Revising Beliefs on the Basis of Evidence

James P. Delgrande
School of Computing Science
Simon Fraser University
Burnaby, B.C.,
Canada V5A 1S6
jim@cs.sfu.ca

Abstract

Approaches to belief revision most commonly deal with categorical information: an agent has a set of beliefs and the goal is to consistently incorporate a new item of information given by a formula. However, most information about the real world is not categorical. In revision, one may circumvent this fact by assuming that, in some fashion or other, an agent has elected to accept a formula ϕ, and the task of revision is to consistently incorporate ϕ into its belief corpus. Nonetheless, it is worth asking whether probabilistic information and noncategorical beliefs may be reconciled with, or even inform, approaches to revision. In this paper, one such account is presented. An agent receives uncertain information as input, and its probabilities on (a finite set of) possible worlds are updated via Bayesian conditioning. A set of formulas among the noncategorical beliefs is identified as the agent's categorical belief set. The effect of this updating on the belief set is examined with respect to its appropriateness as a revision operator. We show that few of the classical AGM belief revision postulates are satisfied by this approach. Most significantly, though not surprisingly, the success postulate is not guaranteed to hold. However it does hold after a sufficient number of iterations. As well, it proves to be the case that in revising by a formula consistent with the agent's beliefs, revision does not correspond to expansion. Postulates for iterated revision also examined, and it proves to be the case that most such postulates also do not hold. On the other hand, limiting cases of the presented approach correspond to specific approaches to revision that have appeared in the literature.

1 Introduction

In all but the simplest of circumstances, an agent's knowledge of a domain will be incomplete and inaccurate. Consequently, an agent will need to change its beliefs in response to receiving new information. *Belief revision* is the area of knowledge representation that is concerned with how an agent may incorporate new information about a domain into its set of beliefs. It is assumed that the agent has some corpus of beliefs K which are accepted as being true, or holding in the domain of application. A new formula ϕ is given, which the agent is to incorporate into its set of beliefs. Since consistency is to be maintained wherever possible, if ϕ conflicts with K some beliefs will have to be dropped from K before ϕ can be added. It is generally accepted that there is no single "best" revision operator, and different agents may have different revision functions. However, revision functions are not arbitrary, but are usually regarded as being guided, or characterised, by various *rationality criteria*. The original and best-known approach to belief revision is called the *AGM approach* [Alchourrón *et al.*, 1985; Gärdenfors, 1988], named after the developers of this framework.

There are of course other approaches for revising an agent's beliefs, most notably and obviously via probability theory and using Bayesian conditioning (e.g. [Pearl, 1988; Gärdenfors, 1988]). In this case, formulas are held with attached (subjective) probabilities. New evidence is received, also with an attached probability, and the agent's corpus of beliefs is modified via updating the associated probabilities.

Superficially, these approaches may be seen as broadly addressing the same problem since they both address the change in an agent's beliefs in the presence of new information. Yet it can also be argued that the two approaches can be seen as addressing quite different problems. In the case of belief revision, an agent *accepts* a certain set of beliefs as categorically holding, and another categorical belief is to be consistently incorporated into this set. This approach is fundamentally *qualitative*, since sentences of the language can be partitioned into those that are accepted and those that are not.[1] In the case of updating via Bayesian conditioning, beliefs are generally not held with certainty, but rather with varying levels or degrees of confidence. The task then is to modify these degrees of confidence, expressed as probabilities, as new evidence is received. Hence this approach is fundamentally *quantitative*.

The issue of whether (or when) beliefs are best treated as certain or uncertain is an important one, but one that we don't get into here. Certainly, compelling arguments can be made for either view: On the one hand, most people if pressed would agree that their beliefs are not fully certain, but rather are held with varying degrees of confidence. Thus, most people would be confident that their car would be left where they parked it (if they drove to work, for example) but would admit that there was a small chance that it wasn't there, perhaps having been stolen or towed. In fact, if pressed, it seems plausible that many, if not most, people would agree that the probability that the sun will rise tomorrow morning is not 1, but rather is something less – maybe extremely close to 1 but nonetheless less by some finite amount. On the other hand, most people *act* as though their beliefs are true, and generally describe or regard their beliefs in day-to-day affairs as simply being true. Hence if asked where their car was, most people would reply that it was where they had parked it, not that it was likely there or very probably there.

These considerations reflect a broader division in Artificial Intelligence and perhaps in science as a whole, between what has been called the *probabilists* on the one hand and *logicists* on the other. Henry Kyburg put it as follows:

> There are two fundamentally distinct ways of thinking about thinking about the world. One is contemplative; the other oriented toward action. One seeks pure knowledge; the other is pragmatic. One leads to hedged claims; the other leads

[1] There is more to the story, as discussed later, but the main point is that the focus rests on an agent's *belief set*.

to categorical claims in a hedged way. [Kyburg, 1994, p. 3]

As to the last point, a categorical, albeit hedged, belief is one that may be altered or withdrawn on the basis of new evidence; i.e. it may be the subject of belief revision. However, since categorical beliefs arise from noncategorical, hedged claims (and so uncertain evidence), it is an interesting question to ask whether the latter approach may inform the former. That is, an interesting question is whether an underlying non-categorical approach, here based on subjective probabilities, may have something to say about the categorical approach of belief revision. Another way of phrasing this is to first observe that evidence about the real world is generally uncertain (as well as incomplete and inaccurate), and so it is of interest to examine how such a setting may be reconciled with the assumptions underlying the area of belief revision.

To address this question, we begin with a simple model of an agent's beliefs. Probabilities are associated with possible worlds, characterising the agent's subjective knowledge. To ease the development, and because nothing of interest is lost with respect to the goals of the paper, the set of possible worlds is assumed to be finite. The agent's accepted, categorical beliefs are characterised by the least set of possible worlds, such that the set contains those worlds of highest probability such that the sum of the probabilities over those worlds exceeds a given threshold. As new, uncertain information is received, the probabilities attached to worlds are modified and the set of accepted beliefs correspondingly changes. One can then examine the properties of belief change with respect to the accepted beliefs from the point of view of classical AGM belief revision. It proves to be the case that, not surprisingly, only a subset of the AGM revision postulates are satisfied. For example, in a revision of a belief set K by ϕ, in the AGM approach ϕ is believed. In the approach at hand, ϕ is not necessarily believed following revision of K by ϕ, but it is believed to be possible, in that $\neg\phi$ is not believed. As well, after some number of iterations of revision by ϕ, ϕ will come to be believed. Further, if a formula ϕ is consistent with a belief set K, revision of K by ϕ may not correspond to the addition of ϕ to K. We also examine the approach with respect to iterated revision, specifically the Darwiche and Pearl [1997] and Jin and Thielscher [2007] accounts of iterated revision. Here it proves to be the case that only one of the Darwiche/Pearl postulates hold, and the Jin and Thielscher independence postulate does not hold. On the other hand, two extant approaches to belief revision prove to be closely related to instances of the approach developed here.

The next section provides requisite background material: terminology is introduced, belief revision is reviewed, and related work is surveyed. Section 3 describes the updating of probabilities in terms of probabilities on possible worlds. The following section motivates and defines the notion of *epistemic state* as used in the paper. Section 5 describes belief revision in this framework, including properties of the resulting revision operator and a comparison to related work. Section 6 gives a brief summary.

2 Background

2.1 Formal Preliminaries

We assume a propositional language $\mathcal{L}$ over a finite set of atomic sentences, $\mathcal{P} = \{a, b, \ldots\}$, closed under the usual connectives $\neg$, $\wedge$, $\vee$, and $\supset$, and with the classical consequence relation $\vdash$. $\mathcal{C}n(A)$ is the set of logical consequences of a formula or set of formulas A; that is $\mathcal{C}n(A) = \{\phi \in \mathcal{L} \mid A \vdash \phi\}$. $\top$ stands for some arbitrary tautology and $\bot$ is defined to be $\neg\top$. Given two sets of formulas A and B, $A + B$ denotes the *expansion* of A by B, that is $A + B = \mathcal{C}n(A \cup B)$. Expansion of a set of formulas A by a formula ϕ is defined analogously. Two sentences ϕ and ψ are *logically equivalent*, written $\phi \equiv \psi$, iff $\phi \vdash \psi$ and $\psi \vdash \phi$. This also extends to sets of formulas.

A propositional *interpretation* (also referred to as a *possible world*) is a mapping from $\mathcal{P}$ to {true, false}. The set of all interpretations of $\mathcal{L}$ is denoted by $W_{\mathcal{L}}$. A *model* of a sentence ϕ is an interpretation w that makes ϕ true according to the usual definition of truth, and is denoted by $w \models \phi$. For $W \subseteq W_{\mathcal{L}}$, we also write $W \models \phi$ if $w \models \phi$ for every $w \in W$. For a set of sentences A, $Mod(A)$ is the set of all models of A. For simplicity, $Mod(\{\phi\})$ is also written as $Mod(\phi)$. Conversely, given a set of possible worlds $W \subseteq W_{\mathcal{L}}$, we denote by $\mathcal{T}(W)$ the set of sentences which are true in all elements of W; that is $\mathcal{T}(W) = \{\phi \in \mathcal{L} \mid w \models \phi \text{ for every } w \in W\}$.

A total preorder $\preceq$ is a reflexive, transitive binary relation, such that either $w_1 \preceq w_2$ or $w_2 \preceq w_1$ for every w_1, w_2. As well, $w_1 \prec w_2$ iff $w_1 \preceq w_2$ and $w_2 \not\preceq w_1$. $w_1 = w_2$ abbreviates $w_1 \preceq w_2$ and $w_2 \preceq w_1$.[2] Given a set S and total preorder $\preceq$ defined on members of S, we denote by $\min(S, \preceq)$ the set of minimal elements of S in $\preceq$.

Last, let $P : W_{\mathcal{L}} \mapsto [0, 1]$ be a function such that $0 \leq P(w) \leq 1$ and $\sum_{w \in W_{\mathcal{L}}} P(w) = 1$. P is a *probability assignment* to worlds. We distinguish the function $P_\top$ where $P_\top(w) = \frac{1}{|W_{\mathcal{L}}|}$ for every world w. We also include the *absurd assignment* $P_\bot$ among the set of probability assignments, where $P_\bot(w) = 1$ for every world w. $P_\top$ can be used to characterise a state of ignorance for an agent, while $P_\bot$ is a technical convenience that will be used to characterise an inconsistent belief state. Thus, when we come to define an agent's categorical beliefs based on a probability assignment (Definition 4), an agent with associated probability assignment $P_\top$ will believe only the tautologies, whereas one with $P_\bot$ will believe all formulas.

These functions are extended to subsets of $W_{\mathcal{L}}$ by, for $W \subseteq W_{\mathcal{L}}$, $P(W) = \sum_{w \in W} P(w)$. Informally, $P(w)$ is the (subjective) probability that, as far as the agent knows, w is the actual world being modelled; and for $W \subseteq W_{\mathcal{L}}$, $P(W)$ is the probability that the real world is a member of W. As will be later described, the function P can be taken as comprising the major part of an agent's *epistemic state* [Darwiche and Pearl, 1997; Peppas, 2007]. The probability of a formula ϕ then is given by:

$$P(\phi) \;=\; \sum_{w \models \phi} P(w) \;=\; P(Mod(\phi)).$$

[2] As will be subsequently described, relations in a total preorder will be subscripted by an epistemic state. In particular the last relation will be written $=_{\mathcal{K}}$ and so there is no confusion with equality, written as usual as $=$, unsubscripted.

