Revision By Conditional Beliefs

Craig Boutilier
Univ. of British Columbia
Dept. of Computer Science
Vancouver, BC V6T 1Z2
CANADA
cebly@cs.ubc.ca

Moisés Goldszmidt
Rockwell International
444 High Street
Palo Alto, CA 94301
U.S.A.
moises@rpal.rockwell.com

Abstract

Both the dynamics of belief change and the process of reasoning by default can be based on the *conditional belief set* of an agent, represented as a set of "if-then" rules. In this paper we address the open problem of formalizing the dynamics of revising this conditional belief set by *new* if-then rules, be they interpreted as new default rules or new revision policies. We start by providing a purely semantic characterization, based on the semantics of conditional rules, which induces logical constraints on any such revision process. We then introduce logical (syntax-independent) and syntax-dependent techniques, and provide a precise characterization of the set of conditionals that hold after the revision. In addition to formalizing the dynamics of revising a default knowledge base, this work also provides some of the necessary formal tools for establishing the truth of nested conditionals, and attacking the problem of learning new defaults.

Introduction

Consider a child using a single default "typically birds fly", to predict the behavior of birds. Upon learning of the class of penguins and their exceptional nature she considers *revising* her current information about birds to include the information that penguins are birds yet "typically penguins do not fly". This process is different from that usually modeled in approaches to nonmonotonic reasoning and belief revision, where upon discovering that Tweety is a (nonflying) penguin she simply retracts her previous belief that Tweety does fly. Instead, the example above addresses the issue of revising the set of *conditional beliefs*, namely, the default rules that guide the revision of our *factual* beliefs. In this paper we are concerned with the dynamics of such conditional beliefs. Our objective is to characterize how the conditional information in a knowledge base evolves due to the incorporation of the new conditionals, which rules should be given up in case of inconsistency, and what principles guide this process.[1]

One well-known theory addressing the dynamics of factual beliefs is that proposed by Alchourron, Gärdenfors and Makinson (1985; 1988). The *AGM theory* takes epistemic states to be deductively closed sets of (believed) sentences and characterizes how a rational agent should change its set K of beliefs. This is achieved with postulates constraining revision functions $*$, where K^*_A represents the belief set that results when K is revised by A. Unfortunately, the AGM theory does not provide a calculus with which one can realize the revision process or even specify the content of an epistemic state (Boutilier 1992a; Doyle 1991; Nebel 1991). Recent work (Boutilier 1992a; Goldszmidt 1992) shows that AGM revision can be captured by assuming that an agent has a knowledge base (*KB*) containing *subjunctive conditionals* of the form $A \to B$ (where A and B are objective formulae). These conditionals define the agent's belief set and guide the revision process via the *Ramsey test* (Stalnaker 1968): $A \to B$ is accepted iff revision by A results in a belief in B. Such conditionals may be given a probabilistic interpretation (Goldszmidt 1992): each $A \to B$ is associated with a conditional probability statement arbitrarily close to one. They may also be interpreted a statements in a suitable modal logic (Boutilier 1992a). The corresponding logics (and indeed semantics) are identical (Boutilier 1992a), and furthermore there is a strong relation between these conditionals and conditional default rules (Boutilier 1992c; Goldszmidt and Pearl 1992a).

The AGM theory has two crucial limitations. First, the conditionals (or revision policies) associated with K, that determine the form of K^*_A, provide no guidance for determining the conditionals accepted in K^*_A itself. The theory only determines the new *factual* beliefs held after revision. Even if conditionals are contained in K, the AGM theory cannot suggest which conditionals should be retained or retracted in the construction of K^*_A. *Subsequent* revisions of K^*_A can thus be almost arbitrary. Second, the theory pro-

[1]We will not address the important question of why and when an agent decides to revise its conditional beliefs or defaults.

vides no mechanism for revising a belief set with new *conditionals*. Thus, the revision policies of an agent cannot, in general, be changed.[2] This paper provides a solution to this second problem, and extends our recent work on a solution to the first problem (Boutilier 1993; Goldszmidt and Pearl 1992b).

