On the Triviality of High-Order Probabilistic Beliefs

Dov Samet

April, 1997

Abstract

Probabilistic beliefs are studied here as modalities of a propositional modal logic. Several axioms of this logic are examined, and the relations between them are proved, using appropriate possible world semantics. Two axioms concerning high-order probabilistic beliefs are investigated in particular. The first is the triviality axiom, which says that one is certain (that is, has a belief of degree one) of one's own beliefs. The second is the averaging axiom, which states that a first order belief concerning some fact f is the average of all degrees of beliefs concerning f, weighted by the degrees given to them by the second order beliefs. It is shown that one whose beliefs satisfy the averaging axiom must be certain that his beliefs satisfy the triviality axiom.

1 Introduction

The Bayesain paradigm deals with the relation between present (or *posterior*) beliefs of an agent and his former (or *prior*) beliefs. In a nut shell, it claims that beliefs change, or are updated, as a result of acquiring information, and the posterior beliefs are formed by conditioning the prior beliefs on all the information that the agent acquired. There are several expressions, or metaphors, that are used to describe the acquired information. In the economic literature and decision theory agents are said to receive a *signal;* in models of games with incomplete information a player is said to learn his *type*.

This picture can be questioned on two accounts. First, the assumption seems to be that beliefs change as a result of gaining certainty of some facts, be it information or a signal. But experience tells us that sometimes we change our beliefs without being able to specify any relevant facts, of which we became certain. Thus, for example, we may have some prior beliefs concerning the honesty of some person and change it considerably after a short conversation with him, in which no new facts are revealed to us. The change is the result of an impression which we cannot express as a fact that we had learned. How can the Bayesian theory account for such an innocent scenario? How does conditioning work here?

Second, the above description of Bayesianism seems to lack an empirical value. What questionaire should we administer to an agent today in order to verify that he is Bayesian updater of his yesterday's beliefs? It should involve questions concerning yesterday's beliefs of everything that he would learn for sure

and this is not the type of events we would expect agents to have beliefs about.

It is not a direct description of an event

High-order beliefs are beliefs about beliefs, beliefs about beliefs about beliefs and so on. Natural languages, and formal ones if they are not arbitrarily restricted, allow the expression of beliefs concerning one's own beliefs. The study of belief using modal logics, the operator of which is interpreted as a belief operator, takes for granted high-order beliefs. Indeed, it is the very nature of such beliefs that is one of the most important foci of attention in these logics.

The beliefs studied here are high-order *probabilistic* beliefs—ones that have numerical degrees, and obey the rules of probability. They are studied using precisely the same tools that are used to study non-quantitative beliefs, namely, the syntax and semantics of modal logic.

The use of modal logic to study probabilistic beliefs, high-order beliefs in particular, is far from being commonplace.[1] Students of modal logics, on one

[1]See Skyrms [16], p. 29–36, for a survey of the views of subjective probability theorists concerning high-order beliefs. The lack of any comparison to the modal logic of non-quantitative beliefs exemplifies our claim.

hand, are usually unaware that the logic of probabilistic beliefs is a particular instance of such logics.[2] On the other hand, logicians who are interested in probability usually view it as part of the semantics of some language, or as a part of deduction theory, but not as modalities which are part of the objective language. Even probability theorists who accept probabilistic statement as part of the syntax seldom resort to the semantics of modal logic.[3]

The syntax of non-quantitative beliefs is based on one modality, *belief*, which is used in the construction of sentences of the form "f is believed".[4] The syntax of probabilistic belief has a host of modalities, *belief with degree* p, for each p between 0 and 1, which are used to construct sentences of the form "f is believed with degree p".[5]

The semantics of both non-quantitative and probabilistic beliefs is that of possible worlds. Non-quantitative belief is modeled semantically by specifying for each possible world ω the propositions, i.e., subsets of possible worlds, which are believed at ω. Using this semantics, truth conditions for sentences can be easily defined. The sentence *"f is believed"* is true in the possible world ω when the proposition F, corresponding to the sentence f, is believed at ω.

The semantics of probabilistic beliefs is similarly defined. Instead of dividing, for each ω, the set of all propositions into those which are believed at ω and those which are not, a number is now attached to each proposition, which is the degree to which it is believed at ω. Moreover, these numbers are defined in such a way as to form a probability measure on the propositions. In short, a probability distribution is associated with each possible world.[6]

[2] Exceptions are a few computer scientists and game theorists who are interested, by the nature of their subject matter, in high-order beliefs of several agents. See Fagin and Halpern [3] and Aumann [1]. However, their interest in the properties of high-order beliefs of one person is limited. They assume that these beliefs are trivial in the sense defined in the sequel.

[3] Jeffrey [7], p. 126 described a semantic model for high-order probabilistic beliefs in a footnote, but never made any use of it.

[4] In order to simplify the formulation, I use here and later the passive form "... is believed" instead of the more appropriate form "x believes...", where x is the name of the person whose beliefs are being modeled.

[5] By adopting modalities that read "f is believed with a degree *at least* p", countably many modalities, for rational p's only, suffice. See footnote 10.

[6] The use of probabilistic semantic models like this can be traced back to Harsanyi [4], although his presentation does not make any use of syntax, or reference to modal logic. See the discussion in subsection 6.1.

Truth conditions are also similarly defined. The sentence *"f is believed with degree p"* is true in the possible world ω when the probability measure associated with ω gives probability p to the proposition F, corresponding to the sentence f.

There is no attempt in this paper to characterize the set of probabilistic models axiomatically. That is, no set of axioms is specified for which these models satisfy a completeness theorem. Instead, I concentrate on a few axioms concerning high-order beliefs, and find the special structure of the semantic models that is required for the axioms to be satisfied. Using the semantic expression of the axioms it is possible to find implication relations among them.

