Belief Revision: A Critique*

Nir Friedman
Computer Science Division
University of California
Berkeley CA 94720.
nir@cs.berkeley.edu
http://www.cs.berkeley.edu/~nir

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

June 20, 1997

Abstract

We examine carefully the rationale underlying the approaches to belief change taken in the literature, and highlight what we view as methodological problems. We argue that to study belief change carefully, we must be quite explicit about the "ontology" or scenario underlying the belief change process. This is something that has been missing in previous work, with its focus on postulates. Our analysis shows that we must pay particular attention to two issues that have often been taken for granted: The first is how we model the agent's epistemic state. (Do we use a set of beliefs, or a richer structure, such as an ordering on worlds? And if we use a set of beliefs, in what language are these beliefs are expressed?) We show that even postulates that have been called "beyond controversy" are unreasonable when the agent's beliefs include beliefs about her own epistemic state as well as the external world. The second is the status of observations. (Are observations known to be true, or just believed? In the latter case, how firm is the belief?) Issues regarding the status of observations arise particularly when we consider *iterated* belief revision, and we must confront the possibility of revising by φ and then by $\neg\varphi$.

Keywords: Belief revision, AGM postulates, Iterated revision.

*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. This work was also supported in part by NSF under grants IRI-95-03109 and IRI-96-25901, by the Air Force Office of Scientific Research under grant F49620-96-1-0323, and by an IBM Graduate Fellowship to the first author. A preliminary version of this paper appeared in L. C. Aiello, J. Doyle, and S. C. Shapiro (Eds.) *Principles of knowledge representation and reasoning : proc. Fifth International Conference (KR '96)*, pp. 421–431, 1996.

1 Introduction

The problem of *belief change*—how an agent should revise her beliefs upon learning new information—has been an active area of research in both philosophy and artificial intelligence. The problem is a fascinating one in part because it is clearly no unique answer. Nevertheless, there is a strong intuition that one wants to make *minimal* changes, and all the approaches to belief change in the literature, such as [Alchourrón, Gärdenfors, and Makinson 1985; Gärdenfors 1988; Katsuno and Mendelzon 1991a], try to incorporate this principle. However, approaches differ on what constitutes a minimal change. This issue has come to the fore with the spate of recent work on *iterated* belief revision (see, for example, [Boutilier 1996a; Boutilier and Goldszmidt 1993; Darwiche and Pearl 1994; Freund and Lehmann 1994; Lehmann 1995; Levi 1988; Williams 1994]).

The approaches to belief change typically start with a collection of postulates, argue that they are reasonable, and prove some consequences of these postulates. Occasionally, a semantic model for the postulates is provided and a representation theorem is proved (of the form that every semantic model corresponds to some belief revision process, and that every belief revision process can be captured by some semantic model). Our goal in this paper is not to introduce yet another model of belief change, but to examine carefully the rationale underlying the approaches in the literature. The main message of the paper is that describing postulates and proving a representation theorem is not enough. While it may have been reasonable when research on belief change started in the early 1980s to just consider the implications of a number of seemingly reasonable postulates, it is our view that it should no longer be acceptable just to write down postulates and give short English justifications for them. While postulates do provide insight and guidance, it is also important to describe what, for want of a better word, we call the underlying *ontology* or scenario for the belief change process. Roughly speaking, this means describing carefully what it means for something to be believed by an agent and what the status is of new information that is received by the agent. This point will hopefully become clearer as we present our critique. We remark that even though the issue of ontology is tacitly acknowledged in a number of papers (for example, in the last paragraph of [Lehmann 1995]), it rarely enters into the discussion in a significant way. We hope to show that ontology must play a central role in all discussions of belief revision.

Our focus is on approaches that take as their starting point the postulates for belief revision proposed by Alchourrón, Gärdenfors, and Makinson (AGM from now on) [1985], but our critique certainly applies to other approaches as well, in particular, Katsuno and Mendelzon's *belief update* [1991b]; see Section 5. The AGM approach assumes that an agent's epistemic state is represented by a *belief set*, that is, a set K of formulas in a logical language $\mathcal{L}$. What the agent learns is assumed to be characterized by some formula φ, also in $\mathcal{L}$; $K * \varphi$ describes the belief set of an agent that starts with belief set K and learns φ.

There are two assumptions implicit in this notation:

- The functional form of $*$ suggests that all that matters regarding how an agent revises her beliefs is the belief set and what is learnt.
- The notation suggests that the second argument of $*$ can be an arbitrary formula in $\mathcal{L}$. But what does it mean to revise by *false*? In what sense can *false* be learnt? More generally, is it reasonable to assume that an arbitrary formula can be learnt in a given epistemic state?

The first assumption is particularly problematic when we consider the postulates that AGM require $*$ to satisfy. These essentially state that the agent is consistent in her choices, in the sense that she acts as though she has an ordering on the strength of her beliefs [Gärdenfors and Makinson 1988; Grove 1988], or an ordering on possible worlds [Boutilier 1994; Grove 1988; Katsuno and Mendelzon 1991b], or some other predetermined manner of choosing among competing beliefs [Alchourrón, Gärdenfors, and Makinson 1985]. However, the fact that an agent's epistemic state is characterized by a collection of formulas means that the epistemic state cannot include information about relative strength of beliefs (as required for the approach of, say, [Gärdenfors and Makinson 1988]), unless this information is expressible in the language. Note that if $\mathcal{L}$ is propositional logic or first-order logic, such information cannot be expressed. On the other hand, if $\mathcal{L}$ contains *conditional* formulas of the form $p > q$, interpreted as "if p is learnt, then q will be believed", then constraints on the relative strength of beliefs can be expressed (indirectly, by describing which beliefs will be retained after a revision).

Problems arise when the language is not rich enough to express relative degrees of strength in beliefs. Consider, for example, a situation where $K = Cl(p \wedge q)$ (the logical closure of $p \wedge q$; that is, the agent's beliefs are characterized by the formula $p \wedge q$ and its logical consequences), and then the agent learns $\varphi = \neg p \vee \neg q$. We can imagine that an agent whose belief in p is stronger than her belief in q would have $K * \varphi = Cl(p)$. That is, the agent gives up her belief in q, but retains a belief in p. On the other hand, if the agent's belief in q is stronger than her belief in p, it seems reasonable to expect that $K * \varphi = Cl(q)$. This suggests that it is unreasonable to take $*$ to be a function if the representation language is not rich enough to express what may be significant details of an agent's epistemic state.

