A Logic for Revision and Subjunctive Queries

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

Abstract

We present a logic for belief revision in which revision of a theory by a sentence is represented using a conditional connective. The conditional is not primitive, but rather defined using two unary modal operators. Our approach captures and extends the classic AGM model without relying on the Limit Assumption. Reasoning about counterfactual or hypothetical situations is also crucial for AI. Existing logics for such subjunctive queries are lacking in several respects, however, primarily in failing to make explicit the epistemic nature of such queries. We present a logical model for subjunctives based on our logic of revision that appeals explicitly to the Ramsey test. We discuss a framework for answering subjunctive queries, and show how integrity constraints on the revision process can be expressed.

Introduction

An important and well-studied problem in philosophical logic, artificial intelligence and database theory is that of modeling theory change or belief revision. That is, given a knowledge base KB, we want to characterize semantically the set that results after learning a new fact α. However, the question of how to revise KB is important not just in the case of changing information or mistaken premises, but also when we want to investigate questions of the form "What if A were true?" A subjunctive conditional $A > B$ is one of the form[1] "If A were the case then B would be true." Subjunctives have been widely studied in philosophy and it is generally accepted that (some variant of) the *Ramsey test* is adequate for evaluating the truth of such conditionals:

> 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. (Stalnaker 1968, p.44)

The connection to belief revision is quite clearly spelled out in this formulation of the Ramsey test: to evaluate a subjunctive conditional $A > B$, we revise our beliefs to include A and see if B is believed. If we take KB to represent some initial state of knowledge, a characterization of subjunctive reasoning must include an account of how to revise KB with new information.

In this paper, we will develop a logic for belief revision using a conditional connective $\stackrel{\text{KB}}{\longrightarrow}$, where $A \stackrel{\text{KB}}{\longrightarrow} B$ is interpreted roughly as "If KB were revised by A, then B would be believed." The connective will not be primitive however; instead it is defined in terms of two unary modal operators, which refer to truth at accessible and *inaccessible* worlds. Our model of revision will satisfy the classic AGM postulates and will be general enough to represent any AGM revision function. However, our approach is based on a very expressive logical calculus rather than extra-logical postulates, and can be used to express natural constraints on the revision process. Furthermore, this is accomplished without reliance on the Limit Assumption. We will then use this logic to develop a framework in which subjunctive queries of a knowledge base can be answered, and show that it improves on existing subjunctive logics and systems in several crucial respects. Finally, we provide a semantic characterization of integrity constraints suitable for this type of subjunctive reasoning.

Revision and the AGM Postulates Recently, work on the logic of theory change has been adopted by the AI community for use in the task of belief revision. By far the most influential approach to revision has been that of Alchourrón, Gärdenfors and Makinson (1985; 1988), which we refer to as the AGM theory of revision. We assume beliefs sets to be deductively closed sets of sentences, and for concreteness we will assume that the underlying logic of beliefs is classical

[1] At least, in "deep structure."

propositional, CPL. We let $\models$ and Cn denote classical entailment and consequence, respectively, and use K to denote arbitrary belief sets. If $K = Cn(KB)$ for some finite set of sentences KB, we say K is *finitely specified* by KB.

Revising a belief set K is required when new information must be accommodated with these beliefs. If $K \not\models \neg A$, learning A is relatively unproblematic as the new belief set $Cn(K \cup \{A\})$ seems adequate for modeling this change. This process is known as *expansion* and the expanded belief set is denoted K_A^+. More troublesome is the revision of K by A when $K \models \neg A$. Some beliefs in K must be given up before A can be accommodated. The problem is in determining which part of K to give up, as there are a multitude of choices. Furthermore, in general, there are no logical grounds for choosing which of these alternative revisions is acceptable (Stalnaker 1984), the issue depending largely on context.

Fortunately, there are some logical criteria for reducing this set of possibilities, the main criterion for preferring certain choices being that of *minimal change*. Informational economy dictates that as "few" beliefs as possible from K be discarded to facilitate belief in A (Gärdenfors 1988), where by "few" we intend that, as much as possible, the informational content of K be kept intact. While pragmatic considerations will often enter into these deliberations, the main emphasis of the work of AGM is in logically delimiting the scope of acceptable revisions. To this end, the AGM postulates, given below, are maintained to hold for any reasonable notion of revision (Gärdenfors 1988). We use K_A^* to denote the belief set that results from the revision of K by A and $\perp$ to denote falsity.

(R1) K_A^* is a belief set.

(R2) $A \in K_A^*$.

(R3) $K_A^* \subseteq K_A^+$.

