A simple modal logic for belief revision: Extended Abstract

Giacomo Bonanno
Department of Economics,
University of California,
Davis, CA 95616-8578, USA
e-mail: gfbonanno@ucdavis.edu

August 23, 2005

The notions of static belief and of belief revision have been extensively studied in the literature. However, there is a surprising lack of uniformity in the two approaches. In the philosophy and logic literature the notion of static belief has been studied mainly within the context of modal logic. On the syntactic side a belief operator B is introduced, with the intended interpretation of $B\phi$ as "the individual believes that ϕ". Various properties of beliefs are then expressed by means of axioms, such as the positive introspection axiom $B\phi \to BB\phi$, which says that if the individual believes ϕ then she believes that she believes ϕ. On the semantic side Kripke structures are used, consisting of a set of states (or possible worlds) Ω together with a binary relation $\mathcal{B}$ on Ω, with the interpretation of $\alpha\mathcal{B}\beta$ as "at state α the individual considers state β possible". The connection between syntax and semantics is then obtained by means of a valuation V which associates with every atomic sentence p the set of states where p is true. The pair $\langle \Omega, \mathcal{B} \rangle$ is called a frame and the addition of a valuation V to a frame yields a model. Rules are given for determining the truth of an arbitrary formula at every state of a model; in particular, the formula $B\phi$ is true at state α if and only if ϕ is true at every β such that $\alpha\mathcal{B}\beta$, that is, if ϕ is true at every state that the individual considers possible at α. A property of the accessibility relation $\mathcal{B}$ is said to correspond to an axiom if every instance of the axiom is true at every state of every model based on a frame that satisfies the property and *vice versa*.

The AGM theory of belief revision, on the other hand, has followed a different path. In this literature beliefs are modeled as sets of formulas in a given syntactic language and the problem that has been studied is how a belief set ought to be modified when new information, represented by a formula ϕ, becomes available. With a few exceptions, the tools of modal logic have not been explicitly employed in the analysis of belief revision.

In the economics and game theory literature, it is standard to represent

Dagstuhl Seminar Proceedings 05321
Belief Change in Rational Agents: Perspectives from Artificial Intelligence, Philosophy, and Economics
http://drops.dagstuhl.de/opus/volltexte/2005/330

beliefs by means of a probability measure over a set of states Ω and belief revision is modeled using Bayes' rule. Let P_0 be the prior probability measure representing the initial beliefs, $E \subseteq \Omega$ an event representing new information and P_1 the posterior probability measure representing the revised beliefs. Bayes' rule says that, if $P_0(E) > 0$, then, for every event A, $P_1(A) = \frac{P_0(A \cap E)}{P_0(E)}$. Bayes' rule thus implies the following, which we call the *Qualitative Bayes Rule*:

$$\text{if } \; supp(P_0) \cap E \neq \varnothing, \text{ then } \; supp(P_1) = supp(P_0) \cap E.$$

where $supp(P)$ denotes the support of the probability measure P.

In this paper we propose a unifying framework for static beliefs and belief revision by bringing belief revision under the umbrella of modal logic and by providing an axiomatization of the Qualitative Bayes Rule in a simple logic based on three modal operators: B_0, B_1 and I, whose intended interpretation is as follows:

$B_0\phi$	initially (at time 0) the individual believes that ϕ
$I\phi$	(between time 0 and time 1) the individual is informed that ϕ
$B_1\phi$	at time 1 (after revising his beliefs in light of the information received) the individual believes that ϕ.

Semantically, it is clear that the Qualitative Bayes Rule embodies the *conservativity principle* for belief revision, according to which "When changing beliefs in response to new evidence, you should continue to believe as many of the old beliefs as possible". The set of all the propositions that the individual believes corresponds to the set of states that she considers possible (in a probabilistic setting a state is considered possible if it is assigned positive probability). The conservativity principle requires that, if the individual considers a state possible and her new information does not exclude this state, then she continue to consider it possible. Furthermore, if the individual regards a particular state as impossible, then she should continue to regard it as impossible, unless her new information excludes *all* the states that she previously regarded as possible. The axiomatization we propose gives a transparent syntactic expression to the conservativity principle.