We could, of course, assume that information about the relative strength of beliefs in various propositions is implicit in the choice of the revision operator $*$, even if it is not contained in the language. This is perfectly reasonable, and also makes it more reasonable that $*$ be a function. However, note that we can then no longer assume that we use the same $*$ when doing iterated revision, since there is no reason to believe that the relative strength of beliefs is maintained after we learn a formula. In fact, in a number of recent papers [Boutilier 1996a; Boutilier and Goldszmidt 1993; Friedman and Halpern 1997a; Williams 1994], $*$ is defined as a function from (epistemic states $\times$ formulas) to epistemic states, but the epistemic states are no longer just belief sets; they include information regarding relative strengths of beliefs. The revision function on epistemic states induces a mapping from (belief sets $\times$ formulas) to belief sets, but at the level of belief sets, the mapping may not be functional; for a belief set K and formula φ, the belief set $K * \varphi$ may depend on what epistemic state induced K. Thus, the effect of $*$ on belief sets may change over time.[1]

There is certainly no agreement on what postulates belief change should satisfy. However, the following two postulates are almost universal:

- $\varphi \in K * \varphi$
- if K is consistent and $\varphi \in K$, then $K * \varphi = K$.

These postulates have been characterized by Rott [1989] as being "beyond controversy". Nevertheless, we argue that they are not as innocent as they may at first appear.

The first postulate says that the agent believes the last thing she learns. Making sense of this requires some discussion of the underlying ontology. It certainly makes sense if we take (as Gärdenfors [1988] does) the belief set K to consist of all formulas that are accepted by the agent (where "accepted" means "treated as true"), and the agent revises by φ only if φ has somehow come to be accepted. However, note that deciding when a formula has come to be accepted is nontrivial. In particular, just observing φ will not in general be enough for accepting φ. Acceptance has a complex interaction with what is already believed. For example, imagine a scientist who believes that heavy objects drop faster than light ones, climbs the tower of Pisa, drops a 5 kilogram textbook and a 500 milligram novel, and observes they hit the ground at the same time. This scientist will probably not accept that the time for an object to fall to the ground is independent of its weight, on the basis of this one experiment (although perhaps repeated experiments may lead her to accept it).

While the acceptance point of view is certainly not unreasonable, the fact that just observing φ is not necessarily enough for acceptance often seems forgotten. It also seems hard to believe that *false* would ever be accepted. More generally, it is far from obvious that in a given epistemic state K we should allow arbitrary consistent formulas to be accepted. Intuitively, this does not allow for the possibility that some beliefs are held so firmly that their negations could never be accepted. (Later we describe an ontology where observations are taken to be known in which in fact some consistent formulas will not be accepted in some epistemic states.)

While the second postulate is perhaps plausible if we cannot talk about epistemic importance or strength of belief in the language, it is less so once we can talk about such things (or if either epistemic importance or strength of belief is encoded in the epistemic state some other way). For suppose that $\varphi \in K$. Why should $K * \varphi = K$? It could well be that being informed of φ raises the importance of φ in the epistemic

[1] Freund and Lehmann [1994] have called the viewpoint that $*$ may change over time the *dynamic* point of view. However, this seems somewhat of a misnomer when applied to papers such as [Boutilier 1996a; Boutilier and Goldszmidt 1993; Friedman and Halpern 1997a; Williams 1994], since there $*$ in fact is static, when viewed as a function on epistemic states and formulas.

ordering, or the agent's strength of belief in φ. If strength of belief can be talked about in the language, then a notion of minimal change should still allow strengths of belief to change, even when something expected is observed. Even if we cannot talk about strength of belief in the language, this observation has an impact on iterated revisions. For example, one assumption made by Lehmann [1995] (his postulate I4) is that if p is believed after revising by φ, then revising by $[\varphi \cdot p \cdot \psi]$—that is, revising by φ then p then ψ—is equivalent to revising by $[\varphi \cdot \psi]$. But consider a situation where after revising by φ, the agent believes both p and q, but her belief in q is stronger than her belief in p. We can well imagine that after learning $\neg p \vee \neg q$ in this situation, she would believe $\neg p$ and q. However, if she first learned p and then $\neg p \vee \neg q$, she would believe p and $\neg q$, because, as a result of learning p, she would give p higher epistemic importance than q. In this case, we would not have $[\varphi \cdot p \cdot (\neg p \vee \neg q)] = [\varphi \cdot (\neg p \vee \neg q)]$. In light of this discussion, it is not surprising that the combination of the second postulate with a language that can talk about epistemic ordering leads to technical problems such as Gärdenfors' *triviality result* [1988].

To give a sense of our concerns here, we discuss two basic ontologies. The first ontology that seems (to us) reasonable assumes that the agent has some knowledge as well as beliefs. We can think of the formulas that the agent knows as having the highest state of epistemic importance. In keeping with the standard interpretation of knowledge, we also assume that the formulas that the agent knows are true in the world. Since agents typically do not have certain knowledge of very many facts, we assume that the knowledge is augmented by beliefs (which can be thought of as defeasible guides to action). Thus, the set of formulas that are known form a subset of the belief set. We assume that the agent observes the world using reliable sensors; thus, if the agent observes φ, then the agent is assumed to know φ. After observing φ, the agent adds φ to his stock of knowledge, and may revise his belief set. Since the agent's observations are taken to be knowledge, the agent will believe φ after observing φ. However, the agent's epistemic state may change even if she observes a formula that she previously believed to be true. In particular, if the formula observed was believed to be true but not known to be true, after the observation it is known. Note that, in this ontology, the agent never observes *false*, since *false* is not true of the world. In fact, the agent never observes anything that contradicts her knowledge. Thus, $K * \varphi$ is defined only for formulas φ that are compatible with the agent's knowledge. Moving to iterated revision, this means we cannot have a revision by φ followed by a revision by $\neg\varphi$. This ontology underlies some of our earlier work [1997, 1997b, 1997a]. As we show here, a variant of Darwiche and Pearl's approach [1994] captures them as well.