(R4) If $\neg A \notin K$ then $K_A^+ \subseteq K_A^*$.

(R5) $K_A^* = Cn(\perp)$ iff $\models \neg A$.

(R6) If $\models A \equiv B$ then $K_A^* = K_B^*$.

(R7) $K_{A \wedge B}^* \subseteq (K_A^*)_B^+$.

(R8) If $\neg B \notin K_A^*$ then $(K_A^*)_B^+ \subseteq K_{A \wedge B}^*$.

Of particular interest are (R3) and (R4), which taken together assert that if A is consistent with K then K_A^* should merely be the expansion of K by A. This seems to reflect our intuitions about informational economy, that beliefs should not be given up gratuitously.

Revision and Subjunctive Conditionals Counterfactuals and subjunctives have received a great deal of attention in the philosophical literature, one classic work being that of Lewis (1973). A number of people have argued that these conditionals have an important role to play in AI, logic programming and database theory. Bonner (1988) has proposed a logic for hypothetical reasoning in which logic programs or deductive databases are augmented with hypothetical implications. Ginsberg (1986) has identified a number of areas in AI in which counterfactuals may play an important role in the semantic analysis of various tasks (e.g., planning, diagnosis). He proposes a system for reasoning about counterfactuals based on the ideas of Lewis. Unfortunately, this model suffers form certain shortcomings, including a sensitivity to the syntactic structure of KB. Jackson (1989) considers the problems with this approach and presents a model-theoretic system for counterfactual reasoning based on the possible models approach to update of Winslett (1990). Again, this system is extra-logical in nature, and is committed to specific minimality criteria.

The systems of Ginsberg and Jackson both take very seriously the idea that counterfactuals are intimately tied to belief revision. However, this connection had not gone unappreciated by the revision community. Gärdenfors (1988) provides an explicit postulate for revision and conditional reasoning based on the Ramsey test. If we assume that conditionals can be part of our belief sets, a concise statement of the Ramsey test is

(RT) $A > B \in K$ iff $B \in K_A^*$.

Gärdenfors also describes a formal semantics for conditionals. Variants of postulates (R1) through (R8), together with (RT), determine a conditional logic based on a "revision style" semantics that corresponds exactly to Lewis's (1973) counterfactual logic VC.

A Conditional for Revision

The Modal Logic CO* We now present a modal logic for revision in which we define a conditional connective $\xrightarrow{\text{KB}}$. $A \xrightarrow{\text{KB}} B$ is read "If (implicit theory) KB is revised by A, then B will be believed." The modal logic CO is based on a standard propositional modal language (over variables $\mathbf{P}$) augmented with an additional modal operator $\overleftarrow{\Box}$. The sentence $\overleftarrow{\Box}\alpha$ is read "α is true at all *inaccessible* worlds" (in contrast to the usual $\Box\alpha$ that refers to truth at accessible worlds). A CO-model is a triple $M = \langle W, R, \varphi \rangle$, where W is a set of worlds with valuation φ and R is an accessibility relation over W. We insist that R be transitive and

connected.[2] Satisfaction is defined in the usual way, with the truth of a modal formula at a world defined as:

1. $M \models_w \Box\alpha$ iff for each v such that wRv, $M \models_v \alpha$.

2. $M \models_w \overleftarrow{\Box}\alpha$ iff for each v such that not wRv, $M \models_v \alpha$.

We define several new connectives as follows: $\Diamond\alpha \equiv_{\rm df} \neg\Box\neg\alpha$; $\overleftarrow{\Diamond}\alpha \equiv_{\rm df} \neg\overleftarrow{\Box}\neg\alpha$; $\overleftrightarrow{\Box}\alpha \equiv_{\rm df} \Box\alpha \wedge \overleftarrow{\Box}\alpha$; and $\overleftrightarrow{\Diamond}\alpha \equiv_{\rm df} \Diamond\alpha \vee \overleftarrow{\Diamond}\alpha$. It is easy to verify that these connectives have the following truth conditions: $\Diamond\alpha$ ($\overleftarrow{\Diamond}\alpha$) is true at a world if α holds at some accessible (inaccessible) world; $\overleftrightarrow{\Box}\alpha$ ($\overleftrightarrow{\Diamond}\alpha$) holds iff α holds at all (some) worlds. The following set of axioms and rules is complete for CO (Boutilier 1991):

K $\Box(A \supset B) \supset (\Box A \supset \Box B)$

K′ $\overleftarrow{\Box}(A \supset B) \supset (\overleftarrow{\Box}A \supset \overleftarrow{\Box}B)$

T $\Box A \supset A$

4 $\Box A \supset \Box\Box A$

S $A \supset \overleftarrow{\Box}\Diamond A$

H $\overleftrightarrow{\Diamond}(\Box A \wedge \overleftarrow{\Box}B) \supset \overleftrightarrow{\Box}(A \vee B)$

Nes From A infer $\overleftrightarrow{\Box}A$.

