Revision Sequences and Nested Conditionals

Craig Boutilier
Department of Computer Science
University of British Columbia
Vancouver, BC, CANADA, V6T 1Z2
email:cebly@cs.ubc.ca

Abstract

The truth conditions for conditional sentences have been well-studied, but few compelling attempts have been made to define means of evaluating iterated or nested conditionals. In particular, most approaches impose very few constraints on the set of conditionals an agent can hold after revision of its belief set. In this paper we describe the method of natural revision that ensures the preservation of conditional beliefs after revision by an objective belief. Our model, based on a simple modal logic for beliefs and conditionals, extends the AGM theory of belief revision to account for sentiences of objective revisions of a belief set. This model of revision ensures that an agent makes as few changes as possible to the conditional component of its belief set. Adopting the Ramsey test, natural revision provides truth conditions for arbitrary right-nested conditionals. We show that the problem of determining acceptance of any such nested conditional can be reduced to acceptance tests for unnested conditionals, indicating that iterated revision can be simulated by virtual updates. We also briefly describe certain reductions to (sometimes tractable) propositional inference, and other informational properties.

1 Introduction

Subjunctive conditionals have recently attracted much attention in the knowledge representation community. It has been pointed out that counterfactuals may play a large role in planning and diagnostic systems (Ginsberg 1986), that subjunctives may be used to capture knowledge base update and revision (Katsuno and Mendelzon 1991; Boutilier 1992b), and that they are intimately related to the conditionals used in default reasoning (Boutilier 1992c; Makinson and Gardenfors 1990). We denote by $A > B$ the subjunctive conditional "If A were the case then B would be true." Various subjunctive logics have been proposed to account for properties of the connective > (Stalnaker 1968; Lewis 1973).

From the point of view of knowledge representation, acceptance conditions for $A > B$ are especially important. Under what conditions should an agent assent to the conditional? A widely endorsed acceptance test for conditionals is the Ramsey test (Stalnaker 1968, p.44):

> First add the antecedent (hypothetically) to your stock of beliefs; second make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is true.

The key step in the Ramsey test is the revision of the belief set. The notion of revision adopted will determine which conditionals are accepted and rejected. Conversely, given a fixed (complete) set of accepted conditionals, the revision function adopted by an agent will also be determined: revising by A is simply a matter of believing those B such that $A > B$ is accepted. Consequently, the study of revision and (subjunctive) conditional logic are virtually the same if one accepts the Ramsey test.

The most prominent theory of belief revision is that put forth by Alchourron, Gardenfors and Makinson (1985) and expounded by Gardenfors (1988). Within this framework many people have explored the connection to conditionals (Gardenfors 1988; Gardenfors 1986; Rott 1989; Boutilier 1992b; Boutilier 1992c). The AGM theory, which we describe in the next section, imposes various constraints on acceptable revision functions. Roughly, a revision function preserves as much information as possible. Unfortunately, the AGM theory has little to say about revision sequences. If we accept the Ramsey test as determining conditional beliefs, this means that the conditionals accepted in the new belief set need not be related to those in the first set. Thus, while information content is "preserved" with respect to objective beliefs, the information content of conditionals is ignored.

It is this problem of preserving conditionals under revision that we investigate here. The semantic model of AGM revision (for propositional belief sets) we describe in the next section orders possible worlds according to their plausibility, or "degree of consistency" with a fixed belief set K. While this ordering guides the selection of a revised belief set K^*_A (which incorporates A into K), this AGM model fails to provide a new ordering suitable for the revision of K^*_A. The goal of this paper is to show how one might use the original ordering to constrain the new ordering, hence revision of the *revised* belief set K^*_A. We propose that the new ordering retain as much of the old ordering as possible, consistent with the AGM postulates. This *minimal change* in the ordering is precisely defined and provides the semantic basis for our model of revision sequences. This approach, dubbed *natural revision,* ensures that a maximal subset of one's conditional beliefs is retained during the revision process. Of course, this is a specific instance of the more general phenomenon of iterated revision captured by general revision systems. However, we shall argue that our model provides a very natural way of extending the concept of "minimal change," the hallmark of the AGM theory, to sequences of revisions and conditional beliefs. Indeed, natural revision provably retains the maximum amount of conditional information consistent with the constraints of the AGM theory. It also provides a logical characterization and computational methods such iteration. Proofs of the main results can be found in the full version of the paper (Boutilier 1992d).

2 A Logic for Revision

In this section we very briefly describe the AGM theory of revision (Alchourrón, Gärdenfors and Makinson 1985; Gärdenfors 1988) and the logic and semantic model for revision proposed by Boutilier (1992b). For more details and motivation we refer to these works.