We can consider a second ontology that has a different flavor. In this ontology, if we observe something, we believe it to be true and perhaps even assign it a strength of belief. But this assignment does not represent the strength of belief of the observation in the resulting epistemic state. Rather, the belief in the observation must "compete" against current beliefs if it is inconsistent with these beliefs. In this ontology, it is not necessarily the case that $\varphi \in K * \varphi$, just as it is not the case that a scientist will necessarily adopt the consequences of his most recent observation into his stock of beliefs (at least, not without doing some additional experimentation). Of course, to flesh out this ontology, we need to describe how to combine a given strength of belief in the observation with the strengths of the beliefs in the original epistemic state. Perhaps the closest parallel in the literature is something like the Dempster-Shafer rule of combination [Shafer 1976], which gives a rule for combining two separate bodies of belief. We believe that this type of ontology deserves further study. We sketch one particular approach to modeling this ontology in Section 5.

The rest of the paper is organized as follows. In Section 2, we review the AGM framework, and point out some problems with it. In Section 3, we consider proposals for belief change and iterated belief change from the literature due to Boutilier [1996a], Darwiche and Pearl [1994], Freund and Lehmann [1994], and Lehmann [1995], and try to understand the ontology implicit in the proposal (to the extent that one can be discerned). In Section 4, we consider the first ontology discussed above in more detail. We conclude with some discussion in Section 5.

2 AGM Belief Revision

In this section we review the AGM approach to belief revision. As we said earlier, this approach assumes that beliefs and observations are expressed in some language $\mathcal{L}$. It is assumed that $\mathcal{L}$ is closed under negation and conjunction, and comes equipped with a consequence relation $\vdash_{\mathcal{L}}$ that contains the propositional calculus and satisfies the deduction theorem. The agent's epistemic state is represented by a belief set, that is, a set of formulas in $\mathcal{L}$ closed under deduction. There is also assumed to be a revision operator $*$ that takes a belief set K and a formula φ and returns a new belief set $K * \varphi$, intuitively, the result of revising K by φ. The following AGM postulates are an attempt to characterize the intuition of "minimal change":

R1. $K * \varphi$ is a belief set

R2. $\varphi \in K * \varphi$

R3. $K * \varphi \subseteq Cl(K \cup \{\varphi\})$

R4. If $\neg\varphi \notin K$ then $Cl(K \cup \{\varphi\}) \subseteq K * \varphi$

R5. $K * \varphi = Cl(\mathit{false})$ if and only if $\vdash_{\mathcal{L}} \neg\varphi$

R6. If $\vdash_{\mathcal{L}} \varphi \Leftrightarrow \psi$ then $K * \varphi = K * \psi$

R7. $K * (\varphi \wedge \psi) \subseteq Cl(K * \varphi \cup \{\psi\})$

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

The essence of these postulates is the following. Revising K by φ gives a belief set (Postulate R1) that includes φ (R2). If φ is consistent with K, then $K * \varphi$ consist precisely of those beliefs implied by the combination of the old beliefs with the new belief

(R3 and R4). Note that it follows from R1–R4 that if $\varphi \in K$, then $K = K * \varphi$. The next two postulates discuss the coherence of beliefs. R5 states that as long as φ is consistent, then so is $K * \varphi$, and 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 $K * \varphi$ then $K * (\varphi \wedge \psi)$ is the result of adding ψ to $K * \varphi$.

The intuitions described by AGM is based on one-step (noniterated) revision. Nevertheless, the AGM postulates do impose some restrictions on iterated revisions. For example, suppose that q is consistent with $K * p$. Then, according to R2 and R3, $(K*p)*q = \mathrm{Cl}(K*p \cup \{q\})$. Using R7 and R8 we can conclude that $(K*p)*q = K*(p \wedge q)$.

There are several representation theorems for AGM belief revision; perhaps the clearest is due to Grove [1988]. We discuss a slight modification, due to Boutilier [1994] and Katsuno and Mendelzon [1991b]: Let an *$\mathcal{L}$-world* be a complete and consistent truth assignment to the formulas in $\mathcal{L}$. Let $\mathcal{W}$ consist of all the $\mathcal{L}$-worlds, and let $\preceq$ be a *ranking*, that is, a total preorder, on the worlds in $\mathcal{W}$. Let $\min_{\preceq}$ consist of all the minimal worlds with respect to $\preceq$, that is, all the worlds w such that there is no w' with $w' \prec w$. With $\preceq$ we can associate a belief set $\mathrm{Bel}(\preceq)$, consisting of all formulas φ that are true in all the worlds in $\min_{\preceq}$. Moreover, we can define a revision operator $*$ on $\mathrm{Bel}(\preceq)$, by taking $\mathrm{Bel}(\preceq) * \varphi$ to consist of all formulas ψ that are true in all the minimal φ-worlds according to $\preceq$. It can be shown that $*$ satisfies the AGM postulates (when its first argument is $\mathrm{Bel}(\preceq)$). Thus, we can define a revision operator by taking a collection of orderings $\preceq_K$, one for each belief set K. To define $K * \varphi$ for a belief set K, we apply the procedure above, starting with the ranking $\preceq_K$ corresponding to K.[2] Furthermore, Grove [1988], Katsuno and Mendelzon [1991b], and Boutilier [1994] show that every belief revision operator satisfying the AGM axioms can be characterized in this way.

This elegant representation theorem also brings out some of the problems with the AGM postulates. First, note that a given revision operator $*$ is represented by a family of rankings, one for each belief set. There is no necessary connection between the rankings corresponding to different belief sets. It might seem more reasonable to have a more global setting (perhaps one global ranking) from which each element in the family of rankings arises.

A second important point is that the epistemic state here is represented not by a belief set, but by a ranking. Each ranking $\preceq$ is associated with a belief set $\mathrm{Bel}(\preceq)$, but it is the ranking that gives the information required to describe how revision is carried out. The belief set does not suffice to determine the revision; there are many rankings $\preceq$ for which the associated belief set $\mathrm{Bel}(\preceq)$ is K. Since the revision process only gives us the revised belief set, not the revised ranking, the representation does not support iterated revision.