This then overloads the function $P(\cdot)$; however, this overloading is benign in that there is no ambiguity in the use of this function. *Conditional probability* is defined, as usual, by

$$P(\phi|\psi) \;=\; \frac{P(\phi \wedge \psi)}{P(\psi)}$$

and is undefined when $P(\psi) = 0$.

2.2 Belief revision

A common approach in addressing belief revision is to provide a set of *rationality postulates* for belief change functions. The *AGM approach* of Alchourrón, Gärdenfors, and Makinson [Alchourrón *et al.*, 1985; Gärdenfors, 1988] provides the best-known set of such postulates.[3] The goal is to describe belief change at the *knowledge level*, that is on an abstract level, independent of how beliefs are represented and manipulated. An agent's beliefs are modelled by a set of sentences, called a *belief set*, closed under the logical consequence operator of a logic that includes classical propositional logic. Thus a belief set K satisfies the constraint:

$$\phi \in K \text{ if and only if } K \text{ logically entails } \phi.$$

K can be seen as a partial theory of the world. $K_\perp$ is the inconsistent belief set (i.e. $K_\perp = \mathcal{L}$).

In *belief revision*, a formula ϕ is to be incorporated into the agent's set of beliefs K so that the resulting belief set is consistent, provided that ϕ is consistent. Since ϕ may be inconsistent with K, revision may also necessitate the removal of beliefs from K in order to retain consistency. In the AGM approach, revision is modeled as a function from belief sets and formulas to belief sets. However, various researchers have argued that it is more appropriate to consider *epistemic states* (also called *belief states*) as objects of revision. An epistemic state $\mathcal{K}$ effectively includes sufficient information to determine how the revision function itself changes following a revision; see [Darwiche and Pearl, 1997] or [Peppas, 2007] for discussions on this topic. The belief set corresponding to belief state $\mathcal{K}$ is denoted $Bel(\mathcal{K})$. As well, we will use the notation $Mod(\mathcal{K})$ to mean $Mod(Bel(\mathcal{K}))$. Formally, a revision operator $*$ maps an epistemic state $\mathcal{K}$ and new information ϕ to a revised epistemic state $\mathcal{K} * \phi$. Then, in the spirit of [Darwiche and Pearl, 1997], the AGM postulates for revision can be reformulated as follows:

(K*1) $Bel(\mathcal{K} * \phi) = \mathcal{C}n(Bel(\mathcal{K} * \phi))$

(K*2) $\phi \in Bel(\mathcal{K} * \phi)$

(K*3) $Bel(\mathcal{K} * \phi) \subseteq Bel(\mathcal{K}) + \phi$

(K*4) If $\neg\phi \notin Bel(\mathcal{K})$ then $Bel(\mathcal{K}) + \phi \subseteq Bel(\mathcal{K} * \phi)$

(K*5) $Bel(\mathcal{K} * \phi)$ is inconsistent, only if $\nvdash \neg\phi$

(K*6) If $\phi \equiv \psi$ then $Bel(\mathcal{K} * \phi) \equiv Bel(\mathcal{K} * \psi)$

[3] See also [Peppas, 2007] for a comprehensive survey of belief change.

(K*7) $Bel(\mathcal{K} * (\phi \wedge \psi)) \subseteq Bel(\mathcal{K} * \phi) + \psi$

(K*8) If $\neg\psi \notin Bel(\mathcal{K} * \phi)$ then $Bel(\mathcal{K} * \phi) + \psi \subseteq Bel(\mathcal{K} * (\phi \wedge \psi))$

That is, the result of revising $\mathcal{K}$ by ϕ is an epistemic state in which ϕ is believed in the corresponding belief set ((K*1), (K*2)); whenever the result is consistent, the revised belief set consists of the expansion of $Bel(\mathcal{K})$ by ϕ ((K*3), (K*4)); the only time that $Bel(\mathcal{K})$ is inconsistent is when ϕ is inconsistent ((K*5)); and revision is independent of the syntactic form of the formula for revision ((K*6)). The last two postulates deal with the relation between revising by a conjunction and expansion: revision by a conjunction corresponds to revision by one conjunct followed by expansion by the other, whenever the final result thus obtained is consistent.[4] Motivation for these postulates can be found in [Gärdenfors, 1988]. A dual operator, called contraction, is similarly defined, so that for a contraction of ϕ from K, denoted $K \dot{-} \phi$, the result is a belief set in which ϕ is not believed. See [Gärdenfors, 1988] for the set of contraction postulates and its relation with revision.

Several constructions have been proposed to characterise belief revision. Katsuno and Mendelzon [1991] (but see also [Grove, 1988]) have shown that a universal scheme for constructing an AGM revision operator can be given where any epistemic state $\mathcal{K}$ can induce a total preorder on the set of possible worlds. Formally, for epistemic state $\mathcal{K}$, a *faithful ranking* on $\mathcal{K}$ is a total preorder $\preceq_{\mathcal{K}}$ on the possible worlds $W_{\mathcal{L}}$, such that for any possible worlds $w_1, w_2 \in W_{\mathcal{L}}$:

1. If $w_1, w_2 \models Bel(\mathcal{K})$ then $w_1 =_{\mathcal{K}} w_2$
2. If $w_1 \models Bel(\mathcal{K})$ and $w_2 \not\models Bel(\mathcal{K})$, then $w_1 \prec_{\mathcal{K}} w_2$

Intuitively, $w_1 \preceq_{\mathcal{K}} w_2$ if w_1 is at least as plausible as w_2 according to the agent. The first condition asserts that all models of the agent's knowledge are ranked equally, while the second states that the models of the agent's knowledge are lowest in the ranking.

It follows directly from the results of [Katsuno and Mendelzon, 1991] that a revision operator $*$ satisfies $(\mathcal{K}*1)$–$(\mathcal{K}*8)$ iff there exists a faithful ranking $\preceq_{\mathcal{K}}$ for an arbitrary belief state $\mathcal{K}$, such that for any sentence ϕ:

$$Bel(\mathcal{K} * \phi) = \begin{cases} \mathcal{L} & \text{if } \vdash \neg\phi \\ \mathcal{T}(\min(Mod(\phi), \preceq_{\mathcal{K}})) & \text{otherwise} \end{cases}$$

Thus in the case where ϕ is satisfiable, the belief set corresponding to $\mathcal{K} * \phi$ is characterised by the least ϕ models in the ranking $\preceq_{\mathcal{K}}$.

The AGM postulates do not address properties of iterated belief revision. This has led to the development of additional postulates for iterated revision; the best-known approach is that of Darwiche and Pearl [1997]. They propose the following postulates, adapted according to our notation:

(C1) If $\psi \vdash \phi$, then $Bel((\mathcal{K} * \phi) * \psi) = Bel(\mathcal{K} * \psi)$

(C2) If $\psi \vdash \neg\phi$, then $Bel((\mathcal{K} * \phi) * \psi) = Bel(\mathcal{K} * \psi)$

[4] I thank a referee for suggesting this wording.

(C3) If $\phi \in Bel(\mathcal{K} * \psi)$, then $\phi \in Bel((\mathcal{K} * \phi) * \psi)$

(C4) If $\neg\phi \notin Bel(\mathcal{K} * \psi)$, then $\neg\phi \notin Bel((\mathcal{K} * \phi) * \psi)$

Darwiche and Pearl show that an AGM revision operator satisfies each of the postulates (C1)–(C4) iff the way it revises faithful rankings satisfies the respective conditions:

(CR1) If $w_1, w_2 \models \phi$, then $w_1 \preceq_{\mathcal{K}} w_2$ iff $w_1 \preceq_{\mathcal{K}*\phi} w_2$

(CR2) If $w_1, w_2 \not\models \phi$, then $w_1 \preceq_{\mathcal{K}} w_2$ iff $w_1 \preceq_{\mathcal{K}*\phi} w_2$

(CR3) If $w_1 \models \phi$ and $w_2 \not\models \phi$, then $w_1 \prec_{\mathcal{K}} w_2$ implies $w_1 \prec_{\mathcal{K}*\phi} w_2$

(CR4) If $w_1 \models \phi$ and $w_2 \not\models \phi$, then $w_1 \preceq_{\mathcal{K}} w_2$ implies $w_1 \preceq_{\mathcal{K}*\phi} w_2$

Thus postulate (C1) asserts that revising by a formula and then by a logically stronger formula yields the same belief set as simply revising by the stronger formula at the outset. The corresponding semantic condition (CR1) asserts that in revising by a formula ϕ, the relative ranking of ϕ worlds remains unchanged. The other postulates and semantic conditions can be interpreted similarly; see [Darwiche and Pearl, 1997] for more on motivation and interpretation of these conditions.

Subsequently, other approaches for iterated revision have been proposed, including [Boutilier, 1996; Nayak *et al.*, 2003; Jin and Thielscher, 2007]. For example, Jin and Thielscher [2007] have proposed the so-called Postulate of Independence:

(Ind) If $\neg\phi \notin Bel(\mathcal{K}*\psi)$ then $\phi \in Bel((\mathcal{K}*\phi)*\psi)$

Postulate (Ind) strengthens both (C3) and (C4). Thus, Jin and Thielscher's suggested set of postulates consists of (C1), (C2), and (Ind). They also give a necessary and sufficient condition for an AGM revision operator to satisfy (Ind):

(IndR) If $w_1 \models \phi$ and $w_2 \models \neg\phi$, then $w_1 \preceq_{\mathcal{K}} w_2$ implies $w_1 \prec_{\mathcal{K}*\phi} w_2$.

Again, this seems plausible: In the case not covered by (CR3) and (CR4), if two worlds are equally ranked, and ϕ is true at one world but not the other, then a revision by ϕ is sufficient to distinguish the relative rankings of the worlds.

2.3 Related Work

In probability theory and related approaches, there has of course been work on incorporating new evidence to produce a new probability distribution. The simplest means of updating probabilities is via conditionalisation: If an agent holds ϕ with probability q, and so $P(\phi) = q$, and the agent learns ψ with certainty, then one can define the updated probability $P'(\phi)$ via

$$P'(\phi) \;=\; P(\phi|\psi) \;=\; P(\phi \wedge \psi)/P(\psi).$$

[Gärdenfors, 1988] in fact discusses conditionalisation as a form of *expansion* for probability functions.

Of course an agent may not learn ψ with certainty, but rather may change its probability assignment to ψ from $P(\psi)$ to a new value $P'(\psi)$. The question then is how probabilities assigned to other formulas should be modified. Jeffrey [1983] proposes that for proposition ϕ, the new probability should be given by what has come to be known as Jeffrey's Rule for updating probabilities:

$$P'(\phi) \;=\; P(\phi|\psi)P'(\psi) + P(\phi|\neg\psi)(1 - P'(\psi)).$$

So $P'(\psi) = q$ means that the agent has learned that the probability of ψ is q. In particular, if the probability of ψ is further updated to $P''(\psi)$ but it turns out that $P''(\psi) = P'(\psi)$, then the distributions P' and P'' will coincide.

