# Characterizing the Common Prior Assumption*

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

May 21, 2000

**Abstract**

Logical characterizations of the common prior assumption (CPA) are investigated. Two approaches are considered. The first is called *frame distinguishability*, and is similar in spirit to the approaches considered in the economics literature. Results similar to those obtained in the economics literature are proved here as well, namely, that we can distinguish finite spaces that satisfy the CPA from those that do not in terms of disagreements in expectation. However, it is shown that, for the language used here, no formulas can distinguish infinite spaces satisfying the CPA from those that do not. The second approach considered is that of finding a sound and complete axiomatization. Such an axiomatization is provided; again, the key axiom involves disagreements in expectation. The same axiom system is shown to be sound and complete both in the finite and the infinite case. Thus, the two approaches to characterizing the CPA behave quite differently in the case of infinite spaces.

*Work supported in part by NSF under grant IRI-96-25901 and by the Air Force Office of Scientific Research under grant F49620-96-1-0323.

# 1 Introduction

The *common prior assumption* (CPA) is one that, up until quite recently, was almost an article of faith among economists. This assumption says that differences in beliefs among agents can be completely explained by differences in information. Essentially, the picture is that agents start out with identical prior beliefs (the common prior) and then condition on the information that they later receive. If their later beliefs differ, it must thus be due to the fact that they have received different information.

The CPA has played a prominent role in economic theory. Harsanyi [1968] showed that a game of incomplete information could be reduced to a standard game of imperfect information with an initial move by nature iff individuals could be viewed as having a common prior over some state space. Aumann [1976] showed that individuals with a common prior could not "agree to disagree"; that is, if their posteriors were derived from a common prior and they had common knowledge of their posterior probabilities of a particular event, these posteriors would have to be the same.

The CPA has come under a great deal of scrutiny recently. (See, for example, the exchange between Gul [1998] and Aumann [1998]; see [Morris 1995] for an overview.) In an effort to try to understand the implications of the CPA better, there have been a number of attempts to characterize the CPA. Of most relevance here are the results of Bonanno and Nehring [1999], Feinberg [1995, 2000], Morris [1994], and Samet [1998], who all showed that, in finite spaces, the common prior could be characterized by a disagreement in expectations, in a sense explained below. Feinberg [2000] extended this result to infinite spaces that satisfied a certain compactness condition, and also showed that this compactness condition was necessary.

This paper continues these efforts. I characterize the CPA using traditional tools from modal logic, and compare these characterizations to those used in the economics literature. In the process, I highlight the role of the language used in getting a characterization. Feinberg [2000] showed how to characterize the CPA in *syntactic terms*, essentially using a logic with operators for knowledge and probability. I use a much richer language here, one introduced in [Fagin and Halpern 1994], which has operators for knowledge, common knowledge, and probability. Feinberg's language is weaker than that used here in two significant respects. The first is that it does not include an operator for common knowledge. To get around this, his characterization involves infinite sets of formulas. The second is that the operators in his language do not allow us to express expectation. In particular, this means that disagreement in expectation cannot be expressed. Feinberg gets around this by an ingenious construction that involves adding coin tosses to the description of the world, in order to construct a more complex model. In this model, disagreement in expectation is converted to disagreement between two agents about the probability of an event, and this fact can be expressed in Feinberg's language. By using a richer language, the need for this construction is completely obviated.

However, characterizing the CPA involves more than just language. It depends on what counts as a characterization. I consider two quite different characterizations here.

One is called *frame distinguishability*, and is very similar in spirit to the types of characterization considered in the economics literature. Not surprisingly, the results I obtain for frame distinguishability are quite similar to those obtained in the economics literature (and much the same techniques are used). In particular, I show that finite frames (essentially, finite spaces) that satisfy the CPA can be distinguished from those that do not in terms of disagreements in expectation. However, there are no formulas in the language considered here that can distinguish infinite spaces satisfying the CPA from those that do not.

The second type of characterization I consider is that of finding a sound and complete axiomatization. I provide such an axiomatization; again, the key axiom involves disagreements in expectation. The same axiom system is shown to be sound and complete both in the finite and the infinite case. Thus, the distinction between finite and infinite spaces vanishes when we consider axiomatizations. Roughly speaking, this can be understood as saying that the language is too weak to distinguish finite from infinite spaces (despite being much stronger than that considered by Feinberg).