We begin with the semantics. A *frame* is a quadruple $\langle \Omega, \mathcal{B}_0, \mathcal{B}_1, \mathcal{I} \rangle$ where Ω is a set of *states* and $\mathcal{B}_0$, $\mathcal{B}_1$, and $\mathcal{I}$ are binary relations on Ω, whose interpretation is as follows:

$\alpha\mathcal{B}_0\beta$	at state α the individual initially (at time 0) considers state β possible
$\alpha\mathcal{I}\beta$	at state α, state β is compatible with the information received
$\alpha\mathcal{B}_1\beta$	at state α the individual at time 1 (in light of the information received) considers state β possible.

Let $\mathcal{B}_0(\omega) = \{\omega' \in \Omega : \omega\mathcal{B}_0\omega'\}$ denote the set of states that, initially, the individual considers possible at state ω. Define $\mathcal{I}(\omega)$ and $\mathcal{B}_1(\omega)$ similarly.By *Qualitative Bayes Rule* (QBR) we mean the following property:

$$\forall \omega \in \Omega, \; if \; \mathcal{B}_0(\omega) \cap \mathcal{I}(\omega) \neq \varnothing \; then \; \mathcal{B}_1(\omega) = \mathcal{B}_0(\omega) \cap \mathcal{I}(\omega). \tag{QBR}$$

Thus QBR says that if at a state the information received is consistent with the initial beliefs − in the sense that there are states that were considered possible initially and are compatible with the information − then the states that are considered possible according to the revised beliefs are precisely those states.

On the syntactic side we consider a modal propositional logic based on three operators: B_0, B_1 and I whose intended interpretation is as explained in Section 1. The formal language is built in the usual way from a countable set S of atomic propositions, the connectives $\neg$ (for "not") and $\vee$ (for "or") and the modal operators.

The connection between syntax and semantics is given by the notion of model. Given a frame $\langle \Omega, \mathcal{B}_0, \mathcal{B}_1, \mathcal{I} \rangle$, a *model* is obtained by adding a *valuation* $V : S \to 2^{\Omega}$ (where 2^{Ω} denotes the set of subsets of Ω, usually called *events*) which associates with every atomic proposition $p \in S$ the set of states at which p is true. The truth of an arbitrary formula at a state is then defined inductively as follows ($\omega \models \phi$ denotes that formula ϕ is true at state ω; $\|\phi\|$ is the truth set of ϕ, that is, $\|\phi\| = \{\omega \in \Omega : \omega \models \phi\}$):

if q is an atomic proposition, $\omega \models q$ if and only if $\omega \in V(q)$,
$\omega \models \neg\phi$ if and only if $\omega \not\models \phi$,
$\omega \models \phi \vee \psi$ if and only if either $\omega \models \phi$ or $\omega \models \psi$ (or both),
$\omega \models B_0\phi$ if and only if $\mathcal{B}_0(\omega) \subseteq \|\phi\|$,
$\omega \models B_1\phi$ if and only if $\mathcal{B}_1(\omega) \subseteq \|\phi\|$,
$\omega \models I\phi$ if and only if $\mathcal{I}(\omega) = \|\phi\|$.

Remark 1 *Note that, while the truth conditions for $B_0\phi$ and $B_1\phi$ are the standard ones, the truth condition of $I\phi$ is unusual in that the requirement is $\mathcal{I}(\omega) = \|\phi\|$ rather than merely $\mathcal{I}(\omega) \subseteq \|\phi\|$.*

We say that a formula ϕ is *valid in a model* if $\omega \models \phi$ for all $\omega \in \Omega$, that is, if ϕ is true at every state. A formula ϕ is *valid in a frame* if it is valid in every model based on that frame. Finally, we say that a property of frames is *characterized by* (or characterizes) an axiom if (1) the axiom is valid in any frame that satisfies the property and, conversely, (2) whenever the axiom is valid in a frame, then the frame satisfies the property.

