Conservative Belief Change

James P. Delgrande
School of Computing Science
Simon Fraser University
Burnaby, B.C.
Canada V5A 1S6
jim@cs.sfu.ca

Abhaya C. Nayak
Department of Computing
Macquarie University
NSW, 2109
Australia
abhaya@ics.mq.edu.au

Maurice Pagnucco
School of CSE and National ICT Aust.
The University of New South Wales
Sydney, NSW, 2052
Australia
morri@cse.unsw.edu.au

Abstract

A standard assumption underlying traditional accounts of belief change is the *principle of minimal change*, that an agent's belief state should be modified minimally to incorporate new information. In this paper we introduce a novel account of belief change in which the agent's belief state is modified minimally to incorporate *exactly* the new information. Thus a revision by $p \vee q$ will result in a new belief state in which $p \vee q$ is believed, but a stronger proposition (such as $p \wedge q$) is not, regardless of the initial form of the belief state. This form of belief change is termed *conservative belief change* and corresponds to a Gricean interpretation of the input formula. We investigate belief revision in this framework, and provide a representation result between a set of postulates characterising this form of belief change and a construction in terms of systems of spheres. This approach is extended to that of belief revision with respect to a specified context. Last, we show how this approach resolves a longstanding problem in belief revision.

1 Introduction

An agent interacting with an external environment will need to maintain its stock of beliefs in the face of new information. Such belief change is not arbitrary, but rather is usually taken to be guided by various *rationality criteria*. One of the most widely accepted rationality criteria is the *principle of minimal change*: that a belief state is modified minimally to incorporate new information. This principle has many guises (Makinson 1993; Rott 2000). Perhaps the most evident way in which a change in belief can be said to be minimal is in terms of standard constructions such as systems of spheres (Grove 1988) or epistemic entrenchment (Gärdenfors & Makinson 1988), i.e., orderings of logical interpretations or orderings of sentences.

In this paper we introduce an account of belief change in which "minimal change" is taken with respect to the new information. We examine an account of belief change in which *all* we wish to accept is the new information itself—no more, no less. This is reminiscent of the Gricean principle of *Conversational Implicature*, in particular, *The Maxim of Quantity*, that in interpreting a speaker we should assume that the speaker means no more, and no less, than what she says. Our approach ensures that, in a sense to be specified, *exactly* the sentence accepted as evidence is incorporated. Thus for example if an agent believed that $p \wedge q$ was true and was subsequently informed that $p \vee q$ was the case, then in the resulting belief state the agent would believe only $p \vee q$. This is in contrast with the standard account, in which revision by an implied sentence has no effect; here this has the effect of weakening the belief state. It proves to be the case that a modified knowledge base in this new approach is a *conservative extension* (see Section 3) of the sentence for belief change; consequently we term this *conservative* belief change. We provide a characterisation of this form of belief change in terms of a set of postulates and in terms of a construction involving systems of spheres (i.e., orderings over logical interpretations); a representation result establishes a correspondence between these characterisations. This approach is extended to a context-dependent approach to belief change. A revision of a belief set is now with respect to a context (a set of atoms) and a formula; intuitively the idea is that for a revision, the given formula is *exactly* what is known about the provided context.

The paper is organised as follows. In the next section we give some motivating examples that highlight particular aspects of our proposal; as well we examine related work. Section 3 provides the necessary background material. In Section 4 we outline our proposed method of belief change. Section 5 discusses the significance of these results and Section 6 presents our conclusions. Proofs of theorems are given in an extended version of this paper, as is the treatment of belief update in this approach.

2 Motivation and Examples

The following example illustrates the traditional account of integrating new information in accord with the principle of minimal change. We write $K * \alpha$ to denote the belief set resulting from the revision of K by the sentence α.

Example 1 (Exclusive disjunctive revision) *Leslie and Robin are two students who share a flat above yours. They are independent and have their own circles of friends. One evening, you believe that both are out of town,* $K \equiv \neg l \wedge \neg r$. *However, you hear unmistakable sounds of domestic activity. You modify your beliefs minimally to account for this new information, and so you conclude just that one of them* *has gone out, i.e.* $K * (l \vee r) \equiv (l \leftrightarrow \neg r)$.[1]

To be sure, this result is not dictated by the standard revision postulates (see Section 3). However, if the assumption that Leslie and Robin are independent and have their own circles of friends is taken seriously, and coded into the system of spheres accompanying the belief set K, then this result becomes the most plausible minimal change.[2] This phenomenon recurs in the standard distance-based approaches to update of (Winslett 1990; Forbus 1989), as well as in the belief revision counterparts. The next example illustrates that these results aren't always desirable.