It may seem strange at first that a language not rich enough to provide a distinguishing test can still completely characterize all the properties of a notion of interest in that language. But this phenomenon is actually quite familiar in other areas of mathematics. There is a well-known complete axiomatization of the real numbers with addition and multiplication due to Tarski [1951]. Nevertheless there are nonstandard models of the reals that satisfy the same axioms, so the language cannot distinguish the standard models from the nonstandard models. (This observation is in fact the basis for the whole enterprise of nonstandard analysis [Davis 1977].) Essentially, I show that, just as there are nonstandard models of the reals that satisfy all the properties of the reals, there are "nonstandard" models that satisfy all the properties of the CPA expressible in the (rather rich) language considered here yet do not satisfy the CPA.

A natural question to ask is which of the two types of characterization I consider is more appropriate. That, of course, depends on the application. If we are interested in testing whether the CPA holds in a given space, this is a question essentially about frame distinguishability. As it happens, if a finite space does not satisfy the CPA, there is a single formula that will be true in that space that is not true in any space that satisfies the CPA. Moreover, that formula is one that the agents themselves know to be true, so not only can the modeler make the distinction, the agents themselves can.[1] On the other hand, suppose rather than being given a particular space, all that the modeler is given is a finite collection $\Sigma$ of facts about the space. (For example, $\Sigma$ may give information about the agents' knowledge and beliefs.) Note that $\Sigma$ will in general not determine a single space; there may be a number of spaces compatible with $\Sigma$. The modeler may then be interested in knowing what follows from the CPA together with $\Sigma$ as opposed to just from $\Sigma$ alone. That is, what extra conclusions follow from the CPA, given $\Sigma$. This is a question that can be answered using a complete axiomatization—frame distinguishability

[1] There are some subtle computational issues here though; see Section 3.1.

is of no help at all. Thus, for example, the axiomatization can be viewed as providing, among other things, an exact characterization of the extent to which the CPA implies no common knowledge of disagreement.

The rest of this paper is organized as follows. In Section 2, I carefully define the language considered and its semantics. In Section 3, I consider the two types of characterizations. I also consider what happens if the common knowledge operator is not in the language. In this case, I show that there are no new consequences of the CPA. This result is similar in spirit to, but different from, one of Lipman [1997]. Lipman showed that there are some (albeit weak) consequences of the CPA, even without common knowledge in the language. The differences in our results are attributable to a small but significant difference in our definitions of the CPA in the case when there are information sets with prior probability 0; see Section 3 for details. I conclude in Section 4 with some discussion of these results.

Most proofs can be found in the appendix.

# 2 Syntax and Semantics

To reason formally about knowledge and probability, the standard approach in the literature in philosophy and mathematics, which has also been adopted in computer science, starts with a language (the *syntax*). Of course, there is some flexibility in exactly what language should be chosen. Since I want to reason about knowledge, common knowledge, and probability here, I use the syntax first defined in [Fagin and Halpern 1994], that lets us reason explicitly about all these notions. This choice of language (particularly the assumption that the language includes common knowledge) has nontrivial consequences for the results of this paper, as we shall see. I return to the issue of the choice of language in Section 4; for now I just focus on this language (occasionally without common knowledge).

Suppose we consider a system with $n$ agents, say $1, \ldots, n$, and we have a nonempty set $\Phi$ of primitive propositions about which we wish to reason. (Think of these primitive propositions as representing basic events such as "agent 1 went left on his last move".) We take $\mathcal{L}_n^{K,C,pr}$ to be the least set of formulas that includes $\Phi$ and is closed under the following construction rules:[2] If $\varphi, \varphi', \varphi_1, \ldots, \varphi_m$ are formulas in $\mathcal{L}_n^{K,C,pr}$, then so are $\neg\varphi$, $\varphi \wedge \varphi'$, $K_i\varphi$, $i = 1, \ldots, n$, (which is read "agent $i$ knows $\varphi$"), $C\varphi$ ("$\varphi$ is common knowledge"), and $a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) > b$, where $a_1, \ldots, a_m, b$ are rational numbers, ($pr_i(\varphi)$ is read "the probability of $\varphi$ according to agent $i$"). Let $\mathcal{L}_n^{K,pr}$ consist of all the formulas in $\mathcal{L}_n^{K,C,pr}$ that do not mention the $C$ operator.