We now introduce three axioms that, together, provide a characterization of the Qualitative Bayes Rule.

$$\textit{QUALIFIED ACCEPTANCE:} \quad (I\phi \wedge \neg B_0\neg\phi) \to B_1\phi.$$

This axiom says that if the individual is informed that ϕ ($I\phi$) and he initially considered ϕ possible (that is, it is not the case that he believed its negation: $\neg B_0\neg\phi$) then he accepts ϕ in his revised beliefs. That is, information that is not surprising is believed.

The next axiom says that if the individual receives non-surprising information (i.e. information that does not contradict his initial beliefs) then he continues to believe everything that he believed before:

$$\textit{PERSISTENCE:}\ \ (I\phi \wedge \neg B_0\neg\phi) \to (B_0\psi \to B_1\psi).$$

The third axiom says that beliefs should be revised in a minimal way, in the sense that no new beliefs should be added unless they are implied by the old beliefs and the information received:

$$\textit{MINIMALITY:}\ \ (I\phi \wedge B_1\psi) \to B_0(\phi \to \psi).$$

The Minimality axiom is not binding (that is, it is trivially satisfied) if the information is surprising: suppose that at a state, say α, the individual is informed that ϕ ($\alpha \models I\phi$) although he initially believed that ϕ was *not* the case ($\alpha \models B_0\neg\phi$). Then, for every formula ψ, the formula $(\phi \to \psi)$ is trivially true at every state that the individual initially considered possible ($\mathcal{B}_0(\alpha) \subseteq \|\phi \to \psi\|$) and therefore he initially believed it ($\alpha \models B_0(\phi \to \psi)$). Thus the axiom restricts the new beliefs only when the information received is not surprising, that is, only if $(I\phi \wedge \neg B_0\neg\phi)$ happens to be the case.

Proposition 2 *The Qualitative Bayes Rule (QBR) is characterized by the conjunction of the three axioms Qualified Acceptance, Persistence and Minimality (that is, if a frame satisfies QBR then the three axioms are valid in it and - conversely - if the three axioms are valid in a frame then the frame satisfies QBR).*

We now provide a sound and complete logic for belief revision. Because of the non-standard validation rule for the information operator I, we need to add the universal or global modality A. The interpretation of $A\phi$ is "it is globally true that ϕ". As before, a frame is a quadruple $\langle \Omega, \mathcal{B}_0, \mathcal{B}_1, \mathcal{I} \rangle$. To the validation rules discussed above we add the following:

$$\omega \models A\phi \text{ if and only if } \|\phi\| = \Omega.$$

We denote by $\mathfrak{L}$ the logic determined by the following axioms and rules of inference.

AXIOMS:

1. All propositional tautologies.
2. Axiom K for B_0, B_1 and A (note the absence of an analogous axiom for I):

$$B_0\phi \wedge B_0(\phi \to \psi) \to B_0\psi \quad (\mathrm{K}_0)$$
$$B_1\phi \wedge B_1(\phi \to \psi) \to B_1\psi \quad (\mathrm{K}_1)$$
$$A\phi \wedge A(\phi \to \psi) \to A\psi \quad (\mathrm{K}_A)$$

3. S5 axioms for A:

$$A\phi \to \phi \quad (\mathrm{T}_A)$$
$$\neg A\phi \to A\neg A\phi \quad (5_A)$$

4. Inclusion axioms for B_0 and B_1 (note the absence of an analogous axiom for I):

$$A\phi \to B_0\phi \quad (\mathrm{Incl}_0)$$
$$A\phi \to B_1\phi \quad (\mathrm{Incl}_1)$$