The main issue studied here is the extent to which high-order beliefs are restricted by lower-order beliefs. High-order beliefs are *trivial* when they are completely determined by lower order ones. In such a case, a description of the lowest-order beliefs, those that concern only non-epistemic facts, suffices for the full account of all orders of beliefs, and therefore is a complete description of a mental state.

Let us start by recalling some well-known facts about non-quantitative beliefs. There are two axioms of propositional modal logic, interpreted as logic of belief, that trivialize high-order beliefs in the above mentioned sense. The first, sometimes labeled *positive introspection*, is:

If f is believed, then it is believed that f is believed.

This axiom alone is not enough to determine high-order beliefs, and does not render them trivial. The second axiom, called by some *negative introspection*, is:

If f is not *believed, then it is believed that f is not believed.*

Together, the two axioms, which describe the ability of one to have the right beliefs about his own beliefs or lack of them, make trivial the task of specifying high-order beliefs, once the lower ones are given.[7] The famous $S5$ logic includes these two axioms. The syntactic triviality of high-order beliefs of this logic is reflected in the simple structure of its semantic models.

Consider now probabilistic beliefs. Triviality of high-order probabilistic beliefs requires, analogously to the non-quantitative case, that statements

[7]More than these two axioms is required to imply determination of high-order belief by low-order ones, but these two axioms are the only ones that involve high-order beliefs.

expressing beliefs with a certain degree are correctly believed. As beliefs now come with a degree tag, it is rather obvious that correct beliefs are, in this case, ones of degree 1. Hence the *triviality axiom*,

> *If f is believed with degree p, then it is believed with degree 1 that f is believed with degree p.*

What is shown here is that although the triviality axiom resembles the positive introspection axiom, in not using negation, it is stronger in the following sense. If the triviality axiom is true in a given possible world, then the following counterpart of the negative introspection axiom also holds true in this world:

> *If f is* not *believed with degree p, then it is believed with degree 1 that f is not believed with degree p.*

The characterization of the simple structure of the models in which the triviality axiom is true (that is, the models for which the axiom is true in each of their worlds) is as follows. Consider the partition of such a model into propositions according to the probability associated with the worlds. Then, the probability associated with the worlds in one of these propositions has all its mass concentrated on that proposition. Such models are named here, *partition models*. The analogy of partition models to the $S5$ models is made clearer in Section 2.

Some students of epistemic logic find it hard to forego positive introspection, for they see it as part of the very meaning of belief. For them, the possibility that f is believed while this fact itself is not, seems an abuse of the notion of belief.

The richer structure of probabilistic beliefs makes the foregoing on triviality even more bizarre, but at the same time enables explanations that are impossible for the non-quantitative belief. If f is believed with degree p, and this fact itself is not certain, that is, it is not believed with degree 1, then it is believed in a degree less than 1. But this means that a proposition "f is believed in degree q", with $q \neq p$ is believed with a positive degree. In other words, lack of triviality means not only that one is not certain of one's own beliefs, but that one believes (with a positive degree) in one's wrong beliefs. But if one is uncertain about one's beliefs concerning f, and considers possible (that is ascribes a positive degree to) many of these beliefs, what makes one of them the true one in the believer's eye?

This last question can be given a somewhat satisfactory answer. The first-order belief is the average of all possible first-order beliefs, weighted by the degree given to them by the high-order beliefs. That is, if p_i are different probability numbers and q_i are probability numbers that sum to 1, then:

> *If it is believed with degree q_i that f is believed with degree p_i, for all i, then f is believed with degree $\sum q_i p_i$.*

This axiom is called here, *averaging*. It is easy to see that it is implied by the triviality axiom. Skyrms [15] and Jeffrey [7], while rejecting triviality, consider averaging as an essential assumption.

One of my main purposes is to show that although averaging does not imply triviality, it implies that triviality is certain. That is, one whose beliefs satisfy averaging may fail to have trivial high-order beliefs, but he must be certain that they are trivial. In other words, averaging implies that for each proposition f and number p:

> *It is certain (i.e., believed with degree 1) that if f is believed with degree p, then it is certain that f is believed with degree p.*

Here again the proof that averaging implies certainty of triviality is carried out by characterizing the models, which are called here *almost-partition models*, in which the averaging axiom is true. The crucial element in this characterization is the observation that a probabilistic model can be formally thought of as a specification of the transition probabilities of a Markov chain,[8] and that the averaging axiom requires that the probability associated with each world is an invariant probability measure of this Markov chain.

The averaging principle is related to another natural principle, which was named *reflection* by van Frassen [17]. This principle requires that high-order beliefs be reflected in the lower ones in the following sense.

> *The degree of belief in f,* given *that f is believed with degree p, is p.*

Skyrms [15] shows that reflection implies averaging. It is shown here, moreover, that the two axioms are equivalent.

The plan of this paper is as follows. The next section describes briefly the syntax and semantics of propositional modal logic with one modality, which

[8]This observation was first made, and used for characterization of common priors, in Samet [13].

is referred to as the logic of non-quantitative belief. This is done only in order to introduce the very similar syntax and semantics of the logic of probabilistic beliefs. Subsection 3.1 is a very short reminder of full introspection of non-quantitative belief. In subsection 3.2 the axiom of triviality for probabilistic belief is introduced which corresponds to full introspection. The semantic equivalent of this axiom is shown. The axiom of averaging is explained in subsection 3.3. Its semantic representation and the implications for triviality are given in the next subsection. The axiom of reflection and its equivalence to averaging are presented in subsection 3.5. In the discussion, in Section 4, the basic idea that underlies Kripke's [9] and Harsanyi's [4] seminal works is described, and it is shown how this idea can be given a form that generalizes many logics of beliefs. Finally, the similarities and differences between the two pairs, prior and posterior beliefs, and first- and second-order beliefs are discussed.