As usual, $\varphi \vee \varphi'$ and $\varphi \Rightarrow \varphi'$ are abbreviations for $\neg(\neg\varphi \wedge \neg\varphi')$ and $\neg\varphi \vee \varphi'$, respectively. In addition, $E^1\varphi$ ("everyone knows $\varphi$") is an abbreviation for $K_1\varphi \wedge \ldots K_n\varphi$ and $E^{m+1}\varphi$

[2]Strictly speaking, I should write $\mathcal{L}_n^{K,C,pr}(\Phi)$, because $\Phi$ is also a parameter of the language, just as $n$ is. However, I omit it here, to simplify the notation.

is an abbreviation for $E^1 E^m \varphi$ ("everyone knows that everyone knows ...that everyone knows—$m+1$ times—$\varphi$"), for $m \geq 1$. Many other abbreviations will be used for reasoning about probability without further comment, such as $pr_i(\varphi) \leq b$ for $\neg(pr_i(\varphi) > b)$, $pr_i(\varphi) \geq b$ for $-pr_i(\varphi) \leq -b$, and $pr_i(\varphi) = b$ for $pr_i(\varphi) \leq b \wedge pr_i(\varphi) \geq b$. Note that we can express simple conditional probabilities such as $pr_i(\varphi|\psi) = 2/3$ by clearing the denominator to get $pr_i(\varphi \wedge \psi) = \frac{2}{3} pr_i(\psi)$.

The operators $K_i$ and $C$ allow us to reason about knowledge and common knowledge, respectively. Formulas such as $a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) > b$ are called *$i$-probability formulas*; they allow us to express a number of notions of interest.[3] Note that by using $i$-probability formulas, we can also describe agent $i$'s beliefs about the expected value of a random variable, provided that the worlds in which the random variable takes on a particular value can be characterized by formulas. For example, suppose that agent 1 wins \$2 if a coin lands heads and loses \$3 if it lands tails. Then the formula $2pr_1(\mathit{heads}) - 3pr_1(\mathit{tails}) > 1$ says that agent 1 believes his expected winnings are at least \$1. This is a much richer language for expressing an agent's beliefs than that used in the relevant literature in economics (for example, [Feinberg 2000]), although the *belief indices* of Bonanno and Nehring [1999] provide a semantic way of expressing yet richer notions.

To assign truth values to formulas in $\mathcal{L}_n^{K,C,pr}$, we need a semantic model. The basic semantic model we use is a *(Kripke) frame (for knowledge and probability for $n$ agents)*. This is a tuple $F = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n)$, where $W$ is a set of possible worlds or states, $\mathcal{K}_1, \ldots, \mathcal{K}_n$ are equivalence relations on $W$, and $\mathcal{PR}_1, \ldots, \mathcal{PR}_n$ are *probability assignments*; $\mathcal{PR}_i$ associates with each world $w$ in $W$ a probability space $\mathcal{PR}_i(w) = (W_{w,i}, \mathcal{X}_{w,i}, \Pr_{w,i})$. Intuitively, $\mathcal{K}_i(w) = \{w' : (w, w') \in \mathcal{K}_i\}$ is the set of worlds that agent $i$ considers possible in world $w$ and $\mathcal{PR}_i(w)$ is the probability space that agent $i$ uses at world $w$. $\mathcal{PR}_i$ must satisfy the following three assumptions.

A1. $W_{w,i} = \mathcal{K}_i(w)$: that is, the sample space at world $w$ consists of the worlds that agent $i$ considers possible at $w$.

A2. If $w' \in \mathcal{K}_i(w)$, then $\mathcal{PR}_i(w) = \mathcal{PR}_i(w')$: at all worlds that agent $i$ considers possible, he uses the same probability space.

A3. $\mathcal{X}_{w,i}$, the set of measurable sets, includes $\mathcal{K}_i(w) \cap \mathcal{K}_j(w')$ for each agent $j$ and world $w' \in \mathcal{K}_i(w)$. Intuitively, each agent's information partitions are measurable.

Apart from minor notational differences, a Kripke frame is the standard model used in the economics literature to capture knowledge and probability (see, for example, [Feinberg 2000]); $\mathcal{K}_i(w)$ is usually called agent $i$'s *information set* at world $w$. In the economics literature, an agent's knowledge is usually characterized by a partition, but this, of course,

[3] Note that the syntax does not allow "mixed" formulas such as $pr_1(\varphi_1) + pr_2(\varphi_2) \geq 1$. There would be no difficulty giving semantics to such formulas, but the results on complete axiomatizations become more difficult if we allow them. Thus, for ease of exposition, they are disallowed here, just as they are in [Fagin and Halpern 1994].

is equivalent to using an equivalence relation.[4] I sometimes describe a relation $\mathcal{K}_i$ by describing the partition it induces.

