Modeling Belief in Dynamic Systems. Part II: Revision and Update*

Nir Friedman
Computer Science Division
387 Soda Hall
University of California
Berkeley, CA 94720
nir@cs.berkeley.edu
url: `http://www.cs.berkeley.edu/~nir`

Joseph Y. Halpern
Computer Science Department
Cornell University
Ithaca, NY 14853
halpern@cs.cornell.edu
url: `http://www.cs.cornell.edu/home/halpern`

February 5, 1998

Abstract

The study of *belief change* has been an active area in philosophy and AI. In recent years two special cases of belief change, *belief revision* and *belief update*, have been studied in detail. In a companion paper [Friedman and Halpern 1997a], we introduce a new framework to model belief change. This framework combines temporal and epistemic modalities with a notion of plausibility, allowing us to examine the change of beliefs over time. In this paper, we show how belief revision and belief update can be captured in our framework. This allows us to compare the assumptions made by each method, and to better understand the principles underlying them. In particular, it shows that Katsuno and Mendelzon's notion of belief update [Katsuno and Mendelzon 1991a] depends on several strong assumptions that may limit its applicability in artificial intelligence. Finally, our analysis allow us to identify a notion of *minimal change* that underlies a broad range of belief change operations including revision and update.

*Some of this work was done while both authors were at the IBM Almaden Research Center. The first author was also at Stanford while much of the work was done. IBM and Stanford's support are gratefully acknowledged. The work was also supported in part by the Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080 and grant F94620-96-1-0323 and by NSF under grants IRI-95-03109 and IRI-96-25901. The first author was also supported in part by an IBM Graduate Fellowship and by Rockwell Science Center. A preliminary version of this paper appears in J. Doyle, E. Sandewall, and P. Torasso (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, 1994, pp. 190–201, under the title "A knowledge-based framework for belief change, Part II: revision and update."

1 Introduction

The study of *belief change* has been an active area in philosophy and artificial intelligence. The focus of this research is to understand how an agent should change her beliefs as a result of getting new information. Two instances of this general phenomenon have been studied in detail. *Belief revision* [Alchourrón, Gärdenfors, and Makinson 1985; Gärdenfors 1988] focuses on how an agent revises her beliefs when she adopts a new belief. *Belief update* [Katsuno and Mendelzon 1991a], on the other hand, focuses on how an agent should change her beliefs when she realizes that the world has changed. Both approaches attempt to capture the intuition that an agent should make minimal changes in her beliefs in order to accommodate the new belief. The difference is that belief revision attempts to decide what beliefs should be discarded to accommodate a new belief, while belief update attempts to decide what changes in the world led to the new observation.

Belief revision and belief update are two of many possible ways of modeling belief change. In [Friedman and Halpern 1997a], we introduce a general framework for modeling belief change. We start with the framework for analyzing knowledge in multi-agent systems, introduced in [Halpern and Fagin 1989], and add to it a measure of plausibility at each situation. We then define belief as truth in the most plausible situations. The resulting framework is very expressive; it captures both time and knowledge as well as beliefs. Having time allows us to reason in the framework about changes in the beliefs of the agent. It also allows us to relate the beliefs of the agent about the future with her actual beliefs in the future. Knowledge captures in a precise sense the non-defeasible information the agent has about the world, while belief captures the defeasible assumptions implied by her plausibility assessment. The framework allows us to represent a broad spectrum of notions of belief change. In this paper, we focus on how, in particular, belief revision and update can be represented.

We are certainly not the first to provide semantic models for belief revision and update. For example, [Alchourrón, Gärdenfors, and Makinson 1985; Grove 1988; Gärdenfors and Makinson 1988; Rott 1991; Boutilier 1992; de Rijke 1992] deal with revision, and [Katsuno and Mendelzon 1991a; del Val and Shoham 1992] deal with update. In fact, there are several works in the literature that capture both using the same machinery [Katsuno and Satoh 1991; Goldszmidt and Pearl 1992], and others that simulate belief revision using belief update [Grahne, Mendelzon, and Rieter 1992; del Val and Shoham 1994]. Our approach is different from most in that we do not construct a specific framework to capture one or both of these belief change paradigms. Instead, we start from a natural framework to model how an agent's knowledge changes over time and add to it machinery that captures a defeasible notion of belief.

We believe that our representation offers a number of advantages, and gives insight into both revision and update. For one thing, we show that both revision and update can be viewed as proceeding by *conditioning* on initial prior plausibilities. Thus, our representation emphasizes the role of conditioning as a way of understanding *minimal change*. Moreover, it shows that that the major differences between revision and update can be understood as corresponding to differences in initial beliefs. For example, revision places full belief on the assumption that the propositions used to describe the world are *static*, and do not change their truth value over time.[1] By way of contrast, update allows for the possibility that propositions change their truth

[1]In the literature, belief revision has been described (by Katsuno and Mendelzon [1991a], for example) as

value over time. However, the family of prior plausibilities corresponding to update have the property that they prefer sequences of events where abnormal events occur as late as possible. The fact that time appears explicitly in our framework allows us to make this precise.

Finally, our representation makes it clear how the intuitions of revision and update can be applied in settings where the postulates used to describe them are not sound. For example, we consider situations where they may be irreversible changes (such as death, or breaking a glass vase), and where the agent may perform actions beyond just making observations. Revision and update, as they stand, cannot handle such situations. As we show, our framework allows us to extend them in a natural way so they do.

The rest of the paper is organized as follows. In Section 2, we give an overview of the framework we introduced in [Friedman and Halpern 1997a]. In Section 3, we give a brief review of belief revision and belief update. In Section 4, we define a specific class of structures that embody assumptions that are common to both update and revision. In Section 5, we describe additional assumptions that are required to capture revision. In Section 6, we describe the assumptions that are required to capture update. In Section 7, we reexamine the differences and similarities between belief revision and update. In Section 8, we consider possible extensions to the setup of revision and update, and discuss how these extensions can be handled in our framework. Finally, in Section 9, we conclude with a discussion of related and future work.

2 The Framework

We now review the framework of Halpern and Fagin [1989] for modeling knowledge, and our extension of it for dealing with belief change. The reader is encouraged to consult [Fagin, Halpern, Moses, and Vardi 1995] for further details and motivation.

2.1 Modeling Knowledge

The framework of Halpern and Fagin was develop to model knowledge in distributed (i.e., multi-agent) systems [Halpern and Fagin 1989; Fagin, Halpern, Moses, and Vardi 1995]. In this paper, we restrict our attention to the single agent case. The key assumption in this framework is that we can characterize the system by describing it in terms of a *state* that changes over time. Formally, we assume that at each point in time, the agent is in one of a possibly infinite set of (local) states. At this point, we do not put any further structure on these states (although, as we shall see from our examples, when we model situations in a natural way, states typically do have a great deal of meaningful structure). Intuitively, this local state encodes the information the agent has observed thus far. There is also an *environment*, whose state encodes relevant aspects of the system that are not part of the agent's local state.

A *global state* is a pair (s_e, s_a) consisting of the environment state s_e and the local state s_a of the agent. A *run* of the system is a function from time (which, for ease of exposition, we

a process of changing beliefs about a *static world*, but this is slightly misleading. In fact, what is important for revision is not that the world is static, but that the propositions used to describe the world are static. For example, "At time 0 the block is on the table" is a static proposition, while "The block is on the table" is not, since it implicitly references the current state of affairs. The assumption that the propositions are static is not unique to belief revision. Bayesian updating, for example, makes similar assumptions.

assume ranges over the natural numbers) to global states. Thus, if r is a run, then $r(0), r(1), \ldots$ is a sequence of global states that, roughly speaking, is a complete description of what happens over time in one possible execution of the system. Given a run r, we can define two functions r_e and r_a that map from time to states of the environment and the agent, respectively, by taking $r_e(m)$ to be the state of the environment in the global state $r(m)$ and $r_a(m)$ to be the agent's local state in $r(m)$. We can thus identify run r with the pair of functions $\langle r_e, r_a \rangle$. We take a *system* to consist of a set of runs. Intuitively, these runs describe all the possible behaviors of the system, that is, all the possible sequences of events that could occur in the system over time.

Given a system $\mathcal{R}$, we refer to a pair (r, m) consisting of a run $r \in \mathcal{R}$ and a time m as a *point*. We say two points (r, m) and (r', m') are *indistinguishable* to the agent, and write $(r, m) \sim_a (r', m')$, if $r_a(m) = r'_a(m')$, i.e., if the agent has the same local state at both points. Finally, an *interpreted* system $\mathcal{I}$ is a tuple $(\mathcal{R}, \pi)$ consisting of a system $\mathcal{R}$ together with a mapping π that associates with each point a truth assignment to a set Φ of primitive propositions. In an interpreted system we can talk about an agent's knowledge: the agent knows φ at a point (r, m) if φ holds in all points (r', m') such that $(r, m) \sim_a (r', m')$. Intuitively, an agent knows φ at (r, m) if φ is implied by the information in the local state $r_a(m)$. We give formal semantics for a language of knowledge (and time and plausibility) in Section 2.3.

Example 2.1: The circuit diagnosis problem has been well studied in the literature (see [Davis and Hamscher 1988] for an overview). Consider a circuit that contains n logical components $c_1, \ldots, c_n$ and k lines $l_1, \ldots, l_k$. The agent can set the values on the input lines of the circuit and observe the values on the output lines. The agent then compares the actual output values to the expected output values and attempts to locate faulty components. Since a single test is usually insufficient to locate the problem, the agent might performs a sequence of such tests.

We want to model diagnosis using an interpreted system. To do so, we need to describe the agent's local state, the state of the environment, and some appropriate propositions for reasoning about diagnosis. Intuitively, the agent's state is the sequence of input-output relations observed, while the environment's state describes the current state of the circuit. This consists of the *failure set*, that is, the set of faulty components of the circuit and the values on all the lines in the circuit. Each run describes the results of a specific series of tests the agent performs and the results she observes. We make two additional assumptions: (1) the agent does not forget what tests were performed and their results, and (2) the faults are persistent and do not change over time.

To make this precise, we define the environment state at a point (r, m) to consist of the failure set at (r, m), which we denote $\mathit{fault}(r, m)$, as well as the values of all the lines in the circuit. We require that the environment state be consistent with the description of the circuit. Thus, for example, if c_1 is an AND gate with input lines l_1 and l_2 and output line l_3, then if $r_e(m)$ says that c_1 is not faulty, then we require that there is a 1 on l_3 if and only if there is a 1 on both l_1 and l_2.[2] We capture the assumption that faults are persistent by requiring that $\mathit{fault}(r, m) = \mathit{fault}(r, 0)$. For our later results, it is useful to describe the agent's observations using our logical language.

[2]Note that this means that we can recover the behavior of the circuit (although not necessarily its exact description) by simply looking at the environment state at a point where there are no failures. Of course, if we could have a yet richer environment state that encodes the actual description of the circuit, but this is unnecessary for the analysis we do here.

Consider the set $\Phi_{diag} = \{f_1, \ldots, f_n, h_1, \ldots, h_k\}$ of primitive propositions, where f_i denotes that component i is faulty and h_i denotes that there is a 1 on line i (that is, line i in a "high" state). An observation is a conjunction of literals of the form h_i and $\neg h_i$. The agent's state at time m is a sequence of m such observations. Formally, we define the agent's state $r_a(m)$ to be $\langle o_1, \ldots, o_m \rangle$, where, intuitively, o_k is the formula describing the input-output relation observed at time k. We use the notation $io(r,k)$ to denote the formula describing the observation made by the agent at the point (r,k). Given this language, we can define the interpretation π_{diag} in the obvious way. We say that an observation o is *consistent* with an environment state $r_e(m)$ if the states of the input/output lines in $r_e(m)$ agree with these in o. The system $\mathcal{R}_{diag}$ consists of all runs r satisfying these requirements in which $io(r,m)$ is consistent with $r_e(m)$ for all times m.

Given the system $(\mathcal{R}_{diag}, \pi_{diag})$, we can examine the agent's knowledge after making a sequence of observations $o_1, \ldots, o_m$. It is easy to see that the agent knows that the fault set must be one with which all the observations are consistent. However, the agent cannot rule out any of these fault sets. Thus, even if all the observations are consistent with the circuit being fault-free, the agent does not know that the circuit is fault-free, since there might be a fault that manifests itself only in configurations that have not yet been tested. Of course, the agent might strongly believe that the circuit is fault-free, but we cannot (yet) express this fact in our formalism. The next section rectifies this problem. ∎

2.2 Plausibility Measures

Most non-probabilistic approaches to belief change require (explicitly or implicitly) that the agent has some ordering over possible alternatives. For example, the agent might have a preference ordering over possible worlds [Boutilier 1994c; Grove 1988; Katsuno and Mendelzon 1991b] or an entrenchment ordering over formulas [Gärdenfors and Makinson 1988]. This ordering dictates how the agent's beliefs change. For example, in [Grove 1988], the new beliefs are characterized by the most preferred worlds that are consistent with the new observation, while in [Gärdenfors and Makinson 1988], beliefs are discarded according to their degree of entrenchment until it is consistent to add the new observation to the resulting set of beliefs. We represent this ordering using *plausibility measures*, which were introduced in [Friedman and Halpern 1995; Friedman and Halpern 1997b]. We briefly review the relevant definitions and results here.

Recall that a probability space is a tuple $(W, \mathcal{F}, \Pr)$, where W is a set of worlds, $\mathcal{F}$ is an algebra of *measurable* subsets of W (that is, a set of subsets closed under union and complementation to which we assign probability), and $\Pr$ is a *probability measure*, that is, a function mapping each set in $\mathcal{F}$ to a number in $[0,1]$ satisfying the well-known probability axioms ($\Pr(\emptyset) = 0$, $\Pr(W) = 1$, and $\Pr(A \cup B) = \Pr(A) + \Pr(B)$, if A and B are disjoint).

Plausibility spaces ares a direct generalization of probability spaces. We simply replace the probability measure $\Pr$ by a *plausibility measure* Pl, which, rather than mapping sets in $\mathcal{F}$ to numbers in $[0,1]$, maps them to elements in some arbitrary partially ordered set. We read $\mathrm{Pl}(A)$ as "the plausibility of set A". If $\mathrm{Pl}(A) \leq \mathrm{Pl}(B)$, then B is at least as plausible as A. Formally, a *plausibility space* is a tuple $S = (W, \mathcal{F}, \mathrm{Pl})$, where W is a set of worlds, $\mathcal{F}$ is an algebra of subsets of W, and Pl maps sets in $\mathcal{F}$ to some domain D of *plausibility values* partially ordered by a relation $\leq_D$ (so that $\leq_D$ is reflexive, transitive, and anti-symmetric). We assume that D is *pointed*: that is, it contains two special elements $\top_D$, and $\perp_D$ such that $\perp_D \leq_D d \leq_D \top_D$ for all $d \in D$; we further assume that $\mathrm{Pl}(W) = \top_D$ and $\mathrm{Pl}(\emptyset) = \perp_D$. As usual, we define the ordering $<_D$ by taking $d_1 <_D d_2$ if $d_1 \leq_D d_2$ and $d_1 \neq d_2$. We omit the subscript D from $\leq_D$, $<_D$, $\top_D$, and $\perp_D$ whenever it is clear from context.

Since we want a set to be at least as plausible as any of its subsets, we require

A1 If $A \subseteq B$, then $\mathrm{Pl}(A) \leq \mathrm{Pl}(B)$.

Some brief remarks on this definition: We have deliberately suppressed the domain D from the tuple S, since for the purposes of this paper, only the ordering induced by $\leq$ on the subsets in $\mathcal{F}$ is relevant. The algebra $\mathcal{F}$ also does not play a significant role in this paper. Unless we say otherwise, we assume $\mathcal{F}$ contains all subsets of interest and suppress mention of $\mathcal{F}$, denoting a plausibility space as a pair (W, Pl).

Clearly plausibility spaces generalize probability spaces. In [Friedman and Halpern 1997b; Friedman and Halpern 1995] we show that they also generalize *belief function* [Shafer 1976], *fuzzy measures* [Wang and Klir 1992], *possibility measures* [Dubois and Prade 1990], *ordinal ranking* (or *κ-ranking*) [Goldszmidt and Pearl 1992; Spohn 1988], *preference orderings* [Kraus, Lehmann, and Magidor 1990; Shoham 1987], and *parameterized probability distributions* [Goldszmidt, Morris, and Pearl 1993] that are used as a basis for Pearl's *ϵ-semantics* for defaults [Pearl 1989].

Our goal is to describe the agent's beliefs in terms of plausibility. To do this, we describe how to evaluate statements of the form $B\varphi$ given a plausibility space. In fact, we use a richer logical language that also allows us to describe how the agent compares different alternatives. This is the logic of conditionals. Conditionals are statements of the form $\varphi \rightarrow \psi$, read "given φ, ψ is plausible" or "given φ, then by default ψ". The syntax of the logic of conditionals is simple: we start with primitive propositions and close off under conjunction, negation and the modal operator $\rightarrow$. The resulting language is denoted $\mathcal{L}^C$.

A *plausibility structure* is a tuple $PL = (W, \mathrm{Pl}, \pi)$, where W is a set of possible worlds, Pl is a plausibility measure on W, and $\pi(w)$ is a truth assignment to primitive propositions. Given a plausibility structure $PL = (W, \mathrm{Pl}, \pi)$, we define $[\![\varphi]\!]_{PL} = \{w \in W : \pi(w) \models \varphi\}$ to be the set of worlds that satisfy φ. We omit the subscript PL, when it is clear from the context. Conditionals are evaluated according to a rule that is essentially the same as the one used by Dubois and Prade [1991] to evaluate conditionals using possibility measures:

- $PL \models \varphi \rightarrow \psi$ if either $\mathrm{Pl}([\![\varphi]\!]) = \perp$ or $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) > \mathrm{Pl}([\![\varphi \wedge \neg\psi]\!])$.

Intuitively, $\varphi \rightarrow \psi$ holds vacuously if φ is impossible; otherwise, it holds if $\varphi \wedge \psi$ is more plausible than $\varphi \wedge \neg\psi$. As we show in [Friedman and Halpern 1997b], this semantics of conditionals also generalizes the semantics of conditionals in κ-ranking [Goldszmidt and Pearl 1992], and PPD structures [Goldszmidt, Morris, and Pearl 1993]. As we also show in [Friedman and Halpern 1997b], this semantics for conditionals generalizes the semantics of preferential structures. As this relationship plays a role in the discussion below, we review the necessary definitions here. A *preferential structure* is a tuple $(W, \prec, \pi)$, where $\prec$ is a partial order on W. Roughly speaking,

$w \prec w'$ holds if w is *preferred* to w'.[3] The intuition [Shoham 1987] is that a preferential structure satisfies a conditional $\varphi \rightarrow \psi$ if all the most preferred worlds (i.e., the minimal worlds according to $\prec$) in $[\![\varphi]\!]$ satisfy ψ. However, there may be no minimal worlds in $[\![\varphi]\!]$. This can happen if $[\![\varphi]\!]$ contains an infinite descending sequence $\ldots \prec w_2 \prec w_1$. What do we do in these structures? There are a number of options: the first is to assume that, for each formula φ, there are minimal worlds in $[\![\varphi]\!]$; this is the assumption actually made in [Kraus, Lehmann, and Magidor 1990], where it is called the *smoothness* assumption. A yet more general definition—one that works even if $\prec$ is not smooth—is given in [Lewis 1973; Boutilier 1994a]. Roughly speaking, $\varphi \rightarrow \psi$ is true if, from a certain point on, whenever φ is true, so is ψ. More formally,

> $(W, \prec, \pi)$ satisfies $\varphi \rightarrow \psi$, if for every world $w_1 \in [\![\varphi]\!]$, there is a world w_2 such that (a) $w_2 \preceq w_1$ (so that w_2 is at least as normal as w_1), (b) $w_2 \in [\![\varphi \wedge \psi]\!]$, and (c) for all worlds $w_3 \prec w_2$, we have $w_3 \in [\![\varphi \Rightarrow \psi]\!]$ (so any world more normal than w_2 that satisfies φ also satisfies ψ).

It is easy to verify that this definition is equivalent to the earlier one if $\prec$ is smooth.

Proposition 2.2: [Friedman and Halpern 1997b] *If $\prec$ is a preference ordering on W, then there is a plausibility measure $\mathrm{Pl}_\prec$ on W such that $(W, \prec, \pi) \models \varphi \rightarrow \psi$ if and only if $(W, \mathrm{Pl}_\prec, \pi) \models \varphi \rightarrow \psi$.*

We briefly describe the construction of $\mathrm{Pl}_\prec$ here, since we use it in the sequel. Given a preference order $\prec$ on W, let D_0 be the domain of plausibility values consisting of one element d_w for every element $w \in W$. We define a partial order on D_0 using $\prec$: $d_v < d_w$ if $w \prec v$. (Recall that $w \prec w'$ denotes that w is preferred to w'.) We then take D to be the smallest set containing D_0 that is closed under least upper bounds (so that every set of elements in D has a least upper bound in D). For a subset A of W, we can then define $\mathrm{Pl}_\prec(A)$ to be the least upper bound of $\{d_w : w \in A\}$. Since D is closed under least upper bounds, $\mathrm{Pl}(A)$ is well defined. As we show in [1997b], this choice of $\mathrm{Pl}_\prec$ satisfies Proposition 2.2.

The results of [Friedman and Halpern 1997b] show that this semantics for conditionals generalizes previous semantics for conditionals. Does this semantics capture our intuitions about conditionals? In the AI literature, there has been little consensus on the "right" properties for defaults (which are essentially conditionals). However, there has been some consensus on a reasonable "core" of inference rules for default reasoning. This core is usually known as the KLM properties [Kraus, Lehmann, and Magidor 1990], and includes such properties as

AND From $\varphi \rightarrow \psi_1$ and $\varphi \rightarrow \psi_2$ infer $\varphi \rightarrow \psi_1 \wedge \psi_2$

OR From $\varphi_1 \rightarrow \psi$ and $\varphi_2 \rightarrow \psi$ infer $\varphi_1 \vee \varphi_2 \rightarrow \psi$

What constraints on plausibility spaces gives us the KLM properties? Consider the following two conditions:

[3] We follow the standard notation for preference here [Kraus, Lehmann, and Magidor 1990], which uses the (perhaps confusing) convention of placing the more likely (or less abnormal) world on the left of the $\prec$ operator. Unfortunately, when translated to plausibility, this will mean $w \prec w'$ holds iff $\mathrm{Pl}(\{w\}) > \mathrm{Pl}(\{w'\})$.

A2 If A, B, and C are pairwise disjoint sets, $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(C)$, and $\mathrm{Pl}(A \cup C) > \mathrm{Pl}(B)$, then $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$.

A3 If $\mathrm{Pl}(A) = \mathrm{Pl}(B) = \perp$, then $\mathrm{Pl}(A \cup B) = \perp$.