Example 2 (Inclusive disjunctive revision) [3] *There are two rooms in a warehouse, one on the left and another on the right. Let* l *and* r *denote the fact that the respective rooms are not empty. You believe that there are a number of boxes (that can fit together in a single room) outside the warehouse and the rooms are empty, and so* $K \equiv \neg l \wedge \neg r$. *You are later informed that it had been raining, and the boxes had been moved inside. You conclude just that the rooms are not empty, i.e.* $K * (l \vee r) \equiv (l \vee r)$.

On the face of it, this example violates the principle of minimal change. As well it conflicts with the aforecited distance-based approaches, which dictate that the result be $l \leftrightarrow \neg r$, that all the boxes are in one room or the other. The next example is expressed in first-order terms, but has an obvious propositional encoding over a finite domain.

Example 3 (Generalised inclusive disjunctive revision) *A robbery has taken place; with no other information, we have* $K \equiv \exists x R(x)$, *that someone is a robber. It is subsequently learned that there were exactly three people* A, B, *and* C *present at the time of the robbery, that is* $\phi = (R(A) \vee R(B) \vee R(C))$. *We conclude that* $K * \phi \equiv (R(A) \vee R(B) \vee R(C))$ *– i.e. the robber(s) constitutes a (nonempty) subset of* $\{A, B, C\}$. *However standard accounts of minimization (e.g., Dalal's revision operator) stipulate that* $K * \phi$ *entail that* R *is true of exactly one of* $\{A, B, C\}$.

Example 1 comprises the standard interpretation of belief revision: the belief set is modified in a minimal fashion so as to incorporate a formula consistently. Examples 2 and 3 comprise a distinct sense for belief change, in which for a revision by a formula ϕ, *exactly* ϕ is to be incorporated into the belief set. Consider $K * (p \vee q)$. If the idea is that all we know about p and q is that $p \vee q$ is true, then we would want the possible combinations of truth values $\{p, q\}$, $\{\neg p, q\}$, and $\{p, \neg q\}$ to be considered possible, and so be consistent with $K * (p \vee q)$. In this example, the atoms p and q supply an implicit context. We can extend this notion by giving an explicit context, where a context (or, subject matter) is comprised of some set of atoms. Thus a revision of belief set K by p in the context $\{p, q\}$, say $K * (\{p, q\}, p)$, would be intended to convey that, of the context $\{p, q\}$, precisely p will be known to be true; in particular, neither q nor $\neg q$ will be believed in the resulting belief set.

This sense is reminiscent of Gricean conversational implicature (Grice 1989) wherein a speaker is required to be maximally informative. Thus if a listener is told that $p \vee q$ is true, then the communicator does not know which of p, q are true; if they did, they would have conveyed the stronger information to the listener. A similar notion has been studied by Lakemeyer and Levesque (see (Lakemeyer & Levesque 2000)) dealing with "only-knowing" or "only-knowing about". These concepts arise in autoepistemic default reasoning where one may want to assert that all an agent knows is ϕ or all that an agent knows about α is ϕ.

An example similar to Example 2 has been addressed in (Herzig & Rifi 1999) in what is there called "the problem of disjunctive input." Their diagnosis is that the problem arises from how disjunction is interpreted. Here, in contrast, we argue that a deeper issue is manifested in these examples involving disjunction and, further, that there is a second, distinct way in which belief change can be interpreted, in which the input formula for belief change expresses all that is known concerning the propositions expressible in the language of this formula. Technically in our approach this will amount to the result of a belief change being a *conservative extension* (Section 3) of the formula to be incorporated into the belief set. As we discuss later, this division is essentially the dual of the revision/update distinction for types of belief change. As well, we show that this distinction provides a resolution to a longstanding problem concerning the recovery postulate in belief revision.

3 Background

The underlying logic will be classical propositional logic. We consider a finitary propositional language $\mathcal{L}$, over a set of atoms, or propositional letters, $\mathbf{P} = \{a, b, c, \ldots\}$, and truth-functional connectives $\neg$, $\wedge$, $\vee$, $\supset$, and $\leftrightarrow$. $\mathcal{L}$ also includes the truth-functional constants $\top$ and $\bot$. To clarify the presentation we shall use the following notational conventions. Upper-case Roman characters (A, $B, \ldots$) denote *consistent conjunctions of literals* from $\mathcal{L}$. Lower-case Greek characters (ϕ, ψ, ξ, $\ldots$) denote arbitrary sentences of $\mathcal{L}$.

An *interpretation* of $\mathcal{L}$ is a function from $\mathbf{P}$ to $\{T, F\}$; M is the set of interpretations of $\mathcal{L}$. A *model* of a sentence ϕ is an interpretation that makes ϕ true, according to the usual definition of truth. A model can be equated with its defining set of literals. $|\phi|_{\mathcal{L}}$ denotes the set of models of sentence ϕ over language $\mathcal{L}$. For interpretation ω we write $\omega \models \phi$ for ϕ is true in ω. For $\phi \in \mathcal{L}$, we will define $\mathcal{L}(\phi)$, the language in which ϕ is expressed, as comprising the minimum set of atoms required to express ϕ, as follows, where ϕ^p_q is

[1] We use $\leftrightarrow$ for *material biconditional* and $\equiv$ for *logical equivalence*.