In contrast, we are interested in the case where we have some underlying proposition, say that a light is on, represented by *on*, and we are given an observation Obs_{on}, where Obs_{on} has an attached probability. Then if the agent receives repeated observations from a sensor that the light is on, the agent's confidence that *on* is true will increase with each positive observation. Details are given in the next section; the main point here is that Bayes' Rule will be more appropriate in this case, where Bayes' Rule is given by:

$$P(\phi|\psi) \;=\; \frac{P(\psi|\phi)P(\phi)}{P(\psi)}.$$

In other related work, [Bovens and Hartmann, 2003] have a somewhat similar goal to our's, in that they use a Bayesian Network [Pearl, 1988] to model multiple reports of a given fact, where the reports come with varying degrees of reliability. Essentially a probabilistic graphical model is used to model a fixed number of information sources and their influence on a "summary" vertex. Analogously, there has been work on updating evidence in other approaches to uncertain reasoning. For example in the Dempster-Shafer approach, a rule for evidence combination is given; subsequent work includes [Fagin and Halpern, 1990; Kulasekere *et al.*, 2004].

Previous research dealing with both AGM-style belief revision and probability is generally concerned with revising a probability function. In such approaches, an agent's belief set K is given by those formulas that have probability 1.0. These formulas with probability 1.0 are referred to as the *top* of the probability function. For a revision $K * \phi$, the probability function is revised by ϕ, and the belief set corresponding to $K * \phi$ is given by the top of the resulting probability function. So such approaches allow the characterisation of not just the agent's beliefs, but also allow probabilities to be attached to non-beliefs. These approaches will be seen to differ from that developed here, in that in the current approach, an agent may accept a formula as true, even though the probability of that formula is less than 1.0.

One difficulty with revising probability functions is the *non-uniqueness problem*, that there are many different probability functions that have K as their top. Lindström and Rabinowicz [1989] consider various ways of dealing with this problem. Boutilier [1995] considers the same general framework, but rather focuses on issues of iterated belief revision. However the approach described herein addresses a different problem: a means of incorporating uncertain information into a given probability function is assumed, and the question addressed is how such an approach may be reconciled with AGM revision, or alternatively, how such an approach may be considered as an instance (or proto-instance) of AGM-style revision. To this end, Gärdenfors [1988, Ch. 5] has also considered an extension of the AGM approach to the revision of probability functions; we discuss this work after our approach has been presented.

In a somewhat different vein, results in belief change have been applied to issues in dealing with probability. [Cozic, 2011] uses intuitions from belief change, namely the difference between revision and update, to inform a proposed solution to a problem in probability and belief dynamics called the *Sleeping Beauty Problem*. [Makinson, 2011] uses results in belief revision to consider the situation in conditionalisation where a formula ϕ is consistent but has zero probability, and yet arguably conditioning on ϕ is meaningful.

With respect to qualitative, AGM-style belief revision, the approach at hand might seem to be an instance of an *improvement operator* [Konieczny and Pino Pérez, 2008]. An improvement operator according to Konieczny and Pino Pérez is a belief change operator where new information isn't necessarily immediately accepted. However plausibility is increased and, after a sufficient number of iterations, the information will come to be believed. Interestingly, as we discuss later, the approach described here differs in significant ways from those of [Konieczny and Pino Pérez, 2008].

The overall setting adopted here is similar to that of [Bacchus *et al.*, 1999]: Agents receive uncertain information, and alter their (probabilistic) beliefs about the world based on this information. However, the goals are quite different. [Bacchus *et al.*, 1999] is concerned with an extension of the situation calculus [Levesque *et al.*, 1998] to deal with noisy sensors. Consequently their focus is on a version of the situation calculus in which the agent doesn't hold just categorical beliefs, but also probabilistic beliefs. The main issue then is how to update these probabilities in the presence of sensing and non-sensing actions. In contrast, the present paper is concerned with the possible role of probabilistic beliefs with respect to a (classical AGM-style) belief revision operator. We further discuss this and other related work once the approach has been presented.

3 Unreliable Observations and Updating Probabilities

Consider the situation in which we have an agent in some domain, and where this agent may make observations concerning this domain. These observations are assumed to be independent. However, the sensors associated with the agent may be unreliable, in that an aspect of the domain may be incorrectly sensed or reported. Thus, an agent may sense whether a light is on or not, and there is some probability that the sensor reports that the light is on when in fact it is not, or that it is off when in reality it is on. This degree of (un)reliability of a sensor is known to the agent. It is assumed also that the domain is static, in that it does not evolve or change over time. The agent has an associated probability assignment to formulas, expressed in terms of a probability assignment to possible worlds, and the task is to update this probability assignment given such possibly-erroneous observations.

To put this in more concrete terms, consider a situation in which an agent observes or senses ϕ with a given probability $q > .5$. That is, ϕ is either true or false in the domain.[5]

[5] The case of non-binary valued sensing is straightforward and adds nothing of additional interest with

We can let Obs_ϕ stand for the fact that the agent's sensor reports ϕ as being true; $\neg Obs_\phi$ then means that the agent's sensor does not report ϕ as being true, and thus reports it as being false. So we can write $P(Obs_\phi|\phi) = q$ to indicate that the reliability of the sensor is q in the case that ϕ is true. Because it simplifies the development and because it has no effect on the results obtained, we make the simplifying assumption that a sensor for ϕ is equally reliable whether or not ϕ is true. Hence, we also have that $P(\neg Obs_\phi|\neg\phi) = q$. We return to this point briefly, once the approach has been presented.

So, under the assumption that a sensor is equally reliable whether or not the sensed formula is true, we have the following. From $P(Obs_\phi|\phi) = q$ we obtain that $P(\neg Obs_\phi|\phi) = 1 - q$, and similarly from $P(\neg Obs_\phi|\neg\phi) = q$ we obtain $P(Obs_\phi|\neg\phi) = 1 - q$. One would expect then that following an observation that ϕ was true, the agent's confidence in ϕ would increase; following a subsequent observation that ϕ was true (say, Obs'_ϕ), one would expect that the agent's confidence would increase further.

However, the probability of a proposition is determined via a probability assignment to possible worlds. So the task at hand is to update probabilities attached to possible worlds, given an observation report such as Obs_ϕ.[6] Consider the situation in which we sense ϕ with probability q, and so $P(Obs_\phi|\phi) = q$. As before, Obs_ϕ is true if ϕ is observed to be true, and false otherwise. For a world $w \in W_\mathcal{L}$, we also have the prior probability assignment $P(w)$.[7] For each $w \in W_\mathcal{L}$ we wish to calculate $P(w|Obs_\phi)$, that is the conditional probability of w given the observation Obs_ϕ.

From Bayes' Rule we have:

$$P(w|Obs_\phi) \;=\; \frac{P(Obs_\phi|w)P(w)}{P(Obs_\phi)} \qquad (1)$$

On the left side of the equality, we wish to determine the (updated) probability of w, given that ϕ is observed. For the numerator on the right hand side, $P(Obs_\phi|w)$ is the probability of observing ϕ given that the agent is in w. If $w \models \phi$, then from our notion of sensor reliability, we have that $P(Obs_\phi|w) = q$. If $w \not\models \phi$, then from our notion of sensor reliability, we have that $P(\neg Obs_\phi|w) = q$, from which we obtain that $P(Obs_\phi|w) = 1 - q$ in this case. That is,

$$\begin{aligned} P(Obs_\phi|w) &= q && \text{if } w \models \phi \\ P(Obs_\phi|w) &= 1 - q && \text{if } w \not\models \phi. \end{aligned}$$

The other part of the numerator in (1), $P(w)$, is simply the (known) probability assignment

respect to the problem at hand; see for example [Bacchus *et al.*, 1999] for how this can be handled.

[6] Recall that we assume a finite language and a finite set of possible worlds. The concluding section discusses this assumption further.

[7] Since a world may be associated with the conjunction of literals true at that world, we do not need to introduce "worlds" as separate entities in what follows.

$Worlds$	P	$P(a,.8)$	$P(a,.8)(a,.8)$	$P(a,.8)(b,.8)$
a, b, c	.150	.240	.2824	.3333
$a, b, \overline{c}$	.150	.240	.2824	.3333
$a, \overline{b}, c$	.100	.160	.1882	.0556
$a, \overline{b}, \overline{c}$	.100	.160	.1882	.0556
$\overline{a}, b, c$	.150	.060	.0176	.0972
$\overline{a}, b, \overline{c}$	.150	.060	.0176	.0972
$\overline{a}, \overline{b}, c$	.100	.040	.0118	.0139
$\overline{a}, \overline{b}, \overline{c}$	.100	.040	.0118	.0139

Table 1: Example of Updating Probabilities of Worlds

to w. For the denominator in (1), we have that

$$\begin{aligned}
P(Obs_\phi) &= \sum_{w \in W_\mathcal{L}} P(Obs_\phi|w)P(w) \\
&= \sum_{w \models \phi} P(Obs_\phi|w)P(w) \ + \ \sum_{w \not\models \phi} P(Obs_\phi|w)P(w) \\
&= \sum_{w \models \phi} (q \times P(w)) \ + \ \sum_{w \not\models \phi} ((1-q) \times P(w)) \\
&= q \times P(Mod(\phi)) \ + \ (1-q) \times P(Mod(\neg\phi)). \\
&= q \times P(\phi) \ + \ (1-q) \times P(\neg\phi).
\end{aligned}$$

This justifies the following definition.

Definition 1 *Let P be a probability assignment to worlds. Let $\phi \in \mathcal{L}$ and $q \in [0,1]$. Let $\eta \ = \ P(\phi) \times q \ + \ P(\neg\phi) \times (1-q)$.*

Define the probability assignment $P(\phi, q)$ by:

$$\begin{aligned}
P(\phi, q) &= P_\perp \qquad\qquad\qquad\qquad \textit{if } \eta = 0; \quad \textit{otherwise:} \\
P(\phi, q)(w) &= \begin{cases} (P(w) \times q)/\eta & \textit{if } w \models \phi \\ (P(w) \times (1-q))/\eta & \textit{if } w \not\models \phi \end{cases}
\end{aligned}$$

Thus, for a given probability function P, the new probability function $P(\phi, q)$ results after sensing ϕ with probability q. It can be observed that $P(\phi, q)(w)$ in Definition 1 corresponds to $P(w|Obs_\phi)$ in (1). In the case where $\eta = 0$, the updated probability assignment involves accepting with certainty (i.e. $q = 1$) an impossible proposition ($P(\phi) = 0$), or rejecting with certainty a necessarily true proposition. In either case, the incoherent state of affairs ($P_\perp$) results. Lastly, Definition 1 is clearly compatible with a sequence of observations. For example the observation that ϕ was true with probability q would be given by $P(\phi, q)$, and the subsequent observation that ψ was true with probability q' would be given by $P(\phi, q)(\psi, q')$.