In this paper we focus on a particular model of *conditional revision* that extends propositional natural revision introduced by Boutilier (1993). The *natural revision* model addresses the problem of determining new conditional beliefs after revision by factual beliefs, and extends the notion of minimal change (characteristic of the AGM theory) to the conditional component of a *KB*. Thus, when a factual revision is applied to *KB*, the revised *KB*$'$ contains as much of the *conditional information* from *KB* as possible. The extension to conditional revision presented here preserves these properties and possesses the crucial property that the beliefs resulting from any sequence of (conditional or factual) updates can be determined using only properties of the original ranking, and tests involving simple (unnested) conditionals.[3]

A model for revising *KB* with new conditional belief (e.g., a rule $C \to D$) is crucial for a number of reasons. The problem of truth conditions for nested conditionals is subsumed by this more general problem. The semantics of conditionals with arbitrary nesting requires an account of revision by *new conditional information*. To test the truth of $(A \to B) \to C$, we must first revise *KB* by $A \to B$ and then test the status of C (Goldszmidt and Pearl 1992b). Also, it is clear that our beliefs do not merely change when we learn new factual information. We need a model that accounts for updating our belief set with new conditional probabilities and new subjunctive conditionals to guide the subsequent revision of beliefs. Given the strong equivalence between conditionals of the type described here and conditional default rules (Boutilier 1992c; Goldszmidt and Pearl 1992a), a model of conditional revision provides an account of updating a *KB* with new default rules. Any specification of how an agent is to learn new defaults must describe how an agent is to incorporate a new rule into its corpus of existing knowledge. Hence, the process we study in this paper is crucial for providing a semantic core for learning new default information.

We first review the basic concepts underlying belief revision. We then describe the basics of conditional belief revision by presenting a set of operations on ranked-models, and an important representation theorem. Finally, we explore a syntax-independent and a syntax-dependent approach to the conditional revision of a *KB*.

Propositional Natural Revision

In this section we briefly review a semantic account of belief revision (we refer the reader to (Gärdenfors 1988; Goldszmidt and Pearl 1992b; Boutilier 1992b) for details). We assume the existence of a deductively closed belief set K over a classical propositional language $\mathbf{L}_{CPL}$. Revising this belief set with a new proposition A is problematic when $K \models \neg A$, for simply adding the belief A will cause inconsistency. To accommodate A, certain beliefs must be given up before A is added. The AGM theory of revision provides a set of constraints on *revision functions* $*$ that map belief sets K into revised belief sets K^*_A. Any theory of revision also provides a theory of conditionals if we adopt the *Ramsey test*. This test states that one should accept the conditional "If A then B" just when $B \in K^*_A$.

A key representation result for this theory shows that changes can be modeled by assuming an agent has an ordering of *epistemic entrenchment* over beliefs: revision always retains more entrenched propositions in preference to less entrenched ones. Grove (1988) shows that entrenchment can be modeled semantically by an ordering of worlds. This is pursued by Boutilier (1992b) who presents a modal logic and semantics for revision. A *revision model* $M = \langle W, \leq, \varphi \rangle$ consists of a set of worlds W (assigned valuations by φ) and an *plausibility* ordering $\leq$ over W. If $v \leq w$ then v is at least as plausible as w. We insist that $\leq$ be transitive and connected (so $w \leq v$ or $v \leq w$ for all v, w). We denote by $\|A\|$ the set of worlds in M satisfying A (those w such that $M \models_w A$). We define the set of most plausible A-worlds to be those worlds in A minimal in $\leq$; so *min*(M, A) is just

$$\{w \in W : M \models_w A, \text{ and } M \models_v A \text{ implies } w \leq v\}$$

We assume that all models are *smooth* in the sense that *min*$(M, A) \neq \emptyset$ for all (satisfiable) $A \in \mathbf{L}_{CPL}$.[4] The *objective belief set* K of a model M is the set of $\alpha \in \mathbf{L}_{CPL}$ such that *min*$(M, \top) \subseteq \|\alpha\|$ (those α true at each *most* plausible world). Such α are believed by the agent. These objective or *factual* beliefs capture the agent's judgements of true facts in the world. They should be contrasted with the conditional beliefs of an agent, described below.

To capture the revision of a belief set K, we define a K-revision model to be any revision model such that

[2] Surprisingly, these two issues have remained largely unexplored, due largely to the Gärdenfors (1988) triviality result, which points to difficulties with the interpretation of conditional belief sets. But these can be easily circumvented (Boutilier 1992c).

[3] A second method of revision is the model of *J-conditioning* (Goldszmidt and Pearl 1992b): when *KB* is updated with a new fact A, the revised *KB*$'$ is determined by Bayesian conditionalization, giving rise to a qualitative abstraction of probability theory (Adams 1975; Goldszmidt 1992). This mechanism preserves the (qualitative) conditional probabilities in *KB* as much as possible and thus guarantees that the relative strength of the conditionals also remains constant. The extension of J-conditionalization to the conditional revision case is explored in the full version of the paper (Boutilier and Goldszmidt 1993).