[2] In this construction, for each belief set K other than the inconsistent belief set, we have $\mathrm{Bel}(\preceq_K) = K$. The inconsistent belief set gets special treatment here.

This suggests that we should consider, not how to revise belief sets, but how to revise rankings. More generally, whatever we take to be our representation of the epistemic state, it seems appropriate to consider how these representations should be revised. We can define an analogue of the AGM postulates for epistemic states in a straightforward way (cf. [Friedman and Halpern 1994]): Taking E to range over epistemic states and $Bel(E)$ to represent the belief set associated with epistemic state E, we have

R1′. $E * \varphi$ is an epistemic state

R2′. $\varphi \in Bel(E * \varphi)$

R3′. $Bel(E * \varphi) \subseteq Cl(Bel(E) \cup \{\varphi\})$

and so on, with the obvious syntactic transformation. In fact, as we shall see in the next section, a number of processes for revising epistemic states have been considered in the literature, and in fact they all do satisfy these modified postulates.

Finally, even if we restrict attention to belief sets, we can consider what happens if the underlying language $\mathcal{L}$ is rich enough to talk about how revision should be carried out. For example, suppose $\mathcal{L}$ includes conditional formulas, and we want to find some ranking $\preceq$ for which the corresponding belief set is K. Not just any ranking $\preceq$ such that $\mathrm{Bel}(\preceq) = K$ will do here. The beliefs in K put some constraints on the ranking. For example, if $p > q$ is in K and $p \notin K$, then the minimal $\preceq$-worlds satisfying p must all satisfy q, since after p is learnt, q is believed. Once we restrict to rankings $\preceq_K$ that are consistent with K then the AGM postulates are no longer sound. This point has essentially been made before [Boutilier 1992; Rott 1989]. However, it is worth stressing the sensitivity of the AGM postulates to the underlying language and, more generally, to the choice of epistemic state.

3 Proposals for Iterated Revision

We now briefly review some of the previous proposals for iterated belief change, and point out how the impact of the observations we have been making on the approaches. Most of these approaches start with the AGM postulates, and augment them to get seemingly appropriate restrictions on iterated revision. This is not an exhaustive review of the literature on iterated belief revision by any stretch of the imagination. Rather, we have chosen a few representative approaches that allow us to bring out our methodological concerns.

3.1 Boutilier's Natural Revision

As we said in the previous section, Boutilier takes the agent's epistemic state to consist of a ranking of possible worlds. Boutilier [1996a] describes a particular revision operator $*_B$ on epistemic states that he calls *natural revision* operator. Natural revision maps a ranking $\preceq$ of possible worlds and an observation φ to a revised ranking $\preceq *_B \varphi$ such that

(a) $\preceq *_B \varphi$ satisfies the conditions of the representation theorem described above—the minimal worlds in $\preceq *_B \varphi$ are precisely the minimal φ-worlds in $\preceq$, and (b) in a precise sense, $\preceq *_B \varphi$ is the result of making the minimal number of changes to $\preceq$ required to guarantee that all the minimal worlds in $\preceq *_B \varphi$ satisfy φ. Given a ranking $\preceq$ and a formula φ, the ranking $\preceq *_B \varphi$ is identical to $\preceq$ except that the minimal φ-worlds according to $\preceq$ have the minimal rank in the revised ranking, while the relative ranks of all other worlds remains unchanged.

Boutilier characterizes the properties of natural revision. Suppose that, starting in some epistemic state, we revise by $\varphi_1, \ldots, \varphi_n$. Further suppose φ_{i+1} is consistent with the beliefs after revising by $\varphi_1, \ldots, \varphi_i$. Then the beliefs after revising by $\varphi_1, \ldots, \varphi_n$ are precisely the beliefs after observing $\varphi_1 \wedge \ldots \wedge \varphi_n$. (More precisely, given any ranking $\preceq$, the belief set associated with the ranking $\preceq *_B \varphi_1 *_B \ldots *_B \varphi_n$ is the same as that associated with the ranking $\preceq *_B (\varphi_1 \wedge \ldots \wedge \varphi_n)$. Note, however, that $\preceq *_B \varphi_1 *_B \ldots *_B \varphi_n \neq \preceq *_B (\varphi \wedge \ldots \wedge \varphi_n)$ in general.) Thus, as long as the agent's new observations are not surprising, the agent's beliefs are exactly the ones she would have had had she observed the conjunction of all the observations. This is an immediate consequence of the AGM postulates, and thus holds for any approach that attempts to extend the AGM postulates to iterated revision.

What happens when the agent observes a formula φ_{n+1} that is inconsistent with her current beliefs? Boutilier shows that in this case the new observation nullifies the impact of the all the observations starting with the most recent one that is inconsistent with φ_{n+1}. More precisely, suppose φ_{i+1} is consistent with the belief after observing $\varphi_1, \ldots, \varphi_i$ for $i \leq n$, but φ_{n+1} is inconsistent with the beliefs after observing $\varphi_1, \ldots, \varphi_n$. Let k be the maximal index such that φ_{n+1} is consistent with the beliefs after learning $\varphi_1, \ldots, \varphi_k$. The agent's beliefs after observing φ_{n+1} are the same as her beliefs after observing $\varphi_1, \ldots, \varphi_k, \varphi_{n+1}$. Thus, the agent acts as though she did not observe $\varphi_{k+1}, \ldots, \varphi_n$.

Boutilier does not provide any argument for the reasonableness of this ontology. In fact, Boutilier's presentation (like almost all others in the literature) is not in terms of an ontology at all; he presents natural revision as an attempt to minimize changes to the ranking. While the intuition of minimizing changes to the ranking seems reasonable at first, it becomes less reasonable when we realize its ontological implications. The following example, due to Darwiche and Pearl [1994], emphasizes this point. Suppose we encounter a strange new animal and it appears to be a bird, so we believe it is a bird. On closer inspection, we see that it is red, so we believe that it is a red bird. However, an expert then informs us that it is not a bird, but a mammal. Applying natural revision, we would no longer believe that the animal is red. This does not seem so reasonable.