A frame does not tell us how to connect the language to the worlds. For example, it does not tell us under what circumstances a primitive proposition $p$ is true. To do that, we need an *interpretation*, that is, a function that associates with each primitive proposition an event, namely, the set of worlds where it is true. The traditional way of capturing this in the logic community is by taking $\pi$ to be a function that associates with each world $w$ a truth assignment to the primitive propositions in $\Phi$; i.e., , $\pi(w)(p) \in \{\mathbf{true}, \mathbf{false}\}$ for each primitive proposition $p \in \Phi$ and each world $w \in W$. A *(Kripke) structure (for knowledge and probability for n agents)* is a tuple $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi)$, where $F = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n)$ is a frame and $\pi$ is an interpretation, with the restriction that

A4. $\mathcal{K}_i(w) \cap [\![p]\!]_M \in \mathcal{X}_{w,i}$ for each primitive proposition $p \in \Phi$, where $[\![p]\!]_M = \{w : \pi(w)(p) = \mathbf{true}\}$ is the event that $p$ is true in structure $M$. Intuitively, this makes $[\![p]\!]_M$ a measurable event at every world.

We say that the structure $M$ is *based on* the frame $F$. Note that there are many structures based on a frame $F$, one for each choice of interpretation.

Kripke structures for knowledge and probability were first considered in [Fagin and Halpern 1994], but A1–A4 were not required in the basic framework. These four requirements correspond to the requirements denoted CONS (for *consistency*), SDP (for *state-determined probability*), and MEAS (for *measurability*) in [Fagin and Halpern 1994].

We can now associate an event with each formula in $\mathcal{L}_n^{K,C,pr}$ in a Kripke structure. We write $(M, w) \models \varphi$ if the formula $\varphi$ is true at world $w$ in Kripke structure $M$; generalizing the earlier notation, we denote by $[\![\varphi]\!]_M = \{w : (M, w) \models \varphi\}$ the event that $\varphi$ is true in structure $M$. We proceed by induction on the structure of $\varphi$, assuming that we have given the definition for all subformulas $\varphi'$ of $\varphi$ and that $[\![\varphi']\!]_M \cap \mathcal{K}_i(w) \in \mathcal{X}_{w,i}$; that is, the event corresponding to each formula must be measurable.

$(M, w) \models p$ (for $p \in \Phi$) iff $\pi(w)(p) = \mathbf{true}$

$(M, w) \models \varphi \wedge \varphi'$ iff $(M, w) \models \varphi$ and $(M, w) \models \varphi'$

$(M, w) \models \neg\varphi$ iff $(M, w) \not\models \varphi$

$(M, w) \models K_i\varphi$ iff $(M, w') \models \varphi$ for all $w' \in \mathcal{K}_i(w)$

$(M, w) \models C\varphi$ iff $(M, w) \models E^k\varphi$ for all $k \geq 1$

[4] Bonanno and Nehring [1999] assume only that the relation is serial, Euclidean, and transitive, which is a weaker assumption than it being an equivalence relation, because they want to model belief rather than knowledge. Otherwise, their formalism is the same.

$(M,w) \models a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) > b$
if $a_1 \Pr_{w,i}([\![\varphi_1]\!]_M \cap W_{w,i}) + \cdots + a_m \Pr_{w,i}([\![\varphi_m]\!]_M \cap W_{w,i}) > b$.

The clause for $K_i\varphi$ captures the intuition that $K_i\varphi$ is true at world $w$ if $\varphi$ is true all the worlds the agent considers possible at $w$, namely $\mathcal{K}_i(w)$; the clause for $C\varphi$ enforces the intuition that common knowledge is equivalent to everyone knows, and everyone knows that everyone knows, .... Finally, the clause for $i$-probability formulas captures the intuition that a formula such as $pr_i(\varphi) + 2pr_i(\psi) > 1$ just says that, according to agent $i$, the probability of $\varphi$ plus twice the probability of $\psi$ is at least 1.