A plausibility space (W, Pl) is *qualitative* if it satisfies A2 and A3. A plausibility structure (W, Pl, π) is qualitative if (W, Pl) is a qualitative plausibility space. In [Friedman and Halpern 1997b], we show that, in a very general sense, qualitative plausibility structures capture default reasoning. More precisely, we show that the KLM properties are sound with respect to a class of plausibility structures if and only if the class consists of qualitative plausibility structures. (We also provide show a weak condition that we show is necessary and sufficient for the KLM properties to be complete.) These results show that plausibility structures provide a unifying framework for the characterization of default entailment in these different logics.

2.3 Plausibility and Knowledge

In [Friedman and Halpern 1997a] we show how plausibility measures can be incorporated into the multi-agent system framework of [Halpern and Fagin 1989]. This allows us to describe the agent' assessment of the time we also introduce conditionals into the logical language in order to reason about these plausibility assessments. We now review the relevant details.

An *(interpreted) plausibility system* is a tuple $(\mathcal{R}, \pi, \mathcal{P})$ where, as before, $\mathcal{R}$ is a set of runs and π maps each point to a truth assignment, and where $\mathcal{P}$ is a *plausibility assignment function* mapping each point (r, m) to a qualitative plausibility space $\mathcal{P}(r, m) = (W_{(r,m)}, \mathrm{Pl}_{(r,m)})$. Intuitively, the plausibility space $\mathcal{P}(r, m)$ describes the relative plausibility of events from the point of view of the agent at (r, m). In this paper, we restrict to plausibility spaces that satisfy two additional assumptions:

- $W_{(r,m)} = \{(r', m') | (r, m) \sim_a (r', m')\}$. Thus, the agent considers plausible only situations that are possible according to her knowledge.
- if $(r, m) \sim_a (r', m')$ then $\mathcal{P}(r, m) = \mathcal{P}(r', m')$. This means that the plausibility space is a function of the agent's local state. [4]

We define a logical language to reason about interpreted systems. The syntax of the logic is simple; we start with primitive propositions and close off under conjunction, negation, the K modal operator ($K\varphi$ says that the agent knows φ), the $\bigcirc$ modal operator ($\bigcirc\varphi$ says that φ is true at the next time step), and the $\rightarrow$ modal operator. The resulting language is denoted $\mathcal{L}^{KPT}$.[5] We recursively assign truth values to formulas in $\mathcal{L}^{KPT}$ at a point (r, m) in a plausibility system $\mathcal{I}$. The truth of primitive propositions is determined by π, so that

$(\mathcal{I}, r, m) \models p$ if $\pi(r, m)(p) = \mathbf{true}$.

[4] The framework presented in [Friedman and Halpern 1997a] is more general than this, dealing with multiple agents and allowing the agent to consider several plausibility spaces in each local state. The simplified version we present here suffices to capture belief revision and update.

[5] It is easy to add other temporal modalities such as *until, eventually, since*, etc. These do not play a role in this paper.

Conjunction and negation are treated in the standard way, as is knowledge: The agent knows φ at (r, m) if φ holds at all points that she cannot distinguish from (r, m). Thus,

$$(\mathcal{I}, r, m) \models K\varphi \text{ if } (\mathcal{I}, r', m') \models \varphi \text{ for all } (r', m') \sim_a (r, m).$$

$\bigcirc\varphi$ is true at (r, m) if φ is true at $(r, m+1)$. Thus,

$$(\mathcal{I}, r, m) \models \bigcirc\varphi \text{ if } (\mathcal{I}, r, m+1) \models \varphi.$$

Finally, we define the conditional operator $\rightarrow$ to describe the agent's plausibility assessment at the current time. Let $[\![\varphi]\!]_{(r,m)} = \{(r', m') \in W_{(r,m)} : (\mathcal{I}, r, m) \models \varphi\}$.

$$(\mathcal{I}, r, m) \models \varphi \rightarrow \psi \text{ if either } \mathrm{Pl}_{(r,m)}([\![\varphi]\!]_{(r,m)}) = \perp \text{ or } \mathrm{Pl}_{(r,m)}([\![\varphi \wedge \psi]\!]_{(r,m)}) > \mathrm{Pl}_{(r,m)}([\![\varphi \wedge \neg\psi]\!]_{(r,m)}).$$

We now define a notion of *belief.* Intuitively, the agent believes φ if φ is more plausible than not. Formally, we define $B\varphi \Leftrightarrow (\mathit{true} \rightarrow \varphi)$.

In [Friedman and Halpern 1997a] we prove that, in this framework, knowledge is an S5 operator, the conditional operator $\rightarrow$ satisfies the usual axioms of conditional logic [Burgess 1981], and $\bigcirc$ satisfies the usual properties of temporal logic [Manna and Pnueli 1992]. In addition, these properties imply that belief is a K45 operator, and the interactions between knowledge and belief are captured by the axioms $K\varphi \Rightarrow B\varphi$ and $B\varphi \Rightarrow KB\varphi$.

Example 2.3: [Friedman and Halpern 1997a] We add a plausibility measure to the system defined in Example 2.1. We define $\mathcal{I}_{diag} = (\mathcal{R}_{diag}, \pi_{diag}, \mathcal{P}_{diag})$, where $\mathcal{P}_{diag}$ is the plausibility assignment we now describe. We assume that failures of individual components are independent of one another. If we also assume that the likelihood of each component failing is the same, and also that this likelihood is small (i.e., failures are exceptional), then we can construct a plausibility measure as follows: If (r', m) and (r'', m) are two points in $W_{(r,m)}$, we say that (r', m) is more plausible than (r'', m) if $|\mathit{fault}(r', m)| < |\mathit{fault}(r'', m)|$, that is, if the failure set at (r', m) consists of fewer faulty components than at (r'', m). We extend these comparisons to sets: $\mathrm{Pl}_{(r,m)}(A) \leq \mathrm{Pl}_{(r,m)}(B)$ if $\min_{(r',m) \in A}(|\mathit{fault}(r', m)|) \geq \min_{(r',m) \in B}(|\mathit{fault}(r', m)|)$; that is, A is less plausible if all the points in A have failure sets of larger cardinality then the minimal one in B. With this plausibility measure, if all of the agent's observations up to time m are consistent with there being no failures, then the agent believes that all components are functioning correctly. On the other hand, if the observations do not match the expected output of the circuit, then the agent considers minimal failure sets that are consistent with her observations. Thus, if the observations are consistent with a failure of c_1, or a failure of c_3, or the combined failure of c_2 and c_7, then the agent believes that either c_1 or c_3 is faulty, but not both.

We now make this more precise. A *failure set* (i.e., a diagnosis) is characterized by a complete formula over $f_1, \ldots, f_n$—that is, one that determines the truth values all these propositions. For example, if $n = 3$, then $f_1 \wedge \neg f_2 \wedge \neg f_3$ characterizes the failure set $\{c_1\}$. We define $D_{(r,m)}$ to be the set of failure sets (i.e., diagnoses) that the agent considers possible at (r, m); that is $D_{(r,m)} = \{f \in F : (\mathcal{I}_{diag}, r, m) \models \neg B \neg f\}$ where F is the set of all possible failure sets.

Belief change in $\mathcal{I}_{diag}$ is characterized by the following proposition.

Proposition 2.4: *If there is some $f \in D_{(r,m)}$ that is consistent with the new observation $io(r, m+1)$, then $D_{(r,m+1)}$ consists of all the failure sets in $D_{(r,m)}$ that are consistent with $io(r, m+1)$. If all $f \in D_{(r,m)}$ are inconsistent with $io(r, m+1)$, then $D_{(r,m+1)}$ consists of all failure sets of cardinality j that are consistent with $io(r, 1), \ldots, io(r, m+1)$, where j is the least cardinality for which there is at least one failure set consistent with these observations.*

Thus, in $\mathcal{I}_{diag}$, a new observation consistent with the current set of most likely explanations reduces this set (to those consistent with the new observation). On the other hand, a surprising observation (one inconsistent with the current set of most likely explanations) has a rather drastic effect. It easily follows from Proposition 2.4 that if $io(r, m+1)$ is surprising, then $D_{(r,m)} \cap D_{(r,m+1)} = \emptyset$, so the agent discards all her current explanations in this case. Moreover, an easy induction on m shows that if $D_{(r,m)} \cap D_{(r,m+1)} = \emptyset$, then the cardinality of the failure sets in $D_{(r,m+1)}$ is greater than the cardinality of failure sets in $D_{(r,m)}$. Thus, in this case, the explanations in $D_{(r,m+1)}$ are more complicated than those in $D_{(r,m)}$. ∎

2.4 Conditioning

In an interpreted system, the agent's beliefs change from point to point because her plausibility space changes. The general framework does not put any constraints on the assignment of plausibility spaces at consecutive epistemic states. If we were thinking probabilistically, we could imagine the agent starting with a prior on the runs in the system. Since a run describes a complete history over time, this means that the agent puts a prior probability on the possible sequences of events that could happen. We would then expect the agent to modify her prior by conditioning on whatever information he has learned. As we show below, this notion of conditioning is strongly related to belief revision and update.

We start by making the simplifying assumption that we are dealing with *synchronous* systems where agents have *perfect recall* [Halpern and Vardi 1989]. Intuitively, this means that the agent knows what the time is and does not forget the observations she has made. Formally, a system is synchronous if $(r, m) \sim_a (r', m')$ only if $m = m'$. In synchronous systems, the agent has perfect recall if $(r', m+1) \sim_a (r, m+1)$ implies $(r', m) \sim_a (r, m)$. Thus, the agent considers run r possible at the point $(r, m+1)$ only if she also considers it possible at (r, m). This means that any runs considered impossible at (r, m) are also considered impossible at $(r, m+1)$: The agent does not forget what she knew.

Just as with probability, we assume that the agent has a prior plausibility measure on runs that describes her prior assessment on the possible executions of the system. As the agent gains knowledge, she updates her prior by conditioning. More precisely, at each point (r, m), the agent conditions her previous assessment on the set of runs considered possible at (r, m). This results in an updated assessment (posterior) of the plausibility of runs. This posterior induces, via a projection from runs to points, a plausibility measure on points. We can think of the agent's posterior at time m as simply her prior conditioned on her knowledge at time m.

Formally, the prior plausibility of the agent is a plausibility measure $\mathcal{P}_a = (\mathcal{R}, \mathrm{Pl}_a)$ over the runs in the system. If A is a set of points, we define $\mathcal{R}(A) = \{r : \exists m((r, m) \in A)\}$ to be the set of runs on which the points in A lie. The agent updates plausibilities by conditioning in $\mathcal{I}$ if the following condition is met:

PRIOR There is prior $\mathcal{P}_a = (\mathcal{R}, \mathrm{Pl}_a)$ such that $\mathrm{Pl}_{(r,m)}(A) \leq \mathrm{Pl}_{(r,m)}(B)$ if and only if $\mathrm{Pl}_a(\mathcal{R}(A)) \leq \mathrm{Pl}_a(\mathcal{R}(B))$, for all runs r, times m, and sets $A, B \in W_{(r,m)}$.

This definition implies that the agent's plausibility assessment at each point is determined, in a straightforward fashion, by her prior.

As shown in [Friedman and Halpern 1997a], in synchronous systems that satisfy PRIOR where agent have perfect recall, we can say even more: the agent's plausibility measure at time $m+1$ is determined by her plausibility measure at time m. To make this precise, if A is a set of points, let $\mathrm{prev}(A) = \{(r,m) : (r, m+1) \in A\}$.

Theorem 2.5: [Friedman and Halpern 1997a]. *Let $\mathcal{I}$ be a synchronous system satisfying PRIOR where agents have perfect recall. Then $\mathrm{Pl}_{(r,m+1)}(A) \leq \mathrm{Pl}_{(r,m+1)}(B)$ if and only if $\mathrm{Pl}_{(r,m)}(prev(A)) \leq \mathrm{Pl}_{(r,m)}(prev(B))$, for all runs r, times m, and sets $A, B \in W_{(r,m+1)}$.*

Thus, in synchronous systems where agent have perfect recall PRIOR implies a "local" rule for update that incrementally changes the agent's plausibility at each step. This local rule consists of two steps. First, the agent's plausibility at time m is projected to time $m+1$ points. Second, time $m+1$ points that are inconsistent with the agent knowledge at $(r, m+1)$ are discarded. This procedure implies that the relative plausibility of two sets of runs does not change unless one of them is incompatible with the new knowledge.

Example 2.6: It is easy to verify that the system $\mathcal{I}_{diag}$ we consider in Example 2.3 satisfies PRIOR. The prior $\mathcal{P}_a$ is determined by the failure set in each run in a manner similar to the construction of $\mathrm{Pl}_{(r,m)}$. That is, R_1 is more plausible than R_2 if there is a run in R_1 with a smaller failure set than all the runs in R_2. ∎

3 Review of Revision and Update

We now present a brief review of belief revision and update.

Belief revision attempts to describe how a rational agent incorporates new beliefs. As we said earlier, the main intuition is that as few changes as possible should be made. Thus, when something is learned that is consistent with earlier beliefs, it is just added to the set of beliefs. The more interesting situation is when the agent learns something inconsistent with her current beliefs. She must then discard some of her old beliefs in order to incorporate the new belief and remain consistent. The question is which ones?

The most widely accepted notion of belief revision is defined by the AGM theory [Alchourrón, Gärdenfors, and Makinson 1985; Gärdenfors 1988]. This theory was originally developed in philosophy of science, where one attempts to understand when a scientist changes her beliefs (e.g., theory of physical laws) in a rational manner. In this context, it seems reasonable to assume that the world is *static*; that is, the laws of physics do not change while the scientist is performing experiments.

Formally, this theory assumes a logical language $\mathcal{L}_e$ over a set Φ_e of primitive propositions with a consequence relation $\vdash_{\mathcal{L}_e}$ that contains the propositional calculus and satisfies the deduction theorem. The agent's epistemic state is represented as a *belief set*—that is, a set of formulas in $\mathcal{L}_e$ closed under deduction. There is also assumed to be a revision operator $\circ$ that takes a belief set A and a formula φ and returns a new belief set $A \circ \varphi$, intuitively, the result of revising A by φ. The following AGM postulates are an attempt to characterize the intuition of "minimal change":

(R1) $A \circ \varphi$ is a belief set

(R2) $\varphi \in A \circ \varphi$

(R3) $A \circ \varphi \subseteq Cl(A \cup \{\varphi\})$[6]

(R4) If $\neg\varphi \notin A$ then $Cl(A \cup \{\varphi\}) \subseteq A \circ \varphi$

(R5) $A \circ \varphi = Cl(\mathit{false})$ if and only if $\vdash_L \neg\varphi$

(R6) If $\vdash_L \varphi \Leftrightarrow \psi$ then $A \circ \varphi = A \circ \psi$

(R7) $A \circ (\varphi \wedge \psi) \subseteq Cl(A \circ \varphi \cup \{\psi\})$

(R8) If $\neg\psi \notin A \circ \varphi$ then $Cl(A \circ \varphi \cup \{\psi\}) \subseteq A \circ (\varphi \wedge \psi)$.

The essence of these postulates is the following. After a revision by φ the belief set should include φ (postulates R1 and R2). If the new belief is consistent with the belief set, then the revision should not remove any of the old beliefs and should not add any new beliefs except these implied by the combination of the old beliefs with the new belief (postulates R3 and R4). This condition is called *persistence*. The next two conditions discuss the coherence of beliefs. Postulate R5 states that the agent is capable of incorporating any consistent belief and postulate R6 states that the syntactic form of the new belief does not affect the revision process. The last two postulates enforce a certain coherency on the outcome of revisions by related beliefs. Basically, they state that if ψ is consistent with $A \circ \varphi$ then $A \circ (\varphi \wedge \psi)$ is just $A \circ \varphi \circ \psi$.

The notion of *belief update* originated in the database community [Keller and Winslett 1985; Winslett 1988]. The problem is how a knowledge base should change when something is learned about the world. For example, suppose that a transaction adds to the knowledge base the fact "Table 7 is in Office 2", which contradicts the previous belief that "Table 7 is in Office 1". What else should change? The intuition that update attempts to capture is that such a transaction describes a change that has occurred in the world. Thus, in our example, update might reason that the table was moved from office 1 to office 2. This example shows that, unlike revision, update does not assume that the world is static.

Katsuno and Mendelzon [1991a] suggest a set of postulates that an update operator should satisfy. The update postulates are expressed in terms of formulas, not belief sets. That is, an update operator $\diamond$ maps a pair of formulas, one describing the agent's current beliefs and the other describing the new observation, to a new formula that describes the agent's updated beliefs. This is not unreasonable, since we can identify a formula φ with the belief set $Cl(\varphi)$. Indeed, if Φ is finite (which is what Katsuno and Mendelzon assume) every belief set A can be associated with some formula φ_A such that $Cl(\varphi_A) = A$, and every formula φ corresponds to a belief set $Cl(\varphi)$. Thus, any update operator induces an operator that maps a belief state and an observation to a new belief state. We slightly abuse notation and use the same symbol

[6] $Cl(A) = \{\varphi | A \vdash_{\mathcal{L}_e} \varphi\}$ is the deductive closure of a set of formulas A.

to denote both type of mappings. We say that a belief set A is *complete* if, for every $\varphi \in \mathcal{L}_e$, either $\varphi \in A$ or $\neg\varphi \in A$. A formula μ is *complete* if $Cl(\mu)$ is complete.

The KM postulates are:

(U1) $\vdash_{\mathcal{L}_e} \mu \diamond \varphi \Rightarrow \varphi$

(U2) If $\vdash_{\mathcal{L}_e} \mu \Rightarrow \varphi$, then $\vdash_{\mathcal{L}_e} \mu \diamond \varphi \Leftrightarrow \mu$

(U3) $\vdash_{\mathcal{L}_e} \neg\mu \diamond \varphi$ if and only if $\vdash_{\mathcal{L}_e} \neg\mu$ or $\vdash_{\mathcal{L}_e} \neg\varphi$

(U4) If $\vdash_{\mathcal{L}_e} \mu_1 \Leftrightarrow \mu_2$ and $\vdash_{\mathcal{L}_e} \varphi_1 \Leftrightarrow \varphi_2$ then $\vdash_{\mathcal{L}_e} \mu_1 \diamond \varphi_1 \Leftrightarrow \mu_2 \diamond \varphi_2$

(U5) $\vdash_{\mathcal{L}_e} (\mu \diamond \varphi) \wedge \psi \Rightarrow \mu \diamond (\varphi \wedge \psi)$

(U6) If $\vdash_{\mathcal{L}_e} \mu \diamond \varphi_1 \Rightarrow \varphi_2$ and $\vdash_{\mathcal{L}_e} \mu \diamond \varphi_2 \Rightarrow \varphi_1$, then $\vdash_{\mathcal{L}_e} \mu \diamond \varphi_1 \Leftrightarrow \mu \diamond \varphi_2$

(U7) If μ is complete then $\vdash_{\mathcal{L}_e} (\mu \diamond \varphi_1) \wedge (\mu \diamond \varphi_2) \Rightarrow \mu \diamond (\varphi_1 \vee \varphi_2)$

(U8) $\vdash_{\mathcal{L}_e} (\mu_1 \vee \mu_2) \diamond \varphi \Leftrightarrow (\mu_1 \diamond \varphi) \vee (\mu_2 \diamond \varphi)$.

The essence of these postulates is as following. After learning φ, the agent believes φ (postulate U1, which is analogous to R2). If φ is already believed, then updating by φ does not change the agent's beliefs (postulate U2, which is a weaker version of R3 and R4). The next two postulates (U3 and U4) deal with coherence of the belief change process. They are analogous to R5 and R6, respectively, with minor differences. Postulates U5 and U6 deal with observations that are related to each other. U5 states that beliefs after learning φ that are consistent with ψ are also believed after learning $\varphi \wedge \psi$. U6 states that if φ_2 is believed after learning φ_1 and φ_1 is believed after learning φ_2, then learning either φ_1 or φ_2 leads to the same belief set. Finally, U7 and U8 deal with decomposition properties of the update operation. U7 states that if μ is essentially a truth assignment to $\mathcal{L}$, then if ψ is believed after learning φ_1 and is also believed after learning φ_2 then it is believed after learning $\varphi_1 \vee \varphi_2$. U8 states that the update of the knowledge base can be computed by independent updates on each sub-part of the knowledge. That is if $\mu = \mu_1 \vee \mu_2$, then we can apply update to each of μ_1 and μ_2, and then combine the results.

4 Belief Change Systems

We want to model belief change—particularly belief revision and belief update—in the framework of systems. To do so, we consider a particular class of systems that we call *belief change systems*. In belief change systems, the agent makes observations about an external environment. Just as is (implicitly) assumed in both revision and update, we assume that these observations are described by formulas in some logical language. We then make other assumptions regarding the plausibility measure used by the agent. We formalize our assumptions as conditions BCS1–BCS5, described below, and say that a system $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ is a *belief change system* if it satisfies these conditions. We denote by $\mathcal{C}^{BCS}$ the set of belief change systems.

Assumption BCS1 formalizes the intuition that our language includes propositions for reasoning about the environment, whose truth depends only on the environment state.

BCS1 The language $\mathcal{L}$ includes a propositional sublanguage $\mathcal{L}_e$ over a set Φ_e of primitive propositions. $\mathcal{L}_e$ contains the usual propositional connectives and comes equipped with a consequence relation $\vdash_{\mathcal{L}_e}$. The interpretation $\pi(r, m)$ assigns truth to propositions in Φ_e in such a way that

(a) $\pi(r, m)$ is consistent with $\vdash_{\mathcal{L}_e}$, that is, $\{p : p \in \Phi_e, \pi(r, m)(p) = \textbf{true}\} \cup \{\neg p : p \in \Phi_e, \pi(r, m)(p) = \textbf{false}\}$ is $\vdash_{\mathcal{L}_e}$ consistent, and

(b) $\pi(r, m)(p)$ depends only on $r_e(m)$ for propositions in Φ_e; that is, $\pi(r, m)(p) = \pi(r', m')(p)$ whenever $r_e(m) = r'_e(m')$.

Part (b) of BCS1 implies that we can evaluate formulas in $\mathcal{L}_e$ with respect to environment states; that is, if $\varphi \in \mathcal{L}_e$ and $r_e(m) = r'_e(m')$, then $(\mathcal{I}, r, m) \models \varphi$ if and only if $(\mathcal{I}, r', m') \models \varphi$. Since the environment is all that is relevant for formulas in $\mathcal{L}_e$, if $\varphi \in \mathcal{L}_e$, we write $s_e \models \varphi$ if $(\mathcal{I}, r, m) \models \varphi$ for some point (r, m) such that $r_e(m) = s_e$.