We take a *belief set* K to be a deductively closed set of sentences in the language $\mathbf{L}_{CPL}$ of classical propositional logic. New beliefs must be added to this set when an agent learns new information. If this new sentence A is consistent with K, then the new belief set $K^+_A = Cn(K \cup \{A\})$ seems appropriate. However, whenever $K \vdash \neg A$, certain beliefs must be given up before adding A to K. Alchourrón, Gärdenfors and Makinson (1985) have proposed that those sentences in K with the least information content be retracted. They propose a set of postulates (K*1)–(K*8) that constrain the behavior of logically acceptable revision functions. We use K^*_A to denote revision of belief set K by A. Two of the key postulates are (K*3) and (K*4)

(K*3) $K^*_A \subseteq K^+_A$; **(K*4)** If $\neg A \notin K$ then $K^+_A \subseteq K^*_A$.

Together these ensure that $K^*_A = K^+_A$ whenever $K \nvdash \neg A$. This reflects the principle of informational economy in the extreme case: if no beliefs need to be given up to accommodate a new belief A, then none should. Postulates (K*7) and (K*8) taken together ensure that this applies to revised belief sets as well:

$$\text{If } \neg B \notin K^*_A \text{ then } (K^*_A)^+_B \subseteq K^*_{A \wedge B}.$$

In (Boutilier 1992b) we provide a modal semantics and logic for AGM revision based on an observation of Grove (1988) that equates the entrenchment or importance of beliefs with an ordering on possible worlds. The modal logic used, called CO*, is based on a standard propositional language augmented with two modal operators $\Box$ and $\overleftarrow{\Box}$. We denote by $\mathbf{L}_B$ this bimodal language and by $\mathbf{L}_{CPL}$ its classical propositional sublanguage. A *CO*-model* is a triple $M = \langle W, R, \varphi \rangle$, where W is a set of worlds with valuation function φ and R is an ordering ("accessibility") relation over W. We insist that R be transitive and connected.[1] Furthermore, every propositional valuation, or logically possible world, must be represented by some world in W. This ensures that $\alpha \in \mathbf{L}_{CPL}$ is satisfied at some world if it is consistent. An axiomatization can be found in (Boutilier 1992b). Space limitations preclude a discussion of the semantics of the primitive modal operators. The defined operators (below) are more crucial here. The satisfaction of $\alpha \in \mathbf{L}_{CPL}$ at a world in a model ($M \models_w \alpha$) is defined in the usual way. $\|\alpha\|$ refers to the set $\{w : M \models_w \alpha\}$ and we define the set of *minimal* α-worlds as $min(M, \alpha) =$

$$\{w \in W : M \models_w \alpha, \text{ and } M \models_v \alpha \text{ implies } vRw \text{ for all } v\}$$

The notion of *cluster* will play a large role in future developments. In any reflexive, transitive Kripke frame, a cluster is any maximal set of mutually accessible worlds (Segerberg 1970); $C \subseteq W$ is a cluster just when wRv for all $v, w \in C$ and no extension $C' \supset C$ has this property. CO*-structures consist of a totally-ordered set of clusters of worlds (see Figure 1).

We can use CO*-models to represent the revision of a theory K (for a more detailed account we refer to (Boutilier 1992b)). The interpretation of R is as follows: wRv iff v is as at least as *plausible* a state of affairs as w given an agent's belief state K. As usual, v is *more plausible* than w iff wRv but not vRw. If v is more plausible than w, loosely speaking v is "more consistent" with the belief set K than w.

We require that those worlds consistent with our belief set K should be exactly those minimal in R. That is, vRw for all $v \in W$ iff $M \models_w K$. This condition ensures that no world is more plausible than any world consistent with K, and that all K-worlds are equally plausible. It is just the epistemically possible worlds (those consistent with K) that should be most plausible. Such models are called *K-revision models* and have as their minimal cluster the set $\|K\|$. This constraint can be expressed in $\mathbf{L}_B$ for any K that is finitely expressible as KB (Boutilier 1992b). This sentence is dubbed $O(KB)$ and is intended to mean we "only know" KB.

Given this structure, we want the set of minimal A-worlds to represent the state of affairs believed when K is revised by A, since these are the most plausible worlds, the ones we are most willing to adopt, given A. In Figure 1, we have a typical K-revision model: each circle represents a cluster of equally plausible worlds, with arrows indicating accessibility between clusters. The minimal cluster consists of all K-worlds, and we have $K \vdash \neg A$. The set of minimal A-worlds is indicated by the shaded region, and this set forms the set of "newly accepted" worlds when K is revised by A. Thus $A > B$ should hold exactly when B is true at each world in the shaded region.[2] This connective is definable in $\mathbf{L}_B$

[1] R is (totally) connected if wRv or vRw for any $v, w \in W$ (this implies reflexivity). This restriction is relaxed in (Boutilier 1992a).

[2] Of course such a *minimal* set of A-worlds may not exist. The definition of the connective is valid in this case as well (Boutilier 1992b), but we ignore this circumstance here for simplicity (see Section 3).