It should be clear that this approach of starting with formulas and associating events with them is not so far removed from the more standard approach in the economics literature of defining knowledge in terms of an operators $\mathsf{K}_1, \ldots, \mathsf{K}_n$ on events, where $\mathsf{K}_i(E) = \{w : \mathcal{K}_i(w) \subseteq E\}$. In particular, it is easy to see that $\mathsf{K}_i([\![\varphi]\!]_M) = [\![K_i\varphi]\!]_M$.

For future reference, it is useful to recall a well-known alternative characterization of common knowledge. We say that world $w'$ is *reachable* from $w$ if there exist worlds $w_0, \ldots, w_m$ such that $w = w_0$, $w' = w_m$ and for all $k < m$, there exists an agent $j$ such that $w_{k+1} \in \mathcal{K}_j(w_k)$. Let $\mathcal{C}(w)$ consist of all the worlds reachable from $w$; $\mathcal{C}(w)$ is called the *component* of $w$. The reachability relation is clearly an equivalence relation; thus, $\mathcal{C}$ partitions the set $W$ of worlds into components. A subset $W' \subseteq W$ is a *component of* $W$ if $W' = \mathcal{C}(w)$ for some $w \in W$.

The following lemma is well known (cf. [Fagin, Halpern, Moses, and Vardi 1995, Lemma 2.2.1]).

**Lemma 2.1:** *$(M,w) \models C\varphi$ iff $(M,w') \models \varphi$ for all $w' \in \mathcal{C}(w)$.*

With this background, we can formalize the CPA. It is simply another constraint on probability assignments.

CP. There exists a probability space $(W, \mathcal{X}_W, \Pr_W)$ such that $\Pr_W(W') > 0$ for all components $W'$ of $W$ and for all $i$, $w$, if $\mathcal{PR}_i(w) = (\mathcal{K}_i(w), \mathcal{X}_{w,i}, \Pr_{w,i})$, then $\mathcal{X}_{w,i} \subseteq \mathcal{X}_W$ and, if $\Pr_W(\mathcal{K}_i(w)) > 0$, then $\Pr_{w,i}(U) = \Pr_W(U|\mathcal{K}_i(w))$ for all $U \in \mathcal{X}_{w,i}$. (There are no constraints on $\Pr_{w,i}$ if $\Pr_W(\mathcal{K}_i(w)) = 0$.)

This formalization of the CP is slightly different from the others in the literature. Bonanno and Nehring [1999], Feinberg [2000], and Samet [1998] do not require the condition that the prior gives each component positive probability. However, this condition is necessary for Aumann's theorem to hold; see Example 2.3. Aumann [1976, 1987] starts with the prior and assumes that the posteriors are obtained from the prior by conditioning on the information of the agents; in our language this means that $\Pr_{w,i}$ is obtained from $\Pr_W$ by conditioning on $\mathcal{K}_i(w)$. In [Aumann 1976], Aumann explicitly assumes that $\Pr_W(\mathcal{K}_i(w)) \neq 0$ for all agents $i$ and worlds $w$. (This assumption is also implicitly made in [Aumann 1987].) While the issue of what happens when the prior gives an information set zero probability is a relatively minor technical nuisance, it turns out to play an important role when considering the impact of the CPA. As mentioned in the Introduction, Lipman [1997] shows that there are still some consequences of the CPA even without common knowledge in the language. However, as shown here, the assumption that $\Pr_W(\mathcal{K}_i(w)) \neq 0$ for all $i, w$ is crucial for Lipman's results. With the weaker assumption that only components need get positive probability, there are in fact *no* consequences of the CPA without common knowledge in the language. This is discussed in more detail in Section 3.

The CPA is far from a weak assumption, as the following example shows.

**Example 2.2:** Consider the frame $F_1$ described in Figure 1. There are four worlds;

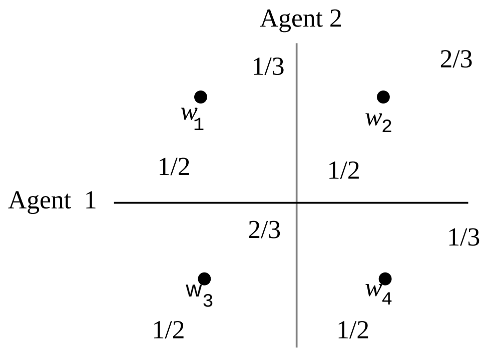


Figure 1: A frame that does not satisfy CP.