One more point is worth observing: As described by Boutilier [1996a], natural revision does not allow revision by *false*. While we could, of course, modify the definition to handle *false*, it is more natural simply to disallow it. This suggests that, whatever ontology is used to justify natural revision, in that ontology, revising by *false* should not make sense.

3.2 Freund and Lehmann's Approach

Freund and Lehmann [1994] stick close to the original AGM approach. They work with belief sets, not more general epistemic states. However, they are interested in iterated revision. They consider the effect of adding just one more postulate to the basic AGM postulates, namely

FL. If $\neg\varphi \in K$, then $K * \varphi = K_\perp * \varphi$,

where $K_\perp$ is the inconsistent belief set, which consists of all formulas.

Suppose $*$ satisfies R1–R8 and FL. Just as with Boutilier's natural revision, if φ_{i+1} is consistent with the beliefs after learning $\varphi_1, \ldots, \varphi_i$ for $i \leq n-1$, then $K * \varphi_1 * \ldots * \varphi_n = K * (\varphi_1 \wedge \ldots \wedge \varphi_n)$. However, if we then observe φ_{n+1}, and it is inconsistent with $K * \varphi_1 \wedge \ldots \wedge \varphi_n$, then $K * \varphi_1 * \ldots * \varphi_{n+1} = K_\perp * \varphi_{n+1}$. That is, observing something inconsistent causes us to retain none of our previous beliefs, but to start over from scratch. While the ontology here is quite simple to explain, as Freund and Lehmann themselves admit, it is a rather severe form of belief revision. Darwiche and Pearl's red bird example applies to this approach as well.

3.3 Darwiche and Pearl's Approach

Darwiche and Pearl [1994] suggest a set of postulates extending the AGM postulates, and claim to provide a semantics that satisfies them. Their intuition is that the revision operator should retain as must as possible certain parts of the ordering among worlds in the ranking. In particular, if w and w' both satisfy φ, then a revision by φ should not change the relative rank of w and w'. Similarly, if both w and w' satisfy $\neg\varphi$, then a revision should not change their relative rank. They describe four postulates that are meant to embody these intuitions:

C1. If $\varphi \vdash \psi$, then $(K * \psi) * \varphi = K * \varphi$

C2. If $\varphi \vdash \neg\psi$, then $(K * \psi) * \varphi = K * \varphi$

C3. If $\psi \in K * \varphi$, then $\psi \in (K * \psi) * \varphi$

C4. If $\neg\psi \notin K * \varphi$, then $\neg\psi \notin (K * \psi) * \varphi$

Freund and Lehmann [1994] point out that C2 is inconsistent with the AGM postulates. This observation seems inconsistent with the fact that Darwiche and Pearl claim to provide an example of a revision method that is consistent with their postulates. What is going on here? It turns out that the issues raised earlier help clarify the situation.

The semantics that Darwiche and Pearl use as an example is based on a special case of Spohn's *ordinal conditional functions* (OCFs) [1988] called *κ-rankings* [Goldszmidt and Pearl 1992]. A κ-ranking associates with each world either a natural number n or ∞, with the requirement that for at least one world w_0, we have $\kappa(w_0) = 0$. We can think of $\kappa(w)$ as the rank of w, or as denoting how surprising it would be to discover that w is the actual world. If $\kappa(w) = 0$, then world w is unsurprising; if $\kappa(w) = 1$, then w is somewhat surprising; if $\kappa(w) = 2$, then w is more surprising, and so on. If $\kappa(w) = \infty$, then w is impossible.[3] OCFs provide a way of ranking worlds that is closely related to, but has a little more structure than the orderings considered by Boutilier. The extra structure makes it easier to define a notion of conditioning.

Given a formula φ, let $\kappa(\varphi) = \min\{\kappa(w) : w \models \varphi\}$; we define $\kappa(\mathit{false}) = \infty$. We say that *$\varphi$ is believed with firmness $\alpha \geq 0$ in OCF κ* if $\kappa(\varphi) = 0$ and $\kappa(\neg\varphi) = \alpha$. Thus, φ is believed with firmness α if φ is unsurprising and the least surprising world satisfying $\neg\varphi$ has rank α. We define $\mathrm{Bel}(\kappa)$ to consist of all formulas that are believed with firmness at least 1.

Spohn defined a notion of conditioning on OCFs. Given an OCF κ, a formula φ such that $\kappa(\varphi) < \infty$, and $\alpha \geq 0$, $\kappa_{\varphi,\alpha}$ is the unique OCF satisfying the property desired by Darwiche and Pearl—namely, if w and w' both satisfy φ or both satisfy $\neg\varphi$, then revision by φ should not change the relative rank of w and w', that is, $\kappa_{\varphi,\alpha}(w) - \kappa_{\varphi,\alpha}(w') = \kappa(w) - \kappa(w')$—such that φ is believed with firmness α in $\kappa_{\varphi,\alpha}$. It is defined as follows:

$$\kappa_{\varphi,\alpha}(w) = \begin{cases} \kappa(w) - \kappa(\varphi) & \text{if } w \text{ satisfies } \varphi \\ \kappa(w) - \kappa(\neg\varphi) + \alpha & \text{if } w \text{ satisfies } \neg\varphi. \end{cases}$$

Notice that $\kappa_{\varphi,\alpha}$ is defined only if $\kappa(\varphi) < \infty$, that is, if φ is considered possible.

Darwiche and Pearl defined the following revision function on OCFs:

$$\kappa *_{DP} \varphi = \begin{cases} \kappa & \text{if } \kappa(\neg\varphi) \geq 1 \\ \kappa_{\varphi,1} & \text{otherwise.} \end{cases}$$

Thus, if φ is already believed with firmness at least 1 in κ, then κ is unaffected by a revision by φ; otherwise, the effect of revision is to modify κ by conditioning so that φ ends up being believed with degree of firmness 1. Intuitively, this means that if φ is not believed in κ, in $\kappa * \varphi$ it is believed, but with the minimal degree of firmness.

