathbf{T}$ that yields for each sentence p its nonmonotonic, or “default” consequences *Inf*(p). With each theory T and belief revision function $*$ we may associate such an inference relation by setting *Inf*(p) = $T * p$. With this definition, it is possible to prove that various belief revision postulates correspond to well-known principles of nonmonotonic reasoning, in the sense that a belief revision function $*$ for a theory T satisfies the belief revision postulate iff the nonmonotonic consequence relation *Inf* associated with $*$ and T satisfies the reasoning principle in question (cf. Rott, 1998, Sect. 3; Maksinon and Gärdenfors, 1991). There is a similar connection between belief revision functions and conditionals, which I consider in some detail in the next subsection.

10.2 Conditionals and belief revision: the Ramsey test

Instead of asking when a conditional $p > q$ is true, we might ask a different question: when should an agent accept a conditional $p > q$? Along the lines developed above, we may argue that the acceptance conditions for $p > q$ are as follows: Given the agent’s current beliefs about the world, the agent should ask herself whether q would obtain if p were true and other circumstances were as much as possible the way that she now believes them to be. In other words, the agent should – for the sake of the argument – incorporate the antecedent p into her beliefs but otherwise change her beliefs as little as possible. If we introduce a belief revision function $*$ to represent minimal belief change, we can state the acceptance condition for a conditional $p > q$ as follows: An agent should accept $p > q$ given her current theory T just in case $T * p \vdash q$.

Since the agent’s current theory T contains exactly the set of assertions that the agent currently accepts – including conditionals – this means that $T \vdash p > q \iff T * p \vdash q$. The proposal to analyze acceptance conditions for conditionals

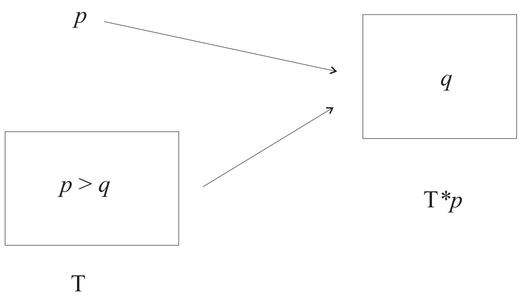

Figure 5. The Ramsey test: $T \vdash p > q$ just in case $T * p \vdash q$

in terms of belief revision originated with Gärdenfors, who credited the basic idea to the logician and philosopher Frank Ramsey. For that reason, belief revision theorists refer to the condition that $T \vdash p > q \iff T * p \vdash q$ as the *Ramsey test* for the acceptability of a conditional. Figure 5 illustrates the Ramsey Test.

For example, let m stand for "the match is struck", and let l stand for "the match lights". Suppose that an agent believes that "if the match is struck, it will light"; formally, $T \vdash m > l$. Then the Ramsey test says that if the agent were to add the information m to his stock of beliefs, then the agent would believe l; formally, $T * m \vdash l$, where $*$ denotes the operation of including the information m as a new belief in a minimal way. Conversely, suppose that upon including the information m, the agent believes l; formally, suppose that $T * m \vdash l$. Then the Ramsey test requires that the agent accept the conditional $m > l$, so that $T \vdash m > l$.

Recall that the Deduction Principle for material implication asserts that $T \vdash p \rightarrow q \iff T + p \vdash q$. Thus the Ramsey test is a kind of Deduction Principle for the conditional $>$ where the belief revision function $*$ plays the role that the expansion function does for material implication. Note also the similarity between the Ramsey test and associating a nonmonotonic consequence relation with a theory and a belief revision operator. We can complete the circle of translations between belief revision, nonmonotonic reasoning and conditionals by connecting conditionals with nonmonotonic consequence via the principle $p > q \iff q \in \mathit{Inf}(p)$.

We will require an extended formulation of the Ramsey test that applies it not just to one given theory T, but to a range of theories that might represent the agent's epistemic state. For that we need a belief revision function that is defined for a range of theories that might represent an agent's belief state, rather than a specific one. Hence the following definition.

Definition 13 (Gärdenfors) *A* belief revision system $\mathscr{K} = \langle \mathbf{K}, * \rangle$ *is a set of theories* $\mathbf{K}$ *in a language* L *with conditionals and a belief revision function* $* : \mathbf{K} \times L \rightarrow \mathbf{K}$ *such that* $\mathbf{K}$ *is closed under expansions. That is, if* $T \in \mathbf{K}$ *and* $p \in L$*, then* $T + p \in \mathbf{K}$*.*

Extending the previous notation, I write $T * p$ for the result of applying a belief revision function $*$ to a theory T and a formula p. Now we are ready to define the connection between the conditionals that an agent accepts with how she revises her beliefs postulated by the Ramsey test.

