Conservative Belief Change: A Gricean Approach

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
The University of New South Wales
NSW, 2052, AUSTRALIA
morri@cse.unsw.edu.au

Abstract

A standard intuition underlying traditional accounts of belief change is the *principle of minimal change*. Briefly put, it states 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 both belief revision and belief update in this framework and discuss both syntax independent as well as syntax sensitive methods. We show how this approach solves a recalcitrant problem in belief revision.

A reasoning entity will need to maintain its stock of beliefs in the face of new information. Such belief change is not arbitrary; rather belief change is generally taken to be guided by various *rationality criteria*. One of the most widely advocated rationality criterion is the *principle of minimal change*: that a belief state is modified minimally to incorporate new information. This principle has many guises [16; 20]. 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 [15; 7] or epistemic entrenchment [4], i.e., orderings of possible worlds or orderings of sentences.

We can also distinguish different fundamental forms of belief change. In the literature at least two are ubiquitous: *revision* and *update*. Intuitively, revision deals with the change that occurs when the reasoner's beliefs about its environment are incomplete and possibly incorrect. Any new information fills in these gaps or rectifies mistaken beliefs. However, the environment is assumed static and does not change. Update on the other hand deals with dynamic environments in which the new information reflects changes brought about by actions that have occurred.

In this paper we introduce an account of belief change that is orthogonal to the notions of revision and update, and 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*, that in interpreting a speaker we should assume that the speaker means no more, and no less, than what she says. We, of course, add as *caveat* that the listener takes the speaker to be sincere (not trying to withhold information) and an expert in the domain of discourse. Our approach ensures that, in a sense to be specified, *exactly* the sentence accepted as evidence is incorporated. A modified knowledge base is a *conservative extension* (see Section 2) of the sentence for belief change; consequently we term this *conservative* belief change. We outline both syntax independent and syntax dependent methods for revision and update.

The paper is organised as follows. In the next section we outline some motivating examples that highlight particular aspects of our proposal; as well we examine related work. Section 2 provides the necessary background material. In Section 3 we outline our proposed method of belief change. Section 4 discusses the significance of these results and Section 5 presents our conclusions.

1 Motivation and Examples

The following example illustrates the traditional account of integrating new information, in accord with the principle of minimal change.

Example 1.1 (Exclusive disjunctive update) *Leslie and Robin are two students who share an apartment above your's. While they get along, they are independent and have their own circles of friends. You initially believe that for the upcoming weekend neither will be in the apartment, say* $K \equiv \neg l \wedge \neg r$. *However, come the weekend you hear muted but unmistakable sounds of domestic activity. You modify your beliefs minimally to account for this new information, and consequently you conclude just that one of them has not gone away for the weekend, i.e.* $K \diamond (l \vee r) \equiv (l \leftrightarrow \neg r)$.[1]

*Also affiliated with National ICT Australia and the ARC Centre of Excellence in Autonomous Systems.

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

To be sure, this result is not dictated by the standard update postulates (see Section 2), but it seems to be the most plausible minimal change, given the information available; as well, this phenomenon recurs in the standard distance-based approaches to update of [21; 3], as well as in the belief revision counterparts. The next example illustrates that these results aren't necessarily desirable all the time.

Example 1.2 (Inclusive disjunctive update) [2] *There are two rooms in a warehouse, on the left and on the right. Let l and r denote the fact that the respective rooms are not empty. There are a number of boxes outside the warehouse but the rooms are initially empty, and so $K \equiv \neg l \wedge \neg r$. It subsequently begins to rain, and the boxes are moved inside. One concludes just that the rooms are not empty, i.e. $K \diamond (l \vee r) \equiv (l \vee r)$.*

This example apparently 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 1.3 (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. Later it is 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 would want to conclude that $K * \phi \equiv (R(A) \vee R(B) \vee R(C))$ – that 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 1.1 comprises the standard interpretation of belief change, whether for revision or update: the knowledge base is to be modified in a minimal fashion so that a formula can be consistently incorporated. Examples 1.2 and 1.3 comprise a distinct sense for belief change. For this second sense, the idea is that for a revision or update by a formula ϕ, *exactly* ϕ is to be incorporated into the knowledge base. 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 $K * (p \vee q)$.

This sense is reminiscent of Gricean conversational implicature [6] 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 Levesque, and Lakemeyer and Levesque (see [14]) dealing with "only-knowing" or "only-knowing about". Lakemeyer [13] also deals with "only-believing". 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.