It is not hard to show that if we take an agent's epistemic state to be represented by an OCF, then Darwiche and Pearl's semantics satisfies all the AGM postulates modified to apply to epistemic states (that is, R1′–R8′ in Section 2), except that revising by *false* is disallowed, just as in natural revision, so that R5′ holds vacuously; in addition, this semantics satisfies Darwiche and Pearl's C1–C4, modified to apply to epistemic states. For example, C2 becomes

C2′. If $\varphi \vdash \neg\psi$, then $\mathrm{Bel}((E * \psi) * \varphi) = \mathrm{Bel}(E * \varphi)$.

[3] Spohn allowed ranks to be arbitrary ordinals, not just natural numbers, and did not allow a rank of ∞, since, for philosophical reasons, he did not want to allow a world to be considered impossible. As we shall see, there are technical advantages to introducing a rank of ∞.

Indeed, as Darwiche and Pearl observe, natural revision also satisfies C1′–C4′, however, it has properties that they view as undesirable. Thus, Darwiche and Pearl's claim that their postulates are consistent with AGM is correct, if we think at the level of general epistemic states. On the other hand, Freund and Lehmann are quite right that R1–R8 and C1–C4 are incompatible; indeed, as they point out, R1–R4 and C2 are incompatible. The importance of making clear exactly whether we are considering the postulates with respect to the OCF κ or the belief set $\mathrm{Bel}(\kappa)$ is particularly apparent here.[4]

The fact that Boutilier's natural revision also satisfies C1′–C4′ clearly shows that these postulates do not capture all of Darwiche and Pearl's intuitions. Their semantics embodies further assumptions. Some of them seem *ad hoc*. Why is it reasonable to believe φ with a *minimal* degree of firmness after revising by φ? Rather than trying to come up with an improved collection of postulates (which Darwiche and Pearl themselves suggest might be a difficult task), it seems to us a more promising approach is to find an appropriate ontology.

3.4 Lehmann's Revised Approach

Finally, we consider Lehmann's "revised" approach to belief revision [1995]. With each sequence σ of observations, Lehmann associates a belief set that we denote $\mathrm{Bel}(\sigma)$. Intuitively, we can think of $\mathrm{Bel}(\sigma)$ as describing the agent's beliefs after making the sequence σ of observations, starting from her initial epistemic state. Lehmann allows all possible sequences of consistent formulas. Thus, he assumes that the agent does not observe *false*. We view Lehmann's approach essentially as taking the agent's epistemic state to be the sequence of observations made, with the obvious revision operator that concatenate a new observation to the current epistemic state. The properties of belief change depend on the function Bel. Lehmann require Bel to satisfy the following postulates (where σ and ρ denote sequences of formulas, and $\cdot$ is the concatenation operator):

I1. $\mathrm{Bel}(\sigma)$ is a consistent belief set

I2. $\varphi \in \mathrm{Bel}(\sigma \cdot \varphi)$

I3. If $\psi \in \mathrm{Bel}(\sigma \cdot \varphi)$, then $\varphi \Rightarrow \psi \in \mathrm{Bel}(\sigma)$

I4. If $\varphi \in \mathrm{Bel}(\sigma)$, then $\mathrm{Bel}(\sigma \cdot \varphi \cdot \rho) = \mathrm{Bel}(\sigma \cdot \rho)$

I5. If $\psi \vdash \varphi$, then $\mathrm{Bel}(\sigma \cdot \varphi \cdot \psi \cdot \rho) = \mathrm{Bel}(\sigma \cdot \psi \cdot \rho)$

I6. If $\neg\psi \notin \mathrm{Bel}(\sigma \cdot \varphi)$, then $\mathrm{Bel}(\sigma \cdot \varphi \cdot \psi \cdot \rho) = \mathrm{Bel}(\sigma \cdot \varphi \cdot \varphi \wedge \psi \cdot \rho)$

I7. $\mathrm{Bel}(\sigma \cdot \neg\varphi \cdot \varphi) \subseteq \mathrm{Cl}(\mathrm{Bel}(\sigma) \cup \{\varphi\})$

We refer the interested reader to [Lehmann 1995] for the motivation for these postulates. As Lehmann notes, the spirit of the original AGM postulates is captured by these postulates. Lehmann views I5 and I7 as two main additions to the basic AGM postulates.

[4] We note that in a recent version of their paper, Darwiche and Pearl [1997] use a similar technique to deal with the inconsistency of C2.

He states that "Since postulates I5 and I7 seem secure, i.e., difficult to reject, the postulates I1–I7 may probably be considered as a reasonable formalization of the intuitions of AGM." Our view is that it is impossible to decide whether to accept or reject postulates such as I5 or I7 (or, for that matter, any of the other postulates) without an explicit ontology. There may be ontologies for which I5 and I7 are reasonable, and others for which they are not. "Reasonableness" is not an independently defined notion; it depends on the ontology. The ontology of the next section emphasizes this point.

4 Taking Observations to be Knowledge

We now consider an ontology where observations are taken to be knowledge. As we said in the introduction, in this ontology, the agent has some (closed) set of formulas that he *knows* to be true, which is included in a larger set of formulas that he *believes* to be true. The belief set can be viewed as the result of applying some nonmonotonic reasoning system grounded in the observations. We can think of there being an ordering on the strength of his beliefs, with the formulas known to be true—the observations and their consequences—having the greatest strength of belief. Because observations are taken to be knowledge, any formula observed is added to the stock of knowledge (and must be consistent with what was previously known). In this ontology, it is impossible to observe *false*. In fact, it is impossible to make any inconsistent sequence of observations. That is, if $\varphi_1, \ldots, \varphi_n$ is observed, then $\varphi_1 \wedge \ldots \wedge \varphi_n$ must be consistent (although it may not be consistent with the agent's original beliefs).

In earlier work [1997, 1997a], we presented one way of formalizing this ontology, based on Halpern and Fagin's [1989] framework of multi-agent systems (see [Fagin, Halpern, Moses, and Vardi 1995] for more details). The key assumption in the multi-agent system 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 some *local state*. 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 tuple (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 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. 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*. If $r(m) = (s_e, s_a)$, we define $r_a(m) = s_a$ and $r_e(m) = s_e$. 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* 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 the primitive propositions.