Definition 14 (Gärdenfors) *A belief revision system* $\mathscr{K} = \langle \mathbf{K}, * \rangle$ satisfies the Ramsey test *iff* $T \vdash p > q \iff T * p \vdash q$, *for all theories* $T \in \mathbf{K}$, *and formulas* $p, q \in L$.

This last definition requires that our basic language L is closed under the conditional connective, that is, if two formulas p, q are in L, then so is the formula $p > q$. In the remainder of this paper, I always assume that the basic language L is closed under the conditional connective.

Consider a belief revision system $\langle \mathbf{K}, * \rangle$ that satisfies the Ramsey test. Suppose we place constraints on the belief revision operator $*$. Then we can ask whether there are axioms governing the conditional $>$ that are valid in $\langle \mathbf{K}, * \rangle$ in the sense that all theories in the system entail them. For example, if the belief revision function $*$ satisfies K*2, then for all theories $T \in \mathbf{T}$ and formulas p, we have that $T * p \vdash p$, and so by the Ramsey test $T \vdash p > p$. Thus in any belief revision system satisfying the Ramsey test and K*2, the basic conditional axiom $p > p$ is accepted in every belief state. And conversely, if $p > p$ is accepted for every formula p and theory T in a belief revision system $\langle \mathbf{K}, * \rangle$ satisfying the Ramsey test, then $*$ satisfies K*2. In this sense the conditional axiom $p > p$ corresponds to the belief revision postulate K*2.

What conditional axioms characterize Pareto-minimal revision functions? Before I give the answer with the next theorem, let's state the exact definition of what it is for a formula to be valid in a belief revision system.[8]

Definition 15 *A formula p is* valid in a belief revision system $\mathscr{K} = \langle \mathbf{K}, * \rangle \iff$ *all theories* $T \in \mathbf{K}$ *entail* p.

Now we are ready to establish which conditional axioms exactly characterize Pareto-minimal theory change.

Theorem 3 *Let* $\mathscr{K} = \langle \mathbf{K}, * \rangle$ *be a belief revision system satisfying the Ramsey test. Then $*$ is a Pareto-minimal belief revision operator* $\iff \mathscr{K}$ *validates*

1. $p > p$, *and*
2. $(p > q) \rightarrow (p \rightarrow q)$, *and*
3. $(p \wedge q) \rightarrow (p > q)$

for all formulas p, q, r.

The proof is in Schulte (1999).

It is worth noting that there is no similar characterization of the AGM postulates in terms of conditional axioms. One problem is that K*2, the postulate

[8] My definition of validity is equivalent to Gärdenfors' (see Schulte, 1999, Sect. 9).

that $T * p = L \iff \vdash \neg p$ is inconsistent with the Ramsey test, as Arló-Costa has shown (Arló-Costa, 1990). The source of the problem is that K*2 requires a revision of the inconsistent theory to be consistent. If we amend K*2 to read $T * p = L \iff T = L$ or $\vdash \neg p$, there is no longer a problem with the Ramsey test. But more fundamentally, in a certain sense made precise by Gärdenfors, the Ramsey test is inconsistent with K*4, the preservation principle Gärdenfors (1988), Ch.7; interestingly, we saw that K*4 is also the main difference between Pareto-minimal and AGM theory revision. [For more discussion of Gärdenfors' result and the Ramsey test, see Gärdenfors (1988, Ch.7), Levi (1988), Arló-Costa (1995), and the references in these papers.]

11 Conclusion

The theory of minimal belief change considers a number of formal models of belief and processes for belief change. A decision-theoretic framework in which we weigh adding and retracting beliefs against each other suggests several natural, precise senses in which a belief change is minimal. Pareto-minimal change appears to be a core constraint on minimal belief change in the sense that belief revisions that are not Pareto-minimal should hardly count as minimal. The main requirement for Pareto-minimality is the AGM postulate K*3 – the revision of a theory T should not be logically stronger than the conjunction of T with the new information. We obtain stronger notions of minimal theory change by adding the preservation principle – do not retract beliefs unless they are inconsistent with the new information – which corresponds to lexicographically giving first priority to avoiding retractions over avoiding additions, or by accepting the full set of AGM postulates. Grove's representation theorem for the AGM postulates shows that they correspond precisely to a rich and plausible model of belief change. The theorem illustrates the interest in formal representations of belief other than identifying beliefs with a logically closed set of sentences, such as ordinal rankings of possible states of affairs. Belief bases – sets of sentences that need not be closed under logical consequence – are another interesting example of an alternative model of belief.