5. Axioms to capture the non-standard semantics for I:

$$(I\phi \wedge I\psi) \to A(\phi \leftrightarrow \psi) \quad (\mathrm{I}_1)$$
$$A(\phi \leftrightarrow \psi) \to (I\phi \leftrightarrow I\psi) \quad (\mathrm{I}_2)$$

RULES OF INFERENCE:

1. Modus Ponens: $\frac{\phi, \quad \phi \to \psi}{\psi}$ (MP)

2. Necessitation for A: $\frac{\phi}{A\phi}$ (Nec_A)

Note that, despite the non-standard validation rule, axiom K for I, namely $I\phi \wedge I(\phi \to \psi) \to I\psi$, is trivially valid in every frame. It follows from the completeness theorem proved below that axiom K for I is provable in $\mathfrak{L}$.

Recall that a logic is *complete* with respect to a class of frames if every formula which is valid in every frame in that class is provable in the logic (that is, it is a theorem). The logic is *sound* with respect to a class of frames if every theorem of the logic is valid in every frame in that class.

Proposition 3 *Logic $\mathfrak{L}$ is sound and complete with respect to the class of all frames $\langle \Omega, \mathcal{B}_0, \mathcal{B}_1, \mathcal{I} \rangle$.*

We are interested in extensions of $\mathfrak{L}$ obtained by adding various axioms. Let $\mathfrak{R}$ ('R' stands for 'Revision') be the logic obtained by adding to $\mathfrak{L}$ the axioms discussed in the previous section:

$$\mathfrak{R} = \mathfrak{L} + \text{ Qualified Acceptance} + \text{Persistence} + \text{Minimality.}$$

Proposition 4 *Logic $\mathfrak{R}$ is sound and complete with respect to the class of frames $\langle \Omega, \mathcal{B}_0, \mathcal{B}_1, \mathcal{I} \rangle$ that satisfy the Qualitative Bayes Rule.*

So far we have not postulated any properties of beliefs, in particular, in the interest of generality, we have not required beliefs to satisfy the KD45 logic. In order to further explore the implications of the Qualitative Bayes Rule, we shall now consider additional axioms:

Consistency of initial beliefs	$B_0\phi \to \neg B_0 \neg\phi$	(D_0)
Positive Introspection of initial beliefs	$B_0\phi \to B_0 B_0\phi$	(4_0)
Self Trust	$B_0(B_0\phi \to \phi)$	(ST)
Information Trust	$B_0(I\phi \to \phi)$	(IT).

Self Trust says that the individual at time 0 believes that his beliefs are correct (he believes that if he believes ϕ then ϕ is true), while Information Trust says that the individual at time 0 believes that any information he will receive will be correct (he believes that if he is informed that ϕ then ϕ is true).

Remark 5 *Since the additional axioms listed above are canonical, it follows from Proposition 4 that if Σ is a set of axioms from the above list, then the logic $\mathfrak{R}+\Sigma$ obtained by adding to $\mathfrak{R}$ the axioms in Σ is sound and complete with respect to the class of frames that satisfy the Qualitative Bayes Rule and the properties corresponding to the axioms in Σ. For example, the logic $\mathfrak{R} + \{D_0, 4_0, ST\}$ is sound and complete with respect to the class of frames that satisfy the Qualitative Bayes Rule as well as seriality, transitivity and secondary reflexivity of $\mathcal{B}_0$.*

It can be shown that No Change ($B_0\phi \wedge I\phi \to (B_1\psi \leftrightarrow B_0\psi)$) is a theorem of $\mathfrak{R} + D_0$. We now discuss some further theorems of extensions of $\mathfrak{R}$. Consider the following axiom:

$$B_0\phi \to B_0 B_1\phi$$

which says that if the individual initially believes that ϕ then she initially believes that she will continue to believe ϕ later.

Proposition 6 *$B_0\phi \to B_0 B_1\phi$ is a theorem of $\mathfrak{R} + 4_0 + ST + IT$.*

Consider now the following axiom which is the converse of the previous one:

$$B_0 B_1\phi \to B_0\phi.$$

This axiom says that if the individual initially believes that later on she will believe ϕ then she must believe ϕ initially.