2 Logics of belief

2.1 Non-quantitative beliefs

The most obvious and straightforward tool for studying high-order non-quantitative beliefs is propositional modal logic. The well-known basic features of such logics are described briefly, in order to demonstrate their similarity to the logic of probabilistic beliefs, to be introduced in the next section.

The alphabet of propositional modal logic consists of: propositional variables $a, b, \dots$; the standard propositional connectors, '$\neg$', '$\wedge$', and '$\rightarrow$' (for negation, conjunction and implication); a modal operator B called the *belief operator*. A *sentence* is any propositional variable, or any one of the following: $\neg f$, $f \wedge g$, $f \rightarrow g$, and Bf, where f and g are sentences.

The semantics for this syntax is that of possible worlds, called for short worlds. Formally, a *model* $\mathbf{M}$ is a triple $(\Omega,\ \Pi,\ [\cdot])$, where Ω is a finite set of worlds, subsets of which are called *propositions*. Beliefs are described by the map Π which associates with each world ω a 0-1 function, denoted by Π_ω, on the set of all propositions.[9] Proposition E is said to be believed at ω when $\Pi_\omega(E) = 1$. Finally, $[\cdot]$ is a function that associates with each propositional

[9]If there exists for each ω a subset E_ω, such that $\Pi_\omega(E) = 1$, if and only if $E_\omega \subseteq E$, then the worlds in E_ω are said to be *accessible* from ω. Such models are the well-known Kripke models.

variable x a proposition $[x]$.

Given a model $\mathbf{M}$, the truth value of each sentence h is uniquely determined at each world ω. Write $\mathbf{M} \models_\omega h$ when h is true at ω. The truth value of sentences is determined as follows. For a propositional variable x, $\mathbf{M} \models_\omega x$ iff $\omega \in [x]$; $\mathbf{M} \models_\omega \neg f$ iff f is not true at ω; $\mathbf{M} \models_\omega f \wedge g$ iff $\mathbf{M} \models_\omega f$ and $\mathbf{M} \models_\omega g$; $\mathbf{M} \models_\omega f \to g$ iff either $\mathbf{M} \models_\omega \neg f$, or $\mathbf{M} \models_\omega g$. Finally, $\mathbf{M} \models_\omega Bf$ iff the set of all worlds in which f is true is believed at ω.

The function $[\cdot]$ can be now extended to all sentences in such a way that for a sentence h the proposition $[h]$ consists of all worlds at which h is true. It is easy to see that $[\neg f] = \neg[f]$ (where $\neg[f]$ is the set theoretic complement of the proposition $[f]$), $[f \wedge g] = [f] \cap [g]$, $[f \to g] = \neg[f] \cup [g]$, and finally, as Bf is true in ω iff $[f]$ is believed at ω,

$$[Bf] = \{\omega \mid \Pi_\omega([f]) = 1\}.$$

The sentence h is said to be true in a model $\mathbf{M}$, which is written in short as $\mathbf{M} \models h$, when h is true at each world. Clearly, h is true in $\mathbf{M}$, iff $[h] = \Omega$.

2.2 Probabilistic beliefs

The syntax required to formulate probabilistic beliefs must, of course, be richer than the one required to describe non-quantitative beliefs. Still, propositional modal logic suffices for this. The same alphabet described in the previous subsection is used, except that instead of the single modal operator B, there are now infinitely many modal operators, B^p, one for each number p, between 0 and 1. The sentence $B^p f$ reads:[10] "f is believed with degree p".

The semantics for this syntax is also that of possible worlds, and is based on the same idea that beliefs should vary with worlds. Beliefs in a given world are described by a probability measure over propositions. Formally, a *model* $\mathbf{M}$ is a triple $(\Omega,\ P,\ [\cdot])$, where Ω is a finite set of worlds. The function P associates with each world ω a probability distribution over Ω, denoted by P_ω. We call this probability the *type* at the world ω. As before, $[\cdot]$ is a function that associates with each propositional variable x, a proposition $[x]$.

[10] The syntax is more expressible when it has modal operators that read "f is believed with a degree *at least* p". Such operators are used in Monderer and Samet [11], Fagin and Halpern [3], Aumann [1]. It is enough to define these operators for rational p only, keeping the alphabet denumerable (See Aumann [1]). The weaker operators suffice here because we are dealing with semantics of finite models only.

Given a model $\mathbf{M}$ and a world ω, the truth value of each sentence h is uniquely determined at ω. Truth values for propositional variables, negations, and conjunctions are determined as in the semantics for B. The sentence $B^p f$ is true at ω iff $P_\omega([f]) = p$, and thus

$$[B^p f] = \{\omega \mid P_\omega([f]) = p\}.$$

The similarity of this definition to that of $[Bf]$ is obvious.

Observe that for $p \neq q$, $B^p f$ and $B^p f$ are disjoint. Thus, the propositions $B^p f$, when f is fixed and p ranges over all probability numbers, is a partition of Ω. Since Ω is finite, there are only finitely many p's for which $[B^p f]$ is not empty.

In some of the proofs it is assumed that there are no redundant worlds in a model. That is, any two worlds in a model are distinguished by some sentence being true in one of them and not in the other. This is an inessential assumption, which is made only to simplify proofs. All the results also hold without the assumption. Formally, the non-redundancy requires that for any two worlds ω and ω' there exists a sentence f such that $\omega \in [f]$, while, $\omega' \notin [f]$. In particular, as models are finite, the conjunction of all sentences that distinguish ω from all other worlds defines ω. That is, if f is this conjunction, then $\{\omega\} = [f]$. It follows then that for each subset F of Ω there is a sentence f such that $F = [f]$.