[2] If we assumed to the contrary that Leslie and Robin always do things together, and reflect this information in the system of spheres, then minimal change of beliefs leads you to believe that neither of them has gone out, i.e., $K*(l \vee r) \equiv (l \wedge r)$, as expected. However both in this example as in the next, we take $l \wedge r$ to be less plausible than either of $l \wedge \neg r$ and $\neg l \wedge r$ in order to examine the problem from the same footing.

[3] It is contentious whether this example illustrates update or revision. We take it to be revision since information about the current state of the world is learned.

the result of substituting atom q everywhere for p in ϕ:

$$\mathcal{L}(\phi) = \{p \in \mathbf{P} \mid \phi^p_\top \not\equiv \phi^p_\bot\} \cup \{\top, \bot\}$$

Thus $\mathcal{L}(p \wedge (q \vee \neg q)) = \mathcal{L}(p) = \{p\}$. This can be extended to sets of sentences in the obvious way. It follows trivially that if $\models \phi \leftrightarrow \psi$ then $\mathcal{L}(\phi) = \mathcal{L}(\psi)$.

We will make use of the notion of a *conservative extension* of one set of sentences by another.

Definition 1 *For* $\Gamma_1 \subseteq \Gamma_2 \subseteq \mathcal{L}$, Γ_2 *is a* conservative extension *of* Γ_1 *iff for every* $\phi \in \mathcal{L}(\Gamma_1)$*, if* $\Gamma_2 \models \phi$ *then* $\Gamma_1 \models \phi$.

Intuitively Γ_2 is a conservative extension of Γ_1 iff Γ_2 extends Γ_1 but tells us nothing more about sentences that are in the language of Γ_1. Γ_2 may entail sentences in its extended language of course but as far as the language which it shares with Γ_1 is concerned, it says no more than Γ_1.

3.1 Belief Revision

A common approach in addressing belief revision has been to provide a set of *rationality postulates* for belief change functions. The *AGM approach* (see (Gärdenfors 1988)) provides the best-known set of such postulates. Belief states are modelled by sets of sentences, called *belief sets*, closed under the logical consequence operator of a logic that includes classical propositional logic. Thus a belief set K satisfies the constraint: $\phi \in K$ if and only if K logically entails ϕ. K can be seen as a partial theory of the world. For belief set K and formula ϕ, $K + \phi$ is the deductive closure of $K \cup \{\phi\}$, the *expansion* of K by ϕ. $K_\bot$ is the inconsistent belief set (i.e. $K_\bot = \mathcal{L}$).

Revision represents the situation in which the new information may be inconsistent with the reasoner's beliefs K and needs to be incorporated in a consistent manner where possible. A revision function $*$ is a function from $2^{\mathcal{L}} \times \mathcal{L}$ to $2^{\mathcal{L}}$ satisfying the following postulates.

$(K * 1)$ $K * \phi$ is a belief set.

$(K * 2)$ $\phi \in K * \phi$.

$(K * 3)$ $K * \phi \subseteq K + \phi$.

$(K * 4)$ If $\neg\phi \notin K$, then $K + \phi \subseteq K * \phi$.

$(K * 5)$ $K * \phi = K_\bot$ iff $\models \neg\phi$.

$(K * 6)$ If $\models \phi \leftrightarrow \psi$, then $K * \phi = K * \psi$.

$(K * 7)$ $K * (\phi \wedge \psi) \subseteq (K * \phi) + \psi$.

$(K * 8)$ If $\neg\psi \notin K * \phi$, then $(K * \phi) + \psi \subseteq K * (\phi \wedge \psi)$.

Motivation for these postulates can be found in (Gärdenfors 1988). A dual operator, called contraction, is similarly defined, so that for a contraction of ϕ from K, denoted $K \dot{-} \phi$, the result is a belief set in which ϕ is not believed. See (Gärdenfors 1988) for the set of contraction postulates.

Various constructions have been proposed to characterise belief revision. We will make reference to Grove's use of a system of spheres (SOS) model for characterizing AGM revision (Grove 1988). A *system of spheres centred on* X is a total preorder on the set of interpretations (or: *possible worlds*), $\leq_{SOS}$, in $\mathcal{L}$ such that for $\omega \in M$ we have that: $\omega \in X$ iff $\omega \leq \omega'$ for all $\omega' \in M$. (That is, X is the least set of worlds in the preorder.) We will often omit the subscript from $\leq_{SOS}$ for readability. Revision is defined for $|K|_{\mathcal{L}} = X$ by

$$|K * \phi|_{\mathcal{L}} = \min_{\leq_{SOS}} \{\omega \in M \mid \omega \models \phi\} \qquad (1)$$

where $\min\{\}$ denotes the minimal models under $\leq$. Grove shows that for every belief revision operator satisfying the AGM postulates there is a system of spheres characterising that operator, and vice versa.