[2] It is contentious whether this example illustrates update or revision. We take it to be update since the domain is dynamic.

An example similar to (1.2) has been addressed in [11] 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 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 2) of the formula to be incorporated in the knowledge base. 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 long-standing problem concerning the recovery postulate in belief revision.

2 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 $\bot$. 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. $Mod_{\mathcal{L}}(\phi)$ denotes the set of models of sentence ϕ over language $\mathcal{L}$; the subscript $\mathcal{L}$ may be dropped if the language is clear. 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 (see [17]). This set of atoms is unique. 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 $\mathcal{L}(\phi) = \mathcal{L}(\psi)$. Also note that if $\models \phi$ then $\mathcal{L}(\phi) = \{\top\}$.

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

Definition 2.1 *For sets of sentences $\Gamma_1 \subseteq \Gamma_2 \subseteq \mathcal{L}$ we have that 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.

2.1 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* [1; 5] provides the best-known set of such postulates. The goal is to describe belief change at the *knowledge level*, that is on an abstract level, independent of how beliefs are represented and manipulated. 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, in a language $\mathcal{L}$. Thus

a belief set is a set K of sentences which 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. $K_\perp$ is the inconsistent belief set (i.e. $K_\perp = \mathcal{L}$). Expansion is intended to be applied when the new information is consistent with the current set of beliefs.

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 $K + \phi \subseteq K * \phi$.

$(K * 5)$ $K * \phi = K_\perp$ 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 [5]. 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 knowledge base in which ϕ is not believed. See [5] 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 [7]. A *system of spheres centred on* X is a total preorder on the set of interpretations, $\leq_{SOS}$, in $\mathcal{L}$ such that for $x \in M$ we have that: $x \in X$ iff $x \leq y$ for all $y \in M$. (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 $Mod_{\mathcal{L}}(K) = X$ by

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

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.

For *revision*, an agent receives new information concerning a static world. In belief *update* [12] an agent receives information about an evolving world. Standards for update are given by the set of *KM update postulates*. An *update* function $\diamond$ is a function from $2^{\mathcal{L}} \times \mathcal{L}$ to $2^{\mathcal{L}}$ satisfying[3].

$(K \diamond 0)$ $K \diamond \phi$ is a belief set.

$(K \diamond 1)$ $\phi \in K \diamond \phi$.

$(K \diamond 2)$ If $\phi \in K$ then $K \diamond \phi = K$.

$(K \diamond 3)$ If $K \not\models \perp$ and $\phi \not\models \perp$ then $K \diamond \phi \neq K_\perp$.

$(K \diamond 4)$ If $\models \phi \leftrightarrow \psi$, then $K \diamond \phi = K \diamond \psi$.

$(K \diamond 5)$ $K \diamond (\phi \wedge \psi) \subseteq (K \diamond \phi) + \psi$.

$(K \diamond 6)$ If $\psi \in K \diamond \phi$ and $\phi \in K \diamond \psi$ then $K \diamond \phi = K \diamond \psi$.

$(K \diamond 7)$ If K is complete then $K \diamond (\phi \vee \psi) \subseteq K \diamond \phi \cup K \diamond \psi$.

$(K \diamond 8)$ $(K_1 \cap K_2) \diamond \phi = (K_1 \diamond \phi) \cap (K_2 \diamond \phi)$.

[3] We give an AGM-style formulation of these postulates (cf. [18]) rather than the formulation by Katsuno and Mendelzon so we can place the two approaches on the same footing.

3 Approach

3.1 Conservative Belief Revision

We use $\hat{*}$ to denote the type of belief revision described in Section 1, 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 knowledge base*, relative to the "subject matter" implicit in 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 knowledge base. In particular, strengthenings of $p \vee q$, such as p or $p \leftrightarrow \neg q$ will not be true in the resulting knowledge base. 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 knowledge base. This restriction does not necessarily hold for sentences not in $\mathcal{L}(\phi)$. The underlying assumption here is that the new information is more reliable than that contained in the current belief set. Moreover, 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 models 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 knowledge base. Clearly, the belief content of the new knowledge base 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.