3 Trivial high-order beliefs

3.1 The non-quantitative case

In the case of non-quantitative beliefs, the triviality of high-order beliefs is formulated by the following two axioms.[11]

(**Positive introspection**) $\quad Bf \rightarrow BBf$

(**Negative introspection**) $\quad \neg Bf \rightarrow B\neg Bf$

These axioms reflect full introspection; one's beliefs, or lack of them, are always observed correctly by one. The axioms play a central role in the logic of knowledge and belief. Combined with several other axioms, they give rise

[11]By an axiom we mean, as usual, an axiom scheme, that is, all the sentences with the specified form.

to the logic $S5$. The conditions under which these axioms, especially negative introspection, are justified have been extensively debated and deliberated in the literature. They are presented here for the sake of comparison to the triviality of probabilistic beliefs, which is known and understood far less.

3.2 The probabilistic case

The *triviality* axiom for probabilistic beliefs says that one is certain of one's own probabilistic beliefs (that is, believes them with degree one). In our syntax this is simply written as,

(**Triviality**) $$B^p f \to B^1 B^p f$$

The triviality axiom resembles the positive introspection axiom for the operator B. In particular, the instances of the axiom with $p = 1$, $B^1 f \to B^1 B^1 f$, have the same form as the positive introspection axiom. But the triviality axiom also implies an axiom in the spirit of negative introspection, in the following sense.

Theorem 1 *If in model* $\mathbf{M}$ *and world* ω, $\mathbf{M} \models_\omega B^p f \to B^1 B^p f$ *for all* p, *then* $\mathbf{M} \models_\omega \neg B^p f \to B^1 \neg B^p f$ *for all* p.[12]

Proof. Suppose that $\mathbf{M} \models_\omega \neg B^p f$. Then, there is $q \neq p$ such that $P_\omega([f]) = q$, and, therefore, $\mathbf{M} \models_\omega B^q f$. Hence, by the stipulation of the theorem, $\mathbf{M} \models_\omega B^1 B^q f$. This means that $P_\omega([B^q f]) = 1$. But obviously $[B^q f] \subseteq [\neg B^p f]$, which implies that $P_\omega([\neg B^p f]) = 1$. This equality translates into $\mathbf{M} \models_\omega B^1 \neg B^q f$, as required. ∎

The structure of models in which the triviality axiom is true is quite simple. This structure is defined next.

Definition 1 *A model* $\mathbf{M} = (\Omega,\ P,\ [\cdot])$ *is a* partition model, *if there exists a partition of* Ω, $E_1, \ldots, E_n$, *such that for each* i, $i = 1, \ldots, n$, *and* $\omega, \omega' \in E_i$,

1. $P_\omega = P_{\omega'}$,
2. $P_\omega(E_i) = 1$.

[12] In particular this theorem implies that if the triviality axiom is true in $\mathbf{M}$, then $\mathbf{M} \models_\omega \neg B^p f \to B^1 \neg B^p f$ for all p.

Note that if P_ω is replaced, in the previous definition, by Π_ω, and Π_ω is monotonic, then the definition obtained is that of an $S5$ model.

Theorem 2 *The triviality axiom is true in model* $\mathbf{M}$*, if and only if* $\mathbf{M}$ *is a partition model.*

Proof. Let $\mathbf{M}$ be a partition model, with the partition $E_1, \ldots, E_n$ of Definition 1. Suppose $\omega \in [B^p f]$, that is, $P_\omega[f] = p$. If $\omega \in E_i$, then by part 1 of the definition, $E_i \subseteq [B^p f]$. Therefore, by part 2 of the definition, $P_\omega([B^p f]) = 1$. Thus, $\omega \in [B^1 B^p f]$, which shows that the triviality axiom is true in $\mathbf{M}$.

Assume now that the triviality axiom is true in the model $\mathbf{M}$. Let $E_1, \ldots, E_n$ be the partition of Ω into equivalence classes of types. That is, for $\omega \in E_i$, E_i includes all the worlds ω' such that $P'_\omega = P_\omega$. It is enough to show that part 2 of the definition holds. Let $\omega \in E_i$ and $j \neq i$. Then, for any $\omega' \in E_j$ $P_\omega \neq P_{\omega'}$, and therefore there is a subset F_j of Ω, such that $P_\omega(F_j) = p_j$ and $P_{\omega'}(F_j) \neq p_j$. Let f_j be a sentence such that $[f_j] = F_j$. Then $E_i \subseteq [B^{p_j} f_j]$, and $E_j \cap [B^{p_j} f_j] = \emptyset$. It follows that, $E_i = [\wedge_{j \neq i} B^{p_j} f_j]$. By the triviality axiom, $E_i \subseteq [B^1 B^{p_j} f_j]$ for all $j \neq i$. In particular, for ω, $P_\omega([B^{p_j} f_j]) = 1$. Hence, $P_\omega([\wedge_{j \neq i} B^{p_j} f_j]) = 1$, that is, $P_\omega(E_i) = 1$, as required. ∎

3.3 Averaging

The rich structure of probabilistic beliefs allows for the weakening of triviality, or full introspection, in a way that cannot be done in the poorer syntax of non-quantitative beliefs. Some students of subjective beliefs, (e.g., Skyrms [15]) insist that it is meaningful, indeed, unavoidable, to abandon the triviality axiom for some weaker assumption on introspection.