To capture the AGM framework, we consider a special class of interpreted systems: We fix a propositional language $\mathcal{L}$. We assume that the agent makes observations, which are characterized by formulas in $\mathcal{L}$, and that her local state consists of the sequence of observations that she has made. We assume that the environment's local state describes which formulas are actually true in the world, so that it is a truth assignment to the formulas in $\mathcal{L}$. As observed by Katsuno and Mendelzon [1991a], the AGM postulates assume that the world is *static*; to capture this, we assume that the environment state does not change over time. Formally, we are interested in the unique interpreted system $(\mathcal{R}^{AGM}, \pi)$ that consists of all runs satisfying the following two assumptions for every point (r, m):

- The environment's state $r_e(m)$ is a truth assignment to the formulas in $\mathcal{L}$ that agrees with π at (r, m) (that is, $\pi(r, m) = r_e(m)$), and $r_e(m) = r_e(0)$.
- The agent's state $r_a(m)$ is a sequence of the form $\langle \varphi_1, \ldots, \varphi_m \rangle$, such that $\varphi_1 \wedge \ldots \wedge \varphi_m$ is true according to the truth assignment $r_e(m)$ and $r_a(m-1) = \langle \varphi_1, \ldots, \varphi_{m-1} \rangle$.

Notice that the form of the agent's state makes explicit an important implicit assumption: that the agent remembers all her previous observations.

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 $(r, m) \sim_a (r', m')$. It is easy to see that, according to this definition, if $r_a(m) = \langle \varphi_1, \ldots, \varphi_m \rangle$, then the agent knows $\varphi_1 \wedge \ldots \wedge \varphi_m$ at the point (r, m): the agent's observations are known to be true in this approach. We are interested in talking about the agent's beliefs as well as her knowledge. To allow this, we added a notion of *plausibility* to interpreted systems in [Friedman and Halpern 1997b]. We consider a variant of this approach here, using OCFs, since it makes it easier to relate our observations to Darwiche and Pearl's framework.

We assume that we start with an OCF κ on runs such that $\kappa(r) \neq \infty$ for any run r. Intuitively, κ represents our prior ranking on runs. Initially, no runs is viewed as impossible. We then associate, with each point (r, m), an OCF $\kappa^{(r,m)}$ on the runs. We define $\kappa^{(r,m)}$ by induction on m. We take $\kappa^{(r,0)} = \kappa$, and we take $\kappa^{(r,m+1)} = \kappa^{(r,m)}_{\varphi_{m+1},\infty}$, where $r_a(m+1) = \langle \varphi_1, \ldots, \varphi_{m+1} \rangle$. Thus, $\kappa^{(r,m+1)}$ is the result of conditioning $\kappa^{(r,m)}$ on the last observation the agent made, giving it degree of firmness ∞. Thus, the agent is treating the observations as knowledge in a manner compatible with the semantics for knowledge in the interpreted system. Moreover, since observations are known, they are also believed.

As we show in [Friedman 1997; Friedman and Halpern 1997a], this framework satisfies the AGM postulates R1′–R8′, interpreted on epistemic states. (Here we take the agent's epistemic state at the point (r, m) to consist of $r_a(m)$ together with $\kappa^{(r,m)}$.) Moreover, we show this framework satisfies an additional postulate, which we call R9′:

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

This postulate captures the intuition that observations are taken to be knowledge, and thus observing φ and then ψ is equivalent to observing $\varphi \wedge \psi$. In fact, we show there that R1′–R9′ characterize, in a precise sense, revision in this framework.

It is easy to verify that the framework also satisfies Darwiche and Pearl's postulates (appropriately modified to apply to epistemic states), except that the contentious C2 is now vacuous, since it is illegal to revise by ψ and then by φ if $\varphi \vdash \neg\psi$.

How does this framework compare to Lehmann's? Like Lehmann's, there is an explicit attempt to associate beliefs with a sequence of revisions. However, we have restricted the sequence of revisions, since we are treating observations as knowledge. It is easy to see that I1–I3 and I5–I7 hold in our framework. However, since we have restricted the sequence of observations allowed, some of these postulates are much weaker in our framework than in Lehmann's. In particular, I7 is satisfied vacuously, since we do not allow a sequence of the form $\sigma \cdot \neg\varphi \cdot \varphi$. On the other hand, I4 is not satisfied in our framework. Our discussion in the introduction suggests a counterexample. Suppose that initially, $\kappa(p \wedge q) = 0$, $\kappa(\neg p \wedge q) = 1$, $\kappa(p \wedge \neg q) = 2$, and $\kappa(\neg p \wedge \neg q) = 3$. Thus, initially the agent believes both p and q, but believes p with firmness 1 and q with firmness 2. If the agent then observes $\neg p \vee \neg q$, he will then believe q but not p. On the other hand, suppose the agent first observes p. He still believes both p and q, of course, but now p is believed with firmness ∞. That means if he then observes $\neg p \vee \neg q$, he will believe q, but not p, violating I4. However, a weaker variant of I4 does hold in our system: if the agent *knows* φ, then observing φ will not change her future beliefs.

5 Discussion

The goal of this paper was to highlight what we see as some methodological problems in much of the literature on belief revision. There has been (in our opinion) too much attention paid to postulates, and not enough to the underlying ontology. An ontology must make clear what the agent's epistemic state is, what types of observations the agent can make, the status of observations, and how the agent goes about revising the epistemic state.

We have (deliberately) not been very precise about what counts as an ontology, and clearly there are different levels of detail that one can provide. Although some papers on belief revision have attempted to provide something in the way of an ontology, the ontology has typically been insufficient detailed to verify the reasonableness of the postulates. For example, Gärdenfors [1988]—whose ontology is far better developed than most—takes belief sets to consist of formulas that are accepted, and revision to be by formulas that are accepted. As we have seen, unless belief sets or the revision operator contain additional information (such as epistemic importance or strengths of beliefs) this ontology will violate R1. On the other hand, unless we are careful about how we add such information, we may well violate some other axioms (such as R3 or R5). In any case, it is the job of the ontology to make completely clear such issues as whether observations are believed to be true or known to be true, and if they are believed, what the strength of belief is. This issue is particularly important if we have epistemic states like rankings that are richer than belief sets. If observations are believed, but not necessarily known, to be true, then it is not clear how to go about revising such a richer epistemic state. With what degree of firmness should the new belief be held? No particular answer seems to us that well motivated. It may be appropriate for the user to attach degrees of firmness to observations, as was done in [Goldszmidt 1992; Williams 1994; Wobcke 1995] (following the lead of Spohn [1988]); we can even generalize to allowing uncertain observations [Dubois and Prade 1992].

