Prolegomenon to a Theory of Conservative Belief Revision

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*. 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.

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 [Makinson, 1993]. 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] i.e., orderings of possible worlds.

In this paper we introduce an account of belief change that is orthogonal to the notion of revision 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. Our approach ensures that, in a sense to be specified, *exactly* the sentence accepted as evidence is incorporated. It proves to be the case that a modified knowledge base is a *conservative extension* (see Section 2) of the sentence for belief change; consequently we term this *conservative* belief change.

1 Motivation and Examples

The following example illustrates the traditional account of integrating new information, in accord with 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]

To be sure, this result is not dictated by the standard postulates 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 (e.g., [Winslett, 1990]), 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) *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 idea here 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 [Grice, 1975] 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 [Lakemeyer and 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

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

[1] $\leftrightarrow$ is *material biconditional* and $\equiv$ is *logical equivalence*.

knows is ϕ or all that an agent knows about α is ϕ. 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.

2 Preliminaries

We consider a finitary propositional language $\mathcal{L}$, over a set of atoms, or propositional letters, $\mathbf{P} = \{a, b, c, \ldots\}$, truth-functional connectives $\neg$, $\wedge$, $\vee$, $\supset$, and $\leftrightarrow$, and truth-functional constants $\top$ and $\bot$. Interpretations and models are defined in the standard way; M is the set of interpretations of $\mathcal{L}$. $Mod_{\mathcal{L}}(\phi)$ 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 $\phi \in \mathcal{L}$, 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 [Parikh, 1999]). Thus $\mathcal{L}(p \wedge (q \vee \neg q)) = \mathcal{L}(p) = \{p\}$. This extends to sets of sentences in the obvious way. It follows that if $\models \phi \leftrightarrow \psi$ then $\mathcal{L}(\phi) = \mathcal{L}(\psi)$, and 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.

A common approach in addressing belief revision has been to provide a set of *rationality postulates* for belief change functions. The *AGM approach* [Gärdenfors, 1988] 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 logical consequence. 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. Expansion is intended to be applied when new information is consistent with current beliefs. $K_\bot$ is the inconsistent belief set (i.e. $K_\bot = \mathcal{L}$).

In *belief revision*, the new information may be inconsistent with the reasoner's beliefs and needs to be incorporated in a consistent manner where possible. See [Gärdenfors, 1988] for the revision postulates. 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, well-founded 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.

3 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" or "context" 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 the 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 a revision may in fact yield a weakening of the knowledge base. This restriction does not necessarily hold for the sentences not in $\mathcal{L}(\phi)$.

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 $\mathcal{L}(\neg x \vee \neg y)$. 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.

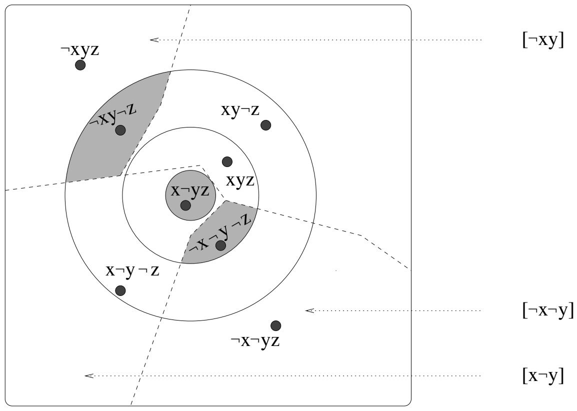

Figure 1: Conservative Revision – Semantics

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 the belief z is lost. In fact, the 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. Given (1), we have the analogous definition for C-revision:

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

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 $\phi \models \psi$.*

We 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 would require $\{p, q\}$, $\{\neg p, q\}$ 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.

Example 3.1 *Consider 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 Definition 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. However, we leave this to future work.

We consider next those postulates satisfied by C-revision.

Theorem 3.3 *Let K be a belief set, ϕ, $\psi \in \mathcal{L}$ and let $\hat{*}$ be defined via (2), then $\hat{*}$ satisfies the following properties:*

$(K \hat{*} 1)$ *$K \hat{*} \phi$ 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 $K \hat{*} \phi \subseteq K \hat{*} (\phi \wedge \psi)$.*

The numbering is intended to reflect correspondences with the 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-style) 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.

There are counterexamples to other AGM postulates.

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

$(K * 4)$ *If $\neg\phi \notin K$, then $K + \phi \subseteq K * \phi$.*
$(K * 7)$ $K * (\phi \wedge \psi) \subseteq (K * \phi) + \psi$.
$(K * 8)$ *If $\neg\psi \notin K * \phi$, then $(K * \phi) + \psi \subseteq K * (\phi \wedge \psi)$.*

4 Conclusion

We have outlined a theory of conservative belief change and presented an analysis of its properties. 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.

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. In a similar way in which we motivate C-revision from standard revision, we can define a notion of C-update from standard update. 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. It is also relatively straightforward to define syntax-dependent versions of both C-revision and C-update. Also we can look at C-contraction operations (both syntax-independent and dependent versions). However we leave this to future work.

References

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

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

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

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

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

[Parikh, 1999] 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.

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