BCS2 is concerned with the form of the agent's local state. The functional form of the revision and update operators suggests that all that matters regarding how an agent changes her beliefs is the agent's belief state and what is learned. This implies that the agent's local state at time $m+1$ should be a function of her local state of time m and the observation made at time m. We in fact make the stronger assumption here that the agent's state consists of the sequence of observations made by the agent. This means that the agent remembers all her past observations. Note that this surely implies that the agent's local state at time $m+1$ is determined by her state at time m and the observation made at time m. We make the further assumption that the observations made by the agent can be described by formulas in $\mathcal{L}_e$. Although this is quite a strong assumption on the expressive power of $\mathcal{L}_e$, it is standard in the literature: Both revision and update assume that observations can be expressed as formulas in the language (see Section 3). These assumptions are formalized in BCS2:

BCS2 For all $r \in R$ and for all m, we have $r_a(m) = \langle o_{(r,1)}, \ldots, o_{(r,m)} \rangle$ where $o_{(r,k)} \in \mathcal{L}_e$ for $1 \leq k \leq m$.

Intuitively, $o_{(r,k)}$ is the observation the agent makes immediately after the transition from time $k-1$ to time k in run r. Thus, it represent what the agent observes about the new state of the system at time k. Note that BCS2 implies that the agent's state at time 0 is the empty sequence in all runs. Moreover, it implies that $r_a(m+1) = r_a(m) \cdot o_{(r,m+1)}$, where $\cdot$ is the append operation on sequences. That is, the agent's state at $(r, m+1)$ is the result of appending to her previous state the latest observation she has made about the system. It is not too hard to show that belief change systems are synchronous and agents in them have perfect recall. (We remark that the agents' local states are modeled in a similar way in the model of knowledge bases presented in [Fagin, Halpern, Moses, and Vardi 1995].)

Clearly we want to reason in our language about the observations the agent makes. Thus, we assume that the language includes propositions that describe the observations made by the agent.

BCS3 The language $\mathcal{L}$ includes a set Φ_{obs} of primitive propositions disjoint from Φ_e such that $\Phi_{obs} = \{\mathit{learn}(\varphi) : \varphi \in \mathcal{L}_e\}$. Moreover, $\pi(r, m)(\mathit{learn}(\varphi)) = \textbf{true}$ if and only if $o_{(r,m)} = \varphi$ for all runs r and times m.

In a system satisfying BCS1–BCS3, we can talk about belief change. The agent's state encodes observations, and we have propositions that allow us to talk about what is observed.

The next assumption is somewhat more geared to situations where observations are always "accepted", so that after the agent observes φ, she believes φ. While this is not a necessary assumption, it is made by both belief revision and belief update. We capture this assumption here by assuming that observations are reliable: the agent observes φ only if the current state of the environment satisfies φ. While this is not the only way of capturing the assumption that observations are accepted, it is perhaps the simplest. In addition, as we shall see, this assumption is consistent with both revision and update, in the sense that we can capture both in systems satisfying it.

BCS4 $(\mathcal{I}, r, m) \models o_{(r,m)}$ for all runs r and times m.

Note that BCS4 implies that the agent never observes *false*. Moreover, it implies that after observing φ, the agent knows that φ is true.

Finally, we assume that belief change proceeds by conditioning. While there are certainly other assumptions that can be made, as we have tried to argue, conditioning is a principled approach that captures the intuitions of minimal change, given the observations. And, as we shall see, conditioning (as captured by PRIOR) is consistent with both revision and update.

BCS5 $\mathcal{I}$ satisfies PRIOR.

Many interesting systems can be viewed as BCS's:

Example 4.1: Consider the systems $\mathcal{I}_{diag,1}$ and $\mathcal{I}_{diag,2}$ of Example 2.1. Are these systems BCSs? Not quite, since π_{diag} is not defined on primitive propositions of the form *learn*(φ), but we can easily embed both systems in a BCS. Let $\mathcal{L}_{diag}$ the propositional language defined over Φ_{diag}, and let Φ^+_{diag} consist of Φ_{diag} together with all the primitive propositions of the form *learn*(φ) for $\varphi \in \mathcal{L}_{diag}$. Let π^+_{diag} be the obvious extension of π_{diag} to Φ^+_{diag}, defined so that BCS3 holds. Then in it is easy to see that $(\mathcal{R}_{diag}, \pi^+_{diag}, \mathcal{P}_{diag,i})$ is a BCS: We take the Φ_e of BCS1 to be Φ_{diag}, and define $\vdash_{\mathcal{L}_{diag}}$ so that it enforces the relationships determined by the circuit layout. Thus, for example, if c_1 is an AND gate with input lines l_1 and l_2 and output line l_3, then we would have $\vdash_{\mathcal{L}_{diag}} \neg f_1 \Rightarrow (h_3 \Leftrightarrow h_1 \wedge h_2)$. It is then easy to see that BCS2–BCS5 hold by our construction. ▮

These definitions set the background for our presentation of belief revision and belief update.

5 Capturing Revision

Revision can be captured by restricting to BCSs that satisfy several additional assumptions. Before describing these assumptions, we briefly review a well-known representation of revision that will help motivate them.

While there are several representation theorems for belief revision, the clearest is perhaps the following [Grove 1988; Katsuno and Mendelzon 1991b]: We associate with each belief set A a set W_A of possible worlds that consists of those worlds where A is true. Thus, an agent whose belief set is A believes that one of the worlds in W_A is the real world. An agent that performs belief revision behaves as though in each belief state A she has a *ranking*, i.e., a total preorder, over all possible worlds such that the minimal (i.e., most plausible) worlds in the ranking are exactly those in W_A. When revising by φ, the agent chooses the minimal worlds satisfying φ in the ranking and constructs a belief set from them. It is easy to see that this procedure for belief revision satisfies the AGM postulates. Moreover, in [Grove 1988; Katsuno and Mendelzon 1991b], it is shown that any belief revision operator can be described in terms of such a ranking.

This representation suggests how we can capture belief revision in our framework. We define $\mathcal{C}^R \subseteq \mathcal{C}^{BCS}$ to be the set of belief change systems $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ that satisfy the conditions REV1–REV4 that we define below.

Revision assumes that the world does not change during the revision process. Formally this implies that propositions in Φ_e do not change their truth value along a run, i.e., $(\mathcal{I}, r, m) \models p$ if and only if $(\mathcal{I}, r, m+1) \models p$ for all $p \in \Phi_e$. This says that the state of the world is the same with respect to the properties that the agent reasons about (i.e., the propositions in Φ_e).

REV1 $\pi(r, m)(p) = \pi(r, 0)(p)$ for all $p \in \Phi_e$ and points (r, m).

The representation of [Grove 1988; Katsuno and Mendelzon 1991a] requires the agent to totally order possible worlds. We put a similar requirement on the agent's plausibility assessment. Recall that BCS5 says that the agent's plausibility is induced by a prior Pl_a; REV2 strengthens this assumption.

REV2 The prior Pl_a of BCS5 is ranked; that is, for all $A, B \subseteq \mathcal{R}$, either $\mathrm{Pl}_a(A) \leq \mathrm{Pl}_a(B)$ or $\mathrm{Pl}_a(B) \leq \mathrm{Pl}_a(A)$, and $\mathrm{Pl}(A \cup B) = \max(\mathrm{Pl}(A), \mathrm{Pl}(B))$.

The representation of [Grove 1988; Katsuno and Mendelzon 1991a] also requires that the agent considers all truth assignments possible. We need a similar condition, except that we want not only that all truth assignments be considered possible, but that they have nontrivial plausibility (i.e., are more plausible than $\perp$) as well.

To make this precise, it is helpful to introduce some notation that will be useful for our later definitions as well. Given a system $\mathcal{I}$ and two sequences $\varphi_1, \ldots, \varphi_k$ and $o_1, \ldots, o_{k'}$ of formulas in $\mathcal{L}_e$, let $\mathcal{R}[\varphi_1, \ldots, \varphi_k; o_1, \ldots, o_{k'}]$ consist of all runs r where for each i with $1 \leq i \leq k$, the formula φ_i is true that (r, i) and the agent observes $o_1, \ldots, o_{k'}$. That is, $\mathcal{R}[\varphi_0, \ldots, \varphi_k; o_1, \ldots, o_{k'}] = \{r \in \mathcal{I} : (\mathcal{I}, r, i) \models \varphi_i, i = 0, \ldots, k, \text{ and } r_a(k') = \langle o_1, \ldots, o_{k'} \rangle\}$. We allow either sequence of formulas to be empty, so, for example, $\mathcal{R}[\varphi; \cdot]$ consists of all runs for which φ is true at the initial state. (Note that if REV1 holds, this means that φ is true in all subsequent states as well.) We use the notation $\mathcal{R}[\varphi_1, \ldots, \varphi_m]$ as an abbreviation for $\mathcal{R}[\varphi_1, \ldots, \varphi_m; \cdot]$.

REV3 If $\varphi \in \mathcal{L}_e$ is consistent, then $\mathrm{Pl}_a(\mathcal{R}[\varphi]) > \perp$.

It might seem that REV1–REV3 capture all of the assumptions made by the representation of [Grove 1988; Katsuno and Mendelzon 1991a]. However, there is another assumption implicit in the way revision is performed in these representations that we must make explicit in our representation, because of the way we have distinguished observing φ (captured by the formula *learn*(φ)) from φ itself. Intuitively, when the agent observes φ, she updates her plausibility assessment by conditioning on φ. This is essentially what we can think of the earlier representations as doing. However, in our representation, the agent does *not* condition on φ, but on the fact that she has observed φ. Although we do require that φ must be true if the agent observes it (BCS4), the agent may in general gain extra information by observing φ.

To understand this issue, consider the following example. Suppose that $\mathcal{R}$ is such that the agent observes p_1 at time (r, m) only if p_2 and q are also true at (r, m), and she observes $p_1 \wedge p_2$ at (r, m) only if q is false. It is easy to construct a BCS satisfying REV1–REV3 that also satisfies these requirements. In this system, after observing p_1, the agent believes p_2 and q. According to AGM's postulate R7 (and also KM's postulate U5) the agent must believe q after observing $p_1 \wedge p_2$. However, in $\mathcal{R}$, the agent believes (indeed knows) $\neg q$ after observing $p_1 \wedge p_2$.[7] Thus, revision and update both are implicitly assuming that the observation of φ does not provide such additional knowledge. The following assumption ensures that this is the case for revision (a more general version will be required for update; see Section 6).

REV4 $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi; o_1, \ldots, o_m])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o_1 \wedge \ldots \wedge o_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi \wedge o_1 \wedge \ldots \wedge o_m])$.

This assumption captures the intuition that observing $o_1, \ldots, o_k$ provides no more information than just the fact that $o_1 \wedge \ldots \wedge o_m$ is true. That is, the agent compares the plausibility of φ and ψ in the same way after conditioning by the observations $o_1, \ldots, o_m$ as after conditioning by the fact that $o_1 \wedge \ldots \wedge o_m$ is true. It easily follows from REV4 and PRIOR that the agent believes ψ after observing $o_1 \wedge \ldots \wedge o_m$ exactly if $o_1 \wedge \ldots \wedge o_m \wedge \psi$ was initially considered more plausible than $o_1 \wedge \ldots \wedge o_m \wedge \neg\psi$. Thus, the agent believes ψ after observing $o_1 \wedge \ldots \wedge o_m$ exactly if initially, she believed ψ conditional on $o_1 \wedge \ldots \wedge o_m$: the observations provide no extra information beyond the fact that each of the o_i's are true.

REV4 is quite a strong assumption. Not only does it say that observations do not give the agent any additional observation (beyond the fact that they are true), it also says that all consistent observations can be made (since if $\varphi \wedge o$ is consistent, we must have $\mathrm{Pl}_a(\mathcal{R}[\varphi; o]) = \mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o]) > \perp$, by REV3 and REV4). We might instead consider using a weaker version of REV4 that says that, provided an observation can be made, it gives no additional information. Formally, this would be captured as

REV4′ If $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) > 0$, then $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi; o_1, \ldots, o_m])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o_1 \wedge \ldots \wedge o_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi \wedge o_1 \wedge \ldots \wedge o_m])$.

The following examples suggests that REV4′ may be more reasonable in practice than REV4. We used REV4 only because it comes closer to the spirit of the requirement of revision that all observations are possible.

Example 5.1: Consider the system $\mathcal{I}_{diag,1}$ described in Example 2.1. As discussed in Example 4.1, this system can be viewed as a BCS. Is it a revision system? It is easy to see that $\mathcal{I}_{diag,1}$ satisfies REV2 and REV3. It clearly does not satisfy REV1, since propositions that describe input/output lines can change their values from one point to the next. However, as we are about to show, a slight variant of $\mathcal{I}_{diag,1}$ does satisfy REV1. A more fundamental problem is that

[7] We stress this does not mean that $p_1 \wedge p_2$ *implies* $\neg q$ in $\mathcal{R}$. There may well be points in $\mathcal{R}$ at which $p_1 \wedge p_2 \wedge q$ is true. However, at such points, the agent would not observe $p_1 \wedge p_2$, since the agent observes $p_1 \wedge p_2$ only if q is false.

$\mathcal{I}_{diag,1}$ does not satisfy REV4. This is inherent in our assumption that the agent never directly observes faults, so that, for example, we have $\mathrm{Pl}_{diag,1}(\mathcal{R}[\cdot; f_1]) = \perp$, while $\mathrm{Pl}_{diag,1}(\mathcal{R}[f_1]) > \perp$. It does, however, satisfy REV4$'$.

To see how to modify $\mathcal{I}_{diag,1}$ so as to satisfy REV1, recall that in the diagnosis task, the agent is mainly interested in her beliefs about faults. Since faults are static in $\mathcal{I}_{diag,1}$, we can satisfy REV1 if we ignore all propositions except $f_1, \ldots, f_n$. Let $\Phi'_{diag} = \{f_1, \ldots, f_n\}$ and let $\mathcal{L}'_{diag}$ be the propositional language over Φ'_{diag}. For every observation o made by the agent regarding the value of the lines, there corresponds a formula in $\mathcal{L}'_{diag}$ that characterizes all the fault sets that are consistent with o. Thus, for every run r in $\mathcal{I}_{diag,1}$, we can construct a run r' where the agent's local state is a sequence of formulas in $\mathcal{L}'_{diag}$. Let $\mathcal{I}'_{diag}$ be the system consisting of all such runs r'. We can clearly put a plausibility assignment on these runs so that $\mathcal{I}_{diag,1}$ and $\mathcal{I}'_{diag}$ are isomorphic in an obvious sense. In particular, the agent has the same beliefs about formulas in $\mathcal{L}'_{diag}$ at corresponding points in the two systems. More precisely, if $\varphi \in \mathcal{L}'_{diag}$, then $(\mathcal{I}'_{diag}, r, m) \models \varphi$ if and only if $(\mathcal{I}_{diag,1}, r, m) \models \varphi$ for all points (r, m) in $\mathcal{I}_{diag,1}$. It is easy to verify that $\mathcal{I}'_{diag}$ satisfies REV1–REV3 and REV4$'$, although it still does not satisfy REV4.

We are not advocating here here using $\mathcal{I}'_{diag}$ instead of $\mathcal{I}_{diag}$—$\mathcal{I}_{diag}$ seems to us a perfectly reasonable way of modeling the situation. Rather, the point is that if we want a BCS to satisfy properties that validate the AGM postulates, we must make some strong, and not always natural, assumptions. ∎

We want to show that a revision operator corresponds to a system in $\mathcal{C}^R$ and vice versa. To do so, we need to examine the beliefs of the agent at each point (r, m). First we note that if $(r, m) \sim_a (r', m')$ then $(\mathcal{I}, r, m) \models B\varphi$ if and only if $(\mathcal{I}, r', m') \models B\varphi$; this is a consequence of the requirement that the agent's plausibility assessment is a function of her local state. Thus, we think of the agent's beliefs as a function of her local state. We use the notation $(\mathcal{I}, s_a) \models B\varphi$ as shorthand for $(\mathcal{I}, r, m) \models B\varphi$ for some (r, m) such that $r_a(m) = s_a$. Let s_a be some local state of the agent. We define the agent's *belief state* at s_a as

$$\mathrm{Bel}(\mathcal{I}, s_a) = \{\varphi \in \mathcal{L}_e : (\mathcal{I}, s_a) \models B\varphi\}.$$

Since the agent's state is a sequence of observations, the agent's state after observing φ is simply $s_a \cdot \varphi$, where $\cdot$ is the append operation. Thus, $\mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi)$ is the belief state after observing φ. We adopt the convention that if the agent can never attain the local state s_a in $\mathcal{I}$, then $\mathrm{Bel}(\mathcal{I}, s_a) = \mathcal{L}_e$. With these definitions, we can compare the agent's belief state before and after observing φ, that is $\mathrm{Bel}(\mathcal{I}, s_a)$ and $\mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi)$.

We start by showing that every AGM revision operator can be represented in $\mathcal{C}^R$.

Theorem 5.2: *Let $\circ$ be an AGM revision operator and let $K \subseteq \mathcal{L}_e$ be a consistent belief state. Then there is a system $\mathcal{I}_{\circ,K} \in \mathcal{C}^R$ such that $Bel(\mathcal{I}_{\circ,K}, \langle\rangle) = K$ and*

$$Bel(\mathcal{I}_{\circ,K}, \langle\rangle) \circ \varphi = Bel(\mathcal{I}_{\circ,K}, \langle\varphi\rangle)$$

for all $\varphi \in \mathcal{L}_e$.

Proof: See Appendix A.1. ∎

Thus, Theorem 5.2 says that we can represent a revision operator $\circ$ in the sense that we have a family of systems $\mathcal{I}_{\circ,K} \in \mathcal{C}^R$, one for each consistent belief state K, such that K is the agent's initial belief state in $\mathcal{I}_{\circ,K}$, and for each formula φ in $\mathcal{L}_e$, the agent's belief state after learning φ is $K \circ \varphi$. Notice that we restrict attention to consistent belief states K. The AGM postulates allow the agent to "escape" from an inconsistent state, so that $K \circ \varphi$ may be consistent even if K is inconsistent. We might thus hope to extend the theorem so that it also applies to the inconsistent belief state, but this is impossible in our framework. If *false* $\in \mathrm{Bel}(\mathcal{I}_{\circ,K}, s_a)$ for some state s_a, and $r_a(m) = s_a$, then $\mathrm{Pl}_{(r,m)}(W_{(r,m)}) = \perp$. Since we update by conditioning, we must have $\mathrm{Pl}_{(r,m+1)}(W_{(r,m+1)}) = \perp$, so the agent's belief state will remain inconsistent no matter what she learns. Although we could modify our framework to allow the agent to escape from inconsistent states, we actually consider this to be a defect in the AGM postulates, not in our framework. To see why, suppose that the agent's belief set is inconsistent at s_a, and $r_a(m) = s_a$. Thus, the agent considers all states in $W_{(r,m)}$ to be completely implausible (since $\mathrm{Pl}_{(r,m)}(W_{(r,m)}) = \perp$). On the other hand, to escape inconsistency, she must have a plausibility ordering over the worlds in $W_{(r,m)}$. These two requirements seem somewhat inconsistent.

Not surprisingly, this inconsistency creates problems for other semantic representations in the literature. For example, Boutilier's representation theorem [1992] states that for every revision operator $\circ$ and belief state K, there is a ranking R such that $\psi \in K \circ \varphi$ if and only if ψ is believed in the minimal φ-worlds according to R. If we examine this theorem, we note that he does not state that the minimal (i.e., most preferred) worlds in R correspond to the belief state K (in the sense that the minimal worlds are precisely those where the formulas in K hold); this would be the analogue of our requiring that $\mathrm{Bel}(\mathcal{I}_{\circ,K}, \langle\rangle) = K$. In fact, if K is $\vdash_{\mathcal{L}_e}$-consistent, the minimal worlds do correspond to K. However, if K is inconsistent, they cannot, since any nonempty ranking induces a consistent set of beliefs. We could state a weaker version of Theorem 5.2 that would correspond exactly to Boutilier's theorem. We presented the stronger result (that does not apply to inconsistent belief states) to bring out what we believe to be a problem with the AGM postulates. See [Friedman and Halpern 1996a] for further discussion of this issue.

Theorem 5.2 shows that, in a precise sense, we can map AGM revision operations to $\mathcal{C}^R$. What about the other direction? The next theorem shows that the first belief change step in systems in $\mathcal{C}^R$ satisfies the AGM postulates.

Theorem 5.3: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$. Then there is an AGM revision operator $\circ_{\mathcal{I}}$ such that*

$$Bel(\mathcal{I}, \langle\rangle) \circ_{\mathcal{I}} \varphi = Bel(\mathcal{I}, \langle\varphi\rangle)$$

for all $\varphi \in \mathcal{L}_e$.

Proof: See Appendix A.1. ∎

We remark that if we used REV4$'$ instead of REV4, then we would be able to prove this result only for those formulas φ that are observable (i.e., for which $\mathrm{Pl}(\mathcal{R}[\varphi]) > \perp$).

Both Theorems 5.2 and 5.3 apply to one-step revision, starting from the initial (empty) state. What happens once we allow iterated revision? In our framework, observations are taken to be known, so if the agent makes an inconsistent sequence of observations, then her belief state will be inconsistent, and (as we observed above) will remain inconsistent from then on, no matter what she observes. This creates a problem if we try to get analogues to Theorems 5.2 and 5.3 for iterated revision. As the following theorem demonstrates, we can already see the problem if we consider one-step revisions from a state other than the initial state.

Theorem 5.4: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ and let $s_a = \langle\varphi_1, \ldots, \varphi_k\rangle$ be a local state in $\mathcal{I}$. Then there is an AGM revision operator $\circ_{\mathcal{I},s_a}$ such that*

$$Bel(\mathcal{I}, s_a) \circ_{\mathcal{I},s_a} \varphi = Bel(\mathcal{I}, s_a \cdot \varphi)$$

for all formulas $\varphi \in \mathcal{L}_e$ such that $\varphi_1 \wedge \ldots \wedge \varphi_k \wedge \varphi$ is consistent.

Proof: See Appendix A.1. ∎

We cannot do better than this. If $\varphi_1 \wedge \ldots \wedge \varphi_k \wedge \varphi$ is inconsistent then, because of our requirements that all observations must be true of the current state of the environment (BCS4) and that propositions are static (REV1), there cannot be any global state in $\mathcal{I}$ where the agent's local state in $s_a \cdot \varphi$. Thus, $\mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi)$ is inconsistent, contradicting R5.