[4] Hence, there exist *most* plausible A-worlds. This is not required, but the assumption does not affect the equivalence below.

min$(M, \top) = \|K\|$. That is, all and only those worlds satisfying the belief set are most plausible. When we revise K by A, we must end up with a new belief set that includes A. Given our ordering, we simply require that the new belief set correspond to the set of most plausible A-worlds. We can define the truth conditions for a conditional connective as

$$M \models_w A \to B \quad \text{iff} \quad \textit{min}(M, A) \subseteq \|B\| \qquad (1)$$

Such *conditional beliefs* characterize the revision policies, hypothetical beliefs or defaults of an agent. Equating $A \to B$ with $B \in K^*_A$, this definition of revision characterizes the same space of revision functions as the AGM theory (Boutilier 1992b).

The AGM theory and the semantics above show how one might determine a new objective belief set K^*_A from a given K-revision model; but it provides no hint as to what new *conditionals* should be held. To do so requires that a new revision model, suitable for K^*_A, be specified. *Natural revision*, proposed by Boutilier (1993), does just this. Given a K-revision model M, natural revision specifies a new model M^*_A suitable for the revision of K^*_A (i.e., a K^*_A-revision model). Roughly, this model can be constructed by "shifting" the set *min*(M, A) to the bottom of the ordering, leaving all other worlds in the same relative relation. This extends the notion of minimal change to the relative plausibility of worlds. To believe A, certainly K^*_A-worlds must become most plausible, but nothing else need change (Boutilier 1993). Hence, natural revision constructs a new ranking to reflect new objective beliefs. With such a ranking one can then determine the behavior of subsequent objective revisions. But no existing model of revision accounts for revision of a ranking to include new conditionals. In the next section we extend natural revision so that new conditional information can be incorporated explicitly in a model.

Conditional Belief Revision: Revising a Model

Given a revision model M, we want to define a new model $M^*_{A\to B}$ that satisfies $A \to B$ but changes the plausibility ordering in M as little as possible. We do this in two stages: first, we define the *contraction* of M so that the "negation" $A \to \neg B$ is not satisfied; then we define the *expansion* of this new model to accommodate the conditional $A \to B$. Let $M = \langle W, \leq, \varphi \rangle$.

Definition 1 The *natural contraction operator* $-$ maps M into $M^-_{A\to B}$, for any simple conditional $A \to B$, where $M^-_{A\to B} = \langle W, \leq', \varphi \rangle$, and:

1. if $v, w \notin \textit{min}(M, A \wedge \neg B)$ then $v \leq' w$ iff $v \leq w$
2. if $w \in \textit{min}(M, A \wedge \neg B)$ then: (a) $w \leq' v$ iff $u \leq v$ for some $u \in \textit{min}(M, A)$; and (b) $v \leq' w$ iff $v \leq u$ for some $u \in \textit{min}(M, A)$

Figure 1 illustrates this process in the principle case, showing how the model $M^-_{A\to B}$ is constructed when $M \models A \to B$. Clearly, to "forget" $A \to B$ we must construct a model where certain minimal A-worlds do not satisfy B. If M satisfies $A \to B$, we must ensure that certain $A \wedge \neg B$-worlds become at least as plausible as the minimal A-worlds, thus ensuring that $A \to B$ is no longer satisfied. Natural contraction does this by making the most plausible $A \wedge \neg B$-worlds just as plausible as the most plausible A-worlds. Simply put, the minimal $A \wedge \neg B$-worlds (the light-shaded region) are shifted to the cluster containing the minimal A-worlds (the dark-shaded region). We have the following properties:[5]

Proposition 1 *Let M be a revision model.*
(1) $M^-_{A\to B} \not\models A \to B$;
(2) If $M \not\models A \to B$ then $M^-_{A\to B} = M$; and
(3) If $M \not\models A \to \neg B$ then $M^-_{A\to B} \models A \not\to B \wedge A \not\to \neg B$.

Theorem 2 *Let M^-_A denote the natural propositional contraction of M by (objective belief) A (as defined in (Boutilier 1993)). Then $M^-_A = M^-_{\top\to A}$.*

Thus, propositional contraction is a special case of conditional contraction.

We define the *expansion* of M by $A \to B$ to be the model $M^+_{A\to B}$ constructed by making the minimal changes to M required to accept $A \to B$. While we do not require that $M \not\models A \to \neg B$ in the following definition, we will only use this definition of expansion for such models.