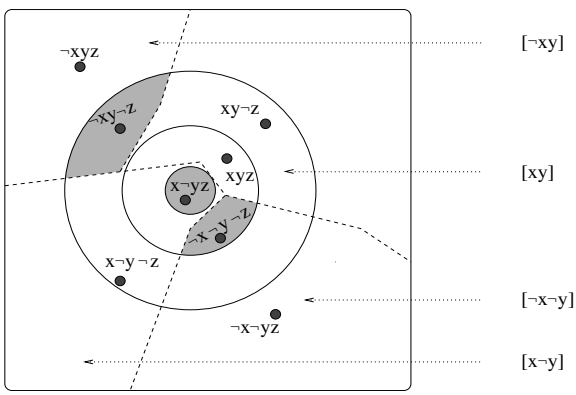

Figure 1: Conservative Revision – Semantics

Now, in determining C-revision, we consider the plausibility of different worlds represented in Figure 1 by the concentric

"rings". 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 can be captured by the beliefs $x \leftrightarrow z$ and $y \vee z$ that are retained from the old knowledge base.

We can formalize this analogously to Grove's system of spheres model for characterizing AGM revision. Revision was defined in (1). We have the analogous definition for C-revision:[4]

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

The overall result is captured by the following theorem:

Theorem 3.1 *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 $\phi \models \psi$.*

An outline of the proof of this theorem follows. Let $\psi \in \mathcal{L}(\phi)$, and $\psi \in K \hat{*} \phi$. We need to show that $\phi \models \psi$. Clearly ψ can be expressed as a disjunction where each disjunct is the conjunction of a maximally consistent subset of literals in $\mathcal{L}(\phi)$. Each of these disjuncts thus corresponds to a model of ϕ over $\mathcal{L}(\phi)$. Models of $K \hat{*} \phi$ can similarly be represented as maximally consistent sets of literals of $\mathcal{L}$. Since $\psi \in K \hat{*} \phi$, it follows that for every model ω of $K \hat{*} \phi$ (over $\mathcal{L}$) there exists a model ω' of ψ (over $\mathcal{L}(\phi)$) such that $\omega' \subseteq \omega$, i.e., every model ω of $K \hat{*} \phi$ is a refinement of some model ω' of ψ. Since $\phi \in K \hat{*} \phi$, we get that every model ω of $K \hat{*} \phi$ is a refinement of some model ω'' of ϕ (over $\mathcal{L}(\phi)$). We need to show that $\phi \models \psi$, i.e., every model of ϕ is also a model of ψ. Pick a model ω_0 of ϕ. From the construction of $K \hat{*} \phi$ it follows that each best/least refinement ω_1 of ω_0 that is an interpretation over $\mathcal{L}$ is a model of $K \hat{*} \phi$. Since $\psi \in K \hat{*} \phi$, it follows that ω_1 is a refinement of some model ω_2 of ψ in $\mathcal{L}(\phi)$. So ω_1 is a refinement of both ω_0 and ω_2, the latter two being interpretations over $\mathcal{L}(\phi)$. The only way that this is possible is if $\omega_0 = \omega_2$, whereby $\omega_0 \models \psi$. But this just says that $K \hat{*} \phi$ is a conservative extension of ϕ. ■

We also obtain the following elementary results.

Theorem 3.2

1. $K \hat{*} \phi \subseteq K * \phi$.
2. *For ϕ a conjunction of literals, $K \hat{*} \phi = K * \phi$.*

This gives rise to the question of whether a specific C-revision function can be captured using the standard definition of revision (1) in some 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 following constraints on the ordering: $\{\neg p, \neg q\} < \{p, q\}, < \{p, \neg q\}$. However, as is easily verified in this case, $K \hat{*} (p \vee q) \equiv p \vee q$. This however 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 3.1 *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 a partial characterisation.

Definition 3.1 *Two systems of spheres, $\leq_1$ and $\leq_2$ are $\hat{*}$-equivalent iff for every sentence $s \in \mathcal{L}$, $K \hat{*}_{\leq_1} s = K \hat{*}_{\leq_2} s$, where $[K]$ is the set of $\leq_{\{1,2\}}$-minimal[5] 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 3.2 *Let $\leq$ be a given SOS. We say an SOS $\leq'$ is a C-transform (conservative 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 3.2 *Assume a language based on atoms $\{p, q, r\}$. Let $\leq$ be: $\{\neg p\neg q\neg r, \neg p\neg qr\} < \{\neg pq\neg r, \neg pqr\} < \{p\neg q\neg r\} < \{p\neg qr\} < \{pq\neg r, pqr\}$. In this example, 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 $\neg pqr <' pq\neg r$*

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

The reader is invited to verify that, given this definition, the two SOS's, SOS_1 and SOS_2 used in Example 3.1 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 3.3 *Two preorders $\leq$ and $\leq'$ are C-transforms of each other iff they are $\hat{*}$-equivalent.*

We consider next those postulates satisfied by C-revision.