MP From $A \supset B$ and A infer B.

For the purposes of revision, we consider the extension of CO based on the class of CO-models in which all propositional valuations are represented in W; that is, $\{f : f \text{ maps } \mathbf{P} \text{ into } \{0,1\}\} \subseteq \{w^* : w \in W\}$.[3] The logic CO*, complete for this class of structures, is the smallest extension of CO containing instances of the following schema:

LP $\overleftrightarrow{\Diamond}\alpha$ for all satisfiable propositional α.

We note that CO*-structures consist of a totally-ordered set of *clusters* of mutually accessible worlds.

[2] R is (totally) connected if wRv or vRw for any $v, w \in W$ (this implies reflexivity). CO was first presented in (Boutilier 1991) to handle the problem of irrelevance in default reasoning.

[3] For all $w \in W$, w^* is defined as the map from $\mathbf{P}$ into $\{0,1\}$ such that $w^*(A) = 1$ iff $w \in \varphi(A)$; in other words, w^* is the valuation associated with w.

Revision as a Conditional A key observation of Grove (1988) is that revision can be viewed as an ordering on possible worlds reflecting an agent's preference on epistemically possible states of affairs. We take this as a starting point for our semantics, based on structures consisting of a set of possible worlds W and a binary accessibility relation R over W. Implicit in any such structure for revision will be some theory of interest or belief set K that is intended as the object of revision. We return momentarily to the problem of specifying K within the structure. The interpretation of R is as follows: wRv iff v is as *plausible* as w given theory K. As usual, v is *more plausible* than w iff wRv but not vRw. Plausibility is a pragmatic measure that reflects the degree to which one would accept w as a possible state of affairs given that belief in K may have to be given up. If v is more plausible than w, loosely speaking, v is "more consistent" with our beliefs than w, and is a preferable alternative world to adopt. This view may be based on some notion of *comparative similarity*, for instance.[4]

We take as minimal requirements that R be reflexive and transitive.[5] Another requirement we adopt in this paper is that of connectedness. In other words, any two states of affairs must be comparable in terms of similarity. If neither is more plausible than the other, then they are equally plausible. We also insist that all worlds be represented in our structures.

Given these restrictions, we can use CO*-models to represent the revision of a theory K. However, arbitrary CO*-models are inappropriate, for we must insist 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. Such a constraint can be expressed in our language as

$$\overleftrightarrow{\Box}(KB \supset (\Box KB \wedge \overleftarrow{\Box}\neg KB)) \qquad (1)$$

for any K that is finitely expressible as KB. This ensures that any KB-world sees every other KB-world ($\overleftarrow{\Box}\neg KB$), and that it sees only KB-worlds ($\Box KB$). All statements about revision are implicitly evaluated with respect to KB. We abbreviate sentence (1) as $O(KB)$ and intend it to mean we "only know" KB.[6] Such models are called *K-revision models*.

Given this structure, we want the set of A-worlds minimal in R to represent the state of affairs believed

[4] See (Lewis 1973; Stalnaker 1984) on this notion.

[5] In (Boutilier 1992) we develop this minimal logic in the context of *preorder revision*.

[6] This terminology is discussed in the next section.

when K is revised by A. These are the most plausible worlds, the ones we are most willing to adopt, given A. Of course such a minimal set may not exist (consider an infinite chain of more and more plausible A-worlds). Still, we can circumvent this problem by adopting a conditional perspective toward revision. Often when revising a belief set, we do not care to characterize the entire new belief state, but only certain consequences of interest of the revised theory (i.e., conditionals).

The sentence $A \xrightarrow{\text{KB}} B$ should be true if, at any point on the chain of decreasing A-worlds, B holds at all more plausible A-worlds (hence, B is true at some hypothetical limit of this chain). We can define the connective as follows:

$$A \xrightarrow{\text{KB}} B \equiv_{\text{df}} \overleftrightarrow{\Box}\neg A \vee \overleftrightarrow{\Diamond}(A \wedge \Box(A \supset B)). \quad (2)$$

This sentence is true in the trivial case when A is impossible, while the second disjunct states that there is some world w such that A holds and $A \supset B$ holds at all worlds still more plausible than w. Thus B holds at the *most* plausible A-worlds (whether this is a "hypothetical" or actual limit). In this manner we avoid the *Limit Assumption* (see below). It is important to note that $\xrightarrow{\text{KB}}$ is a connective in the usual sense, not a *family* of connectives indexed by "KB". Given the Ramsey test, $\xrightarrow{\text{KB}}$ is nothing more than a subjunctive conditional.