the partition induced by $\mathcal{K}_1$ has the equivalence classes $\{w_1, w_2\}$ and $\{w_3, w_4\}$, and the partition induced by $\mathcal{K}_2$ has the equivalence classes $\{w_1, w_3\}$ and $\{w_2, w_4\}$. Whatever two worlds agent 1 considers possible, he ascribes them both probability 1/2. Agent 2, however, thinks that $w_3$ is twice as likely $w_1$ and $w_2$ is twice as likely as $w_4$. It is easy to see that $F_1$ cannot satisfy CP. ∎

Let $\mathcal{F}_n$ consist of all frames for $n$ agents. Let $\mathcal{F}_n^{fin}$ consist of all frames for $n$ agents where the set of worlds is finite and the probability spaces at each point are such that every set is measurable. Let $\mathcal{F}_n^{CP}$ (resp., $\mathcal{F}_n^{CP,fin}$) consist of all frames in $\mathcal{F}_n$ (resp., $\mathcal{F}_n^{fin}$) that satisfy CP. I use $\mathcal{M}_n$, $\mathcal{M}_n^{fin}$, $\mathcal{M}_n^{CP}$, and $\mathcal{M}_n^{CP,fin}$ to denote the corresponding sets of structures.[5]

A formula $\varphi$ is *valid* (resp., *satisfied*) in a Kripke structure $M = (W, \ldots)$ if for all (resp., some) $w \in W$, we have $(M, w) \models \varphi$. A formula is valid (resp., satisfied) in frame $F$ if it is valid in every Kripke structure (resp., satisfied in some Kripke structure) based on $F$. A formula $\varphi$ is valid in a set $\mathcal{M}$ of structures (resp., set $\mathcal{F}$ of frames) if it is valid in every structure $M \in \mathcal{M}$ (resp., every frame $F \in \mathcal{F}$). It is easy to check that a formula is valid in a set $\mathcal{F}$ of frames iff it is valid in the set $\mathcal{M}$ of all structures based on the frames in $\mathcal{F}$.

[5]Technically, these are not sets but *classes*; they are too large to be sets. I ignore the distinction here.

To the extent that there has been consideration of formulas and structures that satisfy them in the economics literature, the focus has been on what has been called the *canonical structure* or *canonical model*. This is essentially a universal structure, which has the property that if a formula is satisfiable at all, it is satisfied at some world in the canonical structure. This was introduced in the economics literature by Aumann [1989], although the basic idea is well known in the modal logic community, and seems to have been introduced independently by Kaplan [1966], Makinson [1966], and Lemmon and/or Scott [Lemmon 1977]. The canonical model has the property that every structure can be embedded in it, in a precise sense.

This may suggest that all we need to consider is the canonical model. While a case for this can be made if we do not have common knowledge in the language, the canonical model construction fails if we add common knowledge to the language, because of the infinitary nature of common knowledge (see [Fagin, Halpern, Moses, and Vardi 1995, Section 3.3]). But even ignoring this issue, there are advantages in considering models other than the canonical model, with its uncountable state space. If we are analyzing a simple game, we are clearly far better off conducting the analysis using a model that reflects that game. In any case, for the results in this paper, it is useful to consider not just the canonical model, but the spaces of structures and frames introduced above.

Aumann's [1976] theorem tells us that for all $a$ and $b$, the formula $\neg C(pr_1(\varphi) = a \wedge pr_2(\varphi) = b)$ is valid in $\mathcal{M}_2^{CP}$ if $a \neq b$: agents cannot agree to disagree in the presence of a common prior. It is, however, not valid in $\mathcal{M}_2$. In fact, if $M_1$ is a structure based on the frame $F_1$ of Example 2.2 where $p$ is true at $w_2$ and $w_4$, then we have $C(pr_1(p) = 1/2 \wedge pr_2(p) = 2/3$ is valid in $M_1$. The requirement that the common prior give each component positive measure is necessary for Aumann's result, as the following example shows:[6]

**Example 2.3:** Consider the structure $M = (W, \mathcal{K}_1, \mathcal{K}_2, \mathcal{PR}_1, \mathcal{PR}_2, \pi)$ described in Figure 2, where $W = \{w_1, w_2, w_3\}$ and the partitions induced by $\mathcal{K}_1$ and $\mathcal{K}_2$ are the same; the equivalence classes are $\{w_1, w_2\}$ and $\{w_3\}$. Agents 1 and 2 disagree about the prob-