If the triviality axiom is not true in a model, this means that for some sentence f, $B^p f$ is true at some world ω, while $B^1 B^p f$ is not true at ω. Hence, there is $q < 1$ (possibly, even $q = 0$) such that $B^q B^p f$ is true there. But then, some belief(s) in f, with degree other than p, must be believed with positive degree at ω. Suppose, then, that each of $B^{p_1} f, \ldots, B^{p_n} f$ is believed at ω with positive degree. What then makes $B^p f$ be the true belief at ω? The most natural answer would be that p is just the average of the various p_i's with the degrees given to the beliefs in $B^{p_i} f$. This relation between first- and second-order beliefs is assumed by several authors (e.g., Skyrms [15] p. 162-164, Jeffrey [7]).

This requirement is formulated now under the name *averaging* axiom as follows. For any $k \geq 1$, and probability numbers $p_1, \ldots, p_k$, $q_1, \ldots, q_k$, and r, such that the p_j's are k different numbers, $\sum_{j=1}^{k} q_j = 1$, and $\sum_{j=1}^{k} q_j p_j = r$ (that is, r is the average of the p_j's with respect to the probability measure $q_1, \ldots, q_k$),

$$(\textbf{Averaging}) \qquad \Big(\wedge_{j=1}^{k} B^{q_j} B^{p_j} f\Big) \to B^r f$$

It can be easily seen that the averaging axiom is indeed weaker than the triviality axiom, in the sense that the former is implied by the latter.

Claim 1 *If the triviality axiom is true in world ω, then the averaging axiom is also true there.*

Proof. By definition, when the triviality axiom is true in world ω, then $B^q B^p f$ is true at ω, if either $q = 1$ and $B^p f$ is true at ω, or $q = 0$ and $B^p f$ is not true at ω. Thus, $\wedge_{j=1}^{k} B^{q_j} B^{p_j} f$ can be true at ω, only if for some i, $q_i = 1$, and $q_j = 0$ for all $j \neq i$, and $B^{p_i} f$ is true at ω. In this case $B^r f$ is also true at ω, as $r = p_i$. ∎

3.4 Averaging and triviality

In this section, averaging is explored semantically and it is shown that it is not weaker than triviality by much. Indeed, we will see that one whose beliefs are the average of his higher-order beliefs is certain that his high-order beliefs are trivial. This is stated formally in the next theorem.

Theorem 3 *If the averaging axiom is true in a model* $\mathbf{M}$*, then*

$$\mathbf{M} \models B^1 \Big(B^p f \to B^1 B^p f\Big)$$

for each f and p.

Theorem 3 is proved by characterizing the models in which the averaging axiom is true. For this purpose, the following kind of models is defined.

Definition 2 *A model* $\mathbf{M} = (\Omega,\ P,\ [\cdot])$*, is an* almost-partition model*, if there exists a proper subset (possibly empty) of Ω, E_0, and a partition $E_1, \ldots, E_n$ of $\Omega \setminus E_0$, such that:*

1. *$P_\omega = P_{\omega'}$, for each $\omega, \omega' \in E_i$, when $i \geq 1$,*

2. *$P_\omega(E_i) = 1$, for each $\omega \in E_i$, when $i \geq 1$,*

3. *for each $\omega \in E_0$ there is a probability vector $(a_1, \ldots, a_n)$, such that $P_\omega = \sum_{i=1}^n a_i P_{\omega_i}$ for any $\omega_i \in E_i$ for $i = 1 \ldots, n$.*

Thus, an almost-partition model consists of the sets $E_1 \ldots, E_n$, which have the same structure as a partition model, and of an additional (possibly empty) part E_0. The latter is further restricted. In particular, the probability associated with any world in it has all its mass concentrated on those worlds that belong to the partition part.

Theorem 4 *The averaging axiom is true in model* $\mathbf{M}$*, if and only if* $\mathbf{M}$ *is an almost-partition model.*

Proof. Suppose that the averaging axiom is true in model $\mathbf{M}$. Observe that P can be viewed as a transition probability of a Markov chain, the set of states of which is Ω. Accordingly, we refer now to worlds as states. Recall that an invariant probability of P is a probability m on Ω, such that for each $F \subseteq \Omega$, $m(F) = \sum_{\omega\in\Omega} m(\omega) P_\omega(F)$. We now show that if the averaging axiom is true in $\mathbf{M}$, then P_ω is an invariant probability distribution of P, for each ω.

Let $F \subseteq \Omega$, and choose a sentence f such that $F = [f]$. Let $p_1, \ldots, p_k$ be numbers such that the propositions $[B^{p_1} f], \ldots, [B^{p_k} f]$ form a partition of Ω. Let, $q_1, \ldots, q_k$ be the numbers for which $B^{q_i} B^{p_i} f$ is true at ω for $i = 1, \ldots, k$. That is, $P_\omega([B^{p_i} f] = q_i$. Then, obviously $\sum_{i=1}^k q_i = 1$, and by the averaging axiom $B^r f$ is true at ω, where $r = \sum_{i=1}^k p_i q_i$. Thus, $P_\omega(F) = r$ and hence,

$$\begin{aligned}
P_\omega(F) &= \sum_{j=1}^k p_j P_\omega([B^{p_j} f]) \\
&= \sum_{j=1}^k p_j P_\omega\left(\{\omega' \mid P_{\omega'}([f]) = p_j\}\right) \\
&= \sum_{i=1}^k p_j \sum_{\omega':\ P_{\omega'}([f])=p_j} P_\omega\left(\{\omega'\}\right) \\
&= \sum_{\omega'\in\Omega} P_\omega\left(\{\omega'\}\right) P_{\omega'}(F).
\end{aligned}$$

Let E_0 be the (possibly empty) set of transient states of the Markov chain with transition probability P, and $E_i, \ldots, E_n$ the partition of $\Omega \setminus E_0$ to communicating classes of recurrent states. If $\omega \in E_i$, for $i \geq 1$, then ω communicates only with states in E_i, that is, $P_\omega(E_i) = 1$. Transient states go with probability one to recurrent ones, and therefore for each $\omega \in E_0$, $P_\omega(\cup_{i=1}^n E_i) = 1$. Moreover, the Markov chain, restricted to E_i, with $i \geq 1$, is ergodic and therefore there exists a unique invariant probability m_i on E_i, and the invariant probabilities of P are all probabilities of the form $\sum_{i=1}^n a_i m_i$ for a probability vector $(a_1, \ldots, a_n)$.