There is another problem with trying to get an analogue of Theorem 5.3 for iterated revision, a problem that seems inherent in the AGM framework. Our framework makes a clear distinction between the agent's *epistemic state* at a point (r, m) in $\mathcal{I}$, which we can identify with her local state $s_a = r_a(m)$, and the agent's *belief state* at (r, m), $\mathrm{Bel}(\mathcal{I}, s_a)$, which is the set of formulas she believes. In a system in $\mathcal{C}^R$, the agent's belief state does not in general determine how the agent's beliefs will be revised; her epistemic state does. On the other hand, the AGM postulates assume that revision is a function of the agent's belief state. Now suppose we have a system $\mathcal{I}$ and two points (r, m) and (r, m') on some run $r \in \mathcal{I}$ such that (1) the agent's belief set is the same at (r, m) and (r, m'), that is $\mathrm{Bel}(\mathcal{I}, r_a(m)) = \mathrm{Bel}(\mathcal{I}, r_a(m'))$, (2) the agent observes φ at both (r, m) and (r, m'), (3) $\mathrm{Bel}(\mathcal{I}, r_a(m+1)) \neq \mathrm{Bel}(\mathcal{I}, r_a(m'+1)$. It is not hard to construct such a system $\mathcal{I}$. However, there cannot be an analogue of Theorem 5.3 for $\mathcal{I}$, even if we restrict to consistent sequences of observations. For suppose there were a revision operator $\circ$ such $\mathrm{Bel}(\mathcal{I}, \langle\rangle)) \circ \varphi_1 \circ \cdots \circ \varphi_k = \mathrm{Bel}(\mathcal{I}, \langle\varphi_1, \ldots, \varphi_k\rangle)$ for all $\varphi_1, \ldots, \varphi_k$ such that $\varphi_1 \wedge \ldots \wedge \varphi_k$ is consistent. Then we would have $\mathrm{Bel}(\mathcal{I}, r_a(m+1)) = \mathrm{Bel}(\mathcal{I}, r_a(m)) \circ \varphi = \mathrm{Bel}(\mathcal{I}, r_a(m')) \circ \varphi = \mathrm{Bel}(\mathcal{I}, r_a(m'+1))$, contradicting our assumption.

The culprit here is the assumption that revision depends only on the agent's belief state. To see why this is an unreasonable assumption, consider a situation where at time 0 the agent believes both p and q, but her belief in q is stronger than her belief in p (i.e., the plausibility of q is greater than that of p). We can well imagine that after observing $\neg p \vee \neg q$ at time 1, she would believe $\neg p$ and q. However, if she first observed p at time 1 and then $\neg p \vee \neg q$ at time 2, she would believe p and $\neg q$, because, as a result of observing p, she would assign p greater plausibility than q. Note, however, that the AGM postulates dictate that after an observation that is already believed, the agent does not change her beliefs. Thus, the AGM setup would force the agent to have the same beliefs after learning $\neg p \vee \neg q$ in both situations.

There has been a great deal of work on the problem of *iterated belief revision* [Boutilier 1996; Darwiche and Pearl 1997; Freund and Lehmann 1994; Lehmann 1995; Levi 1988; Williams 1994]). Much of the recent work moves away from the assumption that belief revision depends solely on the agent's belief state. For example the approaches of Boutilier [1996] and Darwiche and Pearl [1997] define revision operators that map (rankings $\times$ formulas) to rankings. Because our framework makes such a clear distinction between epistemic states and belief states, it gives us a natural way of basically maintaining the spirit of the AGM postulates while assuming that revision is a function of epistemic states. Rather than taking $\circ$ to be a function from (belief states $\times$ formulas) to belief states, we take it $\circ$ to be a function from (epistemic states $\times$ formulas) to epistemic states. We can easily modify the AGM postulates to deal with such revision operators on epistemic states. We start by assuming that there is a set of epistemic states and a function $\mathrm{Bel}(\cdot)$ that maps epistemic states to belief states. We then have analogues to each of the AGM postulates, obtained by replacing each belief set by the beliefs in the corresponding epistemic state. For example, we have

(R1′) $E \circ \varphi$ is an epistemic state

(R2′) $\varphi \in \mathrm{Bel}(E \circ \varphi)$

(R3′) $\mathrm{Bel}(E \circ \varphi) \subseteq Cl(\mathrm{Bel}(E) \cup \{\varphi\})$

and so on, with the obvious transformation.[8]

We can get strong representation theorems if we work at the level of epistemic states. Given a language $\mathcal{L}_e$ (with an associated consequence relation $\vdash_{\mathcal{L}_e}$), let $\mathcal{E}_{\mathcal{L}_e}$ consist of all finite sequences of formulas in $\mathcal{L}_e$. Note that we allow $\mathcal{E}_{\mathcal{L}_e}$ to include sequences of formulas whose conjunction is inconsistent. We define revision in $\mathcal{E}_{\mathcal{L}_e}$ in the obvious way: if $E \in \mathcal{E}_{\mathcal{L}_e}$, then $E \circ \varphi = E \cdot \varphi$.

Theorem 5.5: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ whose local states are $\mathcal{E}_{\mathcal{L}_e}$. There is a function $Bel_{\mathcal{I}}$ that maps epistemic states to belief states such that*

- *if s_a is a local state of the agent in $\mathcal{I}$, then $Bel(\mathcal{I}, s_a) = Bel_{\mathcal{I}}(s_a)$, and*
- *$(\circ, Bel_{\mathcal{I}})$ satisfies R1′–R8′.*

Proof: See Appendix A.1. ∎

Notice that, by definition, we have $\mathrm{Bel}_{\mathcal{I}}(\mathcal{I}, \langle\rangle \circ_{\mathcal{I}} \varphi_1 \circ_{\mathcal{I}} \ldots \circ_{\mathcal{I}} \varphi_k) = \mathrm{Bel}_{\mathcal{I}}(\mathcal{I}, \langle \varphi_1, \ldots, \varphi_k \rangle)$, so, at the level of epistemic states, we get an analogue to Theorem 5.3. We remark that to ensure that R5′ holds for $(\circ, \mathrm{Bel}_{\mathcal{I}})$, we need to define $\mathrm{Bel}_{\mathcal{I}}(E)$ appropriately for sequences $E \in \mathcal{E}_{\mathcal{I}}$ whose conjunction is inconsistent.

Theorem 5.5 shows that any system in $\mathcal{C}^R$ corresponds to a revision operator over epistemic states that satisfies the generalized AGM postulates. We would hope that the converse also holds. Unfortunately, this is not quite the case. There are revision operators on epistemic states that satisfy the generalized AGM postulates but do not correspond to a system in $\mathcal{C}^R$. This is because systems in $\mathcal{C}^R$ satisfy an additional postulate:

(R9′) If $\not\vdash_{\mathcal{L}_e} \neg(\varphi \wedge \psi)$ then $\mathrm{Bel}(E \circ \varphi \circ \psi) = \mathrm{Bel}(E \circ \varphi \wedge \psi)$.

[8] The only problematic postulate is R6. The question is whether R6′ should be "If $\vdash_{\mathcal{L}_e} \varphi \Leftrightarrow \psi$ then $\mathrm{Bel}(E \circ \varphi) = \mathrm{Bel}(E \circ \psi)$" or "If $\vdash_{\mathcal{L}_e} \varphi \Leftrightarrow \psi$ then $E \circ \varphi = E \circ \psi$". Dealing with either version is straightforward. For definiteness, we adopt the first alternative here.

We show that R9$'$ is sound in $\mathcal{C}^R$ by proving the following strengthening of Theorem 5.5.

Proposition 5.6: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ whose local states are $\mathcal{E}_{\mathcal{L}_e}$. There is a function $Bel_{\mathcal{I}}$ that maps epistemic states to belief states such that*

- *if s_a is a local state of the agent in $\mathcal{I}$, then $Bel(\mathcal{I}, s_a) = Bel_{\mathcal{I}}(s_a)$, and*
- *$(\circ, Bel_{\mathcal{I}})$ satisfies R1$'$–R9$'$.*

We can prove the converse to Proposition 5.6: a revision system on epistemic states that satisfies the generalized AGM postulates *and R9$'$* does correspond to a system in $\mathcal{C}^R$.

Theorem 5.7: *Given a function $Bel_{\mathcal{L}_e}$ mapping epistemic states in $\mathcal{E}_{\mathcal{L}_e}$ to belief sets over $\mathcal{L}_e$ such that $Bel_{\mathcal{L}_e}(\langle\rangle)$ is consistent and $(Bel_{\mathcal{L}_e}, \circ)$ satisfies R1$'$–R9$'$, there is a system $\mathcal{I} \in \mathcal{C}^R$ whose local states are in $\mathcal{E}_{\mathcal{L}_e}$ such that $Bel_{\mathcal{L}_e}(s_a) = Bel(s_a)$ for each local state s_a in $\mathcal{I}$.*

Proof: See Appendix A.1. ▮

Notice that, by definition, for the system $\mathcal{I}$ of Theorem 5.7, we have $\mathrm{Bel}(\langle\rangle \circ \varphi_1 \circ \ldots \circ \varphi_k) = \mathrm{Bel}(\langle \varphi_1, \ldots, \varphi_k\rangle)$ as long as $\varphi_1 \wedge \ldots \wedge \varphi_k$ is consistent.

6 Capturing Update

Update tries to capture the intuition that there is a preference for runs where all the observations made are true, and where changes from one point to the next along the run are minimized. To capture the notion of "minimal change from world to world", we use a *distance function d* on worlds.[9] Given two worlds w and w', $d(w, w')$ measures the distance between them. Distances might be incomparable, so we require that d map pairs of worlds into a *partially ordered* domain with a unique minimal element 0 and that $d(w, w') = 0$ if and only if $w = w'$.

We start with some preliminary definitions. Let $\mathcal{I}$ be a BCS, and let $s_0, \ldots, s_n$ be a set of environment states in $\mathcal{I}$. We define $[s_0, \ldots, s_n]$ as the set of runs where $r_e(i) = s_i$ for all $0 \le i \le n$. Thus, $[s_0, \ldots, s_n]$ describes a set of runs that share a common prefix of environment states. A prior plausibility space $\mathcal{P}_a = (\mathcal{R}, \mathrm{Pl}_a)$ is *consistent* with a distance measure d if the following holds:

> $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$ if and only if there is some $j < n$ such that $s_k = s'_k$ for all $0 \le k \le j$, $s_{j+1} \neq s'_{j+1}$, and $d(s_j, s_{j+1}) < d(s_j, s'_{j+1})$.

Intuitively, we compare events of the form $[s_0, \ldots, s_n]$ using a lexicographic ordering based on d. Notice that this ordering focuses on the *first* point of difference. Runs with the a smaller change at this point is preferred, even if later there are abnormal changes. This point is emphasized in the borrowed car example below.

Pl_a is *prefix-defined* if the plausibility of an event is uniquely defined by the plausibility of run-prefixes that are contained in it, so that

[9] Katsuno and Mendelzon identify a "world" with a truth assignment to the primitive propositions. For us, this is just the environment state.

$\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_m]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_m])$ if and only if for all $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_m] - \mathcal{R}[\varphi_0, \ldots, \varphi_m]$ there is some $[s'_0, \ldots, s'_m] \subseteq \mathcal{R}[\varphi_0, \ldots, \varphi_m]$ such that $\mathrm{Pl}_a([s'_0, \ldots, s'_m]) > \mathrm{Pl}_a([s_0, \ldots, s_m])$.

Roughly speaking, this requirement states that we compare events by properties of dominance. This property is similar to one satisfied by the plausibility measures that we get from preference ordering using the construction of Proposition 2.2.

We define the set $\mathcal{C}^U$ to consist of BCSs $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ that satisfy the following four requirements UPD1–UPD4. UPD1 says that there are only finitely many possible truth assignments, and that there is a one-to-one map between environment states and truth assignments.

UPD1 The set Φ_e of propositions (of BCS1) is finite and π is such that for all environment states s, s', if $s \neq s'$, then there is a formula $\varphi \in \mathcal{L}_e$ such that $s \models \varphi$ and $s' \models \neg\varphi$.

UPD2–UPD4 are analogues to REV2–REV4. Like REV2, UPD2 puts constraints on the form of the prior, but now we consider lexicographic priors of the form described above.

UPD2 The prior of BCS5 is prefix defined and consistent with some distance measure.

Recall that REV3 requires only that that all truth assignments initially have nontrivial plausibility. In the case of revision, the truth assignment does not change over time, since we are dealing with static propositions. In the case of update, the truth assignment may change over time, so UPD3 requires that all consistent sequences of truth assignments have nontrivial plausibility.

UPD3 If $\varphi_i \in \mathcal{L}_e$, $i = 0, \ldots, k$, are consistent formulas, then $\mathrm{Pl}(\mathcal{R}[\varphi_0, \ldots, \varphi_k]) > \bot$.

Finally, like REV4, UPD4 requires that the agent gain no information from her observations beyond the fact that they are true.

UPD4 $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_{k+1}; o_1, \ldots, o_k]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_{m+1}; o_1, \ldots, o_m])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_m \wedge o_m, \varphi_{m+1}]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \psi_1 \wedge o_1, \ldots, \psi_m \wedge o_m, \psi_{m+1}])$

We remark that in the presence of REV1, UPD4 is equivalent to REV4. We might consider generalized versions of UPD4, where the two sequences of formulas can have arbitrary relative lengths; this version suffices for our purposes. We can also define an analogue UPD4$'$ in the spirit of REV4$'$, which applies only if $\mathrm{Pl}(\mathcal{R}[\varphi_0, \ldots, \varphi_{m+1}; o_1, \ldots, o_m]) > \bot$.

We now show that $\mathcal{C}^U$ corresponds to Katsuno and Mendelzon's notion of update. Recall that Katsuno and Mendelzon define an update operator as mapping a pair of formulas (μ, φ), where μ describes the agent's beliefs and φ describes the observation, to a new formula $\mu \diamond \varphi$ that describes the agent's new beliefs. However, as we discussed in Section 3, when Φ_e is finite, we can also treat $\diamond$ mapping a belief state and a formula to a new belief state. Also recall that $\mathrm{Bel}(\mathcal{I}, s_a)$ is the agent's belief set when her local state is s_a.

Theorem 6.1: *A belief change operator $\diamond$ satisfies U1–U8 if and only if there is a system $\mathcal{I} \in \mathcal{C}^U$ such that*

$$Bel(\mathcal{I}, s_a) \diamond \psi = Bel(\mathcal{I}, s_a \cdot \psi)$$

for all epistemic states s_a and formulas $\psi \in \mathcal{L}_e$.

Proof: See Appendix A.2. ∎

This result immediately generalizes to sequences of updates:

Corollary 6.2: *A belief change operator $\diamond$ satisfies U1–U8 if and only if there is a system $\mathcal{I}_\diamond \in \mathcal{C}^U$ such that for all $\psi_1, \ldots, \psi_k \in \mathcal{L}_e$, we have*

$$Bel(\mathcal{I}_\diamond, s_a) \diamond \psi_1 \diamond \ldots \diamond \psi_k = Bel(\mathcal{I}_\diamond, s_a \cdot \psi_1 \cdot \ldots \cdot \psi_k).$$

These results show that for update, unlike revision, the systems we consider are such that the belief state does determine the result of the update, i.e., if $\mathrm{Bel}(\mathcal{I}, s_a) = \mathrm{Bel}(\mathcal{I}, s'_a)$, then for any φ we get that $\mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi) = \mathrm{Bel}(\mathcal{I}, s'_a \cdot \varphi)$. Roughly speaking, the reason is that the distance measure that determines the prior does not change over time. While this allows us to get an elegant representation theorem, it also causes problems for the applicability of update, as we shall see below.

Note that, since the world is allowed to change, there is no problem if we update by a sequence $\psi_1, \ldots, \psi_k$ of consistent formulas such that $\psi_1 \wedge \ldots \wedge \psi_k$ is inconsistent. There is no requirement that the formulas $\psi_1, \ldots, \psi_k$ be true simultaneously. All that matters is that ψ_i is true at time i. Also note that an update by an inconsistent formula does not pose a problem for our framework. It follows from postulates U1 and U2 that once the agent learns an inconsistent formula (i.e., *false*), she believes *false* from then on.

How reasonable is the notion of update? As the discussion of UPD2 above suggests, it has a preference for deferring abnormal events. This makes it quite similar to Shoham's *chronological ignorance* [1988], and it suffers from some of the same problems. Consider the following story, that we call the *borrowed-car example*.[10] At time 1, the agent parks her car in front of her house with a full fuel tank. At time 2, she is in her house. At time 3, she returns outside to find the car still parked where she left it. Since the agent does not observe the car while she is inside the house, there is no reason for her to revise her beliefs regarding the car's location. Since she finds it parked at time 3, she still has no reason to change her beliefs. Now, what should the agent believe when, at time 4, she notices that the fuel tank is no longer full? The agent may want to consider a number of possible explanations for her time-4 observation, depending on what she considers to be the most likely sequence(s) of events between time 1 and time 4. For example, if she has had previous gas leaks, then she may consider leakage to be the most plausible explanation. On the other hand, if her spouse also has the car keys, she may consider it possible that he used the car in her absence. Update, however, prefers to defer abnormalities, so it will conclude that the fuel must have disappeared, for inexplicable reasons, between times 3 and 4. To see this, note that runs where the car has been taken on a ride have an abnormality

[10] This example is based on Kautz's stolen car story [1986], and is due to Boutilier, who independently observed this problem [private communication, 1993].

at time 2, while runs where the car did not move at time 2 but the fuel suddenly disappeared, have their first abnormality at time 4, and thus are preferred!

We emphasize that the counterintuitive conclusion drawn in this example is not an artifact of our representation, but inherent in the definition of update. We can formalize the example using propositions such as *car-in-lot, fuel-tank-full*, etc. The observation of $\neg$*fuel-tank-full* at time 4 must be explained by some means, and an update operator will explain it in terms of a change that occurred in states consistent with the beliefs at time 3 (i.e., *car-in-lot, fuel-tank-full*). The exact change assumed will depend on the distance function embodied by the update operator. The key point is that update will not go back and revise the earlier beliefs about what happened between times 1 and 2.

So are there any situations where update is reasonable? Of course, this depends on how we interpret "reasonable". We briefly consider one approach here.

In a world w, the agent has some beliefs that are described by, say, the formula φ. These beliefs may or may not be *correct* (where we say a belief φ is correct in a world w if φ is true of w). Suppose something happens and the world changes to w'. As a result of the agent's observations, she has some new beliefs, described by φ'. Again, there is no reason to believe that φ' is correct. Indeed, it may be quite unreasonable to expect φ' to be correct, even if φ is correct. Consider the borrowed-car example. Suppose that while the agent was sitting inside the house, the car was, in fact, taken for a ride. Nevertheless, the most reasonable belief for the agent to hold when she observes that the car is still in the parked after she leaves the house is that it was there all along.

The problem here is that the information the agent obtains at times 2 and 3 is insufficient to determine what happened. We cannot expect all the agent's beliefs to be correct at this point. On the other hand, if she does obtain sufficient information about the change and her beliefs were initially correct, then it seems reasonable to expect that her new beliefs will be correct. But what counts as *sufficient* information?

We say that φ provides *sufficient information* about the change from w to w' if there is no world w'' satisfying φ such that $d(w, w'') < d(w, w')$. In other words, φ is sufficient information if, after observing φ in world w, the agent will consider the real world (w') one of the most likely worlds. Note that this definition is monotonic, in that if φ is sufficient information about the change, then so is any formula ψ that implies φ (as long as it holds at w'). Moreover, this definition depends on the agent's distance function d. What constitutes sufficient information for one agent might not for another. We would hope that the function d is realistic in the sense that the worlds judged closest according to d really are the most likely to occur.

We can now show that update has the property that if the agent has correct beliefs and receives sufficient information about a change, then she will continue to have correct beliefs.

Theorem 6.3: *Let $\mathcal{I} \in \mathcal{C}^U$. If the agent's beliefs at (r, m) are correct and $o_{(r,m)}$ provides sufficient information about the change from $r_e(m)$ to $r_e(m+1)$, then the agent's beliefs at $(r, m+1)$ are correct.*

Proof: Straightforward; left to the reader. ∎

As we observed earlier, we cannot expect the agent to always have correct beliefs. Nevertheless, we might hope that if the agent does (eventually) receive sufficiently detailed information, then she should realize that her beliefs were incorrect. But this is precisely what does *not* happen in the borrowed-car example. Intuitively, once the agent observes that the fuel tank is not full, this should be sufficient information to eliminate the possibility that the car remained in the parking lot. However, it is not. Roughly speaking, this is because update focuses only on the current state of the world, and thus cannot go back and revise beliefs about the past.

The problem here is again due to the fact that belief update is determined only by the agent's belief state and not her epistemic state. Thus, update can only take into account the agent's current beliefs and not other information, such as the sequence of observations that led to these beliefs. In our example, if we limit our attention to beliefs about the car's whereabouts and the fuel tank, then since the agent has the same belief state at time 1 and 3, she must change her beliefs in the same manner at both times. This implies that the observation the fuel tank is not full at time 4 cannot be sufficient information about the past, since a fuel leak might be the most plausible explanation of missing fuel at time 2.

Our discussion of update shows that update is guaranteed to be safe only in situations where there is always enough information to characterize the change that has occurred. While this may be a plausible assumption in database applications, it seems somewhat less reasonable in AI examples, particularly in cases involving reasoning about action.[11]

7 Synthesis

In previous sections we analyzed belief revision and belief update separately. We provided representation theorems for both notions and discussed issues specific to each notion. In this section, we try to identify some common themes and points of difference.

Katsuno and Mendelzon [1991a] focused on the following three differences between revision and update:

1. Revision deals with static propositions, while update allows propositions that are not static.

2. Revision and update treat inconsistent belief states differently. Revision allows the agent to "recover" from an inconsistent state after observing a consistent formula. Update dictates that once the agent has inconsistent beliefs, she will continue to have inconsistent beliefs. As we noted above, it seems that revision's ability to recover from an inconsistent belief set leads to several technical oddities in iterated revision.

3. Revision considers only total preorders, while update allows partial preorders.

We remark that while the restriction to static propositions may seem to be a serious limitation of belief revision, we can always convert a dynamic proposition to a static one by adding time stamps. That is, we can replace a proposition p by a family of propositions p^m that stand for "p is true at time m". Thus, it is possible to use revision to reason about a changing world. (Of course, it would then be necessary to capture connections between propositions of the form p^m, but specific revision operators could certainly do this.)

[11] Similar observations were independently made by Boutilier [1994b], although his representation is quite different from ours.

Table 1: A summary of the restrictions we had to impose to capture revision and update.

Restriction on	Revision	Update
Environment changes	No change	All possible sequences
Initial plausibility	Total preorder	Lexicographic
Belief change	Conditioning	Conditioning

In any case, our framework suggests a different approach to categorizing the differences between revision and update (and other approaches to belief change): focusing on the restrictions that have to be added to basic BCSs to obtain systems in $\mathcal{C}^R$ and $\mathcal{C}^U$, respectively. In particular, we should focus on three aspects of a system:

- How does the environment state change?
- How does the agent form her initial beliefs? What regularities appear in the agent's beliefs at the initial state?
- How does the agent change her beliefs?

Table 1 summarizes the answers to these questions for revision and update. The table highlights the different restrictions imposed by each. Revision puts a severe restriction on changes of the environment (i.e., the environment cannot change during revision), and a rather mild restriction on the agent's prior beliefs (i.e., they must form a total preorder). On the other hand, update allows all sequences of environment states, but requires the agent's prior beliefs to have a specific form. These formal properties match the intuitive description of revision and update given in [Alchourrón, Gärdenfors, and Makinson 1985; Katsuno and Mendelzon 1991b]. However, the explicit representation of time in our framework allows us to make these intuitions precise. Moreover, our framework makes explicit other assumptions made by revision and update. For example, the lexicographic nature of update is not immediately evident from the presentation in [Katsuno and Mendelzon 1991b].