Definition 2 The *natural expansion operator* $+$ maps M into $M^+_{A\to B}$, for any simple conditional $A \to B$, where $M^+_{A\to B} = \langle W, \leq', \varphi \rangle$, and:

1. if $v \notin \textit{min}(M, A \wedge \neg B)$ then $w \leq' v$ iff $w \leq v$
2. if $v \in \textit{min}(M, A \wedge \neg B)$ then:
   (a) if $w \in \textit{min}(M, A \wedge \neg B)$ then $w \leq' v$; and
   (b) if $w \notin \textit{min}(M, A \wedge \neg B)$ then $w \leq' v$ iff $w \leq v$ and there is no $u \in \textit{min}(M, A \wedge B)$ such that $u \leq w$

Figure 1 illustrates this process in the principle case, showing how the model $M^+_{A\to B}$ is constructed when $M \not\models A \to B$. Clearly, to believe $A \to B$ we must construct a model where all minimal A-worlds satisfy B. If M fails to satisfy $A \to B$, we must ensure that the minimal $A \wedge \neg B$-worlds become less plausible than the minimal A-worlds, thus ensuring that $A \to B$ is satisfied. Natural expansion does this by making the most plausible $A \wedge \neg B$-worlds (the dark-shaded region) less plausible than the most plausible A-worlds (the light-shaded region). This leaves us with $A \to B$, but preserves the relative plausibility ranking of all other worlds. In particular, while the set of minimal $A \wedge \neg B$-worlds becomes less plausible than those worlds with which it shared equal plausibility in M, its relationship to more or less plausible

[5] We let $\alpha \not\to \beta$ stand for $\neg(\alpha \to \beta)$.

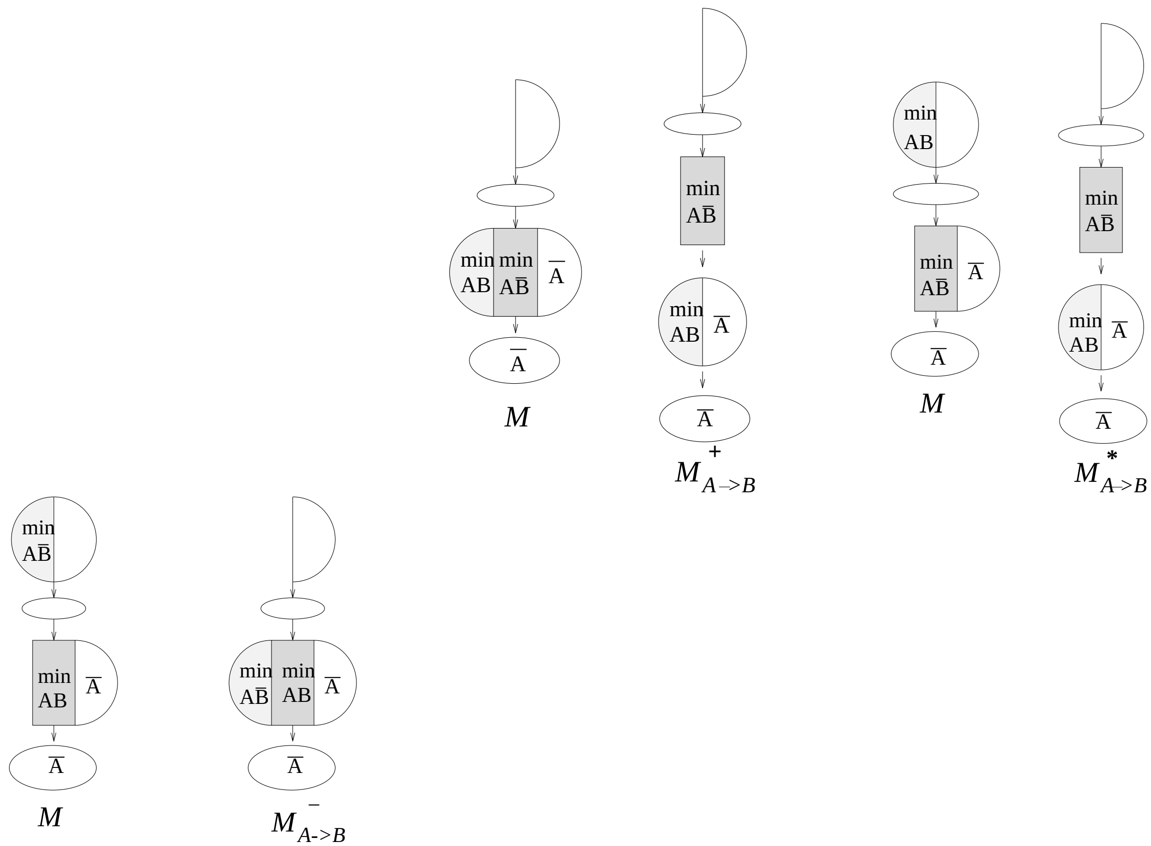

Figure 1: Contraction, Expansion and Revision of a Model

worlds is unchanged. Once again, the idea is that the conditional belief set induced by A should contain only B-worlds, but that all other conditionals should remain unchanged to the greatest extent possible.