Theorem 3.4 *Let K be a belief set, ϕ, $\psi \in \mathcal{L}$ and let $\hat{*}$ be defined via (2), then $\hat{*}$ satisfies:*

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

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

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

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

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

$(K \hat{*} 7)$ *If* $\psi \in K \hat{*} \phi$ *then* $K \hat{*} \phi = K \hat{*} (\phi \wedge \psi)$.

$(K \hat{*} 9)$ *If* $\psi \in \mathcal{L}(\phi)$ *then* $K \hat{*} \phi \models \psi$ *implies* $\phi \models \psi$.

$(K \hat{*} 10)$ *If* $\neg\phi \notin K \hat{*} \psi$ *and* $\neg\psi \notin K \hat{*} \phi$, *then* $K \hat{*} \phi \subseteq K \hat{*} (\phi \wedge \psi)$.

Numbering is intended to reflect correspondences with AGM revision postulates. Postulate $(K \hat{*} 9)$ is new and states that $K \hat{*} \phi$ is a conservative extension of ϕ. Since C-revision behaves the same as (standard, AGM) revision if the formulas involved in a revision are equivalent to sets of literals, AGM postulates 7 and 8 hold in C-revision if ϕ and ψ are equivalent to conjunctions of literals.

Counterexamples to other of the AGM revision postulates are given as follows:

Observation 1 *The following postulates do not hold of $K \hat{*} \phi$.*

$(K * 4)$ *If* $\neg\phi \notin K$, *then* $K + \phi \subseteq K * \phi$.
We have that $K \hat{*} (p \vee q) \equiv p \vee q$ where $K \equiv p$.

$(K * 7)$ $K * (\phi \wedge \psi) \subseteq (K * \phi) + \psi$.
Let $K \equiv \neg p \wedge \neg q$, $\phi = p \vee q$, and $\psi = p$.
Consider the SOS given by $\{\neg p \neg q\} < \{\neg pq\} < \{p \neg q\} < \{pq\}$
When we revise by $\phi \wedge \psi$ that's equivalent to p. So $\mathcal{L}(\phi \wedge \psi)$ is simply $\mathcal{L}(p)$. Hence there are two cells and the result is $p\neg q$. Similarly, $K \hat{*} \phi$ has satisfying worlds $\neg pq, p\neg q, pq$ and so $(K \hat{*} \phi) + \psi$ has satisfying worlds $p\neg q, pq$

$(K * 8)$ *If* $\neg\psi \notin K * \phi$, *then* $(K * \phi) + \psi \subseteq K * (\phi \wedge \psi)$.
Let $K \equiv \neg p \wedge \neg q \wedge \neg r$, $\phi = p \vee q$, and $\psi = r$.
Consider the SOS given by [6] $\{\neg p \neg q \neg r\} < \{\neg pqr, pqr\} < \{p \neg q \neg r\} < \{p \neg qr\}$.
The antecedent condition is true, since $K \hat{*} \phi$ has satisfying worlds $pqr, \neg pqr, p\neg q\neg r$ and so $\neg\psi \notin K \hat{*} \phi$. $(K \hat{*} \phi) + \psi$ has satisfying worlds $\neg pqr, pqr$ whereas $K \hat{*} (\phi \wedge \psi)$ has satisfying worlds $\neg pqr, pqr, p\neg qr$.

[6] Unmentioned interpretations are considered $\leq_{SOS}$-maximal.

3.2 Conservative Belief Update

We can repeat the preceding development for belief update, using the formulation of [12]. For our purposes, they associate a well-founded partial preorder with each interpretation ω, such that ω is the unique minimum in the preorder.[7] The update of K by ϕ, denoted $K \diamond \phi$, is defined by:

$$Mod_{\mathcal{L}}(K \diamond \phi) = \bigcup_{\sigma \in Mod(K)} \min_{\leq_\sigma}\{\omega \in M \mid \omega \models \phi\}. \quad (3)$$

Our definition for C-update is as follows:

$$Mod_{\mathcal{L}}(K \hat{\diamond} \phi) = \bigcup_{\omega_1 \in Mod(K)} \bigcup_{\sigma \in Mod_{\mathcal{L}(\phi)}\phi} \min_{\leq_{\omega_1}}\{\omega \in M \mid \omega \models \sigma\}. \quad (4)$$

We obtain the following postulates.