4 Approach

4.1 Conservative Belief Revision

We use $\hat{*}$ to denote the type of belief revision described in Section 2, called "conservative belief revision" or "C-revision." The idea we wish to capture is that, for $K \,\hat{*}\, \phi$, ϕ *is exactly what will be believed in the resulting belief set*, relative to the "subject matter" ϕ. So for $K \,\hat{*}\, ((p \vee q) \wedge r)$ the idea is that $(p \vee q) \wedge r$ constrains the truth values of atoms in $\{p, q, r\}$, and that exactly $(p \vee q) \wedge r$ will be known about these atoms in the resulting belief set. In particular, strengthenings of $p \vee q$, such as p or $p \leftrightarrow \neg q$ will not be true in the resulting belief set. This will be the case even when K implies p or $p \leftrightarrow \neg q$; hence revision may in fact yield a weakening of the belief set. This restriction does not hold for sentences not in $\mathcal{L}(\phi)$. The assumption is that the new information is more reliable than that contained in the current belief set. Further, the new information completely overrides what was previously believed.

The semantic intuition behind our proposal is easily visualised. In Figure 1 we consider a revision where the underlying language is generated from atoms x, y and z. The agent believes $x \wedge \neg y \wedge z$ and encounters evidence $\neg x \vee \neg y$. Accordingly the interpretations are partitioned into four cells corresponding to the interpretations over the language $\mathcal{L}(\neg x \vee \neg y)$. The best worlds from each of the three cells satisfying $\neg x \vee \neg y$ are chosen to represent the revised belief set. Clearly, the belief content of the new belief set modulo $\mathcal{L}(\neg x \vee \neg y)$ will be exactly $\neg x \vee \neg y$. Beliefs regarding z will depend on extralogical factors, namely the plausibility of different worlds.

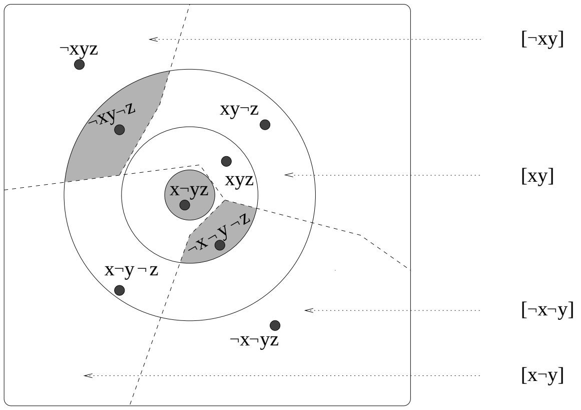

Figure 1: Conservative Revision – Semantics

Now, in determining C-revision, we consider the plausibility of different worlds represented in Figure 1 by the concentric "rings" in the system of spheres model. The worlds that are more centrally located are more plausible. Accordingly, from the $|x\neg y|$ cell, the world $x\neg yz$ is selected, whereas worlds $\neg xy\neg z$ and $\neg x\neg y\neg z$ are selected from the cells $|\neg xy|$ and $|\neg x\neg y|$ respectively. Since some of these selected worlds satisfy z and some $\neg z$, under this plausibility ordering belief z is lost. In fact, new beliefs regarding z are captured by the beliefs $x \leftrightarrow z$ and $y \vee z$. We can formalize this analogously to Grove's system of spheres model for characterizing AGM revision. Revision was defined in (1). We have an analogous definition for C-revision.[4]

$$|K \hat{*} \phi|_{\mathcal{L}} = \bigcup_{\sigma \in |\phi|_{\mathcal{L}(\phi)}} \min_{\leq_{SOS}} \{\omega \in M \mid \omega \models \sigma\}. \qquad (2)$$

A key result is captured by the following theorem:

Theorem 1 *For any belief set K and input sentence ϕ, $K \hat{*} \phi$ is a* conservative extension *of ϕ, i.e., for $\psi \in \mathcal{L}(\phi)$, if $K \hat{*} \phi \models \psi$ then $\phi \models \psi$.*

We also obtain the following results relating this approach to AGM revision.

Theorem 2 *Let $\hat{*}$ be obtained from a systems of spheres $\leq_{SOS}$ and let $*$ be the AGM revision obtained from $\leq_{SOS}$.*

1. *$K \hat{*} \phi \subseteq K * \phi$.*
2. *$K \hat{*} A = K * A$.*[5]

This raises the question of whether a specific C-revision function can be captured using the standard definition of revision (1) in a suitably-constructed system of spheres. In general the answer is negative; for a counterexample, consider where $\mathcal{L} = \{p, q\}$ and we are given a C-revision function such that $K \equiv \neg p \wedge \neg q$ and in which $K \hat{*} p = K \hat{*} (p \wedge q)$. This entails the constraints on the ordering: $\{\neg p, \neg q\} < \{p, q\}, < \{p, \neg q\}$. However, as is easily verified, $K \hat{*} (p \vee q) \equiv p \vee q$. This cannot be obtained by standard revision given the above constraints on the ordering, since it requires $\{p, q\}$, $\{\neg p, q\}$ and $\{p, \neg q\}$ at the same level.