Example: Table 1 provides an example. Let $\mathcal{P} = \{a, b, c\}$. The first column lists possible worlds in terms of an assignment of truth values to atoms, where for readability $\overline{a}$ is used for

$\neg a$. The second column gives an initial probability function, while the next three columns show how P changes under different updates. It can be observed that at the outset $P(a) = .5$, $P(b) = .6$, and $P(c) = .5$. Following an observation of a with reliability .8, we obtain that $P(a) = .8$, $P(b) = .6$, and $P(c) = .5$. If we iterate the process and again observe a with the same reliability, the respective probabilities become $P(a) = .9412$, $P(b) = .6$, and $P(c) = .5$. Essentially the probability of a increases, and the probability of b and c varies depending on the probabilities assigned to individual worlds. In the example they happen to be unchanged, but usually they will not. Last if we first sense a and then b, in both cases with reliability .8 (last column) then we obtain that $P(a) = .7778$, $P(b) = .8610$, and $P(c) = .5$. In the next section we return to this example to see how the agent's set of accepted beliefs changes.

4 Epistemic States

This section presents the notion of *epistemic state* as it is used in the approach. Recall that an epistemic state $\mathcal{K}$ implicitly includes information regarding how the revision function itself changes following a revision. We first discuss intuitions then give the formal details.

4.1 Intuitions

An agent's epistemic state $\mathcal{K}$ is given by a pair (P, c), where P is a probability assignment over possible worlds and c is a *confidence level*. The intent is that the probability function characterises the agent's overall belief state, and that an agent *accepts* a belief represented by a formula just if, in a sense to be described, the probability of the formula exceeds the confidence level c.[8] Thus, the agent accepts a formula if its probability is "sufficiently high". $\mathcal{K}_\perp$ is used to denote the incoherent epistemic state $(P_\perp, c)$.

The notion of *acceptance* adopted here is nonstandard, in that an accepted belief is categorical but its associated probability may be less than 1. This is in contrast to the previous approaches combining probability and revision described in Section 2.3, where the agent's categorical beliefs have probability 1. This also is in contrast with [Bacchus *et al.*, 1999], where non-beliefs have probability 0. In any case, for us an accepted belief is one that is categorical, in that the agent may act under the assumption that it is true, yet it is also "hedged" in that it can be given up following a revision. The issue then becomes one of suitably defining the worlds characterising an agent's accepted beliefs, to which we turn next.

The most straightforward way of defining acceptance is to say that a formula ϕ is accepted just if $P(\phi) \geq c$, for some "suitably high" choice of c. This leads immediately to the *lottery paradox* [Kyburg, 1961]. This problem is usually expressed as follows: There is a lottery, for which a large number of tickets have been sold, and for which there is some (unknown) winner. For any given ticket, it is highly plausible (and so an accepted belief) that the ticket will lose. But this leads to the conclusion that for any ticket, that ticket will lose,

[8] Note that the confidence level c is distinct from the reliability of a sensor described in the previous section, and given by q. The former is part of an agent's epistemic state and is used as a threshold for determining accepted beliefs, while the latter gives the reliability of a sensor.

while there is guaranteed to be a winning ticket. These statements together are inconsistent. The difficulty is that for any $c < 1.0$ one can construct a scenario where $p_1, \ldots, p_n$, along with $\neg p_1 \vee \ldots \vee \neg p_n$ are all accepted. But the set consisting of these formulas is of course inconsistent.

Instead, we define an agent's (categorical, accepted) beliefs in terms of a unique, well-defined set of possible worlds. The formal definition is given in the next subsection; informally, this set is equal to the least set of worlds of greatest probability such that the probability of the set is greater than or equal to c.[9] We have the added proviso that, for worlds w and w', if $P(w) = P(w')$, then w is in this set iff w' is. Since the agent's accepted beliefs are characterised by a unique set of worlds, the lottery paradox doesn't arise.

The assumption that worlds with higher probability are to be preferred to those with lower probability for characterising an agent's beliefs can be justified by two arguments.

1. If an agent had to commit to a single world being the real world, then it would choose a world w for which the probability $P(w)$ was maximum; if it had to commit to n worlds, then correspondingly it would choose the n worlds with highest probability. Similarly, if one were aiming to choose some set of worlds most likely to contain the real world then, for reasons of parsimony, it seems most reasonable to choose worlds of higher probability. Since there is nothing that distinguishes worlds beyond their probability, if $P(w) = P(w')$ then if w is in this set then so is w'.

2. The second argument is related to the principle of *informational economy*: It seems reasonable to assume that, given a set of candidate belief sets, an agent will prefer a set that gives more information over one that gives less.

   This is the case here. In general there will be more than one set of worlds where the probability of the set exceeds c. The set composed of the least number of worlds of maximal probability is generally also the set with the least number of worlds, which in turn will correspond to the belief set with the maximum number of (logically distinct) formulas. So this approach commits the agent to the maximum set of accepted beliefs, where the overall probability of the set exceeds c.[10] Such a set may be said to have the greatest *epistemic content* among the candidate belief sets.

Thus an epistemic state principally consists of a probability function on possible worlds. Via an assumption of maximality of beliefs (or maximal epistemic content), and given the confidence level c, a set of accepted beliefs is defined. So this differs significantly from prior work on revising probability functions, in that in the present approach an accepted formula will generally have an associated probability that is less than 1.0. Arguably this makes sense: for example, I believe that my car is where I left it this morning, in that I act as if this was a true fact even though I don't hold that it is an absolute certainty that the car is where I left it. Moreover, of course, I am prepared to revise this belief if I receive information to the contrary.

[9] A reviewer of the paper suggested an appealing equivalent notion, whereby an agent discards or dismisses from consideration the most unlikely worlds, such that the probability of the remaining set of worlds is not less than the threshold c.

[10] These notions of course make sense only in a finite (under equivalence classes) language, which was assumed at the outset.

4.2 Formal Details

In this subsection we define our notion of epistemic state, and relate it to the faithful rankings that have been used to characterise AGM revision.

Definition 2 $\mathcal{K} = (P, c)$ *is an* epistemic state*, where:*

- P *is probability assignment to possible worlds and*
- $c \in (0, 1]$ *is a* confidence level*.*

As described, an epistemic state characterises the state of knowledge of the agent, both its (contingent) beliefs as well as, implicitly, those beliefs that it would adopt or abandon in the presence of new information. In order to relate this approach to the AGM theory of revision, we need to specify the agent's *belief set* or beliefs about the world at hand. This is most easily done by first defining the worlds that characterise the agent's belief set, and then defining the belief set in terms of these worlds.

Definition 3 *For epistemic state* $\mathcal{K} = (P, c)$*, the set of worlds characterising the agent's belief set,* $Mod(\mathcal{K}) \subseteq W_{\mathcal{L}}$*, is the least set such that:*
If $P = P_{\perp}$ *then* $Mod(\mathcal{K}) = \emptyset$*; otherwise:*

1. $P(Mod(\mathcal{K})) \geq c$*,*
2. *If* $w \in Mod(\mathcal{K})$ *and* $w' \notin Mod(\mathcal{K})$ *then* $P(w) > P(w')$*.*

$Mod(\cdot)$ is uniquely characterised; in particular we have that if $P(w) = P(w')$ then $w \in Mod(\mathcal{K})$ iff $w' \in Mod(\mathcal{K})$.

Definition 4 *For epistemic state* $\mathcal{K}$*, the agent's accepted (categorical) beliefs,* $Bel(\mathcal{K})$*, are given by*

$$Bel(\mathcal{K}) \;=\; \{\phi \mid Mod(\mathcal{K}) \models \phi\} \;=\; \mathcal{T}(Mod(\mathcal{K})).$$

Consequently, an agent accepts a sentence if it is sufficiently likely, and a sentence is considered "sufficiently likely" if it is true in the least set of most plausible worlds such that the probability of the set exceeds the given confidence level. Clearly, $Bel(\mathcal{K})$ is closed under conjunction, and it is also closed under classical consequence.

This then describes the static aspects of an epistemic state. For the dynamic aspects (i.e. revision) it will be useful to distinguish those formulas that are *possible*, in the sense that they are *conceivable*, which is to say they have a non-zero probability. We use $Poss_{\mathcal{K}}(\phi)$ to indicate that, according to the agent, ϕ is possible; that is, there is possible world w such that $w \models \phi$ and $P(w) > 0$. $Poss_{\mathcal{K}}(\cdot)$ has the properties of the uniterated modality $\Diamond$ in the modal logic S5 [Hughes and Cresswell, 1996]. We have the simple consequence:

Proposition 1 *If not* $Poss_{\mathcal{K}}(\phi)$ *then* $\neg\phi \in Bel(\mathcal{K})$

The probability assignment to possible worlds can be used to define a ranking on worlds, where worlds with higher probability are lower in the ranking:

Definition 5 *For given probability assignment P, define $rank_P(w)$ for every $w \in W_{\mathcal{L}}$ by:*

1. *$rank_P(w) = 0$ if $\nexists w'$ such that $P(w') > P(w)$.*
2. *Otherwise, $rank_P(w) = 1 + \max\{rank_P(w') : P(w') > P(w)\}$.*

The set of worlds of rank i is defined by:

$$R_P(i) = \{w \in W_{\mathcal{L}} \mid rank_P(w) = i\}.$$

R_P defines a total preorder over worlds, where lower-ranked worlds are more plausible than higher-ranked worlds. Thus, if c_r is the least number such that

$$\sum_{i=0}^{c_r} P(R_P(i)) \ \geq \ c,$$

then the set of worlds characterising the agent's accepted beliefs is alternately given by

$$Mod(\mathcal{K}) = \begin{cases} \bigcup_{i=0}^{c_r} R_P(i) & \text{if } P \neq P_\perp \\ \emptyset & \text{otherwise} \end{cases}$$

This is slightly different usage from other work in iterated belief change, where an epistemic state is often *equated* with a (qualitative or quantitative) ranking on possible worlds. Rather, here a ranking on possible worlds is *induced* by an epistemic state.

Lastly, we can define a *faithful ranking* (as given in Section 2) to relate the ranking defined here to rankings used in belief revision:

Definition 6 *The* faithful ranking *$\preceq_{\mathcal{K}}$ based on $\mathcal{K}$ is given by:*

1. *If $w_1, w_2 \models Bel(\mathcal{K})$ then $w_1 =_{\mathcal{K}} w_2$*
2. *If $w_1 \models Bel(\mathcal{K})$ and $w_2 \not\models Bel(\mathcal{K})$, then $w_1 \prec_{\mathcal{K}} w_2$*
3. *Otherwise if $rank_P(w_1) \leq rank_P(w_2)$ then $w_1 \preceq_{\mathcal{K}} w_2$.*

Thus, from an epistemic state as given in Definition 2, a corresponding faithful ranking on worlds can be defined in a straightforward manner. The first two conditions stipulate that we have a faithful ranking. The third condition ensures that we have a total preorder that conforms to the probability assignment for those worlds not in $Mod(\mathcal{K})$. It is clear that this faithful ranking suppresses detail found in P. First, quantitative information is lost in going from Definition 2 to Definition 5. Second, gradations in an agent's beliefs are lost: worlds in $Mod(\mathcal{K})$ may have varying probabilities, yet in the corresponding faithful ranking given in Definition 6, all worlds in $Mod(\mathcal{K})$ are ranked equally. Consequently, the notion of epistemic state as defined here is a richer structure than that of a faithful ordering.