The key point to notice in this table is that belief change in both revision and update is done by conditioning. This observation, and the naturalness of conditioning as a notion of change, support our claim that conditioning should be adopted as semantic foundations for minimal change.

8 Extensions

In the preceding sections, we introduced several assumptions that were needed to capture revision and update. Of course, there are other ways of capturing these notions that require somewhat different assumptions. Nevertheless, these assumptions give insight into the underlying choices made, either explicitly or implicitly, in the definition of revision and update. In addition, thinking in terms of such restrictions makes it straightforward to extend the intuitions of revision and update beyond the context where they were originally applied. In this section, we consider a number of such extensions, to illustrate our point.

Knowledge. In many domains of interest, the agent knows that some sequences of observations are impossible. We already saw in the circuit-diagnosis problem that observing failures was impossible. In the context of update, we know that we cannot observe a person die and then be alive, despite the fact that both being dead and being alive are consistent states.

We can easily maintain what we regard as the defining properties of revision and update, as discussed in the previous section: no change in the environment state and a ranked prior in the case of revision, and a lexicographic prior in the case of update, with belief change proceeding by conditioning in both cases. We simply drop REV3 and replace REV4 by REV4$'$ (resp., drop UPD3 and replace UPD4 and UPD4$'$). We remark that this change affects the postulates. For example, consider update. Suppose that the agent considers the possibility that Mr. Bond is dead. If she then observes Mr. Bond alive and well then, according to update, she must account for the new observation by some change from the worlds she previously considered possible. However, there is no transition from worlds in which Mr. Bond is dead that can account for the new observation. Thus, once the agent knows that certain transitions are impossible, some observations (e.g., observing that Mr. Bond is alive) require her to remove from consideration some of the worlds that she previously considered possible. As a consequence, postulate U8 does not hold, since the agent's new beliefs are not determined by a pointwise update at each of the worlds she previously considered possible.

Language of Beliefs. In our analysis of revision and update, we focused on the agent's beliefs about the current state of the environment. Often we are also interested in how the agent changes her beliefs about other types of statements, such as beliefs about future states of the environment, beliefs about other agents' beliefs, and introspective beliefs about her own beliefs. Again, it is straightforward in our framework to deal with an enriched language that lets us express such statements. For example, in [Friedman and Halpern 1994] we examine *Ramsey conditionals*. These are formulas of the form $\varphi > \psi$, which can be read as saying "after learning φ, the agent believes ψ". This formula can be expressed as $learn(\varphi) \Rightarrow B\psi$ in the language $\mathcal{L}^{KPT}$. As is well known, if belief sets include Ramsey conditionals (and not just propositional formulas), then the AGM postulates become inconsistent (at least, provided we have at least three mutually exclusive consistent formulas in the language) [Gärdenfors 1986]. Similar inconsistency results arise when one tries to add other forms of introspective beliefs [Fuhrmann 1989]. In our setting, it is easy to see why the problem arises. Even if we allow belief sets to include nonpropositional formulas, it still seems quite clear that we want to distinguish the propositional formulas from formulas that talk explicitly about an agent's beliefs. For example, it is not clear that we should allow an observation of a formula such as $\varphi > \psi$. What would it mean to observe such a formula? It clearly seems quite different from observing a propositional formula. Nor does it make sense to extend an assumption such as REV1 to arbitrary formulas. While it may be reasonable to restrict to static propositions if we are viewing these as making statements about a relatively stable environment, it seems far less reasonable to assume that formulas that talk about an agent's beliefs will be static, especially when we are trying to model belief change!

Of course, if we allow only propositional formulas to be learned (or observed), and restrict REV1 to propositional formulas, then it is easy to see that all of our results still hold, even if the full language is quite rich; we avoid the triviality result completely.

Observations One of the strongest assumptions made by revision and update involves the treatment of observations. This assumption seems unreasonable in most domains. REV4 and UPD4 essentially assume that the observation that the agent makes is chosen randomly among all formulas consistent with the current state of the world. Suppose that φ says that the agent is outdoors, ψ says that the agent is in the basement, and o_1 says that the basement light is on. We may well have $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o_1]) > \mathrm{Pl}_a(\mathcal{R}[\psi \wedge o_1])$. For example, the agent may hardly ever go to the basement and frequently go outdoors, but her children may often leave the basement light on. Nevertheless, we may also have $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1]) < \mathrm{Pl}_a(\mathcal{R}[\psi; o_1])$, contradicting REV4. Indeed, it may well be impossible for the agent to observe that the basement light is on when she is outdoors, so that $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1]) = \perp$, but this is not permitted according to REV4 or UPD4.

In many domains it is useful to reason about hidden quantities that simply cannot be observed. For example, the event that component c_i is faulty in Example 5.1 is a basic event in our description of the problem, yet it cannot be observed. Similarly, the event where a patient has a disease X or the opponent is planning to capture the queen are useful in reasoning about medical diagnosis and game strategy, yet are not directly observable in practice. Thus, the requirement that all formulas in the language can be observed seems quite unnatural. We note that explicitly modeling sensory input is a standard practice in control theory and stochastic processes (e.g., in hidden Markov chains). In these fields, one models the probability of an observation in various situations. Making an observation increases the probability of situations where that observation is likely to be observation and decreases the probability of situations where it is unlikely. Again, it is straightforward to consider a more detailed model of the observation process in our framework; see [Friedman 1997, Chapter 6].

Actions Our definition of belief change systems essentially assumes that the agent is *passive*. The situation is more complex when the agent can influence the environment. The agent's choice of action interacts with her beliefs. It is clear that after performing an action, the agent should change her beliefs.[12] Moreover, the information content of observations depends on the action the agent has just performed. For example, the agent might consider hearing a loud noise to be surprising. However, it would be expected after the agent pulls the trigger of her gun.

This list of possible extensions is clearly not exhaustive; there are many others that we may want to consider. Nevertheless, these are extensions that seem to be of interest. The main points we want to make here are (1) it is easy to accommodate these extensions in our framework while still maintaining the main characteristics of revision and update, and (2) it is difficult to deal with such extensions if we focus on postulates.

[12] Indeed, an alternative interpretation of the update postulates is that they describe how the agent should update her beliefs after doing the action "achieve φ" [Goldszmidt and Pearl 1992; del Val and Shoham 1992; del Val and Shoham 1993]. However, as these works show, the update postulates are problematic under this interpretation.

9 Conclusion

We have shown how the framework introduced in [Friedman and Halpern 1997a] can be used to capture belief revision and update. Modeling revision and update in the framework also gives us a great deal of further insight into their properties, and emphasizes the role of conditioning as a way of capturing minimal change.

Of course, revision and update are but two points in a wide spectrum of possible types of belief change. Our ultimate goal is to use this framework to understand the whole spectrum better and to help us design belief change operations that overcome some of the difficulties we have observed with revision and update. In particular, we want belief change operations that can handle dynamic propositions, while still being able to revise information about the past.

Our framework suggests how to construct such belief change operations. In this framework, belief change operations can be determined by choosing a plausibility measure that captures the agent's preferences among sequences of worlds. This is the agent's prior plausibility, and captures her initial beliefs about the relative likelihood of runs. As the agent receives information, she changes her beliefs using conditioning. In this paper we show that revision and update correspond to two specific families of priors. Clearly, however, there are prior plausibilities that, when conditioned on a surprising observation, allow the agent to revise some earlier beliefs and to assume that some change has occurred. One obvious problem is that, even if there are only two possible states, there are uncountably many possible runs. How can an agent describe a prior plausibility over such a complex space?

One approach to doing this is based on intuition from the probabilistic settings. In these settings, the standard solution to this problem is to assume that state transitions are independent of when they occur, that is, that the probability of the system going from state s to state s' is independent of the sequence of transitions that brought the system to state s. This *Markov assumption* significantly reduces the complexity of the problem. All that is necessary is to describe the probability of state transitions. In [Friedman and Halpern 1996b; Friedman 1997] we define a notion of plausibilistic independence, and show how to describe priors that satisfy the Markov assumption and the consequences for belief change.

Whether or not this particular approach turns out to be a useful one, it is clear that these are the types of questions we should be asking. As these works show, our framework provides a useful basis for answering them.

Finally, we note that our approach is quite different from the traditional approach to belief change [Alchourrón, Gärdenfors, and Makinson 1985; Gärdenfors 1988; Katsuno and Mendelzon 1991a]. Traditionally, belief change was viewed as an abstract process. Our framework, on the other hand, models the agent and the environment she is situated in, and how both change in time. This allows us to model concrete agents in concrete settings (for example, diagnostic systems are analyzed in [Friedman and Halpern 1997a] and throughout this paper), and to reason about the beliefs and knowledge of such agents. We can then investigate what plausibility ordering induces beliefs that match our intuitions. By gaining a better understanding of such concrete situations, we can better investigate more abstract notions of belief change. More generally, we believe that, when studying belief change, it is important to specify the underlying *ontology*: that is, exactly what scenario underlies the belief-change process. We have specified one such scenario here. While others are certainly possible, we view it as a defect in the

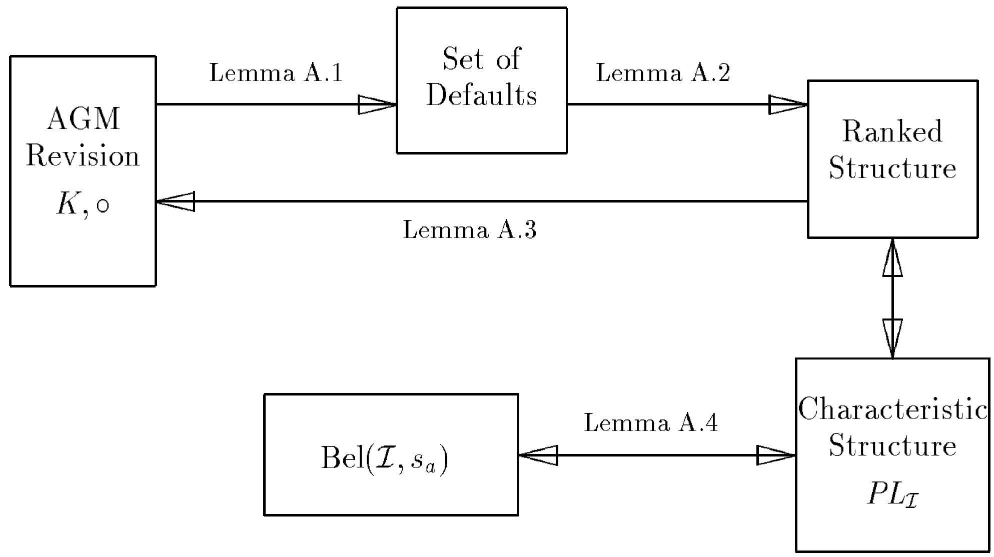

Figure 1: Schematic description of the entities and lemmas involved in the proof of Theorems 5.2 and 5.3.

literature on belief change that the underlying scenario is so rarely discussed. The framework we have introduced here provides a way of making formal what the scenario is. (See [Friedman and Halpern 1996a] for further discussion of this issue.)

Acknowledgments

The authors are grateful to Craig Boutilier, Ronen Brafman, Adnan Darwiche, Moises Goldszmidt, Adam Grove, Alberto Mendelzon, Alvaro del Val, and particularly Daphne Koller and Moshe Vardi, for comments on drafts of this paper and useful discussions relating to this work.

A Proofs

A.1 Proofs for Section 5

We start with the proof of Theorems 5.2 and 5.3. To do this, we need some preliminary definitions and lemmas. Figure 1 shows the general outline of the intermediate representations we use in these proofs. Roughly speaking, we show how to map from a revision operator $\circ$ and a consistent belief set K to a ranking, and similarly how to map from a ranking to an AGM revision operator. These rankings correspond, in a direct way, to priors in systems in $\mathcal{C}^R$, and thus have close connection to the beliefs of the agent in various states.

These mapping between AGM revision operators and rankings are related to the representation theorems of Boutilier [1994c], Grove [1988], and Katsuno and Mendelzon [1991a]. However, the exact details of our representations are different than those of Boutilier, Grove, and Katsuno and Mendelzon. Thus, for completeness we provide the full proofs here.

We start with the mapping from revision operator applied to a specific belief set to a ranking. As an intermediate step we construct a set of defaults as follows. We then will use the results from [Friedman and Halpern 1997b] to construct a ranked plausibility structure that satisfies these defaults.

Lemma A.1: *Let $\circ$ be an AGM revision operator, let $K \subseteq \mathcal{L}_e$ be a consistent belief set, and let*

$$\Delta_{(\circ,K)} = \{\varphi{\rightarrow}\psi : \varphi, \psi \in \mathcal{L}_e,\ \psi \in K \circ \varphi\}.$$

Then the following is true:

- *(a)* $\Delta_{(\circ,K)}$ *is closed under the rules of system* **P**,
- *(b)* $\varphi{\rightarrow}\mathit{false} \notin \Delta_{(\circ,K)}$ *for all consistent* $\varphi \in \mathcal{L}_e$, *and*
- *(c)* $\Delta_{(\circ,K)}$ *satisfies rational monotonicity; that is, if* $\varphi{\rightarrow}\psi \in \Delta_{(\circ,K)}$ *and* $\varphi{\rightarrow}\neg\xi \notin \Delta_{(\circ,K)}$, *then* $\varphi \wedge \xi{\rightarrow}\psi \in \Delta_{(\circ,K)}$.

Proof: We start with part (a):

LLE Assume that $\vdash_{\mathcal{L}_e} \varphi \equiv \varphi'$ and that $\varphi{\rightarrow}\psi \in \Delta_{(\circ,K)}$. Thus, $\psi \in K \circ \varphi$. From R5, it follows that $\psi \in K \circ \varphi'$, and thus $\varphi'{\rightarrow}\psi \in \Delta_{(\circ,K)}$.

RW Assume that $\vdash_{\mathcal{L}_e} \psi \Rightarrow \psi'$ and that $\varphi{\rightarrow}\psi \in \Delta_{(\circ,K)}$. Thus, $\psi \in K \circ \varphi$. Since $K \circ \varphi$ is a belief set, it is closed under logical consequence. In particular, $\psi' \in K \circ \varphi$, and hence $\varphi{\rightarrow}\psi' \in \Delta_{(\circ,K)}$.

REF By R2, $\varphi \in K \circ \varphi$, and thus, $\varphi{\rightarrow}\varphi \in \Delta_{(\circ,K)}$.

AND Assume that $\varphi{\rightarrow}\psi_1, \varphi{\rightarrow}\psi_2 \in \Delta_{(\circ,K)}$. Thus, $\psi_1, \psi_2 \in K \circ \varphi$. Since $K \circ \varphi$ is a belief set, $\psi_1 \wedge \psi_2 \in K \circ \varphi$. Thus, $\varphi{\rightarrow}\psi_1 \wedge \psi_2 \in \Delta_{(\circ,K)}$.

OR Assume that $\varphi_1{\rightarrow}\psi, \varphi_2{\rightarrow}\psi \in \Delta_{(\circ,K)}$. There are two cases. If $K \circ (\varphi_1 \vee \varphi_2)$ is inconsistent, then $\psi \in K \circ (\varphi_1 \vee \varphi_2)$ and thus $\varphi_1 \vee \varphi_2{\rightarrow}\psi \in \Delta_{(\circ,K)}$. If $K \circ (\varphi_1 \vee \varphi_2)$ is consistent, then, by R2, $\varphi_1 \vee \varphi_2 \in K \circ (\varphi_1 \vee \varphi_2)$. Thus, we cannot have both $\neg\varphi_1$ and $\neg\varphi_2$ in $K \circ (\varphi_1 \vee \varphi_2)$. Without loss of generality, assume that $\neg\varphi_1 \notin K \circ (\varphi_1 \vee \varphi_2)$. Using R7 and R8, we get that $K \circ ((\varphi_1 \vee \varphi_2) \wedge \varphi_1) = Cl(K \circ (\varphi_1 \vee \varphi_2) \cup \{\varphi_1\})$. Using R6, we get that $K \circ ((\varphi_1 \vee \varphi_2) \wedge \varphi_1) = K \circ \varphi_1$. Thus, we conclude that $K \circ \varphi_1 = Cl(K \circ (\varphi_1 \vee \varphi_2) \cup \{\varphi_1\})$. Since $\varphi_1{\rightarrow}\psi \in \Delta_{(\circ,K)}$, we have that $\psi \in K \circ \varphi_1$. Thus, we get that $\varphi_1 \Rightarrow \psi \in K \circ (\varphi_1 \vee \varphi_2)$. If $\neg\varphi_2 \notin K \circ (\varphi_1 \vee \varphi_2)$, by similar arguments we get that $\varphi_2 \Rightarrow \psi \in K \circ (\varphi_1 \vee \varphi_2)$. This implies that $(\varphi_1 \vee \varphi_2) \Rightarrow \psi \in K \circ (\varphi_1 \vee \varphi_2)$, and thus $\psi \in K \circ (\varphi_1 \vee \varphi_2)$. On the other hand, if $\neg\varphi_2 \in K \circ (\varphi_1 \vee \varphi_2)$, then, since $\varphi_1 \vee \varphi_2 \in K \circ (\varphi_1 \vee \varphi_2)$, we get that $\varphi_1 \in K \circ (\varphi_1 \vee \varphi_2)$, and thus $\psi \in K \circ (\varphi_1 \vee \varphi_2)$.

CM Assume that $\varphi{\rightarrow}\psi_1, \varphi{\rightarrow}\psi_2 \in \Delta_{(\circ,K)}$. If $K \circ \varphi$ is inconsistent, then using R5 we get that φ is inconsistent. Thus, $\varphi \wedge \psi_1$ is inconsistent, so $\psi_2 \in K \circ (\varphi \wedge \psi_1)$. Now assume that $K \circ \varphi$ is consistent. Since $\varphi{\rightarrow}\psi_1$, we have that $\psi_1 \in K \circ \varphi$. Since $K \circ \varphi$ is consistent, we get that $\neg\psi_1 \notin K \circ \varphi$. Applying R8, we get that $K \circ \varphi \subseteq K \circ (\varphi \wedge \psi_1)$. Since $\varphi{\rightarrow}\psi_2 \in \Delta_{(\circ,K)}$, we have that $\psi_2 \in K \circ \varphi$. Thus, $\psi_2 \in K \circ (\varphi \wedge \psi_1)$. This implies that $(\varphi \wedge \psi_1){\rightarrow}\psi_2 \in \Delta_{(\circ,K)}$.

We now prove part (b). Let $\varphi \in \mathcal{L}_e$ be a consistent formula. Then, using R5, we get that $K \circ \varphi$ is consistent. Thus, $\varphi \rightarrow \mathit{false} \notin \Delta_{(\circ,K)}$.

Finally we prove part (c). Assume that $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$, and $\varphi \wedge \xi \rightarrow \psi \notin \Delta_{(\circ,K)}$. Since $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$, we have that $\psi \in K \circ \varphi$. Now if $\neg\xi \notin K \circ \varphi$, then, using R8, we have that $Cl(K \circ \varphi \cup \{\xi\}) \subseteq K \circ (\varphi \wedge \xi)$. This implies that $\psi \in K \circ (\varphi \wedge \xi)$. However, since we assumed that $\varphi \wedge \xi \rightarrow \psi \notin \Delta_{(\circ,K)}$, we have that $\psi \notin K \circ (\varphi \wedge \xi)$; thus, we get a contradiction. We conclude that $\neg\xi \in K \circ \varphi$. Thus, $\varphi \rightarrow \neg\xi \in \Delta_{(\circ,K)}$. ∎

We now use this result to show that there exists a plausibility structure that corresponds to $\circ$ applied to K.

Lemma A.2: *Let $\circ$ be an AGM revision operator, and let $K \subseteq \mathcal{L}_e$ be a consistent belief set. Then there is a plausibility structure $PL = (W, \mathrm{Pl}, \pi)$ such that Pl is ranked, $PL \models \varphi \rightarrow \psi$ if and only if $\psi \in K \circ \varphi$, and $\mathrm{Pl}([\![\varphi]\!]) > \bot$ for all $\vdash_{\mathcal{L}_e}$-consistent formulas $\varphi \in \mathcal{L}_e$.*

Proof: We use the basic techniques described in the proof of [Friedman and Halpern 1997b, Theorem 8.2]. Let $\Delta_{(\circ,K)}$ be the set of defaults defined by Lemma A.1. We now construct a plausibility space $PL' = (W, \mathrm{Pl}', \pi)$ such that $PL' \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$. We define PL' as follows:

- $W = \{w_V : V \subseteq \mathcal{L}_e \text{ is a maximal } \vdash_{\mathcal{L}_e}\text{-consistent set}\}$,
- $\pi(w_V)(p) = \mathbf{true}$ if $p \in V$, and
- $\mathrm{Pl}'([\![\varphi]\!]) \geq \mathrm{Pl}'([\![\psi]\!])$ if and only if $(\varphi \vee \psi) \rightarrow \varphi \in \Delta_{(\circ,K)}$.

Using [Friedman and Halpern 1997b, Lemma 4.1], we get that $PL' \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$. From Lemma A.1 (c) and and results of [Friedman and Halpern 1997b], it follows that there is a ranked plausibility measure Pl that is *default-isomorphic* to Pl', that is (W, Pl, π) satisfies precisely the same defaults as (W, Pl', π). Let $PL = (W, \mathrm{Pl}, \pi)$.

Since PL is default-isomorphic to PL', we have that $PL \models \varphi \rightarrow \psi$ if and only if $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$. Moreover, using Lemma A.1, we have that $\varphi \rightarrow \psi \in \Delta_{(\circ,K)}$ if and only if $\psi \in K \circ \varphi$. Thus, $PL \models \varphi \rightarrow \psi$ if and only if $\psi \in K \circ \varphi$. Finally, let φ be a $\vdash_{\mathcal{L}_e}$-consistent formula. From Lemma A.1 (b), we get that $\varphi \rightarrow \mathit{false} \notin \Delta_{(\circ,K)}$. Since $\Delta_{(\circ,K)}$ is closed under the rules of system **P**, we conclude that $(\varphi \vee \mathit{false}) \rightarrow \mathit{false} \notin \Delta_{(\circ,K)}$. Thus, $\mathrm{Pl}'([\![\varphi]\!]) \not\leq \bot = \mathrm{Pl}'([\![\mathit{false}]\!])$, and thus $\mathrm{Pl}'([\![\varphi]\!]) > \bot$. Since Pl is default-isomorphic to Pl', we conclude that $\mathrm{Pl}([\![\varphi]\!]) > \bot$. ∎

We now prove the converse to Lemma A.2.