Proposition 3 *Let M be a revision model such that $M \not\models A \to \neg B$. (1) $M^+_{A\to B} \models A \to B$; and (2) If $M \models A \to B$ then $M^+_{A\to B} = M$.*

We can now define revision by a conditional $A \to B$. Briefly, to accept such a conditional we first "forget" $A \to \neg B$ and then "add" $A \to B$.

Definition 3 The *natural revision operator* $*$ maps M into $M^*_{A\to B}$, for any simple conditional $A \to B$, where $M^*_{A\to B} = (M^-_{A\to\neg B})^+_{A\to B}$.

This definition of revision reflects the Levi identity (Levi 1980). Figure 1 illustrates this process in the principle case, showing how the model $M^*_{A\to B}$ is constructed when $M \models A \to \neg B$. Natural revision behaves as expected:

Proposition 4 *Let M be a revision model. (1) $M^*_{A\to B} \models A \to B$; and (2) If $M \models A \to B$ then $M^*_{A\to B} = M$.*

Theorem 5 *Let M^*_A denote the natural propositional revision of M by (objective belief) A (as defined in (Boutilier 1993)). Then $M^*_A = M^*_{\top\to A}$.*

Thus, we can view propositional revision as a special case of conditional revision. We will henceforth take M^*_A as an abbreviation for $M^*_{\top\to A}$.

These results show that natural conditional revision can reasonably be called a revision operator. To show that this revision operator is indeed "natural," we must determine its precise effect on belief in previously accepted conditionals. In particular, we would like a precise characterization the simple conditionals $\alpha \to \beta$ satisfied by the revised model $M^*_{A\to B}$. The following result (Thm. 6) shows that the truth of such conditionals in $M^*_{A\to B}$ is completely determined by the set of simple conditionals satisfied by M. Thus, the truth of an arbitrarily nested conditional under natural revision can be determined by the truth of simple conditionals in our original model. We note that revision models are *complete* in that they satisfy every simple conditional or its negation. We do not require that a conditional *KB* be complete in this sense. We describe how this semantic model can be applied to an incomplete *KB* in the next section.

We now show which conditionals are satisfied by a model $M^*_{A\to B}$. In (Boutilier and Goldszmidt 1993) we also describe similar characterizations of models $M^-_{A\to B}$ and $M^+_{A\to B}$. We begin by noting that if $M \not\models A \to \neg B$, then $M^*_{A\to B} = M^+_{A\to B}$. In particular, if $M \models A \to B$, then $M^*_{A\to B} = M$ and no conditional beliefs are changed. We assume then that $M \models A \to \neg B$, the principle case of revision. We also introduce the notion of *plausibility*: a sentence P is at least as plausible as Q (relative to M) iff the minimal P-worlds are at least as plausible (in the ordering $\leq$) as the minimal Q-worlds. This will be the case exactly when $M \models (P \vee Q) \not\to \neg P$. We write $P <_P Q$ if formula P is more plausible that Q, and $P =_P Q$ if P and Q are equally plausible.[6] To determine whether $\alpha \to \beta$ holds in $M^*_{A\to B}$, we simply need to know how the relative position of worlds in $\|\alpha\|$ is affected by the revision. The relative plausibility of A and α in M is crucial in determining this. If α is more plausible than A, then shifting $A \wedge B$-worlds down cannot affect the most plausible α-worlds. If α is less plausible, the most plausible α-worlds *might* change, but only if there are α-worlds among the most plausible $A \wedge B$-worlds. Finally, there are several different types of changes that can occur if α and A are equally plausible.