We can define for any propositional $A \in \mathbf{L}_{CPL}$, the belief set resulting from revision of K by A as

$$K_A^{*^{\text{M}}} = \{B \in \mathbf{L}_{CPL} : M \models A \xrightarrow{\text{KB}} B\}. \quad (3)$$

Theorem 1 *If M is a K-revision model for any K, then $*^{\text{M}}$ satisfies postulates (R1)–(R8).*

Theorem 2 *Let $*$ be any revision operator satisfying (R1) through (R8). Then there exists a K-revision model, for any theory K, such that $* = *^{\text{M}}$.*

Thus, CO* is an appropriate logical characterization of AGM revision and, in fact, is the first logical calculus of this type suited to the AGM theory. However, the modal approach suggests a number of generalizations of AGM revision, for example, by using CO or dropping connectedness (Boutilier 1992). It also provides considerable expressive power with which we can constrain the revision process in novel ways.

The Limit Assumption Our approach to revision makes no assumption about the existence of minimal A-worlds, which Grove (1988) claims forms an integral part of any model of revision. As Lewis (1973) emphasizes, there is no justification for such an assumption other than convenience. Consider a KB that contains the proposition "I am 6 feet tall." Revising by A = "I am over 7 feet tall" would allow one to evaluate Lewis's classic counterfactual "If I were over 7 feet tall I would play basketball." However, it doesn't seem that there should exist a most plausible A-world, merely an infinite sequence of such worlds approaching the limit world where "I am 7 feet tall" is true. Our model allows one to assume the truth of the counterfactual $A \xrightarrow{\text{KB}} B$ if the consequent B ("I play basketball") is strictly implied (i.e., if $\Box(A \supset B)$ holds) at any world in this sequence of A-worlds. In this respect, we take Lewis's semantic analysis to be appropriate.

Models of revision that rely on the the Limit Assumption (e.g., (Grove 1988; Katsuno and Mendelzon 1991b)) would also make $A \xrightarrow{\text{KB}} B$ true, but for the wrong reasons. It holds vacuously since there are no minimal A-worlds. Of course, $A \xrightarrow{\text{KB}} \neg B$ is true in this case too, which is strongly counterintuitive. How can this be reconciled with Theorems 1 and 2, which show that our notion of revision is equivalent to the AGM version, including those that make the Limit Assumption?

In fact, this points to a key advantage of the modal approach, its increased expressive power. We can easily state that worlds of decreasing height (`ht`), down to my actual height of 6 feet, are more plausible using:[7]

$$\overleftrightarrow{\Box}\forall y > 6[y < x \supset (\texttt{ht(Me)} = x \equiv \overleftarrow{\Box}\texttt{ht(Me)} \neq y)]. \quad (4)$$

Not only can we constrain the revision process directly using conditionals, but also indirectly using such *intensional constraints*. Of course, there must be some AGM operator that has $\beta \in K_\alpha^*$ exactly when our model satisfies $\alpha \xrightarrow{\text{KB}} \beta$, including $B \in K_A^*$ (the basketball counterfactual). But models making the Limit Assumption do not reflect the same structure as our CO*-model. They cannot express nor represent the constraint relating plausibility to height. In order to ensure $B \in K_A^*$ they must violate sentence (4). The modal language also improves on AGM revision in general, where such a constraint can only be expressed by some infinite set of conditions $\beta \in K_\alpha^*$. In (Boutilier 1992) we examine the Limit Assumption and intensional constraints in more detail. In particular, we show how constraints on the entrenchment and plausibility of sentences and beliefs can also be expressed at the object level.