We are currently working on finding alternative ontologies where observations are not taken to be knowledge. One such an ontology, which is motivated by ideas from stochastic processes, is suggested in [Friedman 1997] for the case of update and can be easily adapted for revision as well. (Boutilier [Boutilier 1996b] describes a similar ontology for update.) This ontology explicitly models how "noisy" observations come about. Roughly speaking, we assign prior plausibility to observing a formula φ in a world w. When we observe φ we update our plausibility measure on worlds by considering the plausibility of observing φ in each of these worlds. Of course, we can then consider various assumptions. For example, we might say that we believe observations to be true. That is, our prior plausibility of observing φ in worlds where it is true is higher than the plausibility of observing φ in worlds where it false. We note that this assumption does *not* imply that $\varphi \in \text{Bel}(K * \varphi)$, since we have to combine the plausibility of observing φ with the plausibility of φ before the observation was made. If φ is considered implausible in K, then the single observation of φ might not suffice to make φ more plausible that $\neg\varphi$. (For example, the medieval scientist probably would not change her beliefs about the speed of falling objects after a single experiment.) We hope to report on this work in more detail at a later date.

In this paper we have focused on belief revision. However, the need to clarify the underlying ontology goes far beyond belief revision. Much the same comments can be made for all the work on nonmonotonic logic as well (this point is essentially made in [Halpern 1993]). Not surprisingly, our critique applies to other approaches belief change as well, and in particular to Katsuno and Mendelzon's *belief update* [1991a]. Although the motivation described by Katsuno and Mendelzon is different than that of revision, the discussion of update is stated in terms of postulates about belief states. Thus, for example, the distinction between the agent's belief state and epistemic state arises in update as well: update is defined as function from (belief states $\times$ formulas) to belief states. However, the agent's belief state does not uniquely determine the outcome of update. This is demonstrated, for example, by Katsuno and Mendelzon's semantic characterization that requires the agent to have a ternary relation on possible worlds such that $w_1 <_w w_2$ holds if the agent considers w_1 to be "closer" to w than w_2. Other issues we raise here also apply to update for similar reasons.

The ontology we propose in Section 4, where observations are treated as knowledge, can be applied to update as well. Moreover, the assumption that observations are true is less problematic for iterated update: since update does not assume that propositions are static, we can consider runs where φ is true at one time point, and false at the next one. Thus, assuming that observations are known to be true does not rule out sequences of observations of the form $\varphi, \neg\varphi, \ldots$. In fact, all sequences of consistent observations are allowed. We refer the reader to [Friedman 1997; Friedman and Halpern 1997a] for more details.

It seems to us that many of the intuitions that researchers in the area have are motivated by thinking in terms of observations as known, even if this is not always reflected in the postulates considered. We have examined carefully one particular instantiation of this ontology, that of treating observations as knowledge. We have shown that, in this ontology, some postulates that seem reasonable, such as Lehmann's I4, do not hold. We do not mean to suggest that I4 is "wrong" (whatever that might mean in this context). Rather, it shows that we cannot blithely accept postulates without making the underlying ontology clear. We would encourage the investigation of other ontologies for belief change.

Acknowledgements The authors are grateful to Craig Boutilier, Adnan Darwiche, Adam Grove, and Daniel Lehmann for comments on a draft of this paper and useful discussions relating to this work.

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.), *Principles of Knowledge Representation and Reasoning: Proc. Third International Conference (KR '92)*, pp. 685–696. San Francisco, Calif.: Morgan Kaufmann.

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

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

Boutilier, C. (1996b). A unified model of qualitative belief change: A dynamical systems perspective. Submitted to publication.

Boutilier, C. and M. Goldszmidt (1993). Revising by conditional beliefs. In *Proceedings, Eleventh National Conference on Artificial Intelligence (AAAI '93)*, pp. 648–654.

Darwiche, A. and J. Pearl (1994). On the logic of iterated belief revision. In R. Fagin (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Fifth Conference*, pp. 5–23. San Francisco, Calif.: Morgan Kaufmann.

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

Dubois, D. and H. Prade (1992). Belief change and possibility theory. In P. Gärdenfors (Ed.), *Belief Revision*. Cambridge, U.K.: Cambridge University Press.

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). A knowledge-based framework for belief change. Part II: revision and update. In J. Doyle, E. Sandewall, and P. Torasso (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*, pp. 190–201. San Francisco, Calif.: Morgan Kaufmann.

Friedman, N. and J. Y. Halpern (1997b). Modeling belief in dynamic systems. part I: foundations. *Artificial Intelligence*. Accepted for publication. 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 (1997a). Modeling belief in dynamic systems. Part II: revision and update. Submitted for publication. A preliminary version appears in J. Doyle, E. Sandewall, and P. Torasso, editors. *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.".

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, Calif.: Morgan Kaufmann.

Goldszmidt, M. (1992). *Qualitative probabilities: a normative framework for commonsense reasoning*. Ph. D. thesis, University of California Los Angeles.

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.), *Principles of Knowledge Representation and*

Reasoning: Proc. Third International Conference (KR '92), pp. 661–672. San Francisco, Calif.: Morgan Kaufmann.

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

Halpern, J. Y. (1993). A critical reexamination of default logic, autoepistemic logic, and only knowing. In *Proceedings, 3rd Kurt Gödel Colloquium*, pp. 43–60. Springer-Verlag.

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".

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.), *Principles of Knowledge Representation and Reasoning: Proc. Second International Conference (KR '91)*, pp. 387–394. San Francisco, Calif.: Morgan Kaufmann.

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

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

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

Rott, H. (1989). Conditionals and theory change: revision, expansions, and additions. *Synthese 81*, 91–113.

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

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.

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, Calif.: Morgan Kaufmann.

Wobcke, W. (1995). Belief revision, conditional logic, and nonmonotonic reasoning. *Notre Dame Journal of Formal Logic 36*(1), 55–102.