Theorem 6 *Let $M \models A \to \neg B$ and let $\leq_P$ be the plausibility ordering determined by M. Let $\alpha, \beta \in \mathbf{L}_{CPL}$.*

1. *If $\alpha <_P A$ then $M^*_{A\to B} \models \alpha \to \beta$ iff $M \models \alpha \to \beta$.*
2. *If $\alpha >_P A$ then*
   - *(a) If $M \models A \wedge B \to \neg\alpha$ then $M^*_{A\to B} \models \alpha \to \beta$ iff $M \models \alpha \to \beta$.*
   - *(b) If $M \models A \wedge B \not\to \neg\alpha$ then $M^*_{A\to B} \models \alpha \to \beta$ iff $M \models A \wedge B \wedge \alpha \to \beta$.*
3. *If $\alpha =_P A$ then*
   - *(a) If $M \models A \wedge B \not\to \neg\alpha$ and $M \models \alpha \to A$ then $M^*_{A\to B} \models \alpha \to \beta$ iff $M \models A \wedge B \wedge \alpha \to \beta$.*
   - *(b) If $M \models A \wedge B \not\to \neg\alpha$ and $M \models \alpha \not\to A$ then $M^*_{A\to B} \models \alpha \to \beta$ iff $M \models A \wedge B \wedge \alpha \to \beta$ and $M \models \alpha \wedge \neg A \to \beta$.*
   - *(c) If $M \models A \wedge B \to \neg\alpha$ and $M \models A \wedge \neg B \to \neg\alpha$ then $M^*_{A\to B} \models \alpha \to \beta$ iff $M \models \alpha \to \beta$.*
   - *(d) If $M \models A \wedge B \to \neg\alpha$, $M \models A \wedge \neg B \not\to \neg\alpha$ and $M \models \alpha \to A$ then $M^*_{A\to B} \models \alpha \to \beta$ iff $M \models \alpha \to \beta$.*
   - *(e) If $M \models A \wedge B \to \neg\alpha$, $M \models A \wedge \neg B \not\to \neg\alpha$ and $M \models \alpha \not\to A$ then $M^*_{A\to B} \models \alpha \to \beta$ iff $M \models \alpha \wedge \neg A \to \beta$.*

While this characterization results appears complex, it is rather intuitive, for it captures the interactions caused by the relative plausibility of A and other propositions α. As an example, suppose we believe that a power surge will normally cause a breaker to trip ($S \to B$) and this will prevent equipment damage ($S \to \neg D$); but if the breaker doesn't trip there will be damage ($S \wedge \neg B \to D$). Our characterization shows that, should we learn the breaker is faulty ($S \to \neg B$), we should also change our mind about potential damage, and thus accept $S \to D$. However, information such as $\top \to \neg S$ will continue to be held (the likelihood of a power surge does not change). Hence, our factual beliefs (e.g., $\neg S$) do not change, merely our conditional belief about the breaker: what will happen *if* S.

[6]Plausibility is also induced by the κ-ranking of formulae (Goldszmidt and Pearl 1992b): $P \leq_P Q$ iff $\kappa(P) \leq \kappa(Q)$.

Theorem 6 also shows that the conditionals that hold in the revised model $M^*_{A\to B}$ can be completely characterized in terms of the conditionals in M. This allows us to use the mechanisms and algorithms of Goldszmidt and Pearl (1992b) for computing the new model (Boutilier and Goldszmidt 1993). This also demonstrates that an arbitrary nested conditional sentence (under natural revision) is logically equivalent to a sentence without nesting (involving disjunctions of conditionals). Thus, purely propositional reasoning mechanisms (Pearl 1990) can be used to determine the truth of nested conditionals in a conditional *KB*. Indeed, in many circumstances, a complete semantic model can be represented compactly and reasoned about tractably (Goldszmidt and Pearl 1992b). We explore this in the full paper.

When we revise by $A \to B$ we are indicating a willingness to accept B should we come to accept A. Thus, we might expect that revising by $A \to B$ should somehow reflect propositional revision by B were we to restrict our attention to A-worlds. This is indeed the case. Let $M\backslash\alpha$ denote the model obtained by eliminating all α-worlds from M.

Theorem 7 $(M^*_{A\to B})\backslash\neg A = (M\backslash\neg A)^*_B$

This shows that accepting $A \to B$ is equivalent to accepting B "given" A. Thus, natural revision by conditionals is in fact a conditional form of the propositional natural revision of Boutilier (1993). The only reason the characterization theorem for conditional revision is more complex is the fact that we can "coalesce" partial clusters of worlds, something that can't be done in the propositional case. We also note that $(M^*_{A\to B})\backslash A = M\backslash A$; that is, the relative plausibility of $\neg A$-worlds is unaffected by this revision.