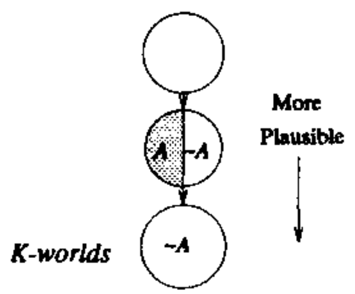

Figure 1: Truth conditions for the conditional

(Boutilier 1992b) and its truth conditions reduce to

$$M \models_w A > B \quad \text{iff} \quad min(M, A) \subseteq \|B\|$$

It is important to note the "global" nature of this connective. If $A > B$ holds at some world in W then it holds at all worlds. The truth of $A > B$ at w does not depend on w, but on the set $min(M, A)$.

For any $A \in \mathbf{L}_{CPL}$ the belief set resulting from revision of K by A, for a giv model M, is

$$K_A^{*^M} = \{B \in \mathbf{L}_{CPL} : M \models A > B\}. \qquad (1)$$

We can show that $*^M$ satisfies the AGM postulates for belief revision and any AGM revision operator has an equivalent formulation as such a $*^M$ (Boutilier 1992b). Thus, we can use the logic CO* to represent the revision of a theory KB, and reason about such a revision, in a manner respecting the AGM postulates.

CO* is a reasonable epistemic logic, as well. We can define a belief modality B in $\mathbf{L}_B$ (Boutilier 1992c) reading $\mathsf{B}(A)$ as "A is believed." This sentence will hold just when A is true at each epistemically possible (minimal) world: $M \models_w \mathsf{B}(A)$ iff $min(M, \top) \subseteq \|A\|$ iff $A \in K$. We note that B behaves according to the usual weak S5 interpretation of belief (Boutilier 1992c), and its truth conditions too are "global." For any CO*-model M, we can define the *objective belief set* (or simply the belief set) associated with it to be its propositional "beliefs":

Def. 1 The *objective belief set* associated with M is $\{\alpha \in \mathbf{L}_{CPL} : M \models \mathsf{B}\alpha\}$.

Naturally, the belief set for any K-revision model is just K. We will be more interested in "subjective beliefs" of a revision model, those beliefs involving certain modal operators (in particular, conditional beliefs).

Def. 2 The *extended belief set* associated with M is $\{\alpha \in \mathbf{L}_B : M \models \mathsf{B}\alpha\}$.

For any CO*-model with belief set K and extended belief set E, we have $K \subseteq E$. While we have clear characterization of the revised belief set K_A^*, it is less clear what form the revised *extended* set E_A^* should take.

3 Natural Revision

3.1 The Problem of Iterated Revision

We say a model is *smooth* iff, for all $A \in \mathbf{L}_{CPL}$, $min(A, M) \neq \emptyset$. Grove (1988) shows that such models are adequate for the representation of AGM revision functions and this result can be applied to our CO* model of revision directly. In what follows we assume all CO*-models are smooth in this sense.

Suppose we have a revision model M suitable for belief set K. When K is revised by A the status of an objective belief B in the new belief set K_A^* is easily determined by inspection of M. B is in K_A^* just when B is true at all most plausible A-worlds in M. (If $E \supseteq K$ is the corresponding extended belief set, this simply means $A > B \in E$.) We can speak of K_A^* being represented by the set $min(M, A)$ because K_A^* contains only *objective* beliefs, and corresponds to this set of worlds in a natural fashion.

It is less clear just what new *subjective* beliefs an agent accepts after revision by A. If $min(M, A)$ is the set of worlds representing an agent's new belief set K_A^*, we might hope that it captures the new extended belief set E_A^*, as well. However, this view is untenable because of the nonextensional truth conditions for conditionals and belief sentences. Unlike an objective sentence, the truth of a conditional $A > B$ or explicit belief sentence $\mathsf{B}(A)$ is not determined by a set of worlds. These can only be evaluated with respect to a complete model, or ordering of worlds, stating the relative plausibility of all worlds. This is due to the global nature of the connectives B and $>$. It should be clear that the original model M is not suitable for this purpose. A K-revision model is suitable for a fixed belief set only (indeed, a fixed extended belief set). It represents one ordering, hence one set of conditionals and one set of beliefs. If we used

$$\{\alpha \in \mathbf{L}_B : M \models_w \alpha \text{ for each } w \in min(M, A)\}$$

as the new extended belief set, one could never give up conditionals or other subjective beliefs, though the objective component K can change drastically. When E is revised by A, the CO*-model used to represent E becomes inadequate. This model is an E-revision model and the representation of E_A^* requires (of course) an E_A^*-revision model.

What are the natural requirements on this new model? When a propositional revision A is received, we want the revision function $*$ to map model M into a new model M_A^* that captures the revised belief set E_A^*. Clearly, K_A^* is uniquely determined by M (in particular, by $min(M, A)$). Naturally, we insist that $K_A^* \subseteq E_A^*$ and that K_A^* form the *entire* objective component of E_A^*. This simply means that M_A^* should be a K_A^*-revision model, or that K_A^* is *only known* in M_A^* (Boutilier 1992b; Levesque 1990). The minimal cluster of worlds in M_A^* should be exactly $min(M, A)$. This is illustrated in Figure 2. Let us dub this constraint the *Basic Requirement* on revision functions as applied to models.

The Basic Requirement: If M is a K-revision model then the K_A^*-revision model M_A^* must be such that $min(M_A^*, \top) = min(M, A)$.