5 Belief Revision in a Probabilistic Framework

We next consider how the approach of the previous sections fits with work in belief revision. A natural way to define the revision of an epistemic state $\mathcal{K} = (P, c)$ by observation ϕ with associated reliability q is to set $\mathcal{K} * (\phi, q) \; = \; (P(\phi, q), c)$. Of course, revision so defined is a ternary function,[11] as opposed to the usual expression of revision as a binary function, $\mathcal{K} * \phi$. There are several ways in which this mismatch may be resolved. First, we could simply regard revision in a probabilistic framework as a ternary function, with the extra argument giving the reliability of the observation. This is problematic, at least with regards to our present aims, since a ternary operator $\mathcal{K} * (\phi, q)$ represents a *quantitative* approach, where the degree of support q of ϕ is somehow taken into account. However, standard AGM revision is *qualitative*, in that for a revision $\mathcal{K} * \phi$, it is the (unqualified) formula ϕ that is a subject of revision. This clash then represents the main issue of this paper: a probabilistic approach is intrinsically quantitative, while standard approaches to belief revision are inherently qualitative. So, in one fashion or another we want to address revision in qualitative terms.

In re-considering revision $*$ as a binary function, the intent is that in expressing $\mathcal{K} * (\phi, q)$ as a binary function $\mathcal{K} * \phi$, we want to study *properties* of the function $*$ without regard to specific values assigned to q. Consequently, we assume that the reliability of a revision is some fixed probability q. Since revision corresponds to the incorporation of *contingent* information, it is reasonable to assume that nothing can be learned with absolute certainty, and so we further assume that $q < 1$.[12] As well, revision by ϕ is intended to *increase* the agent's confidence in ϕ, and so for $\mathcal{K} * \phi$ it is understood that the probability of ϕ is greater than 0.5; this is reflected in the fact that sensors are assumed to be reliable (see Section 3), in that they are correct over 50% of the time. Consequently, in what follows, we assume that the reliability of a revision is a fixed number q in the range $(0.5, 1.0)$. Given that the reliability is fixed, we can drop the probability argument from a statement of revision, and simply write $\mathcal{K} * \phi$. However, for completeness we also later consider the situation where the reliability of observations may vary.

Definition 7 *Let $q \in (0.5, 1.0)$ be fixed. Let $\mathcal{K} = (P, c)$ be an epistemic state and $\phi \in \mathcal{L}$. Define the revision of $\mathcal{K}$ by ϕ by:*

$$\mathcal{K} * \phi \; = \; (P(\phi, q), c)$$

Clearly one needs to know the value of q (along with $\mathcal{K}$ and ϕ) before being able to determine the specific value of $\mathcal{K} * \phi$. However, without knowing the value of q, one can still investigate properties of the class of revision functions, which is our goal here. Other aspects of the definition are discussed below, in the discussion of postulates. However, we first revisit our example of the previous section.

[11] In fact, taking into account the internal structure of an epistemic state, it is a 4-place function, on P, c, ϕ, and q.

[12] It might be pointed out that a tautology can be learned with absolute certainty. However, it can be pointed out in return that a tautology is in fact *known* with certainty, so the probability being 1 or less makes no difference. In any case, we later examine the situation where $q = 1$.

Example (continued): Consider again Table 1, and assume that our initial epistemic state is given by $\mathcal{K} = (P, 0.9)$. At the outset, we have $Bel(\mathcal{K}) = \mathcal{C}n(\top)$. If the probability associated with the world characterised by $\{\overline{a}, \overline{b}, \overline{c}\}$ was .05, with the balance distributed uniformly across other possible worlds, we would have $Bel(\mathcal{K}) = \mathcal{C}n(a \vee b \vee c)$.

We obtain that $Bel(\mathcal{K} * a) = \mathcal{C}n(a \vee b)$, and $Bel(\mathcal{K} * a * a) = \mathcal{C}n(a)$. On the other hand, we obtain $Bel(\mathcal{K} * a * b) = \mathcal{C}n(a \vee b)$ and (not illustrated in Table 1) $Bel(\mathcal{K} * a * b * b) = \mathcal{C}n(b)$. So, not unexpectedly, for repeated iterations, the resulting belief set "converges" toward accepting the iterated formula, with the results being biased by the initial probability distribution.

5.1 Properties of Probability-Based Belief Revision

Recall that $Poss_{\mathcal{K}}(\phi)$ indicates that, according to the agent, ϕ is possible, in that there is a world w such that $w \models \phi$ and $P(w) > 0$. The revision of $\mathcal{K}$ by ϕ, $\mathcal{K} * \phi$, is as given in Definition 7. $\mathcal{K} *^n \phi$ stands for the n-fold iteration of $\mathcal{K} * \phi$, that is:

$$\mathcal{K} *^n \phi \;=\; \begin{cases} \mathcal{K} * \phi & \text{if } n = 1 \\ (\mathcal{K} *^{n-1} \phi) * \phi & \text{otherwise} \end{cases}$$

We obtain the following results corresponding to the AGM revision postulates. For ease of comparison, negative results (i.e. statements of postulates that do not hold) are included below.

Theorem 1 *Let $\mathcal{K}$ be an epistemic state and ϕ, $\psi \in \mathcal{L}$.*

($\mathcal{K} * 1$) $Bel(\mathcal{K} * \phi) = \mathcal{C}n(Bel(\mathcal{K} * \phi))$

($\mathcal{K} * 2a$) *If $Poss_{\mathcal{K}}(\phi)$ then $\phi \in Bel(\mathcal{K} *^n \phi)$ for some $n > 0$*

($\mathcal{K} * 2b$) *If $\mathcal{K} \neq \mathcal{K}_\perp$ and $\phi \in Bel(\mathcal{K})$ then $\phi \in Bel(\mathcal{K} * \phi)$*

($\mathcal{K} * 2c$) *If $\mathcal{K} \neq \mathcal{K}_\perp$ and not $Poss_{\mathcal{K}}(\phi)$ then $Bel(\mathcal{K} * \phi) = Bel(\mathcal{K})$*

($\mathcal{K} * 3$) *The postulate:*

$$Bel(\mathcal{K} * \phi) \subseteq Bel(\mathcal{K}) + \phi$$

does not necessarily hold.

($\mathcal{K} * 4$) *The postulate:*

$$\text{If } \neg\phi \notin Bel(\mathcal{K}) \text{ then } Bel(\mathcal{K}) + \phi \subseteq Bel(\mathcal{K} * \phi)$$

does not necessarily hold.

($\mathcal{K} * 5$) *$Bel(\mathcal{K} * \phi)$ is consistent.*

($\mathcal{K} * 6$) *If $\phi \equiv \psi$ then $Bel(\mathcal{K} * \phi) \equiv Bel(\mathcal{K} * \psi)$*

($\mathcal{K} * 7$) *The postulate:*

$$Bel(\mathcal{K} * (\phi \wedge \psi)) \subseteq Bel(\mathcal{K} * \phi) + \psi$$

does not necessarily hold.

($\mathcal{K} * 8$) *The postulate:*

$$\text{If } \neg\psi \notin Bel(\mathcal{K} * \phi) \text{ then } Bel(\mathcal{K} * \phi) + \psi \subseteq Bel(\mathcal{K} * (\phi \wedge \psi))$$

does not necessarily hold.

Proof: ($\mathcal{K} * 1$) follows directly from Definition 4.

For ($\mathcal{K} * 2a$), it follows from Definition 1 that if $0 < P(\phi) \leq 1$ then $P(\phi) < P_{P(\phi,q)}(\phi) \leq 1.0$, and so that as we iterate a revision by ϕ, the probability of ϕ monotonically increases, with upper bound 1.0. It follows that for some n, $Mod(\mathcal{K} *^n \phi) \subseteq Mod(\phi)$, from which we obtain that for some $n > 0$, $\phi \in Bel(\mathcal{K} *^n \phi)$. ($\mathcal{K} * 2b$) and ($\mathcal{K} * 2c$) have the prerequisite condition that $\mathcal{K}$ is not the incoherent epistemic state. ($\mathcal{K} * 2b$) is obvious, and indicates that once ϕ is accepted it remains accepted under further revisions by ϕ, under the proviso that $Poss_{\mathcal{K}}(\phi)$ holds. For ($\mathcal{K} * 2c$), if $\mathcal{K} \neq \mathcal{K}_\perp$ then if there are no ϕ-worlds with non-zero probability, then Definition 1 can be seen to leave the probability function unchanged.

For ($\mathcal{K} * 3$) consider where $\mathcal{P} = \{a, b\}$, and $\mathcal{K} = (P, 0.97)$ and where:

$$\begin{array}{ll} P(\{a,b\}) = .96 & P(\{a,\neg b\}) = .02, \\ P(\{\neg a,b\}) = .01 & P(\{\neg a,\neg b\}) = .01 \end{array}$$

Hence $Bel(\mathcal{K})$ is characterised by $\{a, b\}$, $\{a, \neg b\}$, i.e. $Bel(\mathcal{K}) = \mathcal{C}n(a)$, and so $Bel(\mathcal{K}) + a = \mathcal{C}n(a)$.

If we revise by a with confidence .8, we get

$$\begin{array}{ll} P(a,.9)(\{a,b\}) \approx .9746 & P(a,.9)(\{a,\neg b\}) \approx .0203 \\ P(a,.9)(\{\neg a,b\}) \approx .0025 & P(a,.9)(\{\neg a,\neg b\}) \approx .0025 \end{array}$$

Thus $Bel(\mathcal{K})$ is characterised by $\{a, b\}$, i.e. $Bel(\mathcal{K}) = \mathcal{C}n(a \wedge b) \neq \mathcal{C}n(a) = Bel(\mathcal{K}) + a$.

For ($\mathcal{K} * 4$) it is possible to have formulas ϕ and ψ such that ϕ and ψ are logically independent, $Bel(\mathcal{K}) = \mathcal{C}n(\phi)$ and $Bel(\mathcal{K} * \psi) = \mathcal{C}n(\psi)$, thus contradicting the postulate. To see this, consider where $\mathcal{P} = \{a, b\}$, and $\mathcal{K} = (P, 0.9)$ and where:

$$\begin{array}{ll} P(\{a,b\}) = .46 & P(\{a,\neg b\}) = .46, \\ P(\{\neg a,b\}) = .06 & P(\{\neg a,\neg b\}) = .02 \end{array}$$

Hence $Bel(\mathcal{K})$ is characterised by $\{a, b\}$, $\{a, \neg b\}$, i.e. $Bel(\mathcal{K}) = \mathcal{C}n(a)$, and so $Bel(\mathcal{K}) + b = \mathcal{C}n(a \wedge b)$.

If we revise by b with confidence .9, we get

$$\begin{array}{ll} P(b,.9)(\{a,b\}) \approx .802 & P(b,.9)(\{a,\neg b\}) \approx .089 \\ P(b,.9)(\{\neg a,b\}) \approx .105 & P(b,.9)(\{\neg a,\neg b\}) \approx .004 \end{array}$$

Since $P(b,.9)(\{a,b\}) + P(b,.9)(\{\neg a,b\}) > .9 = c$, we get that $Bel(\mathcal{K} * b)$ is characterised by $\{a,b\}$, $\{\neg a,b\}$ and so $Bel(\mathcal{K}) = \mathcal{C}n(b)$.