I considered in some detail the process of belief contraction, and the relationships between contraction and revision established by the Levi and Harper identities. I provided necessary and sufficient conditions for when a revision function can be defined via the Levi identity, and for when a contraction function can be defined via the Harper identity. For the Levi identity, the condition is essentially that the revision function must satisfy K*3, which also characterizes Pareto-minimal belief change.

Belief revision theory takes a highly abstract and general perspective. As a result, it potentially applies to a wide range of situations in addition to revising beliefs, such as revising legal codes, and updating knowledge and data bases. The drawback of this generality is that it is difficult to find strong principles that are correct in all potential domains of application, and so the constraints offered by the theory, even the full set of AGM postulates, are far from determining specific belief revision procedures. More guidance may come from considering specific rankings or metrics for possible worlds (e.g., Chou and Winslett, 1994), from examining the learning power of revision operators (Kelly, 1998; Martin and Osherson, 1998), or from knowledge of particular domains, for example rationality assumptions in game theory (Stalnaker, 1996).

A major research topic in modern logic is the close relationship between belief revision and other logical formalisms, such as nonmonotonic reasoning and the logic of conditionals (statements of the form "if p, then q"). The Ramsey test establishes a connection between belief revision postulates and axioms for reasoning about conditionals that allows us to prove exact correspondences between them. It turns out that Pareto-minimal belief revision corresponds exactly to a group of well-known conditional axioms.

Belief revision theory, together with nonmonotonic and conditional logics, is a highly developed and fruitful area of modern logic. Even though we do not yet have a well-developed application in economics, the combination of these logical tools with the methods of economics promises to expand our understanding of belief change in strategic interactions. It was the aim of this paper to put in place some of the main concepts and techniques of belief revision theory for future applications.

12 Proofs

Lemma 1 *Let T_1, T_2 be two theories. Then $T_1 \cap T_2$ is a theory.*

Proof. Let q be a formula in $\text{Cn}(T_1 \cap T_2)$. By Monotonicity we have that $\text{Cn}(T_1 \cap T_2) \subseteq \text{Cn}(T_1)$ and that $\text{Cn}(T_1 \cap T_2) \subseteq \text{Cn}(T_2)$, so it follows that $q \in \text{Cn}(T_1)$ and $q \in \text{Cn}(T_2)$. Since T_1 and T_2 are theories, we have that $q \in T_1 \cap T_2$. Hence $\text{Cn}(T_1 \cap T_2) \subseteq T_1 \cap T_2$, and thus Monotonicity applied to $T_1 \cap T_2$ establishes that $\text{Cn}(T_1 \cap T_2) = T_1 \cap T_2$.□

Lemma 2 *Let T_1, T_2 be two theories. For any formula p, $(T_1 \cap T_2) + p = (T_1 + p) \cap (T_2 + p)$.*

Proof. ($\subseteq$) Let q be a formula in $(T_1 \cap T_2) + p$. Then by Deduction $T_1 \cap T_2 \vdash p \rightarrow q$, and so by Lemma 1, $p \rightarrow q \in (T_1 \cap T_2)$. Hence $p \rightarrow q$ is in T_1 and in T_2, and so by Modus Ponens, $q \in T_1 + p$ and $q \in T_2 + p$, as required.

($\supseteq$) Let q be a formula in $T_1 + p \cap T_2 + p$. Then by Deduction $T_1 \vdash p \rightarrow q$ and $T_2 \vdash p \rightarrow q$. Since T_1 and T_2 are theories, this entails that $p \rightarrow q \in (T_1 \cap T_2)$. Hence by Modus Ponens, $q \in (T_1 \cap T_2) + p$, as required.□

Lemma 3 *Let $\dot{-}$ be a belief contraction function for a theory T, and let $*$ be the function associated with $\dot{-}$. Then*

1. *$*$ is a belief revision function for T, and

2. *if $\dot{-}$ is consistent, then $*$ is consistent. (In other words, if $\dot{-}$ satisfies $K^{-}4$, then $*$ satisfies K*5).*