In fact, from a purely logical perspective, this is probably all we want to say about M_A^*. If one changes an objective belief, it is impossible in general to predict what becomes of one's conditionals. This model of iterated revision is captured by Gärdenfors's (1988) *belief revision systems*, although not in this semantic fashion. A severe

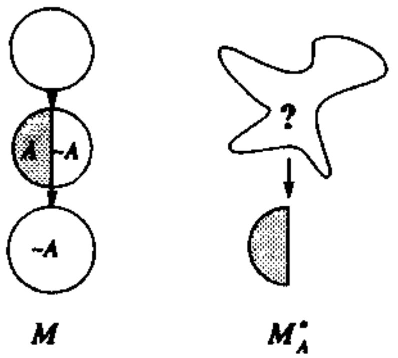

Figure 2: General constraints on the revised model

drawback of such a general model is that just about all of the ordering information, capturing an agent's conditional beliefs and judgements of entrenchment, is (potentially) lost in this mapping (see Figure 2). There is something unsatisfying about this model. The ordering relation R is intended to reflect the informational content or importance of beliefs. When certain beliefs must be given up, it seems natural to try to keep not only important beliefs, but as much of the *ordering* a possible. A revision should not usually change one's opinion of the relative importance of most sentences.

3.2 The Semantics of Natural Revision

Instead of arbitrary mappings from M to M^*_A, we will propose a class of natural mappings that preserve as much ordering information as can be expected. This determines the class of *natural revision functions,* that tend to preserve the entrenchment information and conditional beliefs found in an extended belief set. It is important to note that the model we propose is *not* completely general, for it permits only a subset of those revision functions (on extended sets) allowed by the arbitrary mappings above. However, it is a very natural subset, suitable for determining the result of propositional revision sequences, or the truth of right-nested conditionals, when the general model has little to offer.

The conditionals accepted by an agent are determined by its ordering of plausibility. If we insist that revision preserve as much of this ordering as possible, then, for the most part, the relative entrenchment and plausibility of sentences (hence conditional beliefs) will remain intact. Let $M = \langle W, R, \varphi \rangle$ *be* the revision model reflecting some extended belief set E. Given a propositional revision A of E (or the associated K), we must find a revision model $M^*_A = \langle W, R', \varphi \rangle$ such that R' reflects the minimal mutilation of R.

If $w \in min(M, A)$, by the Basic Requirement w must be minimal in R'', and these must be the only minimal worlds in R'. For any such w the relationships $wR'v$ and $vR'w$ are completely determined by membership of v in $min(M, A)$, independently of their relationship in R. Figure 2 illustrates this. For w, v not in $min(M, A)$, this picture leaves $wR\,v$ completely unspecified. If R is to be left intact to the largest possible extent then the most compelling specification is to insist that $vR'w$ iff vRw.

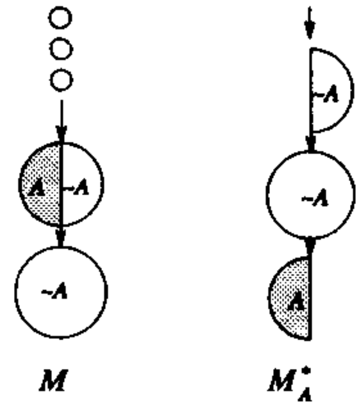

Figure 3: Natural revision of a model

This has the effect of leaving R unaltered except as indisputably required by the Basic Requirement. Such a move is illustrated in Figure 3. We dub such a mapping on revision models the ***natural revision operator***, and now describe the revision function it induces on the associated belief and extended belief sets.

Def. 3 Let $M = \langle W, R, \varphi \rangle$ be a revision model. The *natural revision operator* $*$ maps M into M^*_A, for any $A \in \mathbf{L}_{CPL}$, where $M^*_A = \langle W, R', \varphi \rangle$, and: a) if $v \in min(M, A)$ then $wR'v$ for all $w \in W$ and $vR'w$ iff $w \in min(M, A)$; and b) if $v, w \notin min(M, A)$ then $vR'w$ iff vRw.

Def. 4 Let E be the extended belief set associated with the revision model M. The *natural revision function* associated with M is $*$, defined for all $A \in \mathbf{L}_{CPL}$ as:
$E^*_A = \{B \in \mathbf{L}_B : M^*_A \models \mathsf{B}(B)\}$
Let $K \subseteq E$ be the objective part of E. The natural revision of K, K^*_A, is the restriction of E^*_A to $\mathbf{L}_{CPL}$;
$K^*_A = \{B \in \mathbf{L}_{CPL} : M^*_A \models \mathsf{B}(B)\}$

Clearly, the natural revision function is simply the AGM operator determined by M, when restricted to K.

Proposition 1 *Let M be a K-revision model. The AGM function $*^M$ and the natural revision function $*$ determined by M are identical (when restricted to $\mathbf{L}_{CPL}$).*

Notice that this extends the AGM model of revision. The revised extended set E^*_A is defined using the updated revision model M^*_A and incorporates non-objective beliefs, such as conditionals and nested belief sentences. Had we simply defined E^*_A to be those sentences true in M at the minimal A-worlds, we would have run into the problem discussed in Section 2, namely the fact that a model M can only model a fixed set of beliefs and conditionals.