While a given system of spheres determines a unique C-revision (as constructed by (2)), the converse in general does not hold. The following example demonstrates this point.

Example 4 *Consider two SOS's: SOS$_1$: $\ldots < xyz < x\neg y\neg z$ and SOS$_2$: $\ldots < x\neg y\neg z < xyz$, where the $\ldots$ in the orderings represent an identical subsequence. The C-revision based on these SOS's (using (2)) exhibit identical behaviour since no cell of any partition based on a sub-language of $\{x, y, z\}$ will pick up exactly the set $\{xyz, x\neg y\neg z\}$.*

Thus we notice an asymmetry between the classical AGM account of belief revision and C-revision. An AGM revision operation $*$, given a fixed belief set K, determines a unique system of spheres. On the other hand, the C-revision operation, given a fixed belief set K, corresponds to a class of systems of spheres. It is of interest to characterise the class of systems of spheres that a given C-revision operation $\hat{*}$ determines. We have such a characterisation:

Definition 2 *Two systems of spheres, $\leq_1$ and $\leq_2$ are $\hat{*}$-equivalent iff for every sentence $\phi \in \mathcal{L}$, $K \hat{*}_{\leq_1} \phi = K \hat{*}_{\leq_2} \phi$, where $|K|$ is the set of $\leq_{\{1,2\}}$-minimal*[6] *worlds and $\hat{*}_{\leq_1}$ and $\hat{*}_{\leq_2}$ are defined from $\leq_1$ and $\leq_2$ using (2).*

Our goal is to characterise in formal terms the set of SOS's that are $\hat{*}$-equivalent to a given SOS. Toward this end, we offer the following construction:

Definition 3 *Let $\leq$ be a given SOS. We say an SOS $\leq'$ is a C-transform of $\leq$ iff the former can be constructed from the latter in the following manner: (1) Consider any two worlds ω and ω'. If there is a consistent set S of literals over $\mathcal{L}$ such that both $\omega \models \bigwedge(S)$ and $\omega' \models \bigwedge(S)$, and ω is $\leq$-minimal among all worlds satisfying $\bigwedge(S)$, then $\omega \leq \omega'$ iff $\omega \leq' \omega'$ (note that since $\bigwedge(\omega) \equiv \omega$ we obtain a reflexive $\leq'$); and (2) After obtaining all those constraints on $\leq'$, we complete it as we wish to get a total preorder $\leq'$.*

It is easily verified that C-transformation is a symmetric relation, i.e., if $\leq'$ is a C-transform of $\leq$, then $\leq$ is a C-transform of $\leq'$. The following is a simple example illustrating this construction:

Example 5 *Assume a language based on atoms $\{p, q, r\}$. Let $\leq$ be: $\{\neg p\neg q\neg r, \neg p\neg qr\} < \{\neg pq\neg r, \neg pqr\} < \{p\neg q\neg r\} < \{p\neg qr\} < \{pq\neg r, pqr\}$. If we compare worlds $pq\neg r$ and $p\neg qr$, the only relevant conjuncts are p and $\top$. Since neither of these worlds are $\leq$-minimal either in $|\top|$ (all worlds) or $|p|$ (worlds satisfying p), no particular constraint on $\leq'$ is generated by this comparison. On the other hand, if we compare $pq\neg r$ and $\neg pqr$, we notice that the relevant conjuncts are q and $\top$. Since, among worlds satisfying q, we have $\neg pqr$ as one of the $\leq$-minimal elements, and also $\neg pqr < pq\neg r$ it follows that $\neg pqr <' pq\neg r$*

The reader is invited to verify that, given this definition, the two SOS's, SOS$_1$ and SOS$_2$ used in Example 4 are actually C-transforms of each other. This suggests that there is a close connection between the two concepts, C-transformation and $\hat{*}$-equivalence. In fact the following theorem shows that C-transformation actually captures the set of total preorders that are $\hat{*}$-equivalent to each other.

Theorem 3 *Two preorders $\leq$ and $\leq'$ are C-transforms of each other iff they are $\hat{*}$-equivalent.*

We consider next the properties that characterise C-revision functions.

4.2 Postulates

Recall that upper-case Roman characters (A, $B, \ldots$) denote *consistent conjunctions of literals* from $\mathcal{L}$, and lower-case Greek characters (ϕ, ψ, ξ, $\ldots$) denote arbitrary sentences of $\mathcal{L}$. A *C-revision* function is a function $\hat{*} : 2^{\mathcal{L}} \times \mathcal{L} \to 2^{\mathcal{L}}$ satisfying the following postulates.