Revising a Conditional Knowledge Base

If a conditional *KB* contains a complete set of simple conditionals (i.e., defines a unique revision model) we can use the definitions above to compute the revised *KB*. Often we may use techniques to complete a *KB* as well (Pearl 1990). In practice, however, *KB* will usually be an incomplete set of premises or constraints. We propose the following method of *logical revision*. Since *KB* is not complete, it is satisfied by each of a set $\|KB\|$ of revision models, each of these a "possible" ranking for the agent. When a new conditional $A \to B$ is learned, revision proceeds in the following way. If there are elements of $\|KB\|$ that satisfy $A \to B$, these become the new possible rankings for the agent.[7] In this case we have $KB^*_{A\to B} \equiv KB \cup \{A \to B\}$. If this is not the case, each possibility in $\|KB\|$ must be rejected. To do this, we revise each ranking in $\|KB\|$ and consider the result of this revision to be the set of new possibilities. $KB^*_{A\to B}$ is then

$$\{C \to D : M^*_{A\to B} \models C \to D \text{ for all } M \in \|KB\|\}$$

The breaker example above exemplifies this approach. Clearly, we do not want to resort to generating all models of *KB*. Fortunately, our representation theorem allows us to use any logical calculus for simple conditionals alone to determine the set of all such consequences. A simple conditional $\alpha \to \beta$ will be in the logical revision of *KB* iff the appropriate set of simple conditionals (from Theorem 6) is derivable from *KB* (e.g., one may use the calculus of (Boutilier 1992b; Goldszmidt and Pearl 1991)).

The main problem with an approach based on logical revision is that it is extremely cautious. A direct consequence of this cautious behavior is that the syntactic structure of *KB* is lost: it plays no role in the revision process.[8] For instance, the revisions of either of $\{A \to B, A \to C\}$ or $\{A \to B \wedge C\}$ by $A \to \neg B$ are identical. Yet, in some cases, conditional revision of the first set should yield a *KB* equivalent to $\{A \to \neg B, A \to C\}$ simply because $A \to \neg B$ conflicts only with $A \to B$. Yet logical revision forces into consideration models in which $A \to C$ is given up as well. This is not unreasonable, in general,[9] but the syntactic structure may also be used in revision.

The strategy we propose isolates the portion of *KB* inconsistent with the new rule $A \to B$, which will be denoted by KB_I, and then applies logical revision to KB_I alone. Letting $KB_J = KB - KB_I$, the revised set $KB^*_{A\to B}$ is the union of KB_J and the logically revised KB_I (with $A \to B$). We first introduce the notion of consistency:

Definition 4 A set *KB* is *consistent* iff there exists at least one model M such that, for each $A \to B \in KB$, $min(M, A) \subseteq \|B\|$ and $min(M, A) \neq \emptyset$.

A conditional $A \to B$ is *tolerated* by the set $\{C_i \to D_i\}$, $1 \le i \le n$ iff the propositional formula $A \wedge B \bigwedge_{i=1}^{i=n}\{C_i \supset D_i\}$ is satisfiable. The notion of toleration constitutes the basis for isolating the inconsistent set of *KB*. A set containing a rule tolerated by that set will be called a *confirmable* set. The following theorem presents necessary and sufficient conditions for consistency (Goldszmidt and Pearl 1991):

Theorem 8 *KB is consistent iff every nonempty subset $KB' \subseteq KB$ is confirmable.*

Given *KB*, a subset KB_m is *minimally unconfirmable* iff KB_m is unconfirmable, but every nonempty subset $KB'_m \subseteq$

[7] This is is equivalent to asking if $KB \cup \{A \to B\}$ is consistent (see Def. 4 and Thm. 8).

[8] E.g., Nebel (1991) has advocated syntax-dependent revision.

[9] Indeed, this is exactly analogous to the generality of the AGM theory. Given $K = Cn\{A, B\}$, it is not known whether $B \in K^*_{\neg A}$ or not. Logically, the possibility of a connection between A and B exists, and should be denied or stated (or assumed) explicitly.

KB$_m$ is confirmable.[10] Finally, a set *KB*$_I$ is a *minimal complete inconsistent set* (MCI) with respect to *KB* iff it is the union of all minimally unconfirmable subsets for *KB*. Thus, *KB*$_I$ contains only the conditionals in *KB* that are responsible for the inconsistencies in *KB*. In a syntax-directed revision of *KB* we are primarily interested in uncovering the conditionals in the original *KB* that are still valid after the revision process. The S operator below serves this purpose (note that S is built on top of a logical revision process). Given a set *KB*, and a simple conditional $A \to B$, let $S(KB, A \to B)$ denote the set of conditionals $C \to D$ such that: (1) $C \to D \in KB$ and (2) $KB^L_{A\to B} \models C \to D$, where $KB^L_{A\to B}$ denotes the logical revision of *KB* by $A \to B$. We define the syntactic revision of *KB* by $A \to B$ as follows:

Definition 5 Let *KB* be consistent, and let $A \to B$ be a simple conditional. Let KB'_I denote the MCI of $KB \cup \{A \to B\}$, $KB_I = KB'_I - \{A \to B\}$, and $KB_J = KB - KB_I$. The syntactic revision of *KB* by $A \to B$, written $KB^*_{A\to B}$, will be $KB^*_{A\to B} = S(KB_I, A \to B) \cup KB_J \cup \{A \to B\}$.