Lemma A.3: *Let $PL = (W, \mathrm{Pl}, \pi)$ be a ranked plausibility structure such that $\pi(w)$ is $\vdash_{\mathcal{L}_e}$-consistent for all worlds w, and $PL \not\models \varphi \rightarrow \mathit{false}$ for all $\vdash_{\mathcal{L}_e}$-consistent formulas $\varphi \in \mathcal{L}_e$; let $K = \{\varphi \in \mathcal{L}_e : PL \models \mathit{true} \rightarrow \varphi\}$. Then there is an AGM revision operator $\circ$ such that $\psi \in K \circ \varphi$ if and only if $PL \models \varphi \rightarrow \psi$.*

Proof: Let $\circ$ be some belief change operation such that $K \circ \varphi = \{\psi : PL \models \varphi \rightarrow \psi\}$. Since this requirement constrains only the result of applying $\circ$ to K, we can assume without loss of generality that $\circ$ satisfies the AGM postulates when applied to belief sets other than K. Thus, we need prove only that $\circ$ satisfies the AGM postulates for revision applied to K. (Note that the proofs for R3 and R4 follow from the proofs for R7 and R8, respectively.)

R1 Since PL is qualitative, we have that $\{\psi : PL \models \varphi \rightarrow \psi\}$ is a belief set, that is, closed under logical consequences.

R2 Axiom C1 implies that $PL \models \varphi \rightarrow \varphi$. Thus, $\varphi \in K \circ \varphi$.

R5 By our assumptions, if φ is $\vdash_{\mathcal{L}_e}$-consistent, then $\mathrm{Pl}([\![\varphi]\!]) > \bot$, and thus $PL \not\models \varphi \rightarrow \mathit{false}$. On the other hand, if φ is not $\vdash_{\mathcal{L}_e}$-consistent, then $[\![\varphi]\!] = \emptyset$, and thus $\mathrm{Pl}([\![\varphi]\!]) = \bot$. We conclude that $\mathrm{Pl}([\![\varphi]\!]) = \bot$ if and only if $\vdash_{\mathcal{L}_e} \neg\varphi$. This implies that $PL \models \varphi \rightarrow \mathit{false}$ if and only if $\vdash_{\mathcal{L}_e} \neg\varphi$. Thus, $K \circ \varphi = Cl(\mathit{false})$ if and only if $\vdash_{\mathcal{L}_e} \neg\varphi$.

R6 Assume that $\vdash_{\mathcal{L}_e} \varphi \Leftrightarrow \varphi'$. Then, by our assumption, $\pi(w)(\varphi) = \pi(w)(\varphi')$. Thus, $[\![\varphi \wedge \psi]\!] = [\![\varphi' \wedge \psi]\!]$ for all formulas $\psi \in \mathcal{L}_e$. We conclude that $PL \models \varphi \rightarrow \psi$ if and only if $PL \models \varphi' \rightarrow \psi$. This implies that $K \circ \varphi = K \circ \varphi'$.

R7 There are two cases: either $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) = \bot$ or $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) > \bot$. If $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) = \bot$, then $\varphi \wedge \psi$ is inconsistent. According to R2, we have that $\varphi \in K \circ \varphi$. Thus, $\varphi \wedge \psi \in Cl(K \circ \varphi \cup \{\psi\})$. This implies that $Cl(K \circ \varphi \cup \{\psi\})$ contains *false*, and thus $K \circ (\varphi \wedge \psi) \subseteq Cl(K \circ \varphi \cup \{\psi\})$. If $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) > \bot$, let $\xi \in K \circ (\varphi \wedge \psi)$. We now show that $\xi \in Cl(K \circ \varphi \cup \{\psi\})$. This will show that $K \circ (\varphi \wedge \psi) \subseteq Cl(K \circ \varphi \cup \{\psi\})$. Since $\xi \in K \circ (\varphi \wedge \psi)$, we get that $PL \models (\varphi \wedge \psi) \rightarrow \xi$. Since $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) > \bot$, we get that $\mathrm{Pl}([\![\varphi \wedge \psi \wedge \xi]\!]) > \mathrm{Pl}([\![\varphi \wedge \psi \wedge \neg\xi]\!])$. Then we have that $\mathrm{Pl}([\![\varphi \wedge (\psi \Rightarrow \xi)]\!]) > \mathrm{Pl}([\![\varphi \wedge \neg(\psi \Rightarrow \xi))]\!])$, since $(\varphi \wedge \psi \wedge \xi) \Rightarrow (\varphi \wedge (\psi \Rightarrow \xi))$ and $(\varphi \wedge \neg(\psi \Rightarrow \xi)) \Rightarrow (\varphi \wedge \psi \wedge \neg\xi)$. This also implies that $\mathrm{Pl}([\![\varphi]\!]) > \bot$. Thus, $PL \models \varphi \rightarrow (\psi \Rightarrow \xi)$. So, $(\psi \Rightarrow \xi) \in K \circ \varphi$, and thus $\xi \in Cl(K \circ \varphi \cup \{\psi\})$.

R8 Assume that $\neg\psi \notin K \circ \varphi$. Let $\xi \in Cl(K \circ \varphi \cup \{\psi\})$. We now show that $\xi \in K \circ (\varphi \wedge \psi)$. This will show that $Cl(K \circ \varphi \cup \{\psi\}) \subseteq K \circ (\varphi \wedge \psi)$. Let $A = [\![\varphi \wedge \neg\psi]\!]$, $B = [\![\varphi \wedge \psi \wedge \xi]\!]$, and $C = [\![\varphi \wedge \psi \wedge \neg\xi]\!]$. It is easy to verify that these sets are pairwise disjoint. Since $\varphi \wedge (\psi \Rightarrow \xi) \equiv (\varphi \wedge \neg\psi) \vee (\varphi \wedge \psi \wedge \xi)$ and $(\varphi \wedge \neg(\psi \Rightarrow \xi)) \equiv (\varphi \wedge \psi \wedge \neg\xi)$, we conclude that $[\![\varphi \wedge (\psi \Rightarrow \xi)]\!] = A \cup B$, and $[\![\varphi \wedge \neg(\psi \Rightarrow \xi)]\!] = C$. Since $\xi \in Cl(K \circ \varphi \cup \{\psi\})$, we have that $(\psi \Rightarrow \xi) \in K \circ \varphi$. This means that $PL \models \varphi \rightarrow (\psi \Rightarrow \xi)$. Thus, either $\mathrm{Pl}([\![\varphi]\!]) = \bot$ or $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(C)$. If $\mathrm{Pl}([\![\varphi]\!]) = \bot$, then according to A1, we get that $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) = \bot$. Thus, $PL \models (\varphi \wedge \psi) \rightarrow \xi$ vacuously, and $\xi \in K \circ (\varphi \wedge \psi)$ as desired.

Now assume that $\mathrm{Pl}(A \cup B) > \mathrm{Pl}(C)$. Since Pl is ranked, it satisfies A4′ and A5′. According to A5′, we get that either $\mathrm{Pl}(A) > \mathrm{Pl}(C)$ or $\mathrm{Pl}(B) > \mathrm{Pl}(C)$. Assume that $\mathrm{Pl}(A) > \mathrm{Pl}(C)$ and $\mathrm{Pl}(B) \not> \mathrm{Pl}(C)$. Then, using A4′, we get that $\mathrm{Pl}(A) > \mathrm{Pl}(B)$. Applying A2, we get that $\mathrm{Pl}(A) > \mathrm{Pl}(B \cup C)$. However since $A = [\![\varphi \wedge \neg\psi]\!]$ and $B \cup C = [\![\varphi \wedge \psi]\!]$, this implies that $\neg\psi \in K \circ \varphi$, which contradicts our assumption. Thus, we conclude that $\mathrm{Pl}(B) > \mathrm{Pl}(C)$. Since $B = [\![\varphi \wedge \psi \wedge \xi]\!]$ and $C = [\![\varphi \wedge \psi \wedge \neg\xi]\!]$, we get that $PL \models (\varphi \wedge \psi) \rightarrow \xi$, and thus $\xi \in K \circ (\varphi \wedge \psi)$.

R3 and **R4** Our definition of $\circ$ implies that $K \circ \mathit{true} = K$. According to R6, we have that $K \circ (\mathit{true} \wedge \varphi) = K \circ \varphi$. Combining these two facts, we get that R3 and R4 are special cases of R7 and R8, respectively.

∎

These results show how to map between ranked plausibility structures and AGM revision operators. We now relate systems in $\mathcal{C}^R$ and ranked plausibility structures. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^R$. Recall that REV2 requires that the prior of $\mathcal{I}$ be a ranking. Thus, we can construct a ranked plausibility structure where worlds are runs in $\mathcal{R}$. We define the *characteristic structure* of $\mathcal{I}$ to be $PL_{\mathcal{I}} = (\mathcal{R}, \mathrm{Pl}_a, \pi_{\mathrm{Pl}_{\mathcal{I}}})$, where Pl_a is the agent's prior over runs and $\pi_{\mathrm{Pl}_{\mathcal{I}}}(r)(p) = \pi(r,0)(p)$ for all $p \in \Phi_e$. Note that $[\![\varphi]\!]_{PL_{\mathcal{I}}} = \mathcal{R}[\varphi]$.

We now use $PL_{\mathcal{I}}$ to describe the beliefs of the agent in each local state.

Lemma A.4: *Let $\mathcal{I} \in \mathcal{C}^R$ and let $s_a = \langle o_1, \ldots, o_m \rangle$. Then $\varphi \in Bel(\mathcal{I}, s_a)$ if and only if $PL_{\mathcal{I}} \models (\bigwedge_{i=1}^m o_i) \rightarrow \varphi$. (By convention, if $m = 0$, we take $(\bigwedge_{i=1}^m o_i)$ to be* **true***.)*

Proof: Let $\mathcal{I} \in \mathcal{C}^R$ and let $s_a = \langle o_1, \ldots, o_m \rangle$. There are two cases: either s_a is a local state in $\mathcal{I}$, or it is not.

If s_a is a local state in $\mathcal{I}$, suppose that $r_a(m) = s_a$. Note that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$ if and only if $\mathrm{Pl}_{(r,m)}([\![\varphi]\!]_{(r,m)}) > \mathrm{Pl}_{(r,m)}([\![\neg\varphi]\!]_{(r,m)})$. Recall that, according to the definition of conditioning, $\mathrm{Pl}_{(r,m)}(\cdot)$ is isomorphic to $\mathrm{Pl}_a(\cdot|\mathcal{R}[\cdot; o_1, \ldots, o_m])$. Thus, $\mathrm{Pl}_{(r,m)}([\![\varphi]\!]_{(r,m)}) > \mathrm{Pl}_{(r,m)}([\![\neg\varphi]\!]_{(r,m)})$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi] \mid \mathcal{R}[\cdot; o_1, \ldots, o_m]) > \mathrm{Pl}_a(\mathcal{R}[\neg\varphi] \mid \mathcal{R}[\cdot; o_1, \ldots, o_m])$. Using C1, this is true if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) > \mathrm{Pl}_a(\mathcal{R}[\neg\varphi; o_1, \ldots, o_m])$. Using REV4, this is true if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge \bigwedge_{i=1}^m o_i]) > \mathrm{Pl}_a(\mathcal{R}[\neg\varphi \wedge \bigwedge_{i=1}^m o_i])$. We get that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge \bigwedge_{i=1}^m o_i]) > \mathrm{Pl}_a(\mathcal{R}[\neg\varphi \wedge \bigwedge_{i=1}^m o_i])$. This implies that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$ if and only if $PL_{\mathcal{I}} \models (\bigwedge_{i=1}^m o_i) \rightarrow \varphi$.

If s_a is not a local state in $\mathcal{I}$, then $\mathcal{R}[\cdot; o_1, \ldots, o_m] = \emptyset$, and by definition $\mathrm{Pl}_a(\mathcal{R}[\cdot; o_1, \ldots, o_m]) = \bot$. Using C1 and REV4, we get that $PL_a(\mathcal{R}[\bigwedge_{i=1}^m o_i]) = \bot$, and thus $PL_{\mathcal{I}} \models (\bigwedge_{i=1}^m o_i) \rightarrow \varphi$ for all $\varphi \in \mathcal{L}_e$. Since s_a is not a local state in $\mathcal{I}$, by definition $\mathrm{Bel}(\mathcal{I}, s_a) = \mathcal{L}_e$. Hence, we can conclude that $\varphi \in \mathrm{Bel}(\mathcal{I}, s_a)$ if and only if $PL_{\mathcal{I}} \models (\bigwedge_{i=1}^m o_i) \rightarrow \varphi$. ∎

We now show that given a ranked plausibility structure PL we can construct a system whose characteristic structure is default-isomorphic to PL.

Lemma A.5: *Let $PL_K = (W_K, \mathrm{Pl}_K, \pi_K)$ be a plausibility space that satisfies the conditions of Lemma A.3. Then there is a system $\mathcal{I} \in \mathcal{C}^R$ such that $PL_{\mathcal{I}} = PL_K$.*

Proof: Let $PL_K = (W_K, \mathrm{Pl}_K, \pi_K)$ be a plausibility space that satisfies the conditions of Lemma A.3. For each world $w \in W_K$ and sequence of observations $o_1, o_2, \ldots$, let $r^{w,o_1,o_2,\cdots}$ be the run defined so that $r_e^{w,o_1,o_2,\cdots}(m) = w$ and $r_a^{w,o_1,o_2,\cdots}(m) = \langle o_1, \ldots, o_m \rangle$ for all m. Let $\mathcal{R} = \{r^{w,o_1,o_2,\cdots} : \pi_k(w)(o_i) = \mathbf{true} \text{ for all } i\}$. Define π so that $\pi(r,m)(p) = \pi_K(r_e(m))(p)$ for $p \in \Phi_e$, and so that $\pi(r,m)(\mathit{learn}(\varphi)) = \mathbf{true}$ if $o_{(r,m)} = \varphi$ for $\varphi \in \mathcal{L}_e$. Finally, define the prior plausibility Pl_a so that $\mathrm{Pl}_a(R) = \mathrm{Pl}_K(\{w : \exists r \in R(w = r_e(0))\}$. It is easy to check that this definition implies that $\mathrm{Pl}_a(\mathcal{R}[\varphi]) = \mathrm{Pl}_K([\![\varphi]\!]_{PL_K})$. Thus, $PL_{\mathcal{I}} = PL_K$. Since Pl_K is a ranking, Pl_a is also a ranking and thus qualitative.

We now verify that the resulting interpreted system is indeed in $\mathcal{C}^R$. It is easy to check that $\mathcal{I}$ is a belief change system; that is, it satisfies BCS1–BCS5. The construction is such that $r_e(m) = r_e(0)$ for all runs r and times m. Thus, $\mathcal{I}$ satisfies REV1. Since the prior Pl_a is a ranking, this system also satisfies REV2. Lemma A.2 implies that if φ is a consistent formula, then $\mathrm{Pl}_K([\![\varphi]\!]_{PL_K}) > \bot$. This implies that $\mathrm{Pl}_a(\mathcal{R}[\varphi]) > \bot$, and thus the system satisfies REV3. Finally, it is easy to show that $\mathrm{Pl}_a(\mathcal{R}[\varphi; o_1, \ldots, o_m]) = \mathrm{Pl}_a(\mathcal{R}[\varphi \wedge o_1 \wedge \ldots \wedge o_m]) = \mathrm{Pl}_K([\![\varphi \wedge o_1 \wedge \ldots \wedge o_m]\!]_{PL_K})$. Thus, the system satisfies REV4. ∎

We are finally ready to prove Theorem 5.2.

Theorem 5.2: *Let $\circ$ be an AGM revision operator and let $K \subseteq \mathcal{L}_e$ be a consistent belief set. Then there is a system $\mathcal{I}_{(\circ,K)} \in \mathcal{C}^R$ such that $Bel(\mathcal{I}_{(\circ,K)}, \langle\rangle) = K$ and*

$$Bel(\mathcal{I}_{(\circ,K)}, \langle\rangle) \circ \varphi = Bel(\mathcal{I}_{(\circ,K)}, \langle\varphi\rangle)$$

for all $\varphi \in \mathcal{L}_e$.

Proof: Let $\circ$ be an AGM revision operator and let $K \subseteq \mathcal{L}_e$ be a consistent belief set. By Lemmas A.2 and A.5, there is a system $\mathcal{I}_{(\circ,K)} = (\mathcal{R}_{(\circ,K)}, \pi_{(\circ,K)}, \mathcal{P}_{(\circ,K)}) \in \mathcal{C}^R$ such that $PL_{\mathcal{I}_{(\circ,K)}} \models \varphi \rightarrow \psi$ if and only if $\psi \in K \circ \varphi$. Our construction is such that $\psi \in K \circ \varphi$ if and only if $PL_{\mathcal{I}_{(\circ,K)}} \models \varphi \rightarrow \psi$. Using Lemma A.4, we get that $PL_{\mathcal{I}_{(\circ,K)}} \models \varphi \rightarrow \psi$ if and only if $\psi \in \mathrm{Bel}(\mathcal{I}_{(\circ,K)}, \langle\varphi\rangle)$. Thus, $K \circ \varphi = \mathrm{Bel}(\mathcal{I}_{(\circ,K)}, \langle\varphi\rangle)$.

Finally, we show $\mathrm{Bel}(\mathcal{I}_{(\circ,K)}, \langle\rangle) = K$. We start by showing that $K \circ \mathit{true} = K$. Using R3, we get that $K \circ \mathit{true} \subseteq Cl(K \cup \{\mathit{true}\}) = K$. Since K is consistent, by R4, $Cl(K \cup \{\mathit{true}\}) \subseteq K \circ \mathit{true}$. Thus, $K \circ \mathit{true} = K$. By Lemma A.4, we have that $\mathrm{Bel}(\mathcal{I}, \langle\rangle) = \mathrm{Bel}(\mathcal{I}, \langle \mathit{true}\rangle)$. Since $\mathrm{Bel}(\mathcal{I}, \langle\mathit{true}\rangle) = K \circ \mathit{true}$, we conclude that $\mathrm{Bel}(\mathcal{I}_{(\circ,K)}, \langle\rangle) = K$. ∎

We next prove Theorem 5.3.

Theorem 5.3: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$. Then there is an AGM revision operator $\circ_{\mathcal{I}}$ such that*

$$Bel(\mathcal{I}, \langle\rangle) \circ_{\mathcal{I}} \varphi = Bel(\mathcal{I}, \langle\varphi\rangle)$$

for all $\varphi \in \mathcal{L}_e$.

Proof: Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P})$ be a system in $\mathcal{C}^R$. It is easy to verify that $PL_{\mathcal{I}}$ satisfies the conditions of Lemma A.3 with $K = \mathrm{Bel}(\mathcal{I}, \langle\rangle)$. This lemma implies that there is a revision operator $\circ_{\mathcal{I}}$ such that $\psi \in K \circ_{\mathcal{I}} \varphi$ if and only if $PL_{\mathcal{I}} \models \varphi \rightarrow \psi$. Using Lemma A.4, we have that $\psi \in \mathrm{Bel}(\mathcal{I}, \langle\varphi\rangle)$ if and only if $\mathrm{Pl}_{\mathcal{I}} \models \varphi \rightarrow \psi$. Thus, we have that $K \circ_{\mathcal{I}} \varphi = \mathrm{Bel}(\mathcal{I}, \langle\varphi\rangle)$ for all formulas φ. ∎

Theorem 5.4: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ and $s_a = \langle o_1, \ldots, o_m\rangle$ be a local state in $\mathcal{I}$. Then there is an AGM revision operator $\circ_{\mathcal{I},s_a}$ such that*

$$Bel(\mathcal{I}, s_a) \circ_{\mathcal{I},s_a} \varphi = Bel(\mathcal{I}, s_a \cdot \varphi)$$

for all formulas $\varphi \in \mathcal{L}_e$ such that $o_1 \wedge \ldots o_m \wedge \varphi$ is consistent.

Proof: The structure of the proof is similar to that of Theorem 5.3. As in that proof, we construct a ranked plausibility structure and use Lemma A.3 to find an AGM revision operator. The main difference is that after observing $\varphi_1, \ldots, \varphi_k$, some events are considered impossible. Lemma A.3, however, requires that all possible formulas are assigned a positive plausibility. We overcome this problem by assigning a "fictional" positive plausibility to all non-empty events that are ruled out by the previous observations.

We proceed as follows. Let d_0 be a new plausibility value that is less plausible than all positive plausibilities in Pl_a; that is, if $\mathrm{Pl}_a(A) > \perp$, then $\mathrm{Pl}_a(A) > d_0$. Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^R$; let $s_a = \langle o_1, \ldots, o_m\rangle$. We define $PL = (\mathcal{R}, \mathrm{Pl}, \pi_{PL_{\mathcal{I}}})$, where Pl is such that $\mathrm{Pl}([\![\varphi]\!]) =$

$\max(\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge \bigwedge_{i=1}^m o_i]), d_0)$ for all consistent formulas φ. This definition implies that if φ is consistent with $\bigwedge_{i=1}^m o_i$, then $\mathrm{Pl}([\![\varphi]\!]_{PL}) = \mathrm{Pl}_a(\mathcal{R}[\varphi_1 \wedge \bigwedge_{i=1}^m o_i])$.

We now prove that if φ is consistent with $\bigwedge_{i=1}^m o_i$, then $PL \models \varphi \rightarrow \psi$ if and only if $PL_{\mathcal{I}} \models (\varphi \wedge \bigwedge_{i=1}^m o_i) \rightarrow \psi$.

For the "if" part, assume that $PL_{\mathcal{I}} \models (\varphi \wedge \bigwedge_{i=1}^m o_i) \rightarrow \psi$. Since φ is consistent with $\bigwedge_{i=1}^m o_i$ it follows, from REV3, that $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge (\bigwedge_{i=1}^m o_i])) > \perp$. Thus, $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \psi]) > \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi]) \geq \perp$. Thus, $\varphi \wedge \psi$ is consistent with $\bigwedge_{i=1}^m o_i$. This implies that $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) = \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \psi] > \max(d_0, \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi]) = \mathrm{Pl}([\![\varphi \wedge \neg\psi]\!])$. We conclude that $PL \models \varphi \rightarrow \psi$.

For the "only if" part, assume that $PL_{\mathcal{I}} \not\models (\varphi \wedge (\bigwedge_{i=1}^m o_i)) \rightarrow \psi$. This implies that $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \psi]) \not> \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi])$. Since Pl_a is a ranking, it follows that $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i))\psi]) \leq \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi])$. Since $\perp < \mathrm{Pl}_a(\mathcal{R}[\varphi \wedge (\bigwedge_{i=1}^m o_i)]) = \max(\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \psi]), \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi]))$, we have that $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=1}^m o_i)) \wedge \neg\psi]) > \perp$. We conclude that $\mathrm{Pl}([\![\varphi \wedge \neg\psi]\!]) \geq \mathrm{Pl}([\![\varphi \wedge \psi]\!])$. Thus, $PL \not\models \varphi \rightarrow \psi$.