For $(\mathcal{K} * 5)$ it can be seen from the definitions that if $\mathcal{K} \neq \mathcal{K}_\perp$, then there will be worlds with a non-zero probability, and so $Mod(\mathcal{K}) \neq \emptyset$ in Definition 3, and so $Bel(\mathcal{K})$ is well defined in Definition 4 and specifically $Bel(\mathcal{K}) \neq \mathcal{L}$. In particular, in the case of a revision by an inconsistent formula ϕ, we get that $\mathcal{K} * \phi = \mathcal{K}$: All ϕ worlds (of which there are none) share in the probability q, and all $\neg\phi$ worlds share in the probability $1-q$, where we have that $1-q > 0$. The result is normalised, leaving the probabilities unchanged. If $\mathcal{K} = \mathcal{K}_\perp$, then in Definition 1 we get that $\eta \neq 0$, $0 < q < 1$, and so the probability assignment $P(\phi, q) \neq P_\perp$, and so in Definition 3 we obtain that $Mod(\mathcal{K}) \neq \emptyset$.

Postulate $(\mathcal{K} * 6)$ holds trivially, but by virtue of the fact that the reliability of an observation of ϕ is the same as that of ψ.

$(\mathcal{K} * 7)$ doesn't hold for the same reason $(\mathcal{K} * 3)$ doesn't. Substituting $\top$ for ϕ in $(\mathcal{K} * 7)$ in fact yields $(\mathcal{K} * 3)$.

$(\mathcal{K} * 8)$ doesn't hold for the same reason that $(\mathcal{K} * 4)$ doesn't. Substituting $\top$ for ϕ in $(\mathcal{K} * 8)$ yields $(\mathcal{K} * 4)$. ■

Discussion The weaker version of postulate (K*2), given by $(\mathcal{K} * 2a)$, means that an agent will accept that ϕ is true after a sufficient number of iterations (or "reports") of ϕ. Hence, despite the absence of other AGM postulates, the operator $*$ counts as a revision operator, since the formula ϕ will eventually be accepted, provided that it is possible. Note that if a formula ϕ is not possible, then from our earlier (non-revision) result

If not $Poss_{\mathcal{K}}(\phi)$ then $\neg\phi \in Bel(\mathcal{K})$

together with $(\mathcal{K} * 5)$, we have that ϕ can never be accepted. As well, if ϕ is accepted, it will continue to be accepted following revisions by ϕ $(\mathcal{K} * 2b)$. This last point would seem to be obvious, but is necessitated by the absence of a postulate of success and the absence of $(\mathcal{K} * 4)$. In the case that a formula ϕ is deemed to be not possible, but the agent is not in the incoherent state $\mathcal{K}_\perp$, $(\mathcal{K} * 2c)$ shows that revising by ϕ leaves the agent's belief set unchanged. Contrary to this, as mentioned earlier, [Makinson, 2011] discusses the case where the probability of a contingent formula may be zero, but where it may nonetheless be meaningful to condition on that formula; in fact he uses results in AGM revision to inform just such conditionalisation.

Neither $(\mathcal{K} * 3)$ nor $(\mathcal{K} * 4)$ are guaranteed to hold, and so there is no relation between expansion and revision, even in the case where the formula for revision is consistent with the belief set. The counterexample to $(\mathcal{K} * 3)$ illustrates a curious aspect of the approach: In the counterexample we have that $Bel(\mathcal{K}) = \mathcal{C}n(a)$ yet $Bel(\mathcal{K} * a) = \mathcal{C}n(a \wedge b)$. In revising by a, the probability of worlds given by $\{a,b\}$, $\{a,\neg b\}$ both increase, but that of $\{a,b\}$ increases to such an extent that its probability exceeds the confidence level c, and so it alone characterises the agent's set of accepted beliefs. We discuss this behaviour later, once we have finished presenting the approach as a whole.

The counterexample to $(\mathcal{K} * 4)$ illustrates another interesting point: not only does the postulate fail but, unlike the success postulate $(\mathcal{K} * 2)$, it may fail over any number of iterations. For the example provided, the probability of the world given by $\{a,b\}$ will converge to something just over .88, which is below the given confidence level of $c = 0.9$. Since $\mathcal{C}n(a, b)$ is the result of expansion in the example, this shows that $\mathcal{C}n(a, b)$ will never come to be accepted in this case. Similar remarks hold for $(\mathcal{K} * 8)$.

It can be noted that $K_\perp$ plays no interesting role in revisions; this is reflected by $(\mathcal{K} * 5)$, which asserts that no revision can yield $\mathcal{K}_\perp$.[13] Hence an epistemic state can be inconsistent only if the original assignment of probabilities to worlds is the absurd probability assignment $P_\perp$. Any subsequent revision will have $Bel(\mathcal{K}_\perp * \phi) \neq \mathcal{L}$. In particular if ϕ is $\perp$ then $P_\perp(\phi, q) = P_\top$ and so $Bel(\mathcal{K}_\perp * \perp) = \mathcal{C}n(\top)$.

Postulate $(\mathcal{K} * 5)$ appears to be remarkably strong, in that it imposes no conditions on the original epistemic state $\mathcal{K}$ or the formula for revision. If one begins with the inconsistent epistemic state $\mathcal{K}_\perp$, then revision is defined as being the same as a revision of the epistemic state of complete ignorance $P_\top$. This is pragmatically useful: from $\mathcal{K}_\perp$, if one revises by a formula ϕ where $Pr(\phi) \neq 0$, then analogous to the AGM approach, one arrives at a consistent belief state. This also goes beyond the AGM postulate $(K * 5)$, since if ϕ is held to be impossible (i.e. there are no worlds with nonzero probability in which ϕ is true), then there will be worlds in which $\neg\phi$ is true and with nonzero probability, and so revision yields meaningful results, in particular yielding the epistemic state with probability assignment $P_\top$.

It can be noted that if one allows revision with probability $q = 1$, then Postulate $(\mathcal{K} * 5)$ may be violated: In the approach, if a contingent formula ϕ is such that $P(\phi) = 0$, then the revision $\mathcal{K} * \phi$, where $q = 1$, can be seen to yield the incoherent epistemic state $\mathcal{K}_\perp$.

Postulate $(\mathcal{K} * 6)$ holds trivially, but by virtue of the fact that the reliability of an observation of ϕ is the same as that of ψ. This assumption is, of course, limiting, and in the case where observations may be made with differing degrees of reliability, the postulate would not hold. It can be noted that in the case where observations may be made with differing degrees of reliability, the postulate can be replaced by the weaker version:

If $\phi \equiv \psi$ then $Bel(\mathcal{K} * \phi) \subseteq Bel(\mathcal{K} * \psi)$ or $Bel(\mathcal{K} * \psi) \subseteq Bel(\mathcal{K} * \phi)$.

A final point is that in Definition 1 it was assumed that the reliability of a sensor for ϕ was the same whether ϕ was true or not. It is trivially verified that the postulates that hold (viz. $(\mathcal{K} * 1)$, $(\mathcal{K} * 2a)$, $(\mathcal{K} * 2b)$, $(\mathcal{K} * 2c)$, $(\mathcal{K} * 5)$, and $(\mathcal{K} * 6)$) will also hold if a sensor has differing reliability depending on whether ϕ is true or false.

5.2 Variants of the Approach

We next examine three variants of the approach. In the first, observations are made with certainty. This variation can be seen to coincide with an extant approach in belief revision. As well it can be seen to have close relations to Gärdenfors' revision of probability functions [Gärdenfors, 1988]; a discussion of the relation with this latter work is deferred to the next section. In the second variant, observations are made with near certainty; again this variant corresponds with an extant approach in belief revision. In the last variant, informally,

[13] We might of course declare by fiat that $Bel(\mathcal{K} * \phi) = \mathcal{K}_\perp$ when $\vdash \neg\phi$ if we were interested in trying to satisfy various AGM postulates. However, since our goal is to see how a probabilistic account fits with AGM, $(\mathcal{K} * 5)$ is the appropriate expression of the postulate.

possible worlds that characterise an agent's beliefs are retained after a revision if there is no reason to eliminate them.

Certain Observations Consider where observations are certain, and so the (binary) revision $\mathcal{K}*\phi$ corresponds to $\mathcal{K}*(\phi, 1.0)$. Clearly, we have that if $P(w) = 0$, then $P(\phi, 1.0)(w) = 0$ for any ϕ; that is, if a world had probability 0, then no observation is going to alter this probability. As well, if $w \models \neg\phi$ then $P(\phi, 1.0)(w) = 0$. So in a revision by ϕ with certainty, any $\neg\phi$ world will receive probability 0, and by the previous observation, this probability of 0 will remain fixed after subsequent revisions.

Thus in this case, revision is analogous to a form of *expansion*, but with respect to the epistemic state $\mathcal{K}$. So following a revision by ϕ, all $\neg\phi$ worlds are discarded from the derived faithful ranking. This corresponds to revision in the approach of [Shapiro *et al.*, 2000], where their account of revision is embedded in an account of reasoning about action. For their revision, a plausibility ordering over worlds is given at each world.[14] Observations are assumed to be correct in this approach, and observation of ϕ means that $\neg\phi$ is impossible in the current world, and so all $\neg\phi$ worlds are removed from the ordering. This also means that an observation of ϕ followed by $\neg\phi$ yields the inconsistent epistemic state. This result may be justified by the argument that, if ϕ is observed with certainty, then if the world does not change, then it is *impossible* for $\neg\phi$ to be observed. [Shapiro *et al.*, 2000] show that in their approach, postulates (K*1)-(K*4), and (K*6) are satisfied.

Near-Certain Observations Consider next where we define the binary revision $\mathcal{K} * \phi$ as $\mathcal{K} * (\phi, 1.0 - \epsilon)$, where ϵ is "sufficiently small" compared to the probabilities assigned by P. If the minimum and maximum values in the range of P are min_P and max_P, then "sufficiently small" would mean that $max_P \times \epsilon < min_P \times (1.0 - \epsilon)$. Thus, in revision by ϕ, it is guaranteed that ϕ will be an accepted belief. Moreover, the reliability of the observation is high enough so that no $\neg\phi$ world will have probability greater than any ϕ world. Thus for $P' = P(\phi, 1.0 - \epsilon)$ we would have that for $w \models \phi$, $w' \not\models \phi$, that $P'(w) > P'(w')$. This yields the approach of lexicographic revision [Nayak, 1994] in which, in revising by ϕ, every ϕ world is ranked below every $\neg\phi$ world in the faithful ranking, but the relative ranking of ϕ worlds (resp. $\neg\phi$ worlds) is retained. In this approach, all AGM revision postulates hold.

Retaining Confirmed Possible Worlds The approach as given is clearly belief revision, since under reasonable conditions a formula will become accepted. However, it has notable weaknesses compared to the AGM approach; in particular the postulates $(\mathcal{K} * 3)$, $(\mathcal{K} * 4)$, $(\mathcal{K}*7)$, and $(\mathcal{K}*8)$ all fail. In the case of $(\mathcal{K}*4)$ and $(\mathcal{K}*8)$ this seems unavoidable. However, an examination of $(\mathcal{K}*3)$ and $(\mathcal{K}*7)$ shows a curious phenomenon underlying the approach. Consider $(\mathcal{K} * 3)$: In the counterexample presented, the agent believed that the real world was among the set of worlds $\{ \{a, b\}, \{a, \neg b\} \}$. On revising by a, the agent believed that the real world was among the set of worlds $\{ \{a, b\} \}$, which is to say, that $\{a, b\}$ was the real world. But what this means is that $\{a, \neg b\}$ was considered to be one of the worlds that characterised the agent's accepted beliefs, but on receiving confirmatory evidence (viz.