Consider now $\omega \in E_i$ for $i \geq 1$. We have shown for such ω, that P_ω is invariant, and that $P_\omega(E_i) = 1$. Thus $P_\omega = m_i$. Hence, part 1 and 2 of Definition 2 are satisfied. Also, for $\omega \in E_0$, P_ω must be of the form $\sum_{i=1}^n a_i m_i = \sum_{i=1}^n a_i P_{\omega_i}$, which proves part 3 of Definition 2.

To show the converse, suppose that $\mathbf{M}$ is an almost-partition model. Consider first $\omega \in E_i$ for some $i \geq 1$. Then, by parts 1 and 2 of Definition 2, it follows, precisely as in Theorem 2, that the triviality axiom is true at ω. But then by Claim 1 the averaging axiom is true at ω.

Assume now that $\omega \in E_0$, and that an instance of the antecedent of the averaging axiom, $\wedge_{j=1}^k B^{q_j} B^{p_j} f$, is true at ω. Then $P_\omega([B^{p_j} f]) = q_j$, for $j = 1, \ldots k$. Choose $\omega_i \in E_i$ for each $i \geq 1$; then, by part 3 of Definition 2, $q_j = \sum_{i=1}^n a_i P_{\omega_i}([B^{p_j} f])$. Denote $q_{ji} = P_{\omega_i}([B^{p_j} f])$. Thus $\wedge_{j=1}^k B^{q_{ji}} B^{p_j} f$ is true at ω_i. Note that if $a_i > 0$, then $\sum_{j=1}^k q_{ji} = 1$, because otherwise, there would be p different from all p_j's, and $q > 0$, such that $P_{\omega_i}([B^p f]) = q$, and hence, as $a_i > 0$, $P_\omega([B^p f]) > 0$, contrary to the fact that the q_j's sum to one. Therefore, since we have shown that the averaging axiom is true at ω_i, $B^{r_i} f$ is true at ω_i where $r_i = \sum_{j=1}^k q_{ji} p_j$. Hence, $P_{\omega_i}([f]) = r_i$, and $P_\omega([f]) = \sum_{i=1}^n a_i P_{\omega_i}([f]) = r$. ∎

Proof of Theorem 3. Suppose the averaging axiom is true in $\mathbf{M}$. Then by Theorem 4, $\mathbf{M}$ is an almost-partition model. By parts 1 and 2 of Definition 2 it follows, as in Theorem 2, that the triviality axiom is true in each $\omega \in E_i$ for $i \geq 1$. Thus, $E_i \subseteq [B^p f \to B^1 B^p f]$. But, for such ω, $P_\omega(E_i) = 1$, and therefore $B^1(B^p f \to B^1 B^p f))$ is true at ω.

If $\omega \in E_0$, then $P_\omega(\cup_{i=1}^n E_i) = 1$, by part 3 of Definition 2. But, by the above inclusion, $\cup_{i=1}^n E_i \subseteq [B^p f \to B^1 B^p f)]$. Thus $B^1(B^p f \to B^1 B^p f))$ is true at ω. ∎

3.5 Reflection

Once higher-order beliefs are admitted, one can condition his beliefs, concerning a certain fact f, on his beliefs concerning this fact. The *reflection* axiom, thus named by van Frassen [17], says that the degree of belief in f is reflected in the conditioned belief. More specifically, suppose that f is believed with degree p, that is, $B^p f$ is true. Then the reflection axiom requires that the belief in f *given* $B^p f$ is of degree p.

In order to formulate this axiom in our syntax we should get rid of the conditional probability, as we have no operator for such conditioning. For this, note that the degree of belief in f given $B^p f$ is the degree of belief in f *and* $B^p f$ divided by the degree of belief in $B^p f$. Thus, by the reflection axiom, this quotient must be p. Equivalently, the degree of belief in $f \wedge B^p f$ should be p times the degree of belief in $B^p f$. In other words, if the degree of belief in $B^p f$ is q, then the degree of belief in $f \wedge B^p f$ should be pq. This is precisely what the following sentence says.

(**Reflection**) $$B^q B^p f \rightarrow B^{qp}(f \wedge B^p f)$$

Skyrms [15], shows that the reflection axiom implies the averaging axiom. Using the results of the previous subsection we can show that they are equivalent.