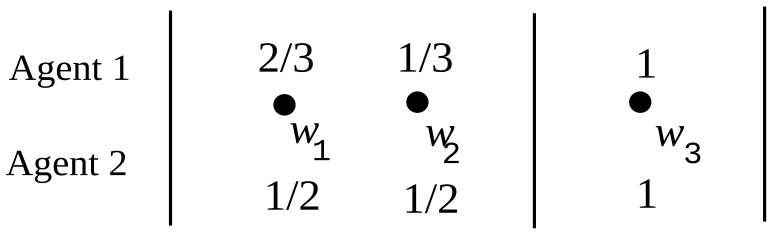


Figure 2: A structure with disagreement in probability in one component.

abilities in the first component. According to agent 1, $w_1$ gets probability 2/3 (so $w_2$ get probability 1/3); according to agent 2, $w_1$ and $w_2$ get equal probability. Suppose $\pi$ is

[6]I think Bart Lipman for bringing this issue—and this example—to my attention.

such that $p$ is true at $w_1$ and false at the other two worlds. Then it is easy to see that $(M, w_1) \models C(pr_1(p) = 2/3 \wedge pr_2(p) = 1/2)$, so we have a disagreement in probability. However, if we drop the requirement in CP that $\Pr_W$ must give each component positive probability, then there would be a common prior in this case: it would assign $w_3$ probability 1 and the other two worlds probability 0. ∎

# 3 Characterizing the CPA

In this section, I consider two approaches to characterizing the CPA. The first is in the spirit of the approaches taken in the economics literature (although it has analogues in the modal logic literature too), while the second involves finding a sound and complete axiomatization. In Section 4, I discuss in more detail what the definitions tell us, in light of the results.

## 3.1 Frame Distinguishability

Frame distinguishability essentially asks whether there is a test (expressed as a set of formulas) that allows us to distinguish the frames satisfying a certain property from ones that do not.

**Definition 3.1:** A set $\mathcal{A}$ of formulas *distinguishes* a collection $\mathcal{F}$ of frames from another collection $\mathcal{F}'$ if (a) every formula in $\mathcal{A}$ is valid in $\mathcal{F}$ and (b) if $F \in \mathcal{F}'$, then some formula in $\mathcal{A}$ is not valid in $F$. ∎

Typically the set $\mathcal{A}$ of formulas in Definition 3.1 consists of all instances of some axiom and the set $\mathcal{F}$ is the set of frames satisfying a certain property (like the CPA). Note that this definition is given in terms of frames, not structures; this is necessary for the technical results to hold.

My results on frame distinguishability parallel those of Feinberg [2000]: we cannot distinguish frames that satisfy the CPA from those that do not, but we can distinguish *finite* frames satisfying the CPA from those that do not. To do this, we might hope to use the axiom characterizing Aumann's "no disagreement" theorem, $\neg C(pr_i(\varphi) = a \wedge pr_j(\varphi) = b)$ for $a \neq b$. While this axiom in valid in $\mathcal{F}_n^{CP}$ (and hence $\mathcal{F}_n^{CP,fin}$), it is not strong enough to distinguish $\mathcal{F}_n^{CP,fin}$ from $\mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$. As Feinberg [2000] points out, there are frames in $\mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$ that satisfy every instance of this axiom, simply because $C(pr_i(\varphi) = a)$ does not hold for any choice of $a$. For example, if we slightly modify the probabilities in the frame $F_1$ of Example 2.2 (for example, changing agent 2's probability so that the probability of $w_3$ is $2/3 + \epsilon$ for some small $\epsilon$ (so that the probability of $w_1$ is $1/3 - \epsilon$), then the only formulas for which agent 2's probabilities are common knowledge are *true* and *false*. Thus, $\neg C(pr_1(\varphi) = a \wedge pr_2(\varphi) = b)$ for $a \neq b$ trivially holds. It follows that we need something stronger than disagreement in probability to characterize the CPA.

Consider the following axiom.