[14] In fact, at each *situation* [Levesque *et al.*, 1998]. However, for us the difference is immaterial.

a revision by a), this world was dropped from the characterising set. But arguably if w may be the actual world according to the agent, and the agent learns ϕ where $w \models \phi$, then it seems that the agent should still consider w as possibly being the actual world.

The reason for this phenomenon is clear: The probability of some worlds (in the example, given by $\{a, b\}$) becomes large enough following revision so that the other worlds (viz. $\{a, \neg b\}$), are no longer required in making up $Mod(\mathcal{K})$. So, to counteract this phenomenon, it seems reasonable to adopt a principle that if a world is considered "possible" by an agent, then it remains "possible" after confirmatory evidence. To this end, the approach can be modified as follows: One now needs to keep track of worlds considered possible by the agent, where these are the worlds characterising the agent's contingent beliefs. So an epistemic state now would be a triple (P, c, B) where $B \subseteq W_{\mathcal{L}}$ characterises the agent's belief set. Thus after revising by ϕ, B would be given by:

$$Mod(P(\phi, q)) \;\cup\; (Mod(\mathcal{K}) \cap Mod(\phi)).$$

It is easily shown that in this case postulates $(\mathcal{K} * 3)$ and $(\mathcal{K} * 7)$ in addition hold.

5.3 Iterated Belief Revision

Turning next to iterated revision, it proves to be the case that three of the Darwiche-Pearl postulates for iterated revision fail to hold, as does the the Jin-Thielscher Independence Postulate. However, as we will see, the reason that these postulates do not hold is not a result of the probabilistic approach per se, but rather is a result of the expression of a faithful ranking (and so, implicitly, a belief set) in terms of probabilities on possible worlds.

Theorem 2 *Let $\mathcal{K}$ be an epistemic state with associated revision operator $*$. Then $\mathcal{K}$ satisfies (C3). $\mathcal{K}$ does not necessarily satisfy (C1), (C3), (C4), and (Ind).*

Proof: For the postulates that do not hold, we consider (C1) in detail; other postulates fail for analogous reasons. For a counterexample to (C1), let $\mathcal{P} = \{a, b\}$, and $\mathcal{K} = (P, 0.9)$, and where:

$$P(\{a, b\}) = .85 \qquad P(\{a, \neg b\}) = .06,$$
$$P(\{\neg a, b\}) = .05 \qquad P(\{\neg a, \neg b\}) = .04$$

$Bel(\mathcal{K})$ is characterised by $\{a, b\}$, $\{a, \neg b\}$ and so $Bel(\mathcal{K}) = \mathcal{C}n(a)$. If we subsequently revise by a with confidence .7, we get

$$P(a, .7)(\{a, b\}) \approx .891 \qquad P(a, .7)(\{a, \neg b\}) \approx .063$$
$$P(a, .7)(\{\neg a, b\}) \approx .023 \qquad P(a, .7)(\{\neg a, \neg b\}) \approx .018$$

Thus $Bel(\mathcal{K} * a) = Bel(\mathcal{K}) = \mathcal{C}n(a)$.

If we next revise again by a with confidence .7, we get

$$P(a, .7)(\{a, b\}) \approx .917 \qquad P(a, .7)(\{a, \neg b\}) \approx .065$$
$$P(a, .7)(\{\neg a, b\}) \approx .010 \qquad P(a, .7)(\{\neg a, \neg b\}) \approx .008$$

Since $P(a, .9)(\{a, b\}) > .9 = c$, we get that $Bel(\mathcal{K} * a * a)$ is characterised by $\{a, b\}$ alone. Hence we get $Bel(\mathcal{K} * a) \neq Bel(\mathcal{K} * a * a)$, thereby violating (C1).[15]

Similar constructions can be formulated to show that (C2), (C4) and (IndR) do not necessarily hold. We omit the details.

For (C3), we obtain that the semantic condition (CR3) holds: If $w_1 \models \phi$ and $w_2 \not\models \phi$, then $w_1 \prec_{\mathcal{K}} w_2$ implies that $P(w_1) < P(w_2)$ from which it follows that $P(\phi, q)(w_1) < P(\phi, q)(w_2)$ and so $w_1 \prec_{\mathcal{K}*\phi} w_2$. By the same argument as [Darwiche and Pearl, 1997, Theorem 13], we get that (C3) is satisfied. ∎

It is worth considering why most of the iteration postulates fail. Interestingly, for the semantic conditions, (CR1) – (CR4) and (IndR), if expressions of the form $w_1 \prec_{\mathcal{K}} w_2$ are replaced by expressions of the form $P(w_1) \leq P(w_2)$, then the modified conditions hold in the current approach. That is, it is easily verified that all of the following hold:

Theorem 3

(PCR1) *If* $w_1, w_2 \models \phi$, *then* $P(w_1) \leq P(w_2)$ *iff* $P(\phi, q)(w_1) \leq P(\phi, q)(w_2)$.

(PCR2) *If* $w_1, w_2 \not\models \phi$, *then* $P(w_1) \leq P(w_2)$ *iff* $P(\phi, q)(w_1) \leq P(\phi, q)(w_2)$.

(PCR3) *If* $w_1 \models \phi$ *and* $w_2 \not\models \phi$, *then* $P(w_1) < P(w_2)$ *implies* $P(\phi, q)(w_1) < P(\phi, q)(w_2)$.

(PCR4) *If* $w_1 \models \phi$ *and* $w_2 \not\models \phi$, *then* $P(w_1) \leq P(w_2)$ *implies* $P(\phi, q)(w_1) \leq P(\phi, q)(w_2)$.

(PIndR) *If* $w_1 \models \phi$ *and* $w_2 \not\models \phi$, *then* $P(w_1) \leq P(w_2)$ *implies* $P(\phi, q)(w_1) < P(\phi, q)(w_2)$.

Proof: Straightforward from Definition 1. ∎

The problem is that our faithful ranking (Definition 6) doesn't preserve the ordering given by P. In particular, if $w_1,\ w_2 \in Mod(\mathcal{K})$ then $w_1 =_{\mathcal{K}} w_2$ in the derived faithful ranking, while most often we will have $P(w_1) \neq P(w_2)$. Essentially, in moving from values assigned via P to the faithful ranking, gradations (given by probabilities) among worlds in $Mod(\mathcal{K})$ are lost. That is, in a sense, the probabilistic approach provides a finer-grained account of an epistemic state than is given by a faithful ranking on worlds, in that models of the agent's belief set also come with gradations of belief.

5.4 Relation with Other Work

Other Approaches to Revision We have already discussed the relation of the present approach to those described in [Shapiro *et al.*, 2000] and [Nayak, 1994] in Section 5.2.

The work in belief change that is closest to that described here is [Konieczny and Pino Pérez, 2008], which discusses *improvement operators*. Recall that an improvement operator according to Konieczny and Pino Pérez is a belief change operator in which new information isn't necessarily immediately accepted, but where the plausibility is increased; and so after a sufficient number of iterations, the information will come to be believed. The general idea

[15] In terms of (CR1), we have $\{a, b\} =_{\mathcal{K}*a} \{a, \neg b\}$ but $\{a, b\} \prec_{\mathcal{K}*a*a} \{a, \neg b\}$, thereby violating (CR1).

of their approach then seems similar to results obtained here. As well, in both approaches the success postulate does not necessarily hold, so new information is not necessarily immediately accepted. However, beyond failure of the success postulate, the approaches have quite different characteristics.

In the postulate set following,[16] $\circ$ is an improvement operator, and $\times$ is defined by: $\mathcal{K} \times \phi = \mathcal{K} \circ^n \phi$ where n is the first integer such that $\phi \in Bel(\mathcal{K} \circ^n \phi)$.

(I1) There exists n such that $\phi \in Bel(\mathcal{K} \circ^n \phi)$

(I2) If $\neg\phi \notin Bel(\mathcal{K})$ then $Bel(\mathcal{K} \times \phi) \equiv Bel(\mathcal{K}) + \phi$

(I3) $Bel(\mathcal{K} \circ \phi)$ is inconsistent, only if $\nvdash \neg\phi$

(I4) If $\phi_i \equiv \psi_i$ for $1 \leq i \leq n$ then $Bel(\mathcal{K} \circ \phi_1 \circ \ldots \circ \phi_n) \equiv Bel(\mathcal{K} \circ \psi_1 \circ \ldots \circ \psi_n)$

(I5) $Bel(\mathcal{K} \times (\phi \wedge \psi)) \subseteq Bel(\mathcal{K} \times \phi) + \psi$

(I6) If $\neg\psi \notin Bel(\mathcal{K} \times \phi)$ then $Bel(\mathcal{K} \times \phi) + \psi \subseteq Bel(\mathcal{K} \times (\phi \wedge \psi))$

To show that the approaches are independent, it suffices to compare $(\mathcal{K} * 3)/(\mathcal{K} * 4)$ with (I2). According to (I2), after a sufficient number of iterations of an improvement operator, the resulting belief set will correspond to expansion of the original belief set by the formula in question. However, there are cases in which neither $(\mathcal{K} * 3)$ nor $(\mathcal{K} * 4)$ are satisfied in our approach regardless of the number of iterations. Similar comments apply to $(\mathcal{K} * 7)$ and $(\mathcal{K} * 8)$ on the one hand, and (I5) and (I6) on the other. The need for the extended postulate for irrelevance of syntax for epistemic states (I4) was noted in [Booth and Meyer, 2006]. In the present approach $(\mathcal{K} * 6)$ suffices.

Revision of Probability Functions [Gärdenfors, 1988][Ch. 5] addresses, among other belief change operators, the revision of probability functions. Let P denote some probability function and let $P_\perp$ be the *absurd* probability function in which $P_\perp(\phi) = 1$ for all formulas ϕ. Gärdenfors defines the *expansion* of P by a formula ϕ, $P + \phi$, to be the conditionalisation of P on ϕ. Then revision is a function from the class of probability functions and formulas to the class of probability functions. The following postulates are given:

(P*1) If P is a probability function and ϕ a sentence, then $P * \phi$ is a probability function.

(P*2) $(P * \phi)(\phi) = 1$

(P*3) If $\phi \equiv \psi$ then $P * \phi = P * \psi$

(P*4) $P * \phi \neq P_\perp$ iff not $\vdash \neg\phi$.

(P*5) If $P(\phi) > 0$ then $P * \phi = P + \phi$

[16][Konieczny and Pino Pérez, 2008] use the formulation of revision in [Katsuno and Mendelzon, 1991], in which the result of revision is a formula, not a belief set. We retain the numbering of [Konieczny and Pino Pérez, 2008], but rephrase their postulates in terms of belief sets. Most pertinently, postulates (K*3) and (K*4) correspond to (I2), while postulates (K*7) and (K*8) correspond to (I5) and (I6) respectively.