If we are to extend the Ramsey test to include nested conditionals, the truth conditions for statements $A > \beta$ must be recast in this framework. For M to satisfy $A > \beta$, we must have $\beta \in E^*_A$ for the natural revision function $*$. For $\beta \in \mathbf{L}_{CPL}$ these truth conditions will be identical to those provided in Section 2. Thus, our new truth conditions for $>$ based on the Ramsey test will form a "conservative extension" of the old definition. However, for arbitrary $\beta \in \mathbf{L}_B$, especially sentences like $B > C$, the meaning of $>$ given in Section 2 is inadequate since it refers to truth at worlds in $min(M, A)$. To evaluate $A > (B > C)$ we must test $B > C$ in M_A^*, not at $min(M, A)$. The connective $>$ must be introduced as primitive. The *conditional language* $\mathbf{L}_C$ is the extension of $\mathbf{L}_B$ with primitive connective $>$. We have, for $A \in \mathbf{L}_{CPL}$ and $B \in \mathbf{L}_C$,

$$M \models_w A > B \quad \text{iff} \quad M_A^* \models \mathsf{B}(B) \tag{2}$$

Now we have a conditional connective whose truth conditions are specified directly by the Ramsey test. While this certainly provides us with a new logic, (seemingly) requiring a new axiomatization, we will soon see that these truth conditions can, in fact, be captured in the bimodal language using only the original version of the connective $>$ (defined in $\mathbf{L}_B$).

Notice that the truth of $A > B$ is unspecified for $A \notin \mathbf{L}_{CPL}$. Natural revision functions are suitable only for sequences of *propositional* revisions. Only the *right-nesting* of conditionals is meaningfully sanctioned in this framework, e.g., $A > (B > C)$ where $A, B, C \in \mathbf{L}_{CPL}$. A sentence $(A > B) > C$ has an unspecified truth value for it asks if C is believed when a knowledge base is revised to include $A > B$. This framework does not specify how to revise a knowledge base with non-objective sentences, though this problem is addressed within the natural framework in (Boutilier and Goldszmidt 1993).

3.3 Properties of Single Revisions

In this section, we examine the behavior of the natural revision model when a single propositional revision is effected. We assume throughout that we have a revision model M determining a belief set K and extended set E. Natural revision is intended to change as little of the ordering information R as possible; consequently, as many conditional beliefs as possible should be retained in the move from E to E_A^*. We now examine the structure of E_A^* and show that this is indeed the case. Since we are only interested in single revisions at this point, we restrict our attention to the *simple conditionals* in E_A^*, of the form $B > C$ where $B, C \in \mathbf{L}_{CPL}$. We note:

Proposition 2 *Let M be a K-revision model where $A \in K$. Then $M_A^* = M$.*

Updating by a sentence already in a belief set not only causes no change in the belief set K, as required by the AGM postulates, but also leaves the revision model M (and the extended set E) intact.

Suppose we revise a belief set $K \subseteq E$ by A as specified by model M. We wish to determine the set of conditionals $B > C$ contained in the revised extended set E_A^*. We consider two cases, $\neg B \notin K_A^*$ and $\neg B \in K_A^*$.

Consider the first situation where $\neg B \notin K_A^*$; that is, $M \models \neg(A > \neg B)$. This means that there is some B-world among the set $min(M, A)$ of minimal A-worlds in M (the shaded region of Figure 3). Clearly then we have that $min(M_A^*, B) = min(M, A \wedge B)$; so $M_A^* \models B > C$ iff $M \models A \wedge B > C$. Therefore, whenever $K_A^* \nvdash \neg B$, a conditional $B > C$ is in E_A^* iff $A \wedge B > C$ is in E. Notice that the status of $B > C$ in E has no bearing on its acceptance or rejection in E_A^*. This behavior is exactly in accordance with the AGM postulates (K*7) and (K*8). Any AGM revision function must ensure that subsequent (iterated) consistent revisions are treated in the same manner as uniterated consistent revisions (i.e., as expansions).

The second situation arises when $\neg B \in K_A^*$; that is, $M \models A > \neg B$. When K (or E) is revised by A, $\neg B$ is in the resulting belief set. This is true exactly when no B-world is contained in $min(M, A)$ (again, the shaded region). Now, $M_A^* \models B > C$ just in case the set $min(M_A^*, B)$ contains only C-worlds. Since $\neg B \in K_A^*$, the set $min(M_A^*, B)$ is not contained in the minimal cluster of M_A^*. However, all worlds outside the minimal cluster stand in exactly the same relation as they do in M. Therefore $min(M_A^*, B) = min(M, B)$ and it follows that $M_A^* \models B > C$ iff $M \models B > C$. For conditionals $B > C$ whose antecedents are not made plausible by the acceptance of A (i.e., $K_A^* \nvdash \neg B$), $B > C$ is in E_A^* iff $B > C$ is in E. Since nothing forces the conditional to be abandoned when A is accepted, it is retained. We can summarize these considerations in the following theorem and equivalent corollaries.