Note that in the case where $A \to B$ is consistent with respect to *KB*, $KB^*_{A\to B}$ will be simply the union of the original *KB* and the new conditional $A \to B$. Also, the syntactic revision of $\{A \to B, A \to C\}$ by $A \to \neg B$ will be the set $\{A \to \neg B, A \to C\}$ since $KB_I = \{A \to B\}$. In the breaker example above, the revision of $\{S \to B, S \wedge \neg B \to D, S \to \neg D\}$ by $S \to \neg B$ will yield $\{S \to \neg B, S \wedge \neg B \to D\}$ which entails the conditional $S \to D$ (as in the case of logical revision). Given that the revision of KB^* is based on Theorem 6 and notions of propositional satisfiability (i.e., toleration), the resulting set of conditionals can be computed effectively. The major problem in terms of complexity is the uncovering of the MCI set KB_I which seems to require an exponential number of satisfiability tests.

Concluding Remarks

We have provided a semantics for revising a conditional *KB* with new conditional beliefs in a manner that extends both the AGM theory and the propositional natural revision model. Our results include a characterization theorem, providing computationally effective means of deciding whether a given conditional holds in the revised model. We have also provided a syntactic characterization for the revision of a *KB*. We remark that, as in the case of proposals for objective belief revision (including the AGM theory), we make no claims or assumptions about the complex process by which an agent decides to incorporate a new conditional belief (or default rule) into its corpus of knowledge. We merely provide the formal means to do so.

Conditional belief revision defines a semantics for arbitrary nested conditionals as proposed in (Goldszmidt and Pearl 1992b), extending the semantics for right-nested conditionals studied in (Boutilier 1993). By describing the process by which an agent can assimilate new information in the form of conditionals, conditional belief revision is proposed as a basis for the learning of new default rules.

We note that the same techniques can be used to model revision by conditionals in a way that respects the probabilistic intuitions of J-conditioning. Analogues of each of the main results for natural revision are shown in the full paper (Boutilier and Goldszmidt 1993). We also explore other mechanisms for revising a *KB* and the relationship of our models to probabilistic conditionalization and imaging. We discuss further constraints on the revision process to reflect a causal interpretation of the conditional sentences.

Acknowledgements: Thanks to Judea Pearl for helpful comments. This research was supported by NSERC Grant OGP0121843 and by Rockwell International Science Center.

[10] If *KB* is consistent, then KB_m is the empty set.

References

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

Alchourrón, C., Gärdenfors, 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. 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 *Proc. of AAAI-92*, pages 609–615, San Jose.

Boutilier, C. 1992c. Normative, subjunctive and autoepistemic defaults: Adopting the Ramsey test. In *Proc. of KR-92*, pages 685–696, Cambridge.

Boutilier, C. 1993. Revision sequences and nested conditionals. In *Proc. of IJCAI-93*, pages 519–525, Chambery, FR.

Boutilier, C. and Goldszmidt, M. 1993. Revising by conditionals. Technical report, University of British Columbia, Vancouver. (Forthcoming).

Doyle, J. 1991. Rational belief revision: Preliminary report. In *Proc. of KR-91*, pages 163–174, Cambridge.

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

Goldszmidt, M. 1992. Qualitative probabilities: A normative framework for commonsense reasoning. Technical Report R-190, University of California, Los Angeles. Ph.D. thesis.

Goldszmidt, M. and Pearl, J. 1991. On the consistency of defeasible databases. *Artificial Intelligence*, 52:121–149.

Goldszmidt, M. and Pearl, J. 1992a. Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions. In *Proc. of KR-92*, pages 661–672, Cambridge.

Goldszmidt, M. and Pearl, J. 1992b. 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.

Levi, I. 1980. *The Enterprise of Knowledge*. MIT Press, Cambridge.

Nebel, B. 1991. Belief revision and default reasoning: Syntax-based approaches. In *Proc. of KR-91*, pages 417–428, Cambridge.

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.

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.