Theorem 5 *The reflection axiom is true in a model* $\mathbf{M}$*, if and only if the averaging axiom is true in* $\mathbf{M}$*.*

Proof. Suppose the reflection axiom is true in $\mathbf{M}$, and let $\bigwedge_{j=1}^k B^{q_j} B^{p_j} f$ be an instance of the antecedent of the averaging axiom which is true in world ω. As the p_j's are different numbers, the propositions $[B^{p_j} f]$ are disjoint in pairs, and therefore $P_\omega(\cup_{j=1}^k [B^{p_j} f]) = \sum_{j=1}^k q_j = 1$. By reflection, $B^{q_j p_j}(f \wedge B^{p_j} f)$ is true in ω for each j. Hence, $P_\omega([f] \cap [B^{p_j} f]) = q_j p_j$. Thus, $P_\omega([f]) = P_\omega\left(f \cap (\cup_{j=1}^k [B^{p_j} f])\right) = P_\omega(\cup_{j=1}^k (f \cap [B^{p_j} f])) = \sum_{j=1}^k q_j p_j$, which show that the consequence of the averaging axiom is true in ω

Conversely, suppose that the triviality axiom is true in $\mathbf{M}$, and let $B^q B^p f$ be an instance of the antecedent of the reflection axiom which is true in world ω. By Theorem 4, the model $\mathbf{M}$ is almost-partition model with the sets $E_0, \ldots, E_n$, as described in Definition 2. Assume first, that $\omega \in E_i$ for $i \geq 1$. If $B^p f$ is true in ω, then q must be 1, and $P_\omega([f] \cap [B^p f]) = P_\omega([f]) = p$, which shows that the consequence of the reflection axiom is true in ω. If $B^p f$

is not true in ω, then q must be 0, and $P_\omega([f]\cap[B^pf]) = 0$, which again shows that the consequence of the reflection axiom is true in ω.

Suppose now that $\omega \in E_0$. Then, for $\omega_i \in E_i$, $i = 1\ldots,n$, $q = P_\omega([B^pf]) = \sum_{i=1}^n a_i P_{\omega_i}([B^pf])$. But $P_{\omega_i}([B^pf])$ is 1 when B^pf is true in ω_i, and 0 otherwise. Therefore, $q = \sum_{i:P_{\omega_i}([B^pf])=1} a_i$. Now, $P_\omega([f] \cap [B^pf]) = \sum_{i=1}^n a_i P_{\omega_i}([f] \cap [B^pf])$, and $P_{\omega_i}([f] \cap [B^pf])$ is p when $P_{\omega_i}([B^pf]) = 1$, and 0 otherwise. Therefore, $P_\omega([f] \cap [B^pf]) = \sum_{i:\ P_{\omega_i}([B^pf])} a_i p = pq$, as required. ∎

4 Discussion

4.1 Set-theoretic presentation of belief

The familiar set-theoretic model of probability, introduced by Kolmogorov [8] in 1933, can serve, as is, as a set-theoretic presentation of belief. The model consists of a set, the subsets of which—called events, or propositions—are the objects of beliefs. A probability measure associates with each event a number which can be thought of as the degree of belief in that event. This model can be generalized by allowing functions that, unlike probability measures, are not additive. It is possible to give up the numerical values associated with events, and describe non-quantitative beliefs by designating a family of events which are the ones that are believed.

In all these set-theoretic models, beliefs are about events, but there is no sensible way to identify beliefs with events. Therefore, such models cannot be used in cases that beliefs themselves are the objects of beliefs.

Two scholars, Kripke [9] and Harsanyi [4], set out, in the 60's, to construct set-theoretic models that overcome this obstacle. Though their works differ very much in detail, as well as in motivation, the main idea is exactly the same. *Starting with a given set, each element of the set is associated with beliefs over subsets of the set.* Once this is done it is possible to identify each particular belief with the subset of all the elements with which such a belief is associated. This idea, which is ingenious in its simplicity, proved to be extremely fruitful. Kripke's paper marked a turning point in modal logic, and the theory of games with incomplete information started with Harsanyi's definition of types.

Kripke's interest was not limited to beliefs. His purpose was to provide semantic models for propositional modal logic and to prove a completeness theorem for this semantics. In the syntax of modal logic—we will refer to its modality as belief—the belief operator is applied to sentences, which may themselves contain belief operators. Thus, if one is to use a set-theoretic model, in which propositions are the objects of beliefs, then one needs to be able to identify beliefs with propositions. Kripke achieved this by equipping the set with a binary relation—called the accessibility relation. Looking at it in another way, in Kripke's model each element of the set (a possible world in his terminology) is associated with a subset (of those worlds that are accessible from it). This set, as well as all its supersets, are the propositions that are believed in that world.

Harsanyi sought a way to analyze games with incomplete information, that is, games in which the players are uncertain of the parameters of the game they are playing. Players act according to what they believe about their environment, but this includes in particular what they believe about other players' beliefs. Unbothered by syntactical questions, Harsanyi was seeking a set-theoretic model in which the probabilistic beliefs of a player about probabilistic beliefs of other players are expressible, that is, one in which probabilistic beliefs can be identified with events. In his model, each element of the set, called a state, is associated with the *types* of the players, where a player's type is a specification of a probability measure over the set.[13]

Thus, Harsanyi and Kripke used the same idea of associating with each world, or state, beliefs about subsets. They differ only in the characteristics of the beliefs associated with each element of the model: a set of accessible worlds in Kripke's model; a probability measure in Harsanyi's model. But even this difference is not really as large as it appear at first glance. It is possible to define a general set-theoretic model of beliefs, as it is done here and in Heifetz and Samet [6], of which Kripke's and Harsanyi's models are special cases.

In this general model, beliefs are represented, in a given set of worlds, by associating with each world a real-valued function on propositions. Restrictions on the allowed functions and the way they vary with worlds define the characteristics of the modality, and if such models are used as semantics for some syntax, then these restrictions define appropriate axioms and inference rules. There is one syntax that fits all such models. It is one with denumerably many modal operators B^p for rational p which stand for degree of belief

[13] Harsanyi's original formulation was slightly different. For each player there is a set of types and the model is the product of these type sets. Thus, a state is a vector of types. Each type of a player is then associated with a probability measure.

at least p.[14] Given a set of axioms, some of these operators may become redundant and can be eliminated from the syntax.