Theorem 3 *Let M be a revision model, let $*$ be the natural revision operator and let $A, B, C \in \mathbf{L}_{CPL}$. (a) If $M_A^* \models \mathsf{B}\neg B$ then $M_A^* \models B > C$ iff $M \models B > C$. (b) If $M_A^* \not\models \mathsf{B}\neg B$ then $M_A^* \models B > C$ iff $M \models A \wedge B > C$.*

Corollary 4 *(a) If $K_A^* \vdash \neg B$ then $C \in (K_A^*)_B^*$ iff $C \in K_B^*$. (b) If $K_A^* \nvdash \neg B$ then $C \in (K_A^*)_B^*$ iff $C \in K_{A \wedge B}^*$.*

Corollary 5 *(a) If $A > \neg B \in E$ then $B > C \in E_A^*$ iff $B > C \in E$. (b) If $A > \neg B \notin E$ then $B > C \in E_A^*$ iff $A \wedge B > C \in E$.*

These results precisely characterize the conditionals that will be preserved in a revised extended belief set. Each shows that the sentences accepted in the new revision model or belief state can be determined by appeal to the original model or belief state. Theorem 3 shows that the conditional belief set captured by M_A^* can be determined by the conditional beliefs of M. Furthermore, it demonstrates that natural revision preserves as much conditional information in the revised belief set as is consistent with the AGM postulates. The conditionals removed from E when constructing E_A^* are only those compelled by postulates (K*7) and (K*8). These are just those conditionals $B > C$ whose antecedent B is consistent with K_A^*. This is reflected in clause b) of the theorem. However, as indicated by clause a), the remaining set of conditionals (or negated conditionals) in E_A^* coincides precisely with the conditional information in the original extended set E. Thus, no AGM revision function could preserve more conditional information than the natural revision function.

Corollary 4 shows that the sequence of two revisions applied to K can be reduced to a single revision, requiring no iterated revision, and that the test to establish which condition holds also requires no iterated revision. Similarly, Corollary 5 shows that the revised extended belief set E_A^* and the nested conditionals in E are determined by the unnested conditionals in E. These properties play a vital role in our characterization of revision sequences in terms of single updates.

3.4 Revision Sequences

For any revision sequence $A_1, \ldots A_n$, let us say that revision A_k is *compatible* within the sequence if it consistent with the result of previous revisions, i.e, if $\neg A_k \notin ((K^*_{A_1})^*_{A_2} \cdots)^*_{A_{k-1}}$. The sequence is compatible if each element $A_2, \ldots A_n$ is compatible. In the previous section we saw that a two-element sequence can be reduce to a single propositional revision: $(K^*_A)^*_B = K^*_{A \wedge B}$ if B is compatible with A, and $(K^*_A)^*_B = K^*_B$ if B is incompatible. This analysis can be extended to arbitrary sequences as well, allowing us to compute single revisions that characterize revision sequences:

Def. 5 Revision sequence $A_1, \ldots A_n$ is *characterized* by the sentence α iff $((K^*_{A_1})^*_{A_2} \cdots)^*_{A_n} = K^*_\alpha$.

Due to space limitations, a detailed motivation and analysis of the following definitions and results cannot be provided. We refer to the full paper (Boutilier 1992d).

We first note that compatible sequences are reducible to single updates.

Theorem 6 *If* $A_1, \ \ldots A_n$ *is compatible then* $((K^*_{A_1})^*_{A_2} \cdots)^*_{A_n} = K^*_{A_1 \wedge \cdots A_n}$.

We also note that a sequence is compatible iff $K^*_{A_1} \not\vdash \neg(A_2 \wedge \cdots A_n)$. We should note, however, that while the objective set $((K^*_{A_1})^*_{A_2} \cdots)^*_{A_n}$ is the equivalent to $K^*_{A_1 \wedge \cdots A_n}$, this is not true for the revision model $((M^*_{A_1})^*_{A_2} \cdots)^*_{A_n}$ or the extended belief set $((E^*_{A_1})^*_{A_2} \cdots)^*_{A_n}$. The revised model is constructed by first moving the set $min(M, A_1)$ to the bottom of the model, then moving the set $min(M^*_{A_1}, A_2)$ to the bottom and so on. Because the sequence is compatible, the final minimal cluster is precisely $min(M, A_1 \wedge \cdots A_n)$ as it would be in the model $M^*_{A_1 \wedge \cdots A_n}$. However, in the model $((M^*_{A_1})^*_{A_2} \cdots)^*_{A_n}$ each revision has left a "residual trace" on the model (Boutilier 1992d) with as many as n new clusters added to M, whereas $M^*_{A_1 \wedge \cdots A_n}$ has only one new cluster. This complicates somewhat our analysis of incompatible revisions.