[4] Since ϕ is a formula and by definition finite, a model of ϕ over the language of ϕ will be finite; hence we are justified in conflating the model σ with a formula, viz. a conjunction of literals.

[5] Recall that formulas A, B, $\ldots$, are conjunctions of literals by convention.

[6] The $\leq_1$-minimal worlds and the $\leq_2$-minimal worlds are both equal to $|K|$ otherwise they are not appropriate for revising K.

$(K \hat{*} 1)$ $K \hat{*} \phi$ is a belief set

$(K \hat{*} 2)$ $\phi \in K \hat{*} \phi$

$(K \hat{*} 3)$ $K \hat{*} A \subseteq K + A$

$(K \hat{*} 4)$ If $\neg A \notin K$, then $K + A \subseteq K \hat{*} A$

$(K \hat{*} 5)$ $K \hat{*} \phi = K_\perp$ iff $\models \neg\phi$.

$(K \hat{*} 6)$ If $\models \phi \leftrightarrow \psi$, then $K \hat{*} \phi = K \hat{*} \psi$

$(K \hat{*} 7)$ $K \hat{*} (A \wedge B) \subseteq (K \hat{*} A) + B$

$(K \hat{*} 8)$ If $\neg B \notin K \hat{*} A$, then $(K \hat{*} A) + B \subseteq K \hat{*} (A \wedge B)$.

$(K \hat{*} 9)$ If $\phi \not\models \perp$, then there is an $A \not\models \perp$ such that, $A \models \phi$ and for all B, $\neg\phi \notin K \hat{*} B$ implies $A \wedge B \models K \hat{*} B$

$(K \hat{*} 10)$ If $A \models \phi$ and $\mathcal{L}(A) \subseteq \mathcal{L}(\phi)$ then $K \hat{*} \phi \subseteq K \hat{*} A$.

$(K \hat{*} 11)$ If $\neg A \notin K \hat{*} \phi$, then there is a C such that $C \models \phi$ and $\neg A \notin K \hat{*} C$, and $\mathcal{L}(C) \subseteq \mathcal{L}(\phi)$.

Postulates $(K \hat{*} 1)$ – $(K \hat{*} 8)$ are the standard AGM postulates with the proviso that postulates $(K \hat{*} 3)$, $(K \hat{*} 4)$, $(K \hat{*} 7)$ and $(K \hat{*} 8)$ are restricted to consistent conjunctions of literals. Postulate $(K \hat{*} 9)$ compensates for the weakening of these postulates. It states that for every consistent formula ϕ there is a stronger conjunction of literals A capable of accounting for any other C-revision, consistent with ϕ, by a conjunction of literals.[7] It is possible to show that the general, AGM, versions of postulates $(K * 3)$ and $(K * 7)$ follow from these postulates. Postulate $(K \hat{*} 10)$ says that if $A \models \phi$, then the only reason for $K \hat{*} \phi$ to not be included in $K \hat{*} A$ is because the language of A is outside the minimum language of ϕ. Postulate $(K \hat{*} 11)$ essentially states the converse: the only reason that a conjunction of literals A is in $K \hat{*} \phi$ is that there is some conjunction of literals C (in fact, a prime implicant of ϕ) such that A is in $K \hat{*} C$. It is possible in fact to rephrase $(K \hat{*} 10)$ and $(K \hat{*} 11)$ in terms of prime implicants:

$(K \hat{*} 10')$ If $A \models \phi$ and $K \hat{*} \phi \not\subseteq K \hat{*} A$, then there is a literal L such that $A \models L$, and L is neither entailed nor contradicted by any prime implicant of ϕ.

$(K \hat{*} 11')$ If $\neg A \notin K \hat{*} \phi$, then $\exists C$ such that $C \models \phi$ and $\neg A \notin K \hat{*} C$, and for all literals L, if $C \models L$, then L is either entailed or contradicted by a prime implicant of ϕ.

The following postulate is also of interest.

$(K \hat{*} 12)$ $K \hat{*} A$ is the largest theory satisfying postulates $(K \hat{*} 1)$ – $(K \hat{*} 10)$.

It is possible to show that this postulate is equivalent to $(K \hat{*} 11)$ in the presence of postulates $(K \hat{*} 1)$ – $(K \hat{*} 10)$.