Harsanyi models are obtained when these functions are required to be probability measures.[15] It is possible to enlarge the set of models by requiring only that these functions be monotonic and have values in the interval $[0, 1]$. Beliefs described in this way are the Dempster-Shafer [14] *belief functions* (Shafer [14]). Another set of allowed functions includes those that take as values only 0 and 1. In such a case it is possible to describe the function, equivalently, by associating with each world the set of propositions on which the function at this world obtains the value 1. Such models are called *minimal models* by Chellas [2]. If the set of these propositions is further restricted to be the family of supersets of some subset of worlds (subset of accessible worlds), then we have a Kripke model.

The general approach depicted here offers a unified treatment of many kinds of epistemic modalities. Highlighting the similarities between probabilistic and non-quantitative beliefs, in this paper, demonstrates the advantage of the unified approach. Following is another example of the benefits of this approach.

When one considers a family of semantic, set-theoretic, models, one of the first questions that comes to mind is that of the relation between these models. Which maps preserve their structure? What kind of universal models do they have? These questions are hardly asked at all in the literature of modal logic, the main interest of which is the relation between syntax and semantics. The literature that followed Harsanyi's work was more sensitive to these questions, starting with the seminal paper of Mertens and Zamir [10]. The set up of Harsanyi's spaces, in which beliefs are presented by real valued functions lent is itself easily to studying and answering these questions. The presentation of beliefs of all sorts in terms of real-valued functions makes it possible to formulate a general theory of universal spaces, applicable to any class of semantics models of belief. Several results in this direction can be found in Heifetz and Samet [5] and [6].

[14]In this paper we preferred to use operators that stand for degree of belief p. See footnote 10.

[15]Harsanyi also required that for each player the model be a partition model, in the sense of this paper.

4.2 Prior and posterior beliefs

Axioms similar to the averaging and reflection are studied in Samet [12] for a model which has prior and posterior beliefs. In that paper, the axiom analogous to averaging requires that when prior beliefs are averaged, weighted by posterior ones, the result is the correct prior belief. The axiom analogous to reflection requires that the conditioning of a posterior belief on a prior one yields the prior belief. Thus the pair prior posterior beliefs in that paper corresponds to the pair first- second-order beliefs here.

There is a big difference, though, formally and conceptually, between the two pairs. First- and second-order beliefs belong to one family of modalities (belief with degree p); prior beliefs and posterior beliefs belong to two different families (prior belief with degree p, and posterior belief with degree p). The difference between first- and second-order beliefs is, in some sense, a syntactic difference which disappears in the semantics. Semantically, beliefs are given by the association of each world with one probability measure, which represents beliefs of all orders. The difference between prior and posterior belief is semantic. In a model which hosts both kind of beliefs, each world is associated with two probability measures, one for prior beliefs, and one for posterior beliefs.

The relation between prior and posterior beliefs is evidently temporal, and conditioning is an updating procedure which involves temporal change of beliefs. Nothing of the sort applies to the pair first and second order beliefs. These beliefs are simultaneous as is evident from their semantics. The reflection axiom expresses a consistency requirement on beliefs, rather than any temporal procedure of updating.

One should not expect, therefore, that the implications of averaging will be the same in the two different setups. Indeed, here the averaging axiom corresponds to almost-partition models. The analogous axiom in Samet [12] (which is called there *invariance*) does not correspond to almost-partition models with respect to the prior beliefs. Yet it is shown there that an axiom stronger than reflection (which is called there *Bayesianism*) corresponds to almost-partition models with respect to the prior beliefs.

References

[1] Aumann, J. R., (1989), "Notes on Interactive Epistemology," Cowles Foundation for Research in Economics working paper.

[2] Chellas, B. F., (1980) *Modal Logic, an Introduction,* Cambridge University Press, Cambridge.

[3] Fagin, R. and J. Y. Halpern, (1988), "Reasoning about Knowledge and Probability," in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge,* Morgan Kaufmann, California.

[4] Harsanyi, J. C., (1967-68), Games with Incomplete Information Played by Bayesian Players, parts I,II, & III, *Man. Sc.*, 14, pp. 159-182, 320-334, 486-502.

[5] Heifetz, A. and D. Samet, (1996) "Topology-Free Typology of Belief," working paper.

[6] Heifetz, A. and D. Samet, (1996) "Knowledge Spaces with Arbitrarily High Rank," working paper.

[7] Jeffrey, R., (1992) *Probability and the Art of Judgment,* Cambridge University Press.

[8] Kolmogorov, A., (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung,* Berlin.

[9] Kripke, S. A., (1963), "Semantical Analysis of Modal Logic I. Normal Modal Propositional Calculi," *Zeitschrift fúr Mathematiche Logik und Grundlagen der Mathematik,* 9, p. 67-96.

[10] Mertens, J.F. and S. Zamir, (1985), "Formulation of Bayesian Analysis for Games with Incomplete Information," *Int. Jour. Game Theory* 14, pp. 1-29.

[11] Monderer, D., and D. Samet, (1989), "Approximating Common Knowledge with Common Belief," *Games and Economic Behavior* 1, pp. 170-190.

[12] Samet, D., (1996), "Looking Backwards Looking Inwards: Priors and Introspection," working paper.

[13] Samet, D., (1996), “Iterated Expectations and Common Priors,” working paper.

[14] Shafer, G. A., (1979), *A Mathematical Theory of Evidence,* Princeton University Press.

[15] Skyrms, B., (1980), *Casual Necessity,* Yale University Press.

[16] Skyrms, B., (1984) *Pragmatics and Empiricism,* Yale University Press.

[17] van Frassen, B., (1984), “Beliefs and the Will” *J. Phil.*, 81, 235-256.