Suppose that an incompatible revision A_{n+1} is added to this sequence. Since $\neg A_{n+1} \in ((M^*_{A_1})^*_{A_2} \cdots)^*_{A_n}$, the set $min(((M^*_{A_1})^*_{A_2} \cdots)^*_{A_n}, A_{n+1})$ does not lie within the minimal cluster of $((M^*_{A_1})^*_{A_2} \cdots)^*_{A_n}$. Furthermore, for any $k < n$, if the sequence $A_1, \ldots A_k, A_{n+1}$ is not compatible, then no A_{n+1}-worlds can lie in the cluster formed when revision by A_k took place (nor for any A_j, $j > k$). However, if $A_1, \ldots A_k, A_{n+1}$ is compatible, some A_{n+1}-world must lie within the cluster of $A_1 \wedge \cdots A_k$-worlds representing $K^*_{A_1 \wedge \cdots A_k}$.

For any revision A_j in the sequence $A_1, \ldots A_n$, its *most recent compatible revision* is defined as the A_k, $k < j$, such that

$$k = \max\{i : i < j \text{ and } \neg A_j \notin ((K^*_{A_1})^*_{A_2} \cdots)^*_{A_i}\}$$

This is the most recent revision (before A_j) in the sequence that did not force rejection of A_j. If A_j is compatible in the sequence, then clearly A_{j-1} is its most recent compatible revision. If the set above is empty, we say A_j has no such compatible revision.

Theorem 7 *Let* $A_1, \ldots A_n$ *be a revision sequence with one incompatible update* A_n. *Then* $((K^*_{A_1})^*_{A_2} \cdots)^*_{A_n} = K^*_{A_1 \wedge \cdots A_k \wedge A_n}$ *where* A_k *is the most recent compatible update for* A_n. *If there is no such* A_k *then* $((K^*_{A_1})^*_{A_2} \cdots)^*_{A_n} = K^*_{A_n}$.

Thus, we need only look back to find the most recent revision that allows the "possibility" of A_n and conjoin A_n to its characterizing sentence. Putting together Theorems 6 and 7 with the obvious inductive argument on the number of incompatible updates, we obtain the main result of this section.

Theorem 8 *For any revision sequence* $A_1, \ldots A_n$, *there is some subset* $S \subseteq \{A_1, \cdots, A_n\}$ *such that* $((K^*_{A_1})^*_{A_2} \cdots)^*_{A_n} = K^*_A$ *and* $A = \wedge S$.

Corollary 9 *For any revision sequence* $A_1, \ldots A_n$, *there is some subset of these updates* $S \subseteq \{A_1, \cdots, A_n\}$ *such that* $((E^*_{A_1})^*_{A_2} \cdots)^*_{A_{n-1}} \models A_n > B$ *iff* $E \models A > B$, *and* $A = \wedge S$.

This result is given its constructive character by Theorem 6, but it seems to suggest that one must keep track of a characterizing sentence $s(A_i)$ for each subsequence $A_1, \ldots A_i$. In fact, the critical sentences are only those corresponding to incompatible revisions in the sequence. Every other characterizing sentence $s(A_i)$ is simply the conjunction of subsequent revisions to the *most recent incompatible revision* (Boutilier 1992d). Taken together, these theorems show that one may implement a procedure that tests for membership of B in a multiply-revised belief set $((K^*_{A_1})^*_{A_2} \cdots)^*_{A_n}$ using only an "oracle" that answers requests of the form "Is $\beta \in K^*_\alpha$?" for $\alpha, \beta \in \mathbf{L}_{CPL}$. Furthermore, the characterizing sentences $s(A_j)$ that need to be recorded are only those capturing some incompatible revision (see the full paper for this algorithm). In essence, iterated revision can be accomplished *virtually*, without having to construct the sequence of new belief sets, and without having to change the ordering of entrenchment or plausibility associated with K. One need only find the appropriate characteristic sentence, and test a simple conditional belief with that antecedent.

4 Concluding Remarks

We have presented a model for belief revision that extends the AGM theory in a manner that accounts for iterated revision. The natural revision model preserves the maximal amount of conditional information, and has the property that the result of any revision *sequence* can be calculated using only the AGM revision function on the original belief set K. This reduction is bears some resemblance to suggestions in the literature equating a nested conditional $A > (B > C)$ with $A \wedge B > C$ (Adams 1975; Levi 1988). If A and B are compatible this is precisely the behavior of natural revision. If A and B are incompatible, our model behaves differently, allowing a sentence like $A > (\neg A > C)$ to have meaningful truth conditions. But it also extends this idea to "defeasibly," not just logically, incompatible sentences.

Another important property of natural revision, reflecting its commitment to information preservation is the following (see (Boutilier 1992d)):

Theorem 10 *Let M be a K-revision model and $A_1, \ldots A_n$ a revision sequence. If $i \leq j$ then $((K^*_{A_1})^*_{A_2} \cdots)^*_{A_j} \not\subset ((K^*_{A_1})^*_{A_2} \cdots)^*_{A_i}$.*

Thus, a revision sequence $A_1, \ldots A_n$ causes a non-decreasing change in "information" in a belief state. No belief set further along in the revision sequence can be smaller than an earlier belief set. This suggests that, as we process a revision sequence, our revision model becomes more and more *informationally complete*. Given "enough" revisions, a model approaches the point where each cluster becomes a single world (since clusters are only broken apart by revision, not put together)- then K and K^*_A become complete theories.