When reasoning about the revision of a knowledge base KB, we require a background theory with the sentence $O(KB)$, which implicitly constrains the conditional connective to refer to KB, and a set of conditional assertions from which we can derive new revision asser-

[7] We assume the obvious first order extension of CO-models, a partial theory of $<$ over (say) the rationals, etc.

tions. For instance, to take an example from default reasoning, if one asserts

$$\{\texttt{bird} \overset{\text{KB}}{\longrightarrow} \texttt{fly}, \texttt{penguin} \overset{\text{KB}}{\longrightarrow} \neg\texttt{fly}, \texttt{penguin} \overset{\text{KB}}{\longrightarrow} \texttt{bird}\}$$

then the conclusion $\texttt{bird} \wedge \texttt{penguin} \overset{\text{KB}}{\longrightarrow} \neg\texttt{fly}$ can be derived. Indeed this should be the case as penguins are a specific subclass of birds and properties associated with them should take precedence over those associated with birds. Beliefs in KB can influence revision as well. If we take KB to be $\{A \supset B, C \supset D\}$ (where A, B, C, D are distinct atoms) then from $O(KB)$ we can infer, for instance, $A \overset{\text{KB}}{\longrightarrow} B$ and $A \vee C \overset{\text{KB}}{\longrightarrow} B \vee D$. We can also derive $\neg(A \overset{\text{KB}}{\longrightarrow} C)$ since revision by A does not force acceptance of C. Other derived theorems include (see (Boutilier 1992) for more examples):

- $(A \overset{\text{KB}}{\longrightarrow} B) \wedge (A \overset{\text{KB}}{\longrightarrow} C) \supset (A \overset{\text{KB}}{\longrightarrow} B \wedge C)$
- $(A \overset{\text{KB}}{\longrightarrow} C) \wedge (B \overset{\text{KB}}{\longrightarrow} C) \supset (A \vee B \overset{\text{KB}}{\longrightarrow} C)$
- $(A \overset{\text{KB}}{\longrightarrow} B) \supset ((A \wedge B \overset{\text{KB}}{\longrightarrow} C) \supset (A \overset{\text{KB}}{\longrightarrow} C))$
- $(A \overset{\text{KB}}{\longrightarrow} B) \wedge (A \overset{\text{KB}}{\longrightarrow} C) \supset (A \wedge B \overset{\text{KB}}{\longrightarrow} C)$
- $(A \overset{\text{KB}}{\longrightarrow} C) \wedge \neg(A \wedge B \overset{\text{KB}}{\longrightarrow} C) \supset A \overset{\text{KB}}{\longrightarrow} \neg B$

We now turn our attention to a systematic framework for representing knowledge about belief revision.

A Framework for Subjunctive Queries

Let KB be as usual a set of beliefs representing our knowledge of the world. We also expect there to be some conditional beliefs among these that constrain the manner in which we are willing to revise our (objective) beliefs. These take the form $\alpha \overset{\text{KB}}{\longrightarrow} \beta$ (or $\alpha > \beta$ when we intend Lewis' connective), and will be referred to as *subjunctive premises*. By a *subjunctive query* we mean something of the form "If A were true, would B hold?" In other words, is $A > B$ a consequence of our beliefs and subjunctive premises?

Given the connection between VC and belief revision, and assuming the Ramsey test is an appropriate truth test for subjunctives, it would appear that VC is exactly the logical calculus required for formulating subjunctive queries. However, we have misrepresented the Gärdenfors result to a certain degree; in fact, his semantics does not account for the postulate of consistent revision (R4). It is excluded because it results in *triviality* (see (Gärdenfors 1988)) and, together with the other postulates, is far too strong to be of use. Because (R4) is unaccounted for in VC, it is inadequate for the representation of certain subjunctive queries.

Example Suppose $KB = \{B\}$, a belief set consisting of a single propositional letter. If we were to ask "If A then B?" intuitively we would expect the answer `YES`, when A is some distinct atomic proposition. With no constraints (such as $A > \neg B$), the postulate of consistent revision should hold sway and revising by A should result in $KB' = \{A, B\}$. Hence, $A > B$ should be true of KB. Similarly, $\neg(A > C)$ should also be true of KB for any distinct atom C.

In VC there is no mechanism for drawing these types of conclusions. At most one could hope to assert B as a premise and derive $A > B$ or $\neg(A > C)$, but neither of $B \vdash_{VC} A > B$ or $B \vdash_{VC} \neg(A > C)$ is true, nor should they be. It *should* be the case that if A is *consistent with our beliefs* that $A > B$ holds, but merely asserting B doesn't carry this force. When B is a premise we mean "B is believed;" but this does not preclude the possibility of A, $\neg A$, C, or anything else being believed. When $KB = \{B\}$ we intend something stronger, that "B is *all* that is believed." Because B is the only sentence in KB, we convey the added information that, say, neither A nor $\neg A$ is believed. In Levesque's (1990) terminology, we *only know* KB.