Proposition 7 *$B_0 B_1\phi \to B_0\phi$ is a theorem of $\mathfrak{R} + ST + IT$.*

Putting together Propositions 6 and 7 we obtain the following corollary.

Corollary 8 *$B_0\phi \leftrightarrow B_0B_1\phi$ is a theorem of $\mathfrak{R} + 4_0 + ST + IT$.*

We now extend the framework to deal with iterated revision. We model, at every state, only the information that is actually received by the individual and do not model how the individual would have modified his beliefs if he had received a different piece of information. Thus we cannot compare the revised beliefs the individual holds after receiving information ϕ with the beliefs he would have had if he had been informed of both ϕ and ψ. On the other hand, it is possible in our framework to model the effect of receiving first information ϕ and then information ψ. Indeed, any sequence of pieces of information can be easily modeled. In order to do this, we need to add a time index to the belief and information operators. Thus, for $t \in \mathbb{N}$ (where $\mathbb{N}$ denotes the set of natural numbers), we have a belief operator B_t representing the individual's beliefs at time t. In order to avoid confusion, we attach a double index $(t, t+1)$ to the an information operator, so that $I_{t,t+1}$ represents the information received by the individual between time t and time $t+1$. Thus the intended interpretation is as follows:

$B_t\phi$	at time t the individual believes that ϕ
$I_{t,t+1}\phi$	between time t and time $t+1$ the individual is informed that ϕ
$B_{t+1}\phi$	at time $t+1$ (in light of the information received between t and $t+1$) the individual believes that ϕ.

Let $\mathcal{B}_t$ and $\mathcal{I}_{t,t+1}$ be the associated binary relations. The iterated version of the qualitative Bayes rule then is the following simple extension of QBR: $\forall\omega \in \Omega, \forall t \in \mathbb{N}$,

$$\textit{if } \mathcal{B}_t(\omega) \cap \mathcal{I}_{t,t+1}(\omega) \neq \varnothing \textit{ then } \mathcal{B}_{t+1}(\omega) = \mathcal{B}_t(\omega) \cap \mathcal{I}_{t,t+1}(\omega). \qquad \text{(IQBR)}$$

The iterated Bayes rule plays an important role in game theory, since it is the main building block of two widely used solution concepts for dynamic (or extensive) games, namely Perfect Bayesian Equilibrium and Sequential Equilibrium. The idea behind these solution concepts is that, during the play of the game, a player should revise his beliefs by using Bayes' rule "as long as possible". Thus if an information set has been reached that had positive prior probability, then beliefs at that information set are obtained by using Bayes' rule (with the information being represented by the set of nodes in the information set under consideration). If an information set is reached that had zero prior probability, then new beliefs are formed more or less arbitrarily, but from that point onwards these new beliefs must be used in conjunction with Bayes' rule, unless further information is received that is inconsistent with those revised beliefs. This is precisely what IQBR requires.

Within this more general framework, a simple adaptation of Propositions 2 and 4 yields the following result:

Proposition 9 *(1) The Iterated Qualitative Bayes Rule (IQBR) is characterized by the conjunction of the following three axioms:*

Iterated Qualified Acceptance:	$(\neg B_t \neg\phi \wedge I_{t,t+1}\phi) \rightarrow B_{t+1}\phi$
Iterated Persistence:	$(\neg B_t \neg\phi \wedge I_{t,t+1}\phi) \rightarrow (B_t\psi \rightarrow B_{t+1}\psi)$
Iterated Minimality	$(I_{t,t+1}\phi \wedge B_{t+1}\psi) \rightarrow B_t(\phi \rightarrow \psi)$.

(2) The logic obtained by adding the above three axioms to the straightforward adaptation of logic $\mathfrak{L}$ *to a multi-period framework is sound and complete with respect to the class of frames that satisfy the Iterated Qualitative Bayes Rule.*