We need not have complete information to reason about natural revision. From a set of premises we can reason in CO* about "constraints" on natural revision, without perhaps determining complete theories. However, there are methods of "completing" an incomplete set of conditionals, for instance, Pearl's (1990) System Z. In (Boutilier 1991) we show how such a completion can be expressed compactly in CO*. In the full paper, we show that if a revision model has a finite number of clusters each corresponding to a propositional theory (as in System Z), then only propositional (sometimes tractable) inference is required to compute natural revision.

Extensions of this work we are currently exploring include the application of this model to more quantitative types of conditionals (e.g., probabilistic degrees of belief and the *J-conditionalization* of (Goldszmidt and Pearl 1992)). We are also looking at applications to diagnosis, where a stream of observations must be reconciled with expected knowledge. Simply conjoining observations is not feasible if they conflict. Planning also requires a notion of revision as an agent must change its beliefs in response to new information, and changes in the world, requiring that we account for belief *update* (Winslett 1988; Katsuno and Mendelzon 1991). Finally, arbitrary nesting is explored in (Boutilier and Goldszmidt 1993), where we extend *natural revision* to include new conditional information, how one can revise an extended belief set to include new conditional information $A > B$.

Acknowledgements

Thanks to Veronica Becher, Moises Goldszmidt and Mirek Truszczynski for discussing these ideas with me. This research was supported by NSERC Research Grant OGP0121843.

References

Adams, E. W. 1975. *The Logic of Conditionals.* D.Reidel, Dordrecht.

Alchourron, C, Gardenfors, P., and Makinson, D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic,* 50:510-530.

Boutilier, C. 1991. Inaccessible worlds and irrelevance: Preliminary report. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence,* pages 413-418, Sydney.

Boutilier, C. 1992a. Conditional logics for default reasoning and belief revision. Technical Report KRR-TR-92-1, University of Toronto, Toronto. Ph.D. thesis.

Boutilier, C. 1992b. A logic for revision and subjunctive queries. In *Proceedings of the Tenth National Conference on Artificial Intelligence,* pages 609-615, San Jose.

Boutilier, C. 1992c. Normative, subjunctive and autoepistemic defaults: Adopting the Ramsey test. In *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning,* pages 685-696, Cambridge.

Boutilier, C. 1992d. Sequences of revisions: On the semantics of nested conditionals. Technical Report 92-24, University of British Columbia, Vancouver.

Boutilier, C. and Goldszmidt, M. 1993. Conditional belief revision: Model-based and syntactic approaches. In *Proceedings of the Eleventh National Conference on Artificial Intelligence,* Washington, D.C. (to appear).

Gardenfors, P. 1986. Belief revisions and the Ramsey test for conditionals. *The Philosophical Review,* 95:81-93.

Gardenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, Cambridge.

Ginsberg, M. L. 1986. Counterfactuak. *Artificial Intelligence,* 30(l):35-79.

Goldszmidt, M. and Pearl, J. 1992. Reasoning with qualitative probabilities can be tractable. In *Proceedings of the Eighth Conference on Uncertainty in AI,* pages 112-120, Stanford.

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

Katsuno, H. and Mendelzon, A. O. 1991. On the difference between updating a knowledge database and revising it. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning,* pages 387-394, Cambridge.

Levesque, H. J. 1990. All I know: A study in autoepistemic logic. *Artificial Intelligence,* 42:263-309.

Levi, I. 1988. Iteration of conditionals and the Ramsey test. *Synthese,* 76(1):49-81.

Lewis, D. 1973. *Counterfactuals.* Blackwell, Oxford.

Makinson, D. and Gardenfors, P. 1990. Relations between the logic of theory change and nonmonotonic logic. In Fuhrmann, A. and Morreau, M., editors, *The Logic of Theory Change,* pages 185-205. Springer-Verlag, Berlin.

Pearl, J. 1990. System Z: A natural ordering of defaults with tractable applications to default reasoning. In Vardi, M., editor, *Proceedings of Theoretical Aspects of Reasoning about Knowledge,* pages 121-135. Morgan Kaufmann, San Mateo.

Rott, H. 1989. Conditionals and theory change: Revisions, expansions, and additions. *Synthese,* 81(1):91—113.

Segerberg, K. 1970. Modal logics with linear alternative relations. *Theoria,* 36:310-322.

Stalnaker, R. C. 1968. A theory of conditionals. In Harper, W., Stalnaker, R., and Pearce, G., editors, *Ifs,* pages 41-55. D. Reidel, Dordrecht. 1981.

Winslett, M. 1988. Reasoning about action using a possible models approach. In *Proceedings of the Seventh National Conference on Artificial Intelligence,* pages 89-93, St. Paul.