Proof. Let a formula p be given. Part 1 is immediate since $T \dot{-} \neg p$ is a theory, and hence $T \dot{-} \neg p + p$ is well-defined and also denotes a theory. Also, $T' + p \vdash p$ for any theory T', and hence in particular for $T' = T \dot{-} \neg p$. For part 2, we must show that $T * p = L \iff \vdash \neg p$ given that $\dot{-}$ is consistent. If $\vdash \neg p$, it follows from Part 1 that $T * p \vdash p \wedge \neg p$, and thus by Inconsistency, $T * p = L$. Conversely, suppose that $\nvdash \neg p$. Then since $\dot{-}$ is consistent, it follows that $T \dot{-} \neg p \nvdash \neg p$. So by Consistency, $T \dot{-} \neg p + p \neq L$.□

Lemma 4 *Let $*$ be a belief revision function for a theory T, and let $\dot{-}$ be the function associated with $*$. Then*

1. *$\dot{-}$ is a belief contraction function for T, and*
2. *if $*$ is consistent, then $\dot{-}$ is consistent. (In other words, if $*$ satisfies T*5, then $\dot{-}$ satisfies $K^{-}4$).*

Proof. Let a theory T and a formula p be given. For part 1, note that $T^{*}\neg p$ is a theory, and hence by Lemma 1, so is $T \cap T^{*}\neg p$, which by definition is $T \dot{-} p$. Also, by Monotonicity $T \vdash T \cap T * \neg p = T \dot{-} p$. For part 2, we must show that $T \dot{-} p \vdash p \iff \vdash p$. Again by Monotonicity, we have that $T \dot{-} p \vdash p$ if $\vdash p$. Conversely, suppose that $T \dot{-} p \vdash p$. Since $*$ is a belief revision operator, we have that $T * \neg p \vdash \neg p$. By Monotonicity, $T * \neg p \vdash T \dot{-} p$. Since $T \dot{-} p \vdash p$, we have that $T * \neg p \vdash p \wedge \neg p$. Hence $T * \neg p = L$, and since $*$ is consistent, it follows that $\vdash \neg\neg p$, which by Double Negation implies that $\vdash p$. So $\dot{-}$ is a consistent belief contraction function.□

Lemma 5 *Let $\dot{-}$ be a belief contraction function for a theory T with associated belief revision function $*$. Then for all formulas p, $T + p \vdash T * p$.*

Proof. By the definition of $*$, we have that $T * p = T \dot{-} \neg p + p$. Since $T \vdash T \dot{-} \neg p$, it follows by Monotonicity that $T + p \vdash T \dot{-} \neg p + p = T * p$.□

Proposition 1 *Let $*$ be a belief revision function for T that respects double negation. Then the Levi Identity inverts the Harper Identity applied to $*$ $\iff$ for all formulas p, $T + p \vdash T * p$.*

Proof. Let $\dot{-}$ be the belief contraction function harper($*$) defined by $T \dot{-} p = T \cap T * \neg p$. First we have that (a) $T \dot{-} \neg p + p = (T \cap T * \neg\neg p) + p = (T \cap T * p) + p$ by the assumption that $T * \neg\neg p = T * p$. By Lemma 2 we have that $(T \cap T * p) + p = T + p \cap T * p + p$, which is equal to $T + p \cap T * p$ since $T * p$ is a theory entailing p. Together with (a), this shows that (b) $T \dot{-} \neg p + p = T + p \cap T * p$. Thus $T \dot{-} \neg p + p = T * p$ if and only if $T + p \supseteq T * p$; in other words, if and only if $T + p \vdash T * p$.□

Corollary 1 *A belief revision function $*$ for a theory T that respects double negation can be generated by the Levi Identity $\iff$ for all formulas p, $T + p \vdash T * p$.*

Proof. It follows from Lemma 5 that if a belief revision function $*$ can be generated by contraction, then $T + p \vdash T * p$. Conversely, if for all formulas p, it is the case that $T + p \vdash T * p$, then by Proposition 1, applying the Harper Identity to $*$ yields a contraction function $\dot{-}$ that generates $*$.□

Lemma 6 *Suppose that $*$ is a belief revision function for a theory T, and that $\dot{-}$ is the contraction function associated with $*$. Then for all formulas p, $T \dot{-} p + p = T + p$.*

Proof. By the Harper Identity, $T \dot{-} p + p = (T \cap T * \neg p) + p$, which by Lemma 2 is equal to $T + p \cap (T * \neg p + p)$. Since $T * \neg p$ entails $\neg p$, by Consistency $T * \neg p + p = L$ and so $T + p \cap (T * \neg p + p) = T + p$.□

Proposition 2 *Assume that the consequence relation Cn satisfies disjunctive syllogism, and let $\dot{-}$ be a belief contraction function for a theory T that respects double negation. Then the Harper Identity inverts the Levi Identity applied to $\dot{-} \iff$ for all formulas p, $T \dot{-} p + p = T + p$.*

Proof. Let $*$ be the belief revision function levi($\dot{-}$) defined by $T * p = T \dot{-} \neg p + p$.