It is easy to verify that PL is ranked, and satisfies the requirements of Lemma A.3. Thus, there exists a revision operator $\circ_{\mathcal{I},s_a}$ such that $\psi \in K \circ_{\mathcal{I},s_a} \varphi$ if and only if $PL \models \varphi \rightarrow \psi$, where $K = \{\varphi : PL \models \mathit{true} \rightarrow \varphi\}$. Moreover, since for all φ consistent with $\bigwedge_{i=1}^m o_i$ we have that$PL \models \varphi \rightarrow \psi$ if and only if $PL_{\mathcal{I}} \models (\varphi \wedge (\bigwedge_{i=1}^m o_i)) \rightarrow \psi$, then, from Lemma A.4, it follows that $K = \mathrm{Bel}(\mathcal{I}, s_a)$ and that if φ is consistent with $\bigwedge_{i=1}^m o_i$, then $PL \models \varphi \rightarrow \psi$ if and only if $\psi \in \mathrm{Bel}(\mathcal{I}, s_a \cdot \varphi)$. ∎

Theorem 5.5: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ whose local states are $\mathcal{E}_{\mathcal{L}_e}$. There is a function $Bel_{\mathcal{I}}$ that maps epistemic states to belief states such that*

- *if s_a is a local state of the agent in $\mathcal{I}$, then $Bel(\mathcal{I}, s_a) = Bel_{\mathcal{I}}(s_a)$, and*
- *$(\circ, Bel_{\mathcal{I}})$ satisfies R1′–R8′.*

Proof: Roughly speaking, we define $\mathrm{Bel}_{\mathcal{I}}(s_a) = \mathrm{Bel}(\mathcal{I}, s_a)$ when s_a is a local state in $\mathcal{I}$. If s_a is not in $\mathcal{I}$, then we set $\mathrm{Bel}_{\mathcal{I}}(s_a) = \mathrm{Bel}(\mathcal{I}, s')$, where s' is the longest consistent suffix of s_a. We now make this definition precise, and show that the resulting $\mathrm{Bel}_{\mathcal{I}}$ satisfies R1′–R8′.

We proceed as follows. We define a function $f(\cdot)$ that maps sequences of observations to suffixes as follows:

$$f(\langle o_1, \ldots, o_m \rangle) = \begin{cases} \langle\rangle & \text{if } m = 0, \\ \langle \mathit{false} \rangle & \text{if } m > 0 \text{ and } o_m \text{ is inconsistent,} \\ \langle o_k, \ldots, o_m \rangle & \text{otherwise, with } k \leq m \text{ the minimal index s. t. } \not\vdash_{\mathcal{L}_e} \neg(o_k \wedge \ldots \wedge o_m). \end{cases}$$

Aside from the special case where o_m is inconsistent, we simply choose the longest suffix of s_a that is still consistent. We define $\mathrm{Bel}_{\mathcal{I}}(s_a) = \mathrm{Bel}(\mathcal{I}, f(s_a))$. Clearly, if s_a is a local state in $\mathcal{I}$, then $f(s_a) = s_a$, so $\mathrm{Bel}_{\mathcal{I}}(s_a) = \mathrm{Bel}(\mathcal{I}, s_a)$.

We now have to show that $(\circ, \mathrm{Bel}_{\mathcal{I}})$ satisfies R1′–R8′. The proof outline is as follows. Given a particular state s_a, we construct a ranked plausibility structure that corresponds, in the sense of Lemma A.2, to belief change from s_a. We then use Lemma A.3 to show that belief changes from s_a satisfies the AGM postulates, i.e., R1–R8. Since this is true from any s_a, we get that $\mathrm{Bel}_{\mathcal{I}}$ satisfies R1′–R8′.

Let $s_a = \langle o_1, \ldots, o_m \rangle$. We define a ranked plausibility space that has the following structure. The most plausible events are the ones consistent with $o_1 \wedge \ldots \wedge o_m$. They are ordered according to the prior ranking conditioned on $o_1 \wedge \ldots \wedge o_m$. The next tier of events are those that are inconsistent with $o_1 \wedge \ldots \wedge o_m$ but are consistent $o_2 \wedge \ldots \wedge o_m$. Again, these are ordered according to the prior ranking conditioned on $o_2 \wedge \ldots \wedge o_m$. We continue this way; the last tier consists of all events that are inconsistent with o_m.

Formally, let $PL = (\mathcal{R}, \mathrm{Pl}, \pi_{PL_{\mathcal{I}}})$, where Pl is such that $\mathrm{Pl}([\![\varphi]\!]) \geq \mathrm{Pl}([\![\psi]\!])$ if $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge (\bigwedge_{i=k}^m o_i)]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi \wedge (\bigwedge_{i=k}^m o_i)])$ where $k \leq m+1$ is the greatest integer such that for all $j < k$, φ and ψ are both inconsistent with $\bigwedge_{i=j}^m o_i$. It is easy to see that PL is ranked, and that if φ is consistent, then $\mathrm{Pl}([\![\varphi]\!]) > \perp$.

Let $\varphi \in \mathcal{L}_e$. We now show that $PL \models \varphi {\rightarrow} \psi$ if and only if $\psi \in \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi)$. If φ is inconsistent, then $PL \models \varphi {\rightarrow} \psi$ for all ψ. Moreover, since φ is inconsistent, $f(s_a \cdot \varphi) = \langle \mathit{false} \rangle$, and thus $\mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi) = \mathcal{L}_e$. We conclude that $\varphi {\rightarrow} \psi$ if and only if $\psi \in \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi)$. If φ is consistent, then let $k \leq m+1$ be the greatest integer such that for all $j < k$, φ is inconsistent with $\bigwedge_{i=j}^m o_i$. It is easy to verify that $f(s_a \cdot \varphi) = \langle o_k, \ldots, o_m, \varphi \rangle$. From Lemma A.4, it follows that $\psi \in \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi) = \mathrm{Bel}(\mathcal{I}, \langle o_k, \ldots, o_m, \varphi \rangle)$ if and only if $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \psi]) > \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi])$. We now show that this is the case if and only if $PL \models \varphi {\rightarrow} \psi$. Suppose that $PL_a(\mathcal{R}[(\varphi \wedge \wedge(\bigwedge_{i=k}^m o_i)) \wedge \psi]) > PL_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi])$. Then, clearly, $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \psi]) > \perp$, and thus $\varphi \wedge \psi$ is consistent with $o_k, \ldots, o_m$. Since both $\varphi \wedge \psi$ and $\varphi \wedge \neg\psi$ are inconsistent with $o_j, \ldots, o_m$ for all $j < k$, we have that $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) > \mathrm{Pl}([\![\varphi \wedge \neg\psi]\!])$. On other hand, if $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \psi]) \not> \mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi])$, then since Pl_a is a ranking $PL_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \psi]) \leq PL_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi])$. Moreover, since φ is consistent with $o_k \wedge \ldots \wedge o_m$, we have that $\mathrm{Pl}_a(\mathcal{R}[\varphi \wedge (\bigwedge_{i=k}^m o_i)]) > \perp$. This implies that $\mathrm{Pl}_a(\mathcal{R}[(\varphi \wedge (\bigwedge_{i=k}^m o_i)) \wedge \neg\psi]) > \perp$ and thus $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) \leq \mathrm{Pl}([\![\varphi \wedge \neg\psi]\!])$. We conclude that $PL \models \varphi {\rightarrow} \psi$ if and only if $\psi \in \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi)$.

By Lemma A.3, there is a revision operator $\circ_{s_a}$ that satisfies R1–R8 such that $\psi \in K \circ \varphi$ if and only if $PL \models \varphi {\rightarrow} \psi$. It is not hard to check that this implies that the change from $\mathrm{Bel}_{\mathcal{I}}(s_a)$ to $\mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi)$ satisfies R1′–R8′. ∎

Proposition 5.6: *Let $\mathcal{I}$ be a system in $\mathcal{C}^R$ whose local states are $\mathcal{E}_{\mathcal{L}_e}$. There is a function $Bel_{\mathcal{I}}$ that maps epistemic states to belief states such that*

- *if s_a is a local state of the agent in $\mathcal{I}$, then $Bel(\mathcal{I}, s_a) = Bel_{\mathcal{I}}(s_a)$, and*
- *$(\circ, Bel_{\mathcal{I}})$ satisfies R1′–R9′.*

Proof: We show that the function $\mathrm{Bel}_{\mathcal{I}}$ defined in the proof of Theorem 5.5 satisfies R9′. Let $s_a = \langle o_1, \ldots, o_m \rangle$, and let $\varphi, \psi \in \mathcal{L}_e$ be formulas such that $\not\vdash_{\mathcal{L}_e} \neg(\varphi \wedge \psi)$. Since φ is consistent with ψ, we get that $f(s_a \cdot \varphi \cdot \psi) = \langle o_k, \ldots, o_m, \varphi, \psi \rangle$, where $k \leq m$ is the least integer such that $\varphi \wedge \psi$ is consistent with $o_k, \ldots, o_m$. For the same reason, we get that $f(s_a \cdot \varphi \wedge \psi) = \langle o_k, \ldots, o_m, \varphi \wedge \psi \rangle$. Using Lemma A.4 we immediately get that $\mathrm{Bel}(\mathcal{I}, \langle o_k, \ldots, o_m, \varphi, \psi \rangle) = \mathrm{Bel}(\mathcal{I}, \langle o_k, \ldots, o_m, \varphi \wedge \psi \rangle)$. Thus, we conclude that $\mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi \cdot \psi) = \mathrm{Bel}_{\mathcal{I}}(s_a \cdot \varphi \wedge \psi)$. ∎

Theorem 5.7: *Given a function $Bel_{\mathcal{L}_e}$ mapping epistemic states in $\mathcal{E}_{\mathcal{L}_e}$ to belief sets over $\mathcal{L}_e$ such that $Bel_{\mathcal{L}_e}(\langle\rangle)$ is consistent and $(Bel_{\mathcal{L}_e}, \circ)$ satisfies R1′–R9′, there is a system $\mathcal{I} \in \mathcal{C}^R$ whose local states are in $\mathcal{E}_{\mathcal{L}_e}$ such that $Bel_{\mathcal{L}_e}(s_a) = Bel(\mathcal{I}, s_a)$ for each local state s_a in $\mathcal{I}$.*

Proof: According to Theorem 5.2, there is a system $\mathcal{I}$ such that $\mathrm{Bel}(\mathcal{I}, \langle\rangle) = \mathrm{Bel}_{\mathcal{L}_e}(\langle\rangle)$ and $\mathrm{Bel}(\mathcal{I}, \langle\varphi\rangle) = \mathrm{Bel}_{\mathcal{L}_e}(\langle\varphi\rangle)$ for all $\varphi \in \mathcal{L}_e$.

We now show that $\mathrm{Bel}(\mathcal{I}, s_a) = \mathrm{Bel}_{\mathcal{L}_e}(s_a)$ for local states s_a in $\mathcal{I}$. We prove this by induction on the length m of s_a. For $m \leq 1$, this is true by our choice of $\mathcal{I}$. For the induction case, let $s_a = \langle o_1, \ldots, o_m\rangle$ be a local state in $\mathcal{I}$. Thus,, $o_1 \wedge \ldots \wedge o_m$ is consistent. From R9′, it follows that $\mathrm{Bel}_{\mathcal{L}_e}(\langle o_1, \ldots, o_m\rangle) = \mathrm{Bel}_{\mathcal{L}_e}(\langle o_1, \ldots, o_{m-2}, o_{m-1} \wedge o_m\rangle)$. Using the induction hypothesis, we have that $\mathrm{Bel}_{\mathcal{L}_e}(\langle o_1, \ldots, o_{m-2}, o_{m-1} \wedge o_m\rangle) = \mathrm{Bel}(\mathcal{I}, \langle o_1, \ldots, o_{m-2}, o_{m-1} \wedge o_m\rangle)$. Using Lemma A.4, we get that $\mathrm{Bel}(\mathcal{I}, \langle o_1, \ldots, o_{m-2}, o_{m-1} \wedge o_m\rangle) = \mathrm{Bel}(\mathcal{I}, \langle o_1, \ldots, o_m\rangle)$. Thus, we conclude that $\mathrm{Bel}_{\mathcal{L}_e}(\langle o_1, \ldots, o_m\rangle) = \mathrm{Bel}(\mathcal{I}, \langle o_1, \ldots, o_m\rangle)$. ∎

A.2 Proofs for Section 6

In this section we prove Theorem 6.1. We start by reviewing the semantic representation of Katsuno and Mendelzon for update. Katsuno and Mendelzon show that there is a direct relation between update operators and distance functions. To make this relation precise, we need to introduce some definitions. An *update structure* is a tuple $U = (W, d, \pi)$, where W is a finite set of worlds, d is a distance function on W, and π is a mapping from worlds to truth assignments for $\mathcal{L}_e$ such that

- $\pi(w)$ is $\vdash_{\mathcal{L}_e}$ consistent,
- if $\not\vdash_{\mathcal{L}_e} \neg\varphi$, then there is some $w \in W$ with $\pi(w)(\varphi) = \mathbf{true}$, and
- if $w \neq w'$ then $\pi(w) \neq \pi(w')$ for all $w, w' \in W$.

Given an update structure $U = (W, d, \pi)$, we define $[\![\varphi]\!]_U = \{w : \pi(w)(\varphi) = \mathbf{true}\}$. Katsuno and Mendelzon use update structures as semantic representations of update operators. Given an update structure $U = (W, d, \pi)$ and sets $A, B \subseteq W$, Katsuno and Mendelzon define $\min_U(A, B)$ to be the set of worlds in B that are closest to worlds in A, according to d. Formally, $\min_U(A, B) = \{w \in B : \exists w_0 \in A \forall w' \in B\, d(w_0, w') \not< d(w_0, w)\}$.

Theorem A.6: [Katsuno and Mendelzon 1991b] *A belief change operator $\diamond$ satisfies U1–U8 if and only if there is an update structure $U = (W, \pi, d)$ such that*

$$[\![\varphi \diamond \psi]\!]_U = \min_U([\![\varphi]\!]_U, [\![\psi]\!]_U).$$

Thus the worlds the agent believes possible after updating with ψ are these worlds that are closest to some world considered possible before learning ψ.

We now show that any system in $\mathcal{C}^U$ corresponds to an update structure. Suppose that $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^U$ is such that the set of environment states is $\mathcal{S}_e$ and the prior of BCS5 is consistent with distance function d. Define an update structure $U_{\mathcal{I}} = (S_e, \pi_e, d)$, where for $p \in \Phi_e$, $\pi_e(s_e)(p) = \pi((s_e, s_a))(p)$ for some choice of s_a. By BCS1, the choice of s_a does not matter. It is easy to see that UPD1 ensures that S_e and π_e satisfy the requirements of the definition of update structures. We want to show that belief change in $\mathcal{I}$ corresponds to belief change in $U_{\mathcal{I}}$ in the sense of Theorem A.6. Since Theorem A.6 states that any belief change operation defined by an update structure satisfies U1–U8, this will suffice to prove the "if" direction of Theorem 6.1. To prove the "only if" direction of Theorem 6.1, we show that that for any update structure U, there is a system $\mathcal{I} \in \mathcal{C}^U$ such that $U_{\mathcal{I}} = U$.

We start with preliminary definitions and lemmas for the "if" direction of Theorem 6.1. Let $s_a = \langle o_1, \ldots, o_m \rangle$. We define $\mathit{States}(\mathcal{I}, s_a) = \{s \in \mathcal{S}_e : s \models \xi \text{ for all } \xi \in \mathrm{Bel}(\mathcal{I}, s_a)\}$. Clearly, if φ is such that $\mathrm{Bel}(\mathcal{I}, s_a) = Cl(\varphi)$, then $\mathit{States}(\mathcal{I}, s_a) = [\![\varphi]\!]_{U_{\mathcal{I}}}$. To show that belief change in $\mathcal{I}$ corresponds to belief change in $U_{\mathcal{I}}$ we have to show that

$$\mathit{States}(\mathcal{I}, s_a \cdot \psi) = \min\nolimits_{U_{\mathcal{I}}}(\mathit{States}(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}}).$$

This is proved in Lemma A.9. To prove this lemma, we need some preliminary lemmas.

Lemma A.7: *Let $\mathcal{I} \in \mathcal{C}^U$, and let $s_a = \langle o_1, \ldots, o_m \rangle$. Then $\varphi \in \mathit{Bel}(\mathcal{I}, s_a)$ if and only if $(\mathcal{I}, r, 0) \models (\bigcirc o_1 \wedge \ldots \wedge \bigcirc^m o_m) \rightarrow \bigcirc^m \varphi$ for some run r in $\mathcal{R}$.*

Proof: The proof of this lemma is analogous to the proof of Lemma A.4, using UPD3 and UPD4 instead of REV3 and REV4. We do not repeat the argument here. ∎

We now provide an alternative characterization of $\mathit{States}(\mathcal{I}, s_a)$ in terms of the agent's prior on run-prefixes.

Lemma A.8: *Let $\mathcal{I} \in \mathcal{C}^U$ and let $s_a = \langle o_1, \ldots, o_m \rangle$. Then $s_m \in \mathit{States}(\mathcal{I}, s_a)$ if and only if there is a sequence of states $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_m]$ such that $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$.*

Proof: For the "if" direction, assume that there is a sequence $s_0, \ldots, s_m$ such that $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_m]$, and $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$. By way of contradiction, assume that $s_m \notin \mathit{States}(\mathcal{I}, m)$. Thus, there is a formula $\xi \in \mathrm{Bel}(\mathcal{I}, m)$ such that $s_m \models \neg\xi$. From Lemma A.7 it follows that since $\xi \in \mathrm{Bel}(\mathcal{I}, s_a)$, $(\mathcal{I}, r, 0) \models (\bigcirc o_1 \wedge \ldots \wedge \bigcirc^m o_m) \rightarrow \bigcirc^m \xi$ for some run r in $\mathcal{R}$. From the definition of conditioning it follows that $\mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi]) > \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\xi])$. Since $s_m \models \neg\xi$, we get that $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\xi]$ and that $\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m]$. From A1, it follows that $\mathrm{Pl}_a([s_0, \ldots, s_m]) < \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$, which contradicts our starting assumption. We conclude that $s_m \in \mathit{States}(\mathcal{I}, s_a)$.

For the "only if" direction, assume that $s_m \in \mathit{States}(\mathcal{I}, a)$. Since $\mathcal{S}_e$ is finite and π_e assigns a different truth assignment to each state in $\mathcal{S}_e$, there is a formula $\xi \in \mathcal{L}_e$ that characterizes s_m; that is, $s \models \xi$ if and only if $s = s_m$. Since $s_m \in \mathit{States}(\mathcal{I}, s_a)$, we have that $\neg\xi \notin \mathrm{Bel}(\mathcal{I}, s_a)$. Using Lemma A.7, we get that $(\mathcal{I}, r, 0) \not\models (\bigcirc o_1 \wedge \ldots \wedge \bigcirc^m o_m) \rightarrow \bigcirc^m \neg\xi$ for all runs $r \in \mathcal{R}$. By BCS5, this is true if and only if $\mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_m]) > \perp$ and $\mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi]) \not< \mathrm{Pl}_a(\mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \neg\xi])$. By UPD2, there is a sequence $[s_0, \ldots, s_m] \subseteq \mathcal{R}[\mathit{true}, o_1, \ldots, o_{m-1}, o_m \wedge \xi]$ such that $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m])$

for all $[s'_0, \ldots, s'_m] \subseteq \mathcal{R}[true, o_1, \ldots, o_{m-1}, o_m \wedge \neg\xi]$. Moreover, without loss of generality, we can assume that $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m])$ for all $[s'_0, \ldots, s'_m] \subseteq \mathcal{R}[true, o_1, \ldots, o_{m-1}, o_m \wedge \xi]$, since there are only finitely many such sequences. Thus, by UPD2, $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$. ∎

We can now prove that belief change in $\mathcal{I}$ corresponds to belief change in $U_{\mathcal{I}}$.

Lemma A.9: *Let $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^U$ Then*

$$States(\mathcal{I}, s_a \cdot \psi) = \min_{U_{\mathcal{I}}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}})$$

for all local states s_a and formulas $\psi \in \mathcal{L}_e$.

Proof: Let Pl_a be the prior in $\mathcal{I}$; assume that Pl_a consistent with a distance function d. Let $s_a = \langle o_1, \ldots, o_m \rangle$.

To show that $\min_{U_{\mathcal{I}}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}}) \subseteq States(\mathcal{I}, s_a \cdot \psi)$, suppose that $s \in \min_{U_{\mathcal{I}}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}})$. Thus, there is a state $s_m \in States(\mathcal{I}, s_a)$ such that $d(s_m, s') \not< d(s_m, s)$ for all states s' that satisfy ψ. We want to show that $s \in States(\mathcal{I}, s_a \cdot \psi)$. From Lemma A.8, it follows that, since $s_m \in States(\mathcal{I}, s_a)$, there is a sequence $s_0, \ldots, s_{m-1}$ such that $[s_0, \ldots, s_m] \in \mathcal{R}[true, o_1, \ldots, o_{m-1}, o_m]$ and $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \ldots, o_{m-1}, o_m] - [s_0, \ldots, s_m])$. We now show that $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \ldots, o_m, \psi] - [s_0, \ldots, s_m, s])$. By Lemma A.8, this suffices to show that $s \in States(\mathcal{I}, s_a \cdot \psi)$. Suppose that $[s'_0, \ldots, s'_{m+1}] \subseteq \mathcal{R}[true, o_1, \ldots, o_m, \psi] - [s_0, \ldots, s_m, s]$. If $[s_0, \ldots, s_m] = [s'_0, \ldots, s'_m]$, then we have that $d(s'_m, s'_{m+1}) \not< d(s_m, s)$. Since Pl_a is consistent with d, it follows that $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m, s'_{m+1}])$. If $[s_0, \ldots, s_m] \neq [s'_0, \ldots, s'_m]$, then, since $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m])$ and Pl_a is consistent with d, we have that $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m, s'_{m+1}])$.

Since $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m, s'_{m+1}])$ for all $[s'_0, \ldots, s'_{m+1}] \subseteq \mathcal{R}[true, o_1, \ldots, o_m, \psi] - [s_0, \ldots, s_m, s]$ and Pl_a is prefix-defined, we have that $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \ldots, o_m, \psi] - [s_0, \ldots, s_m, s]$. By Lemma A.8, $s \in States(\mathcal{I}, s_a \cdot \psi)$, as desired.

To show that $States(\mathcal{I}, s_a \cdot \psi) \subseteq \min_{U_{\mathcal{I}}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}})$, suppose that $s \in States(\mathcal{I}, s_a \cdot \psi)$. By Lemma A.8, there is a sequence $s_0, \ldots, s_m$ such that $[s_0, \ldots, s_m, s]) \subseteq \mathcal{R}[true, o_1, \ldots, o_m, \psi]$ and $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \ldots, o_m, \psi] - [s_0, \ldots, s_m, s])$. We want to show that $s_m \in States(\mathcal{I}, s_a)$ and that $d(s_m, s') \not< d(s_m, s)$ for all s' that satisfy ψ. This suffices to prove that $s \in \min_{U_{\mathcal{I}}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}})$.