(P*6) If $(P * \phi)(\psi) > 0$ then $P * (\phi \wedge \psi) = (P * \phi) + \psi$

It is readily verified that if in Definition 1, $P * \phi$ is read as $P(\phi, q)$ and given fixed q, that only the first and third postulates above are satisfied. However, if we consider the variant of the approach discussed earlier in which $q = 1$ we obtain:

Theorem 4 *Let P be a probability assignment and $P(\phi, 1.0)$ a derived probability assignment. Then taking $P * \phi$ as $P(\phi, 1.0)$, we have that postulates (P*1)–(P*3), (P*5) and (P*6) are satisfied as well as:*

(P*4′) *$P * \phi \neq P_\perp$ iff $P(\phi) > 0$.*

Proof: The proof follows directly from the relevant definitions, and is omitted.■

Recall that when we discussed the variant of our approach where $q = 1$ in Section 5.2, the resulting operator was described as being like an *expansion* with regards to a ranking function. In our approach, when the reliability of an observation ϕ is certain, and so $q = 1$, then any $\neg\phi$ worlds are assigned probability 0 and, as discussed earlier, once a world is given a probability of zero, it remains at zero for subsequent revision. This is the reason that the operator is characterised as an expansion.

In contrast, Gärdenfors describes postulates (P*1)–(P*6) as characterising *revision*. The difference can be found in the difference between postulates (P*4) and (P*4′). The Gärdenfors postulates characterise a class of functions, but notably the postulates don't have anything to say when the probability of a formula ϕ is zero but ϕ is consistent. However, it is possible – and, in the interesting case, essential – to have a situation in which $P(\phi) = 0$ but one may meaningfully revise by ϕ. In particular, if ϕ is consistent then even if $P(\phi) = 0$, postulate (P*2) requires that $(P * \phi)(\phi) = 1$ and postulate (P*4) stipulates that the absurd probability assignment does not result. In contrast, with our approach, if ϕ is consistent but $P(\phi) = 0$ then the absurd probability assignment results.

Other Related Work As discussed in Section 2, earlier work dealing specifically with revision and probability has been concerned with revising probability functions. Thus, work such as [Gärdenfors, 1988; Lindström and Rabinowicz, 1989; Boutilier, 1995] deals with extensions to the AGM approach for the revision of probability functions. In such approaches there is an underlying probability function associated with formulas in which the agent's belief set is characterised by formulas with probability 1.0. For a revision $K * \phi$, ϕ represents new *evidence*, and the probability function is revised by ϕ. The belief set corresponding to $K * \phi$ then is the set of propositions with probability 1.0. In contrast, in the approach at hand, an agent's *accepted* beliefs are characterised by a set of possible worlds whose overall probability in the general case will be less than 1.0. In a sense then there is finer granularity with regards the present approach, since the worlds characterising a belief set may have varying probability. As well, for us if a formula ϕ has probability of 1.0, then it cannot be removed by subsequent revisions. A formula is accepted as true if its probability is sufficiently high; however it may potentially be revised away. This then arguably confirms to intuitions, in that if a formula is held with complete certainty then it *should* be immune from revisions.

It was noted that [Bacchus *et al.*, 1999] presents a similar setting in which an agent receives possibly-unreliable observations. However, the concern in [Bacchus *et al.*, 1999] was to update probabilities associated with worlds and then to use this for reasoning about dynamic domains expressed via the situation calculus. The approach at hand employs a similar method for updating probabilities but addresses the question of how this may be regarded as, or used to formulate, an approach to belief revision. Again, the present approach also has finer granularity, in that in [Bacchus *et al.*, 1999] non-beliefs are given by worlds with probability 0; in the approach at hand, non-beliefs are those that fall outside the set of accepted beliefs, and may have non-zero probability. Again, arguably the present approach conforms to intuitions, since if a formula is held to be impossible then it seems it should forever remain outside the realm of revision.

6 Conclusion

We have explored an approach to belief revision based on an underlying model of uncertain reasoning. With few exceptions, research in belief revision has dealt with categorical information in which an agent has a given set of beliefs and the goal is to consistently incorporate a formula into this set of beliefs. A common means of specifying revision semantically is via a ranking function on possible worlds wherein the agent's beliefs are modelled by the least worlds in the ranking. The revision by formula ϕ then is characterised by the least set of ϕ worlds in the ranking. Clearly however, most information about the real world is not categorical, and arguably no non-tautological belief may be held with complete certainty. To accommodate this, one alternative is to adopt a purely probabilistic framework for belief change. However, such a framework ignores the fact that an agent may well *accept* a formula as being true, even if this acceptance is tentative, or hedged in some fashion. So another alternative, and the one followed here, is to begin with a probabilistic framework, but also define a set of formulas that the agent accepts. Revision can then be defined in this framework, and the results of revision on the agent's accepted beliefs examined.

To this end we have assumed that an agent receives uncertain information as input, and the agent's probabilities on possible worlds are updated via Bayesian conditioning. A set of formulas among the (noncategorical) beliefs is identified as the agent's (categorical) belief set. These are determined via the set of most probable worlds, such that the summed probability of the set exceeds a given threshold. The effect of updating on this belief set is examined with respect to its appropriateness as a revision operator. We show that few of the classical AGM belief revision postulates are satisfied by this approach. Most significantly, though not surprisingly, the success postulate is not guaranteed to hold, though it is after a sufficient number of iterations. As well, it also proves to be the case that in revising by a formula consistent with the agent's beliefs, revision does not necessarily correspond to expansion. As another point of interest, of the postulates for iterated revision that we considered, only (C3) holds. The reason for this is that, even though the updating of the probability assignment P satisfies all of the corresponding semantic conditions, the induced faithful ordering $\prec_{\mathcal{K}}$ does not. Last, although the approach shares motivation and intuitions with improvement operators, our results show that the approach does not fall into the category of improvement operators.

An apparent limitation of the approach is that it deals with a finite language. Moreover it is not clear how the results presented here can be generalised to the infinite case. While acknowledging that this is indeed a formal limitation, nonetheless an agent operating in the real world is a finite entity, and has a finite knowledge base. As well, much of current work in belief change follows the Katsuno-Mendelzon [1991] approach wherein an agent's beliefs are finitely representable. Thus, arguably, the assumption of a finite language is not a significant limitation with respect to modelling revision for computational agents.

There are two ways that these results may be viewed with respect to classical AGM-style belief revision. On the one hand, assuming that the approach to dealing with uncertainty and the means of determining an agent's belief set are reasonable, it can be suggested that the current approach provides a revision operator that is substantially weaker than given in the AGM approach and approaches to iterated revision. On the other hand, the AGM approach and approaches to iterated revision have been justified by appeals to rationality, in that it is claimed that *any* rational agent should conform to the AGM postulates and, say, the Darwiche/Pearl iteration postulates. Thus, to the extent that the presented approach is rational, the present approach would appear to undermine the rationale of these approaches, at least in the case of uncertain information.

References

[Alchourrón *et al.*, 1985] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

[Bacchus *et al.*, 1999] F. Bacchus, J.Y. Halpern, and H.J. Levesque. Reasoning about noisy sensors and effectors in the situation calculus. *Artificial Intelligence*, 111(1-2):171–208, 1999.

[Booth and Meyer, 2006] Richard Booth and Thomas Andreas Meyer. Admissible and restrained revision. *Journal of Artificial Intelligence Research*, 26:127–151, 2006.

[Boutilier, 1995] C. Boutilier. On the revision of probabilistic belief states. *Notre Dame Journal of Formal Logic*, 36(1):158–183, 1995.

[Boutilier, 1996] C. Boutilier. Iterated revision and minimal revision of conditional beliefs. *Journal of Logic and Computation*, 25:262–305, 1996.

[Bovens and Hartmann, 2003] Luc Bovens and Stephan Hartmann. *Bayesian Epistemology.* Oxford: Clarendon Press, 2003.

[Cozic, 2011] Mikal Cozic. Imaging and sleeping beauty: A case for double-halfers. *International Journal of Approximate Reasoning*, 52:137–143, 2011.

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

[Fagin and Halpern, 1990] Ronald Fagin and Joseph Y. Halpern. A new approach to updating beliefs. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, UAI '90, pages 347–374. Elsevier Science Inc., 1990.

[Gärdenfors, 1988] P. Gärdenfors. *Knowledge in Flux: Modelling the Dynamics of Epistemic States*. The MIT Press, Cambridge, MA, 1988.

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

[Hughes and Cresswell, 1996] G.E. Hughes and M.J. Cresswell. *A New Introduction to Modal Logic*. Routledge., London and New York, 1996.

[Jeffrey, 1983] Richard Jeffrey. *The Logic of Decision*. University of Chicago Press, second edition, 1983.

[Jin and Thielscher, 2007] Y. Jin and M. Thielscher. Iterated belief revision, revised. *Artificial Intelligence*, 171(1):1–18, 2007.

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

[Konieczny and Pino Pérez, 2008] Sébastien Konieczny and Ramón Pino Pérez. Improvement operators. In G. Brewka and J. Lang, editors, *Proceedings of the Eleventh International Conference on the Principles of Knowledge Representation and Reasoning*, pages 177–186, Sydney, Australia, 2008. AAAI Press.

[Kulasekere *et al.*, 2004] E. C. Kulasekere, K. Premaratne, D. A. Dewasurendra, M. L. Shyu, and P. H. Bauer. Conditioning and updating evidence. *International Journal of Approximate Reasoning*, 36(1):75–108, 2004.

[Kyburg, 1961] H.E. Kyburg, Jr. *Probability and the Logic of Rational Belief*. Wesleyan University Press, 1961.

[Kyburg, 1994] H.E. Kyburg, Jr. Believing on the basis of evidence. *Computational Intelligence*, 10(1):3–20, 1994.

[Levesque *et al.*, 1998] H.J. Levesque, F. Pirri, and R. Reiter. Foundations for the situation calculus. *Linköping Electronic Articles in Computer and Information Science*, 3(18), 1998.

[Lindström and Rabinowicz, 1989] Sten Lindström and Wlodzimierz Rabinowicz. On probabilistic representation of non-probabilistic belief revision. *Journal of Philosophical Logic*, 11(1):69–101, 1989.

[Makinson, 2011] D. Makinson. Conditional probability in the light of qualitative belief change. *Journal of Philosophical Logic*, 40:121–153, 2011.

[Nayak *et al.*, 2003] Abhaya C. Nayak, Maurice Pagnucco, and Pavlos Peppas. Dynamic belief revision operators. *Artificial Intelligence*, 146(2):193–228, 2003.

[Nayak, 1994] A.C. Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Mateo, CA, 1988.

[Peppas, 2007] P. Peppas. Belief revision. In F. van Harmelen, V. Lifschitz, and B. Porter, editors, *Handbook of Knowledge Representation*, pages 317–359. Elsevier Science, San Diego, USA, 2007.

[Shapiro *et al.*, 2000] S. Shapiro, M. Pagnucco, Y. Lesperance, and H. J. Levesque. Iterated belief change in the situation calculus. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *Proceedings of the Seventh International Conference on the Principles of Knowledge Representation and Reasoning*, pages 527–538, San Francisco, 2000. Morgan Kaufmann.