($\Rightarrow$) If $\dot{-}$ is the result of applying the Harper Identity to the belief revision function $*$, it follows from Lemma 6 that for all formulas p, it is the case that $T \dot{-} p + p = T + p$.

($\Leftarrow$) Suppose that it is the case that $T \dot{-} p + p = T + p$. We want to show that $T \dot{-} p = T \cap T * \neg p$. By the definition of $*$, we must show that $T \dot{-} p = T \cap (T \dot{-} \neg\neg p + \neg p)$, which is equal to $T \cap (T \dot{-} p + \neg p)$ if $\dot{-}$ respects double negation. It is easy to see that $T \dot{-} p \subseteq T \cap (T \dot{-} p + \neg p)$. For if q is a formula in $T \dot{-} p$, then $q \in T$ since $T \dot{-} p \subseteq T$, and by Monotonicity $T \dot{-} p + \neg p \vdash q$. For the converse, let q be a formula in $T \cap (T \dot{-} p + \neg p)$. Then $q \in T + p$, and so by hypothesis $q \in T \dot{-} p + p$. Thus by Deduction, $T \dot{-} p \vdash p \rightarrow q$. Also $T \dot{-} p \vdash \neg p \rightarrow q$ since $q \in T \dot{-} p + \neg p$. So if Cn satisfies disjunctive syllogism, then $q \in T \dot{-} p$; since q is an arbitrary formula, this establishes that $T \dot{-} p = T \cap (T \dot{-} p + \neg p)$ and hence that $T \dot{-} p = T \cap T * \neg p$. Since this holds for any formula p, the Harper Identity inverts the Levi Identity for the belief contraction function $\dot{-}$, which was to be shown.□

Corollary 2 *If the consequence relation Cn satisfies disjunctive syllogism, a belief contraction function $\dot{-}$ for a theory T that respects double negation can be generated by the Harper Identity $\iff$ for all formulas p, $T \dot{-} p + p = T + p$.*

Proof. It follows from Lemma 6 that if a belief revision function $\dot{-}$ can be generated by revision, then $T \dot{-} p + p = T + p$. Conversely, if for all formulas p, it is the case that $T + p \vdash T * p$, then by Proposition 2, applying the Levi Identity to $\dot{-}$ yields a revision function $*$ that generates $\dot{-}$.□

References

Alchourrón, C.E., Makinson, D.: The logic of theory change: Contraction functions and their associated revision function. Theoria **48**, 14–37 (1982)

Arló-Costa, H.: Conditionals and monotonic belief revisions: the success postulate. Studia Logica **49**, 557–566 (1990)

Arló-Costa, H.: Epistemic conditionals, snakes and stars. In: Conditionals, from philosophy to computer science. Studies in logic and computation. Oxford: Oxford University Press 1995

Battigalli, P.: Strategic rationality orderrings and the best rationalization principle. Games and Economic Behavior **13**, 178–200 (1996)

Boutilier, C.: Iterated revision and minimal change of conditional beliefs. Journal of Philosophical Logic **25**, 263–305 (1996)

Brewka, G., Dix, J., Konolige, K.: *Nomonotonic Reasoning: an Overview*. Stanford, CA: Center for the Study of Language and Information 1997

Chou, T., Winslett, M.: A model-based belief revision system. Journal of Automated Reasoning **12**, 157–208 (1994)

Darwiche, A., Pearle, J.: On the logic of iterated belief revision. Artificial Intelligence **89**, 1–29 (1997)

Gärdenfors, P.: Knowledge in flux: modeling the dynamics of epistemic states. Cambridge, MA: MIT Press 1988

Gibbard, A., Harper, W.: Counterfactuals and two kinds of expected utility. In: Ifs: conditionals, beliefs, decision, chance, and time. Dordrecht: Reidel 1981

Ginsberg, M.L., Smith, D.E.: Reasoning about action i: a possible worlds approach. In: Ginsberg, M.L. (ed.) Readings in nonmonotonic reasoning. Los Altos: Morgan Kaufmann 1987

Grove, A.: Two modellings for theory change. Journal of Philosophical Logic **17**, 157–170 (1988)