Theorem 3.5 *Let K be a belief set, ϕ, $\psi \in \mathcal{L}$ and let $\hat{\diamond}$ be defined via (4), then $\hat{*}$ satisfies:*

$(K \hat{\diamond} 0)$ $K \hat{\diamond} \phi$ *is a belief set.*

$(K \hat{\diamond} 1)$ $\phi \in K \hat{\diamond} \phi$.

$(K \hat{\diamond} 2)$ *If* $\phi \in K$ *then* $K \hat{\diamond} \phi = K$.

$(K \hat{\diamond} 3)$ *If* $K \not\models \perp$ *and* $\phi \not\models \perp$ *then* $K \hat{\diamond} \phi \neq K_\perp$.

$(K \hat{\diamond} 4)$ *If* $\models \phi \leftrightarrow \psi$, *then* $K \hat{\diamond} \phi = K \hat{\diamond} \psi$.

$(K \hat{\diamond} 5)$ *If* $\psi \in K \hat{\diamond} \phi$ *then* $K \hat{\diamond} \phi = K \hat{\diamond} (\phi \wedge \psi)$.

$(K \hat{\diamond} 7)$ *If* K *is complete then* $K \hat{\diamond} (\phi \vee \psi) \subseteq K \hat{\diamond} \phi \cup K \hat{\diamond} \psi$.

$(K \hat{\diamond} 8)$ $(K_1 \cap K_2) \hat{\diamond} \phi = (K_1 \hat{\diamond} \phi) \cap (K_2 \hat{\diamond} \phi)$.

$(K \hat{\diamond} 9)$ *If* $\psi \in \mathcal{L}(\phi)$ *then* $K \hat{\diamond} \phi \models \psi$ *implies* $\phi \models \psi$.

Again $(K \hat{\diamond} 9)$ states $K \hat{\diamond} \phi$ is a conservative extension of ϕ.

Note that update postulate $(K \diamond 6)$, is a straightforward consequence of $(K \hat{\diamond} 5)$. As well, $(K \diamond 5)$ does not hold for C-update. A counterexample is given by that for $(K * 7)$.

3.3 Context-Dependent C-Revision and C-Update

We defined $\mathcal{L}(\phi)$ as comprising the minimum set of atoms *required* to express ϕ. Hence $\mathcal{L}(p \wedge (q \vee \neg q)) = \{p\}$. Instead we can define $\mathcal{L}(\phi)$ as comprising the set of atoms *mentioned* in ϕ. This gives a syntax-dependent version of the approach denoted $\hat{*}_s$. For example, $K \hat{*}_s (p \vee \neg p)$ would result in a knowledge base in which the truth value of p is unknown. This can be generalised so that we can talk about a revision $K \hat{*} (C, \phi)$ where the intuition is that K is to be revised so that all that is known concerning the context C is ϕ. The context C is a set of atoms such that $\mathcal{L}(\phi) \subseteq C$. We define *context-dependent* (or *-sensitive*) revision as:

$$K \hat{*} (C, \phi) \;=\; K \hat{*}_s \left(\phi \wedge \bigwedge_{p \in C} (p \vee \neg p). \right)$$

As a special case, when the context is the full set of atoms, context-sensitive revision corresponds to full-meet revision:

Theorem 3.6 $K \hat{*} (\mathbf{P}, \phi) \equiv \phi$.

[7] There are two differences with the Grove system of spheres: each interpretation has its own preorder, and the ordering is partial rather than total.

As an example of context sensitive revision, consider the revision of a knowledge base $\{a \wedge b \wedge c \wedge d\}$ by $a \vee b$. Assume that a, b and c stand respectively for Albert, Becky and Charles being involved in a bank robbery, and d stands 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 will depend on extraneous factors. On the other hand, 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. The "syntax dependent" version of C-revision as defined above fulfills this expectation. We can clearly construct a similar context sensitive account of belief update. Context-based forms of update may be found in the literature. For example, dependence-based update [9; 11; 10] (see also [2]).