To show that $s_m \in States(\mathcal{I}, s_a)$, by Lemma A.8, it suffices to show that $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a(\mathcal{R}[true, o_1, \ldots, o_m] - [s_0, \ldots, s_m])$. Let $s'_0, \ldots, s'_m$ be a sequence such that $[s'_0, \ldots, s'_m] \subseteq \mathcal{R}[true, o_1, \ldots, o_m]$. By definition, $[s'_0, \ldots, s'_m, s] \subseteq \mathcal{R}[true, o_1, \ldots, o_m, \psi]$. Thus, from our choice of $s_0, \ldots, s_m$, it follows that $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m, s])$. Since Pl_a is consistent with d, it follows that $\mathrm{Pl}_a([s_0, \ldots, s_m]) \not< \mathrm{Pl}_a([s'_0, \ldots, s'_m])$. Thus, by Lemma A.8, $s_m \in States(\mathcal{I}, s_a)$. To see that $d(s_m, s') \not< d(s_m, s)$ for all s' that satisfy ψ, let $s' \neq s$ be such that $s' \models \psi$. Thus, $[s_0, \ldots, s_m, s'] \subseteq [true, o_1, \ldots, o_m, \psi]$. From our choice of $s_0, \ldots, s_m$, it follows that $\mathrm{Pl}_a([s_0, \ldots, s_m, s]) \not< \mathrm{Pl}_a([s_0, \ldots, s_m, s'])$. Since Pl_a is consistent with d, it follows that $d(s_m, s') \not< d(s_m, s)$. We conclude that $s \in \min_{U_{\mathcal{I}}}(States(\mathcal{I}, s_a), [\![\psi]\!]_{U_{\mathcal{I}}})$. ∎

We now have the tools to prove the "if" direction of Theorem 6.1.

Lemma A.10: *If $\mathcal{I} = (\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^U$, then there is a belief change operator $\diamond$ that satisfies U1–U8 such that*

$$Bel(\mathcal{I}, s_a) \diamond \psi = Bel(\mathcal{I}, s_a \cdot \psi)$$

for all local states s_a and formulas $\psi \in \mathcal{L}_e$.

Proof: Let $\mathcal{I} \in \mathcal{C}^U$. Using the arguments we presented above, it easy to check that $U_{\mathcal{I}}$ is an update structure. By Theorem A.6, there is a belief change operator $\diamond$ that satisfies U1–U8 such that $[\![\varphi \diamond \psi]\!]_{U_{\mathcal{I}}} = \min_{U_{\mathcal{I}}}([\![\varphi]\!]_{U_{\mathcal{I}}}, [\![\psi]\!]_{U_{\mathcal{I}}})$ for all $\varphi, \psi \in \mathcal{L}_e$. From Lemma A.9, it follows that $\mathrm{Bel}(\mathcal{I}, s_a) \diamond \psi = \mathrm{Bel}(\mathcal{I}, s_a \cdot \psi)$. ∎

We now prove the "only if" direction of Theorem 6.1. Suppose that $\diamond$ is a belief change operator that satisfies U1–U8. According to Theorem A.6, there is an update structure $U_\diamond$ that corresponds to $\diamond$. Thus, it suffices to show that there is a system $\mathcal{I}$ such that $U_{\mathcal{I}} = U_\diamond$.

Lemma A.11: *Let $U = (W, d, \pi_U)$ be an update structure. Then there is a system $\mathcal{I} \in \mathcal{C}^U$ such that $U_{\mathcal{I}} = U$.*

Proof: Given the sequences $w_0, w_1, \ldots \in W$ and $o_1, o_2, \ldots \in \mathcal{L}_e$, let $r^{w_0,w_1,\ldots;o_1,o_2,\ldots}$ be the run defined so that $r_e^{w_0,w_1,\ldots;o_1,o_2,\ldots}(m) = w_m$ and $r_a^{w_0,w_1,\ldots;o_1,o_2,\ldots}(m) = \langle o_1, \ldots, o_m\rangle$. Let $\mathcal{R} = \{r^{w_0,w_1,\ldots;o_1,o_2,\ldots} : \pi_U(w_m)(o_m) = \mathbf{true}$ for all $m\}$. Define π such that $\pi(r,m)(p) = \pi_U(r_e(m))(p)$ for $p \in \Phi_e$ and $\pi(r,m)(\mathit{learn}(\varphi)) = \mathbf{true}$ if $o_{(r,m)} = \varphi$ for $\varphi \in \mathcal{L}_e$.

It is clear that $(\mathcal{R}, \pi)$ satisfies BCS1–BCS4 and UPD1. Thus, all that remains to show is that there is a prior plausibility measure Pl_a that satisfies UPD2–UPD4. This will ensure that $(\mathcal{R}, \pi, \mathcal{P}) \in \mathcal{C}^U$.

We proceed as follows. We define a *preferential space* $(\mathcal{R}, \prec)$ where $r \prec r'$ if and only if there is some m such that $r_e(k) = r'_e(k)$ for all $0 \le k \le m$, $r_e(m+1) \ne r'_e(m+1)$, and $d(r_e(m), r_e(m+1)) < d(r'_e(m), r'_e(m+1))$. Recall that $r \prec r'$ denotes that r is preferred over r'. Thus, this ordering is consistent with the comparison of events of the form $[s_0, \ldots, s_n]$ according to UPD2.

Using the construction of Proposition 2.2, there is a plausibility space (R, Pl_a) such that $\mathrm{Pl}_a(A) \ge \mathrm{Pl}_a(B)$ if and only if for all $r \in B - A$, there is a run $r' \in A$ such that $r' \prec r$ and there is no $r'' \in B - A$ such that $r'' \prec r'$. By [Friedman and Halpern 1997b, Theorem 5.5], Pl_a is a qualitative plausibility measure. We now show that it satisfies UPD2–UPD4.

We start with UPD2. To show that Pl_A is consistent with d, we need to show that $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$ if and only if there is some $m < n$ such that $s_k = s'_k$ for all $0 \le k \le m$, and $d(s_m, s_{m+1}) > d(s'_m, s'_{m+1})$. Suppose that $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$. Let r be some run in $[s_0, \ldots, '_n]$. Without loss of generality we can assume that $r_e(m) = r_e(n)$ for all $m > n$. Since $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$, there is a run $r' \in [s'_0, \ldots, s'_n]$ such that $r' \prec r$. By definition, this implies that there is an m such that $r_e(k) = r'_e(k)$ for all $0 \le k \le m$, and $d(r'_e(m), r'_e(m+1)) < d(r_e(m), r_e(m+1))$. We claim that $m < n$. For if $m \ge n$, then $r_e(m+1) = r_e(m)$ by construction, so $d(r_e(m), r_e(m+1)) = d(r_e(m), r_e(m)) \le d(r'_e(m), r'_e(m+1))$ and $r' \not\prec r$, a contradiction. Thus, $s_k = s'_k$ for all $0 \le k \le m$, $d(s'_m, s'_{m+1}) < d(s_m, s_{m+1})$.

For the converse, suppose that there is an $m < n$ such that $s_k = s'_k$ for all $0 \le k \le m$, and $d(s'_m, s'_{m+1}) < d(s_m, s_{m+1})$. Let r' be the run where $r'_e(k) = s'_k$ for $k \le n$, $r'_e(k) = s'_n$

for $k \geq n$, and $o_{(r',k)} = true$ for all k. It follows $r' \prec r$ for all runs $r' \in [s_0, \ldots, s_n]$. Thus, $\mathrm{Pl}_a([s_0, \ldots, s_n]) < \mathrm{Pl}_a([s'_0, \ldots, s'_n])$.

To show that Pl_a is prefix-defined, we must show that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_n])$ if and only if for all $[s_0, \ldots, s_n] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n]$, there is some $[s'_0, \ldots, s'_n] \subseteq \mathcal{R}[\varphi_0, \ldots, \varphi_n]$ such that $\mathrm{Pl}_a([s'_0, \ldots, s'_n]) > \mathrm{Pl}_a([s_0, \ldots, s_n])$. Suppose that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_n])$. Let $[s_0, \ldots, s_n] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n]$. Let $r \in [s_0, \ldots, s_n]$ be a run such that $r_e(m) = r_e(n)$ for all $m \geq n$. Since $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_n])$ there is a run $r' \in \mathcal{R}[\varphi_0, \ldots, \varphi_n]$ such that $r' \prec r$. This implies that there is an m such that $r_e(k) = r'_e(k)$ for all $0 \leq k \leq m$, and $d(r'_e(m), r'_e(m+1)) < d(r_e(m), r_e(m+1))$. As before, we have that $m < n$, and thus $\mathrm{Pl}_a([r'_e(0), \ldots, r'_e(n)]) > \mathrm{Pl}_a([s_0, \ldots, s_n])$. Since $r' \in \mathcal{R}[\varphi_0, \ldots, \varphi_n]$, we also have that $[r'_e(0), \ldots, r'_e(n)] \subseteq \mathcal{R}[\varphi_0, \ldots, \varphi_n]$, as desired.

For the converse, assume that for all $[s_0, \ldots, s_n] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n]$ there is some $[s'_0, \ldots, s'_n] \subseteq \mathcal{R}[\varphi_0, \ldots, \varphi_n]$ such that $\mathrm{Pl}_a([s'_0, \ldots, s'_n]) > \mathrm{Pl}_a([s_0, \ldots, s_n])$. This implies that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) > \mathrm{Pl}_a([s_0, \ldots, s_n])$ for all for all $[s_0, \ldots, s_n] \subseteq \mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n]$. Since there are only finitely many sequences of states of length m, we can apply A2, and conclude that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) > \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_n] - \mathcal{R}[\varphi_0, \ldots, \varphi_n])$. Thus, $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_n]) \geq \mathrm{Pl}_a((\mathcal{R}[\psi_0, \ldots, \psi_n]))$.

For UPD3, recall that the construction of Proposition 2.2 is such that $\mathrm{Pl}_a(R) > \bot$ for all non-empty $R \subseteq \mathcal{R}$. Since, by our construction, the set $\mathcal{R}[\varphi_0, \ldots, \varphi_n]$ is non-empty for all sequences $\varphi_0, \ldots, \varphi_n$ of consistent formulas, UPD3 must hold.

Finally, we consider UPD4. We have to show that $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \ldots, \varphi_{n+1}; o_1, \ldots, o_n]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \ldots, \psi_{n+1}; o_1, \ldots, o_n])$ if and only if $\mathrm{Pl}_a(\mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_n \wedge o_n, \varphi_{n+1}]) \geq \mathrm{Pl}_a(\mathcal{R}[\psi_0, \psi_1 \wedge o_1, \ldots, \psi_n \wedge o_n, \psi_{n+1}])$. By construction, $\mathcal{R}[\varphi_0, \ldots, \varphi_{n+1}; o_1, \ldots, o_n] \subseteq \mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_n \wedge o_n, \varphi_{n+1}]$. On the other hand, for each run $r \in \mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_n \wedge o_n, \varphi_{n+1}]$ there is a run $r' \in \mathcal{R}[\varphi_0, \ldots, \varphi_{n+1}; o_1, \ldots, o_n]$ such that $r'_e(m) = r_e(m)$ for all m, and $o_{(r,m)} = o_m$ for $1 \leq m \leq n$. Since the preference ordering on runs is a function only of the environment states, it is clear that r and r' are compared in the same manner; that is for all r'', $r'' \prec r$ if and only if $r'' \prec r'$, and $r \prec r''$ if and only if $r' \prec r''$. Thus, we conclude that for the purposes of the preference ordering, both $\mathcal{R}[\varphi_0, \varphi_1 \wedge o_1, \ldots, \varphi_n \wedge o_n, \varphi_{n+1}]$ and $\mathcal{R}[\varphi_0, \ldots, \varphi_{n+1}; o_1, \ldots, o_n]$ are compared in the same manner to other sets. It easy to see that this suffices to show that Pl_a satisfies UPD4. ∎

Finally, we can prove Theorem 6.1.

Theorem 6.1: *A belief change operator $\diamond$ satisfies U1–U8 if and only if there is a system $\mathcal{I} \in \mathcal{C}^U$ such that*

$$Bel(\mathcal{I}, s_a) \diamond \psi = Bel(\mathcal{I}, s_a \cdot \psi)$$

for all epistemic states s_a and formulas $\psi \in \mathcal{L}_e$.

Proof: The "if" direction follows from Lemma A.10. For the "only if" direction, assume that $\diamond$ satisfies U1–U8. By Theorem A.6, there is an update structure $U_\diamond$ such that $[\![\varphi \diamond \psi]\!]_{U_\mathcal{I}} = \min_{U_\mathcal{I}}([\![\varphi]\!]_{U_\mathcal{I}}, [\![\psi]\!]_{U_\mathcal{I}})$ for all $\varphi, \psi \in \mathcal{L}_e$. By Lemma A.11, there is a system $\mathcal{I} \in \mathcal{C}^U$ such that $U_\mathcal{I} = U_\diamond$. From Lemma A.9, it follows that $\mathrm{Bel}(\mathcal{I}, s_a) \diamond \psi = \mathrm{Bel}(\mathcal{I}, s_a \cdot \psi)$ for all local states s_a and formulas $\psi \in \mathcal{L}_e$. ∎

References

Alchourrón, C. E., P. Gärdenfors, and D. Makinson (1985). On the logic of theory change: partial meet functions for contraction and revision. *Journal of Symbolic Logic 50*, 510–530.

Boutilier, C. (1992). Normative, subjective and autoepistemic defaults: adopting the Ramsey test. In B. Nebel, C. Rich, and W. Swartout (Eds.), *Proc. Third International Conference on Principles of Knowledge Representation and Reasoning (KR '92)*, pp. 685–696. San Francisco: Morgan Kaufmann.

Boutilier, C. (1994a). Conditional logics of normality: a modal approach. *Artificial Intelligence 68*, 87–154.

Boutilier, C. (1994b). An event-based abductive model of update. In *Proc. Tenth Biennial Canadian Conference on Artificial Intelligence*, pp. 241–248.

Boutilier, C. (1994c). Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence 68*, 33–85.

Boutilier, C. (1996). Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic 25*, 262–305.

Burgess, J. (1981). Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic 22*, 76–84.

Darwiche, A. and J. Pearl (1997). On the logic of iterated belief revision. *Artificial Intelligence 89*, 1–29.

Davis, R. and W. Hamscher (1988). Model-based reasoning: troubleshooting. In H. Shrobe and T. A. A. for Artificial Intelligence (Eds.), *Exploring AI*, pp. 297–346. SF: Morgan Kaufmann.

de Rijke, M. (1992). Meeting some neighbors. Research Report LP-92-10, University of Amsterdam.

del Val, A. and Y. Shoham (1992). Deriving properties of belief update from theories of action. In *Proceedings, Tenth National Conference on Artificial Intelligence (AAAI '92)*, pp. 584–589. Menlo Park, Calif.: AAAI Press.

del Val, A. and Y. Shoham (1993). Deriving properties of belief update from theories of action (II). In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pp. 732–737. San Francisco: Morgan Kaufmann.

del Val, A. and Y. Shoham (1994). A unified view of belief revision and update. *Journal of Logic and Computation 4*, 797–810.

Dubois, D. and H. Prade (1990). An introduction to possibilistic and fuzzy logics. In G. Shafer and J. Pearl (Eds.), *Readings in Uncertain Reasoning*, pp. 742–761. San Francisco, Calif.: Morgan Kaufmann.

Dubois, D. and H. Prade (1991). Possibilistic logic, preferential models, non-monotonicity and related issues. In *Proc. Twelfth International Joint Conference on Artificial Intelligence (IJCAI '91)*, pp. 419–424. San Francisco: Morgan Kaufmann.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1995). *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press.

Freund, M. and D. Lehmann (1994). Belief revision and rational inference. Technical Report TR 94-16, Hebrew University.

Friedman, N. (1997). *Modeling Beliefs in Dynamic Systems.* Ph. D. thesis, Stanford.

Friedman, N. and J. Y. Halpern (1994). Conditional logics of belief change. In *Proc. National Conference on Artificial Intelligence (AAAI '94)*, pp. 915–921. Menlo Park, Calif.: AAAI Press.

Friedman, N. and J. Y. Halpern (1995). Plausibility measures: a user's manual. In P. Besnard and S. Hanks (Eds.), *Proc. Eleventh Conference on Uncertainty in Artificial Intelligence (UAI '95)*, pp. 175–184. San Francisco: Morgan Kaufmann.

Friedman, N. and J. Y. Halpern (1996a). Belief revision: A critique. In L. C. Aiello, J. Doyle, and S. C. Shapiro (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Fifth International Conference (KR '96).* San Francisco: Morgan Kaufmann.

Friedman, N. and J. Y. Halpern (1996b). A qualitative Markov assumption and its implications for belief change. In E. Horvitz and F. Jensen (Eds.), *Proc. Twelfth Conference on Uncertainty in Artificial Intelligence (UAI '96)*, pp. 263–273. San Francisco: Morgan Kaufmann.

Friedman, N. and J. Y. Halpern (1997a). Modeling belief in dynamic systems. part I: foundations. *Artificial Intelligence*. To appear. A preliminary version appears in R. Fagin editor. *Theoretical Aspects of Reasoning about Knowledge: Proc. Fifth Conference*, 1994, pp. 44–64, under the title "A knowledge-based framework for belief change. Part I: foundations".

Friedman, N. and J. Y. Halpern (1997b). Plausibility measures and default reasoning. *Journal of the ACM*. To appear. A preliminary version of this work appeared in *Proc. National Conference on Artificial Intelligence (AAAI '96)*, 1996, pages 1297–1304.

Fuhrmann, A. (1989). Reflective modalities and theory change. *Synthese 81*, 115–134.

Gärdenfors, P. (1986). Belief revision and the Ramsey test for conditionals. *Philosophical Review 91*, 81–93.

Gärdenfors, P. (1988). *Knowledge in Flux.* Cambridge, Mass.: MIT Press.

Gärdenfors, P. and D. Makinson (1988). Revisions of knowledge systems using epistemic entrenchment. In M. Vardi (Ed.), *Proc. Second Conference on Theoretical Aspects of Reasoning about Knowledge*, pp. 83–95. San Francisco: Morgan Kaufmann.

Goldszmidt, M., P. Morris, and J. Pearl (1993). A maximum entropy approach to nonmonotonic reasoning. *IEEE Transactions of Pattern Analysis and Machine Intelligence 15*(3), 220–232.

Goldszmidt, M. and J. Pearl (1992). Rank-based systems: A simple approach to belief revision, belief update and reasoning about evidence and actions. In B. Nebel, C. Rich, and W. Swartout (Eds.), *Proc. Third International Conference on Principles of Knowledge Representation and Reasoning (KR '92)*, pp. 661–672. San Francisco: Morgan Kaufmann.

Grahne, G., A. Mendelzon, and R. Rieter (1992). On the semantics of belief revision systems. In Y. Moses (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Fourth Conference*, pp. 132–142. San Francisco: Morgan Kaufmann.

Grove, A. (1988). Two modelings for theory change. *Journal of Philosophical Logic 17*, 157–170.

Halpern, J. Y. and R. Fagin (1989). Modelling knowledge and action in distributed systems. *Distributed Computing 3*(4), 159–179. A preliminary version appeared in *Proc. 4th ACM Symposium on Principles of Distributed Computing*, 1985, with the title "A formal model of knowledge, action, and communication in distributed systems: preliminary report".

Halpern, J. Y. and M. Y. Vardi (1989). The complexity of reasoning about knowledge and time, I: lower bounds. *Journal of Computer and System Sciences 38*(1), 195–237.

Katsuno, H. and A. Mendelzon (1991a). On the difference between updating a knowledge base and revising it. In J. A. Allen, R. Fikes, and E. Sandewall (Eds.), *Proc. Second International Conference on Principles of Knowledge Representation and Reasoning (KR '91)*, pp. 387–394. San Francisco: Morgan Kaufmann.

Katsuno, H. and A. Mendelzon (1991b). Propositional knowledge base revision and minimal change. *Artificial Intelligence 52*(3), 263–294.

Katsuno, H. and K. Satoh (1991). A unified view of consequence relation, belief revision and conditional logic. In *Proc. Twelfth International Joint Conference on Artificial Intelligence (IJCAI '91)*, pp. 406–412. San Francisco: Morgan Kaufmann.

Kautz, H. A. (1986). Logic of persistence. In *Proceedings, Fifth National Conference on Artificial Intelligence (AAAI '86)*, pp. 401–405. Menlo Park, Calif.: AAAI Press.

Keller, A. M. and M. Winslett (1985). On the use of an extended relational model to handle changing incomplete information. *IEEE Transactions on Software Engineering SE-11*(7), 620–633.

Kraus, S., D. Lehmann, and M. Magidor (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence 44*, 167–207.

Lehmann, D. (1995). Belief revision, revised. In *Proc. Fourteenth International Joint Conference on Artificial Intelligence (IJCAI '95)*, pp. 1534–1540. San Francisco: Morgan Kaufmann.

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

Lewis, D. K. (1973). *Counterfactuals*. Cambridge, Mass.: Harvard University Press.

Manna, Z. and A. Pnueli (1992). *The Temporal Logic of Reactive and Concurrent Systems*, Volume 1. Berlin/New York: Springer-Verlag.

Pearl, J. (1989). Probabilistic semantics for nonmonotonic reasoning: a survey. In R. J. Brachman, H. J. Levesque, and R. Reiter (Eds.), *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pp. 505–516. Reprinted in *Readings in Uncertain Reasoning*, G. Shafer and J. Pearl (eds.), Morgan Kaufmann, San Francisco, Calif., 1990, pp. 699–710.

Rott, H. (1991). Two methods of constructing contractions and revisions of knowledge systems. *Journal of Philosophical Logic 20*, 149–173.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, N.J.: Princeton University Press.

Shoham, Y. (1987). A semantical approach to nonmonotonic logics. In *Proc. 2nd IEEE Symp. on Logic in Computer Science*, pp. 275–279. Reprinted in M. L. Ginsberg (Ed.), *Readings in Nonmonotonic Reasoning*, Morgan Kaufman, San Francisco, Calif., 1987, pp. 227–250.

Shoham, Y. (1988). Chronological ignorance: experiments in nonmonotonic temporal reasoning. *Artificial Intelligence 36*, 271–331.

Spohn, W. (1988). Ordinal conditional functions: a dynamic theory of epistemic states. In W. Harper and B. Skyrms (Eds.), *Causation in Decision, Belief Change, and Statistics*, Volume 2, pp. 105–134. Dordrecht, Netherlands: Reidel.

Wang, Z. and G. J. Klir (1992). *Fuzzy Measure Theory.* New York: Plenum Press.

Williams, M. (1994). Transmutations of knowledge systems. In J. Doyle, E. Sandewall, and P. Torasso (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, pp. 619–629. San Francisco: Morgan Kaufmann.

Winslett, M. (1988). Reasoning about action using a possible models approach. In *Proceedings, Seventh National Conference on Artificial Intelligence (AAAI '88)*, pp. 89–93. Menlo Park, Calif.: AAAI Press.