Proposition 1 *Postulates $(K \hat{*} 11)$ and $(K \hat{*} 12)$ are equivalent given $(K \hat{*} 1)$ – $(K \hat{*} 10)$.*

For a representation result, the soundness of $(K \hat{*} 1)$ – $(K \hat{*} 11)$ is relatively straightforward:

Theorem 4 *Let K be a theory and $\leq_{SOS}$ a system of spheres centred on $|K|_{\mathcal{L}}$. The function $\hat{*}$ induced from $\leq_{SOS}$ via (2) satisfies $(K \hat{*} 1)$ – $(K \hat{*} 11)$.*

For the completeness of the axioms, we proceed in two steps: (i) we consider the special case of consistent conjunctions of literals, for which C-revision reduces to classical AGM revision, and prove the completeness of $(K \hat{*} 1)$ – $(K \hat{*} 9)$, and (ii) extend the results to arbitrary sentences and include $(K \hat{*} 10)$ and $(K \hat{*} 11)$. More precisely, we have:

Theorem 5

1. *Let K be a theory and $\hat{*}$ a revision function satisfying $(K \hat{*} 1)$ – $(K \hat{*} 9)$. There exists a system of spheres $\leq_{SOS}$ centred on $|K|_{\mathcal{L}}$, such that for any consistent conjunction of literals A, $|K \hat{*} A|_{\mathcal{L}} = \min_{\leq_{SOS}}\{\omega \in M \mid \omega \models A\}$*
2. *Let K be a theory and $\hat{*}$ a revision function satisfying $(K \hat{*} 1)$ – $(K \hat{*} 11)$. There exists a system of spheres $\leq_{SOS}$ centred on $|K|_{\mathcal{L}}$, such that $\hat{*}$ is identical to the C-revision function induced from $\leq_{SOS}$.*

4.3 Context-Dependent Revision

We defined $\mathcal{L}(\phi)$ as comprising the least set of atoms *required* to express ϕ. Thus $\mathcal{L}(\phi)$ provides an implicit context for revision, in that following the C-revision $K \hat{*} \phi$, all conjunctions of literals corresponding to models of ϕ over $\mathcal{L}(\phi)$ are satisfied by the resulting belief set. This can be generalised to revision of the form $K \hat{*} (C, \phi)$,[8] such that $\mathcal{L}(\phi) \subseteq C \subseteq \mathbf{P}$, and the intuition is that K is to be revised so that all that is known concerning the context C is ϕ.

As an example of context sensitive revision, consider the revision of a belief set $\{a \wedge b \wedge c \wedge d\}$ by $a \vee b$. Let a, b and c stand respectively for Albert, Becky and Charles being involved in a bank robbery, and d stand for Doug being a geologist. The input $a \vee b$ suggests that the context is the relevant bank robbery, represented by $\{a, b, c\}$. We would expect that after the revision, we would no longer suspect Charles of robbery, and whether or not we would still believe that Doug is a geologist would depend on extraneous factors. Otherwise, if we let c stand for Charles being a nice dad (and a, b and d as before) we would expect that the context is simply $\{a, b\}$, and as a result of the revision, whether or not c would be maintained will depend on extraneous factors.

We can formalize this analogously to C-revision, as in (2) and again using Grove's system of spheres model. For $K \subseteq \mathcal{L}$, $\phi \in \mathcal{L}$, and $\mathcal{L}(\phi) \subseteq C \subseteq \mathbf{P}$ define:

$$|K \hat{*} (C, \phi)|_{\mathcal{L}} = \bigcup_{\sigma \in |\phi|_C} \min_{\leq_{SOS}} \{\omega \in M \mid \omega \models \sigma\}. \qquad (3)$$

We obtain for any belief set K, context C, and input sentence ϕ, that $K \hat{*} (C, \phi)$ is a *conservative extension* of C in which ϕ is true. As well, we obtain the following results:

Theorem 6 *Let $\hat{*}$ and $*$ (representing AGM revision) be obtained from a system of spheres $\leq_{SOS}$.*

1. *$K \hat{*} (C, \phi) \subseteq K * \phi$.*
2. *If $\mathcal{L}(A) = C$ then $K \hat{*} A = K * A$.*
3. *$K \hat{*} (\mathbf{P}, \phi) \equiv \phi$.*

[7] This postulate can be better motivated in terms of models where it guarantees that among each non-empty set of models of ϕ, $|\phi|_{\mathcal{L}}$, there is a non-empty set of "best" models $|A|_{\mathcal{L}}$.

[8] Writing $K \hat{*} (C, \phi)$ as well as $K \hat{*} \phi$ overloads the symbol $\hat{*}$. However the revision operator intended is clear from the use of $\hat{*}$.

The third result is justified by the fact that we work with a finitary language; it states that, as a special case, when the context is the full set of atoms, context-sensitive revision corresponds to full-meet revision.

Clearly C-revision can be defined in terms of context sensitive revision by defining $K \hat{*} \phi$ as $K \hat{*} (\mathcal{L}(\phi), \phi)$. Conversely, context sensitive revision can be defined in terms of C-revision via: $|K \hat{*} (C, \phi)|_{\mathcal{L}} = \bigcup_{A \in |\phi|_C} |K \hat{*} A|_{\mathcal{L}}$.