To *only know* some sentence is to both know (or believe) B and to know nothing more than B. To know B is to restrict one's set of epistemic possibilities to those states of affairs where B is true. If some $\neg B$-world were considered possible an agent could not be said to know B, for the possibility of $\neg B$ has not been ruled out. To know nothing more than B is is to include all possible A-worlds among one's set of epistemic possibilities. Adding knowledge to a belief set is just restricting one's set of epistemic possibilities to exclude worlds where these new beliefs fail, so if some B-world were excluded from consideration, intuitively an agent would have some knowledge other than B that ruled out this world.

In our logic CO* we have precisely the mechanism for stating that we only know a KB. We consider the set of minimal worlds to represent our knowledge of the actual world. Exactly those possible worlds consistent with our beliefs KB are minimal in any KB-revision model. This is precisely what the sentence $O(KB)$ asserts, that KB is believed (since only KB-worlds are minimal) and that KB is all that is believed (since only minimal worlds are KB-worlds).

Returning to the query $A \overset{\text{KB}}{\longrightarrow} B$, this analysis suggests that $B \vdash_{CO*} A \overset{\text{KB}}{\longrightarrow} B$ is not the proper formulation of the query. This derivation is not valid (just as it is not in VC). Rather, we ought to ask if $A \overset{\text{KB}}{\longrightarrow} B$

holds if we *only* know B. In fact, both

$$O(B) \vdash_{CO*} A \xrightarrow{\text{KB}} B \quad \text{and} \quad O(B) \vdash_{CO*} \neg(A \xrightarrow{\text{KB}} C)$$

are legitimate derivations.

This leads to an obvious framework for subjunctive query answering, given a set of beliefs KB. Our knowledge of the world is divided into two components, a set KB of *objective* (propositional) facts or beliefs, and a set S of subjunctive conditionals acting as premises, or constraints on the manner in which we revise our beliefs. To ask a subjunctive query Q of the form $\alpha \xrightarrow{\text{KB}} \beta$ is to ask if β would be true if we believed α, given that our *only* current beliefs about the world are represented by KB, and that our deliberations of revision are constrained by subjunctive premises S.[8] The expected answers `YES`, `NO` and `UNK` (unknown) to Q are characterized as follows.

$$ASK(Q) = \begin{cases} \texttt{YES} & \text{if } \{O(KB)\} \cup S \models_{CO*} Q \\ \texttt{NO} & \text{if } \{O(KB)\} \cup S \models_{CO*} \neg Q \\ \texttt{UNK} & \text{otherwise} \end{cases}$$

Objective queries about the actual state of affairs (or, more precisely, about our beliefs) can be phrased as $\top \xrightarrow{\text{KB}} \beta$ where β is the objective query of interest. It's easy to see that for such a Q

$$ASK(Q) = \texttt{YES} \quad \text{iff} \quad \vdash_{CO*} KB \supset \beta.$$

The ability to express that *only* a certain set of sentences is believed allows us to give a purely logical characterization of subjunctive queries of a knowledge base. The logic VC seems adequate for reasoning from subjunctive premises and for deriving new conditionals, but it cannot account for the influence of factual information on the truth of conditionals in a completely satisfying manner; for it lacks the expressive power to enforce compliance with postulate (R4).[9] The approaches of Ginsberg and Jackson take VC to be the underlying counterfactual logic. Indeed, their approaches (under certain assumptions) satisfy the Lewis axioms. However, they recognize that the ability to only know a knowledge base is crucial for revision and subjunctive reasoning, an expressive task not achievable in VC. Therein lies the motivation for their extra-logical characterizations, and the underlying idea that KB is representable as a set of sentences or set of possible worlds from which we construct new sets in the course of revision. CO* can be viewed as a *logic* in which one can capture just this process.[10]

Integrity Constraints

Often it is the case that only certain states of knowledge, certain belief sets, are permissible. The concept of *integrity constraints*, widely studied in database theory, is a way to capture just such conditions. For a database (or in our case, a belief set) to be considered a valid representation of the world, it must satisfy these integrity constraints. For instance, we may not consider feasible any belief set in which certain commonsense laws of physics are violated; or a database in which there exists some student with an unknown student number may be prohibited.

This apparently straightforward concept actually has several distinct interpretations. Reiter (1990) surveys these and proposes the definition we favor, which essentially asserts that an integrity constraint C should be entailed by KB. The distinguishing characteristic of Reiter's definition is that integrity constraints can be phrased using a modal knowledge operator, which refers to "what is known by the database." We will assume constraints are propositional and that KB satisfies C just when KB entails C.[11]

As emphasized in (Fagin, Ullman and Vardi 1983) and (Winslett 1990), integrity constraints are particularly important when updating a database. Any new database (or belief set) should satisfy these constraints, therefore any reasonable model of update or revision must explicitly account for integrity constraints.