4 Discussion

With respect to semantics, the distinction between standard AGM revision and C-revision is very much 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, $K \diamond \phi$, we consider instead each model of K individually, and for each model of K look for the closest models of ϕ; the update is the union of all such models. Analogous to update, 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. Thus, for revision as given in (1), we can consider the duality of variants given in update (3) and C-revision (2). These are combined to yield C-update (4). 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.

As well, C-belief change is of independent interest. We have already noted that it conforms to a "Grecian" interpretation of the sentence for revision. As well it resolves the problem of disjunction noted by [11], but in a general and syntax-independent setting. To conclude this section, 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. [8] gives the following counterintuitive example: *Consider where K entails that "Cleopatra had a son and a daughter" (say, $s \wedge d$). Information is received that leads me to believe that Cleopatra didn't have a child; so the new belief set is given by $K \dot{-} (s \vee d)$. Then I learn that she had a child; so the new belief set is $(K \dot{-} (s \vee d)) + (s \vee d)$. Recovery says that I now believe $s \wedge d$, that Cleopatra had a son and daughter. Intuitively it seems that I should just believe $s \vee d$, that is, all I know after these changes is that she had a child.* We first note that in the case where $K \models \phi$, recovery can be written equally well as $K \subseteq (K \dot{-} \phi) * \phi$ using AGM revision, since we have $(K \dot{-} \phi) + \phi = (K \dot{-} \phi) * \phi$. Arguably the "Cleopatra" example is best interpreted as involving conservative revision, rather than standard revision—that is, concerning $\{s, d\}$, one learns at most that Cleopatra had a child. Thus under our alternative reading of revision we have possibly $K \not\subseteq (K \dot{-} \phi) \hat{*} \phi$. Specifically, in our example, if we have $K \equiv (s \wedge d)$, then we would expect: $K \dot{-} (s \vee d) \equiv \neg s \wedge \neg d$; $(K \dot{-} (s \vee d)) \hat{*} (s \vee d) \equiv (s \vee d)$. This accords with intuitions.

Second, and more speculatively, belief contraction has been criticised for removing too little information from a knowledge base; at the other extreme, *severe* contraction [19] has been criticised as removing too much information. [8] (echoing Lindström and Rabinowicz) 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.

5 Conclusion

We outlined 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 much more on the content of the new evidence. Our current analysis suggests that the operator we 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.

Acknowledgments

The authors would like to thank the two anonymous referees whose detailed comments helped to considerably improve this paper.

References

[1] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change. *Journal of Symbolic Logic*, **50**:510–530, 1985.

[2] P. Doherty, W. Lukaszewicz, E. Madalinska-Bugaj. The PMA and relativizing change for action update. In *KR'98*, 1998.

[3] K. D. Forbus. Introducing actions into qualitative simulation. In *IJCAI'89*, pages 1273–1278, 1989.

[4] P. Gärdenfors and D. Makinson. Revisions of knowledge systems using epistemic entrenchment. In *TARK'88*, pages 83–96, 1988.

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

[6] H. P. Grice. Logic and Conversation. In *Syntax and Semantics 3: Speech Acts*. Academic Press, 1975.

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

[8] S-O. Hansson. A textbook of belief dynamics: Theory change and database updating. Kluwer, 1999.

[9] A. Herzig, The PMA revisited. In *KR'96*, pp. 40–50, 1996.

[10] A. Herzig, J. Lang, P. Marquis and T. Polacsek. Updates, actions, and planning. In *IJCAI'01*, pp. 8–13, 2001.

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

[12] H. Katsuno and A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In P. Gärdenfors, editor, *Belief Revision*, pages 183–203. CUP, 1992.

[13] G. Lakemeyer. Limited Reasoning in first-order knowledge bases with full introspection, *Artificial Intelligence*, **84**, pp. 209–255, 1996.

[14] G. Lakemeyer and H.J. Levesque. *Logic of Knowledge Bases*. MIT Press, 2000.

[15] D. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts, 1973.

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

[17] R. Parikh. Beliefs, belief revision, and splitting languages. In L. S. Moss, J. Ginzburg, and M. de Rijke, editors, *Logic, Language and Computation, Vol II*, pages 266–278. CSLI Publications, 1999.

[18] P. Peppas and M-A. Williams. Constructive modelings for theory change. *Notre Dame Journal of Formal Logic*, **36**(1):120–133, Winter 1995.

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

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

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