5 Discussion

Semantically, the distinction between standard AGM revision and C-revision is analogous to the distinction between revision and update, and in fact the two distinctions may be seen as duals of each other. For an (AGM) revision, $K * \phi$, we consider the set of all models of K, and revise by selecting the closest models of ϕ to that set. For an update, $K \diamond \phi$, for each model of K we look for the closest models of ϕ; the update is the union of all such models. Analogous, for a C-revision, $K \hat{*} \phi$, we consider each model of ϕ (over $\mathcal{L}(\phi)$), and revise K by this model; the C-revision is the union of all such models. This duality between C- and standard belief change on the one hand, and between revision and update on the other, completes a classification of belief change operators, in terms of whether the models of a knowledge base or formula for change are considered *en masse*, or individually.

C-belief change is also of independent interest. We have already noted that it conforms to a Gricean interpretation of the sentence for revision. As well it resolves the problem of disjunction noted by (Herzig & Rifi 1999), but in a general and syntax-independent setting. To conclude, we discuss two further ways in which C-revision contributes to the overall theory of belief change.

First, the approach provides a resolution to a recalcitrant problem concerning the *recovery postulate* of belief contraction: $K \subseteq (K \dot{-} \phi) + \phi$. Thus if one contracts a belief set by a sentence, and then adds that sentence, no information is lost with respect to the original belief set. (Hansson 1999) gives the following counterintuitive example (paraphrased): *Let K entail that "Cleopatra had a son and a daughter" ($s \wedge d$). New information is received that Cleopatra didn't have a child, expressed by $K \dot{-} (s \vee d)$. Then one learns that she had a child, thus $(K \dot{-} (s \vee d)) + (s \vee d)$. Recovery says that $s \wedge d$ is believed – that Cleopatra had a son and daughter. Intuitively, just $s \vee d$ should be believed, that all that is known after these changes is that she had a child.* We note that if $K \models \phi$ then recovery can be written as $K \subseteq (K \dot{-} \phi) * \phi$. Arguably this example is best interpreted as involving conservative revision – that is, concerning $\{s, d\}$, one learns *at most* that Cleopatra has a child $s \vee d$. Under this reading of revision we have possibly $K \not\subseteq (K \dot{-} \phi) \hat{*} \phi$. In our example, if $K \equiv (s \wedge d)$, then we would expect: $K \dot{-} (s \vee d) \equiv \neg s \wedge \neg d; \quad (K \dot{-} (s \vee d)) \hat{*} (s \vee d) \equiv (s \vee d)$.

Second, and more speculatively, belief contraction has been criticised for removing too little information from a knowledge base; at the other extreme, *severe* contraction (Rott & Pagnucco 1999) has been criticised as removing too much information. (Hansson 1999) proposes that a realistic contraction operator should lie between AGM-style contraction operators and severe contraction operators. As suggested earlier, a corresponding C-contraction operator $\hat{-}$ is easily defined using the Harper Identity or by direct definition. C-contraction is easily shown to lie between AGM and severe contraction. Given that it comes with (arguably) compelling intuitions and a straightforward semantics, C-contraction can be proposed as a "reasonable" intermediate.

6 Conclusion

We have discussed a theory of conservative belief change. The main intuitive motivation for this work stems from an attempt to make the most of the information presented by new evidence that a reasoner acquires. As such, our approach focuses on the content of the new evidence. Our current analysis suggests that the operators we have introduced based on these intuitions possesses some interesting and appealing properties; as well it resolves a problem with the recovery postulate, and may provide a satisfactory contraction operator. Last, the distinction between traditional belief change and C-belief change appears to be a dual to that between revision and update.

Acknowledgements

The authors would especially like to thank Pavlos Peppas for his invaluable help with this paper.

References

Forbus, K. 1989. Introducing actions into qualitative simulation. In *International Joint Conference on Artificial Intelligence*, 1273–1278.

Gärdenfors, P., and Makinson, D. 1988. Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of the Second Conference on Theoretical Aspect of Reasoning About Knowledge*, 83–96.

Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Bradford Books, MIT Press, Cambridge Massachusetts.

Grice, H. P. 1989. *Studies in the Way of Words*. Harvard University Press, Cambridge Massachusetts.

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

Hansson, S. O. 1999. A textbook of belief dynamics: Theory change and database updating. Kluwer Academic Publishers.

Herzig, A., and Rifi, O. 1999. Propositional belief base update and minimal change. *Artificial Intelligence* 115:107–138.

Lakemeyer, G., and Levesque, H. 2000. *The Logic of Knowledge Bases*. Cambridge, Mass.: MIT Press.

Makinson, D. 1993. Five faces of minimality. *Studia Logica* 52:339–379.

Rott, H., and Pagnucco, M. 1999. Severe withdrawal (and recovery). *Journal of Philosophical Logic* 28(5).

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

Winslett, M. 1990. *Updating Logical Databases*. Cambridge University Press.