Halmos, P.: Measure theory. New York: Springer 1974

Hansson, S.O.: Changes of disjunctively closed bases. Journal of Logic, Language and Information **2**, 225–284 (1993)

Hansson, S.O.: Editorial: Belief revision theory today. Journal of Logic, Language and Information **7**(2), 123–126 (1998)

Harper, W.L.: Rational conceptual change. In: Proceedings of the Meeting of the Philosophy of Science Association, Vol. 2, pp. 462–494. East Lansing, MI: Philosophy of Science Association (1975)

Harper, W.L.: Rational belief change, popper functions and counterfactuals. In: Foundations of probability theory, statistical inference, and statistical theories of science, Vol. I, pp. 73–115. Dordrecht: Reidel 1976

Katsuno, H., Mendelzon, A.O.: 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, pp. 387–394. Los Altos, CA: Morgan Kaufmann 1991

Kelly, K.: Iterated belief revision, reliability, and inductive amnesia. Erkenntnis **50**, 11–58 (1999)

Kelly, K., Schulte, O., Hendricks, V.: Reliable belief revision. In: Proceedings of the IX International Joint Congress for Logic, Methodology and the Philosophy of Science. Dordrecht: Kluwer 1995

Keynes, J.M.: A treatise on probability. London: Macmillan 1921

Kraus, S., Lehmann, D., Magidor, M.: Nomonotonic reasoning, preferential method and cumulative logics. Artificial Intelligence **44**, 167–207 (1990)

Levi, I.: The enterprise of knowledge. Cambridge MA: MIT Press 1980

Levi, I.: Truth, fallibiblity and the growth of knowledge. In: Language, logic and method, pp. 153–174. Dordrecht: Reidel 1983

Levi, I.: Iteration of conditionals and the ramsey test. Synthese **76**, 49–81 (1988)

Levi, I.: For the sake of the argument: Ramsey test conditionals, inductive inference, and nomonotonic reasoning. Cambridge: Cambridge University Press 1996

Lewis, D.: Counterfactuals and comparative possibility. In: Ifs: conditionals, belief, decision, chance, and time, pp. 57–86. Dordrecht: Reidel 1981

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

Makinson, D., Gärdenfors, P.: Relations between the logic of theory change and nonmonotonic logic. In: The logic of theory change, pp. 185–205. Berlin Heidelberg New York: Springer 1991

Martin, E., Osherson, D.: Elements of scientific discovery. Cambridge MA: MIT Press 1998

Meyer, T.: Basic infobase change. In: Foo, N. (ed.) Advanced topics in artificial intelligence, pp. 156–167. Berlin Heidelberg New York: Springer 1999

Nayak, A.C.: Iterated belief change based on epistemic entrenchment. Erkenntnis **41**, 353–390 (1994)

Nebel, B.: A knowledge level analysis of belief revision. In: Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning, pp. 301–311. Los Altos: Morgan Kaufmann 1989

Nebel, B.: Base revision operations and schemes: representation, semantics and complexity. In: Proceedings of the 11th European Conference on Artificial Intelligence, pp. 341–345. Berlin Heidelberg New York: Springer 1994

Osborne, M.J., Rubinstein, A.: A course in game theory. Cambridge, MA: MIT Press 1994

Rott, H.: Logic and choice. In: Tark 98: Proceedings of the Seventh Conference on Theoretical Aspects of Rationality and Knowledge, pp. 235–248, San Francisco: Kaufmann 1998

Rott, H.: Two dogmas of belief revision. Journal of Philosophy **97**(9), 503–522 (2000)

Savage, L.: The foundations of statistics. New York: Dover 1954

Schulte, O.: Minimal belief change and the pareto principle. Synthese **118**, 329–361 (1999)

Schulte, O.: Review of martin and osherson's 'elements of scientific inquiry'. The British Journal for the Philosophy of Science **51**, 347–352 (2000)

Spohn, W.: Ordinal conditional functions: A dynamic theory of epistemic states. In: Causation, decision, belief change, and statistics, Vol.2, pp. 105–134. Dordrecht: Reidel 1987

Stalnaker, R.: A theory of conditionals. In: Ifs: conditionals, belief, decision, chance, and time. Dordrecht: Reidel 1981

Stalnaker, R.: Knowledge, belief and counterfactual reasoning in games. Economics and Philosophy **12**, 133–163 (1996)

van Fraassen, B.: Representation of conditional probability. Journal of Philosophical Logic **5**, 417–430 (1976)