Example Let the constraint C, that a department has only one chair, be expressed as

$$\texttt{chair(x,d)} \wedge \texttt{chair(y,d)} \supset \texttt{x=y} \qquad (5)$$

Suppose we update KB with `chair(Ken,DCS)` $\vee$ `chair(Maria,DCS)`, and assume

$$KB = \{\texttt{chair(Derek,DCS)}\}$$

so this new fact in inconsistent with the existing KB (assuming Unique Names). The constraint can not be enforced in the updated KB', for nothing about C says it must be true in the revised state of affairs, even if it is an explicit *fact* in the original KB.

[8] We note that S can be part of KB, lying within the scope of O, but prefer to keep them separate for this exposition (see (**?**)).

[9] In fact, it is not hard to verify that the axioms for VC are each valid in CO* if we replace nonsubjunctive information (say α) by statements to the effect that α is believed, also expressible in CO*; see (**?**)).

[10] Other distinctions exist (see Section 1), e.g., Ginsberg's proposal is syntax-sensitive, while Jackson's system is committed to *specific* minimality criteria.

[11] We can express constraints involving a knowledge modality (see (**?**) for details).

This example illuminates the need for integrity constraints to be expressed intensionally. They refer not only to the actual world, but to *all (preferred) ways in which we view the world*.

We can ensure a revised belief set or database satisfies a constraint C by asserting $\overleftrightarrow{\Box}C$ as a premise in our background theory (on the same level as $O(KB)$). This has the effect of ensuring *any* possible worlds ever considered satisfy C (thus requiring the logic CO). However, this may be too strong an assertion in many applications. Such a statement will force any revision of KB by a fact inconsistent with C to result in the inconsistent belief set $Cn(\bot)$. In certain (maybe most) circumstances, we can imagine the violation of a constraint C ought not force us into inconsistency.

Instead of abolishing $\neg C$-worlds outright, we'd like to say all C-worlds are "preferred" to any world violating the constraint. Such a condition is expressible as $\overleftrightarrow{\Box}(C \supset \Box C)$. To see this, imagine some $\neg C$-world v is more plausible than some C-world w. Then wRv and $M \not\models_w C \supset \Box C$. We denote this formula by WIC. Since we are often concerned with a set of constraints $C = \{C_1, \cdots, C_n\}$, in such a case we use C to denote their conjunction, and

$$WIC = \overleftrightarrow{\Box}(C \supset \Box C) \quad \text{where} \quad C = \bigwedge_{i \leq n} C_i.$$

Definition M is a revision model for K *with weak integrity constraints* C iff M is a revision model for K and $M \models \overleftrightarrow{\Box}(C \supset \Box C)$.

Theorem 3 *Let M be a revision model for K with weak integrity constraints C. Then $K_A^{*\mathrm{M}} \models C$ for all A consistent with C.*

Thus we validate the definition of integrity constraint. If a sentence A is consistent with C it must be the case that revising by A results in a belief set that satisfies C. Of course, this requires that the original belief set K must also satisfy the integrity constraints, not just revised belief sets $K_A^{*\mathrm{M}}$ (imagine revising by $\top$).

Example Let $KB = \{$`chair(Derek,DCS)`$\}$ and C be the previous constraint (5). If we update KB with `chair(Ken,DCS)` $\vee$ `chair(Maria,DCS)`, then from $\{O(KB)\} \cup WIC$ we can derive in CO*

$$\texttt{chair(Ken,DCS)} \vee \texttt{chair(Maria,DCS)} \xrightarrow{\text{KB}} \texttt{chair(Ken,DCS)} \equiv \neg\texttt{chair(Maria,DCS)}.$$

This definition of integrity constraint has the unappealing quality of being unable to ensure that as many constraints as possible be satisfied. For instance, if some update A violates some C_i of C, then revision by A is not guaranteed to satisfy other constraints. In (Boutilier 1992) we introduce *strong* constraints that accomplish this. We can also *prioritize* constraints, assigning unequal weight to some constraints in C. Fagin, Ullman and Vardi (1983) have argued that sentences in a database can have different priorities and that updates should respect these priorities by "hanging on to" sentences of higher priority whenever possible during the course of revision. Consider two constraints asserting that a department has one chair and that the chair of Computer Science (CS) is the only person without a course to teach. It might be that certain information cannot satisfy both of the constraints, but could satisfy either one singly — for example, when we learn that Maria is the chair and Ken has no course load. We may also prefer to violate the constraint that a non-chair faculty member teaches no course, deferring to the fact that CS has only one chair.

Suppose that the set $C = \{C_1, \cdots, C_n\}$ is now an *ordered* set of integrity constraints with C_i having higher priority than C_j whenever $i < j$. We prefer C_i when a conflict arises with C_j. Let P_i denote the conjunction of the i highest priority integrity constraints

$$P_i = C_1 \wedge C_2 \wedge \cdots C_i.$$

The set of *prioritized integrity constraints* is

$$ICP = \{\overleftrightarrow{\Box}(P_i \supset \Box P_i) : i \leq n\}.$$

Definition M is a *revision model for K with prioritized integrity constraints* $C_1, \cdots, C_n$ iff M is a revision model for K and $M \models ICP$.

Theorem 4 *Let M be a revision model for K with prioritized integrity constraints $C_1, \cdots, C_n$. If A is consistent with the conjunction P_j of all constraints $C_i, i \leq j$, then $K_A^{*\mathrm{M}} \models C_i$ for all $i \leq j$.*

Example Let $KB = \{$`chair(Derek,DCS)`$\}$ and $C = \{C_1, C_2\}$, where C_1 is constraint (5) and

$$C_2 = \texttt{chair(x,DCS)} \equiv \texttt{teachnocourse(x)}.$$

From $\{O(KB)\} \cup ICP$ we can derive in CO*

$$\texttt{teachnocourse(Ken)} \wedge \texttt{chair(Maria,DCS)} \xrightarrow{\text{KB}} \texttt{chair(x,DCS)} \equiv \texttt{x = Maria}.$$

Concluding Remarks

We have presented a modal logic for revision and subjunctive reasoning that, unlike current logics, can account for the effect of *only knowing* a knowledge base.

CO* can be viewed as a logical calculus for AGM revision. Furthermore, it characterizes these processes without making the *Limit Assumption*, as advocated by Lewis (1973), and allows integrity constraints to be expressed naturally *within* the logic. In (**?**) we show that CO*, with its ability to reason about knowledge, can be viewed as a generalization of autoepistemic logic, and that our subjunctive bears a remarkable similarity to the *normative conditionals* postulated for default reasoning. Indeed, we show that default reasoning can be viewed as a special case of revision by using our subjunctive. Furthermore, we show that the triviality result of Gärdenfors (1988) is avoided in our logic at the expense of postulate (R4), which we claim cannot be correct for belief sets that include conditionals. The triviality result has also led to a distinction between revision and *update* (Winslett 1990; Katsuno and Mendelzon 1991a), which has also be used to define subjunctives (Grahne 1991). An interesting avenue of research would be to pursue the connection between the two, examining the extent to which update can be captured by unary modal operators, and the extent to which either problem subsumes the other. The generalizations of revision afforded by the modal approach may also apply to update.

Acknowledgements I would like to thank Peter Gärdenfors, Gösta Grahne, Hector Levesque, David Makinson, Alberto Mendelzon and Ray Reiter for their very helpful comments.

References

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.

Bonner, A. J. 1988. A logic for hypothetical reasoning. In *Proc. of AAAI-88*, pages 480–484, St. Paul.

Boutilier, C. 1991. Inaccessible worlds and irrelevance: Preliminary report. In *Proc. of IJCAI-91*, pages 413–418, Sydney.

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

Fagin, R., Ullman, J. D., and Vardi, M. Y. 1983. On the semantics of updates in databases: Preliminary report. In *Proceedings of SIGACT-SIGMOD Symposium on Principles of Database Systems*, pages 352–365.

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

Ginsberg, M. L. 1986. Counterfactuals. *Artificial Intelligence*, 30(1):35–79.

Grahne, G. 1991. Updates and counterfactuals. In *Proc. of KR-91*, pages 269–276, Cambridge.

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

Humberstone, I. L. 1983. Inaccessible worlds. *Notre Dame Journal of Formal Logic*, 24(3):346–352.

Jackson, P. 1989. On the semantics of counterfactuals. In *Proc. of IJCAI-89*, pages 1382–1387, Detroit.

Katsuno, H. and Mendelzon, A. O. 1991a. On the difference between updating a knowledge database and revising it. In *Proc. of KR-91*, pages 387–394, Cambridge.

Katsuno, H. and Mendelzon, A. O. 1991b. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294.

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

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

Reiter, R. 1990. What should a database know? Technical Report KRR-TR-90-5, University of Toronto, Toronto.

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.

Stalnaker, R. C. 1984. *Inquiry*. MIT Press, Cambridge.

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