CP$_2$. If $\varphi_1, \ldots, \varphi_m$ are *mutually exclusive* formulas (that is, if $\neg(\varphi_i \wedge \varphi_j)$ is an instance of a propositional tautology for $i \neq j$), then

$$\neg C(a_1 pr_1(\varphi_1) + \cdots + a_m pr_1(\varphi_m) > 0 \wedge a_1 pr_2(\varphi_1) + \cdots + a_m pr_2(\varphi_m) < 0).$$

Notice that CP$_2$ is really an axiom *scheme*; that is, it represents a set of formulas, obtained by considering all instantiations of $a_1, \ldots, a_m$ and $\varphi_1, \ldots, \varphi_m$. CP$_2$ is valid in a structure $M$ if it is not common knowledge that agents 1 and 2 disagree about the expected value of the random variable which takes value $a_j$ on $[\![\varphi_j]\!]_M$, $j = 1, \ldots, m$. Intuitively, CP$_2$ says that it cannot be common knowledge that agents 1 and 2 have a disagreement in expectation. It is easy to see that disagreements in expectation cannot exist if there is a common prior; Feinberg [1995, 2000] and Samet [1998] showed that the converse also holds in finite spaces.[7] The following theorem just recasts their results in this framework; its proof shows why we need to use frames rather than structures in Definition 3.1.

**Theorem 3.2:** *CP$_2$ distinguishes $\mathcal{F}_2^{CP,fin}$ from $\mathcal{F}_2^{fin} - \mathcal{F}_2^{CP,fin}$.*

**Proof:** See the appendix. ∎

The proof of Theorem 3.2 shows that if $F$ is a finite frame that does *not* satisfy the CPA, there is a single instance $\varphi$ of CP$_2$ which is not valid in $F$. Since this is an epistemic formula, the agents both know, at a given world in $F$, that $\varphi$ is not valid in $F$. Intuitively, this means that not only can the modeler distinguish $F$ from structures that satisfy the CPA, so can the agents.

Does this mean that the agents in a given a finite frame $F$ can tell if $F$ satisfies the CPA? Given infinite time and computational resources, yes. They simply check each of the (countably many) instances of CP$_2$ to see if they are all valid in $F$. If all of them are, then $F$ satisfies the CPA; if not, then $F$ does not satisfy the CPA. This approach is obviously not feasible in practice. There is a better approach: given $F$, as shown by Samet [1998], a prior is compatible with agent $i$'s posteriors in $F$ iff it is in the convex hull of the probability measures $\mathcal{PR}_i(w)$, for the worlds $w \in F$. Standard techniques of computational geometry can be used to compute the convex hull efficiently for each agent $i$ and to check if the two convex sets thus obtained are disjoint (see, for example, [Cormen, Leiserson, and Rivest 1990]). $F$ satisfies the CPA iff the convex hulls are not disjoint.

As Feinberg and Samet show, we can extend this characterization of the CPA in the case of two agents to $n > 2$ agents. Consider the following axiom:

[7] Essentially the same result is proved by Bonanno and Nehring [1999], but they were dealing with belief rather than knowledge, so rather than being equivalences, their $\mathcal{K}_i$ relations were serial, Euclidean, and transitive.

CP$_n$. If $\varphi_1, \ldots, \varphi_m$ are mutually exclusive formulas and $a_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, are rational numbers such that $\sum_{i=1}^n a_{ij} = 0$, for $j = 1, \ldots, m$, then

$$\neg C(a_{11}pr_1(\varphi_1) + \cdots + a_{1m}pr_1(\varphi_m) > 0 \wedge \ldots \wedge a_{n1}pr_n(\varphi_1) + \cdots + a_{nm}pr_n(\varphi_m) > 0).$$

It is easy to see that CP$_2$ is equivalent to the axiom that results from CP$_n$ above when $n = 2$, so I take the liberty of abusing notation and denoting both as CP$_2$.

The following result generalizes Theorem 3.2; its proof is omitted, since it follows from results of Feinberg and Samet in the same way as Theorem 3.2.

**Theorem 3.3:** *CP$_n$ distinguishes $\mathcal{F}_n^{CP,fin}$ from $\mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$, for all $n \geq 2$.*

What happens if the set of worlds is not finite? Feinberg shows by example that we can find structures for which there is no common prior, and yet there is no disagreement in expectation (at least, not by bounded random variables). His counterexample can also be used to show that CP$_2$ does not distinguish $\mathcal{F}_2^{CP}$ from $\mathcal{F}_2 - \mathcal{F}_2^{CP}$. I give his counterexample here (actually, a simplification of it, which suffices for my purposes), since it will be needed to prove the next theorem.

**Example 3.4:** Let $F^* = (W, \mathcal{K}_1, \mathcal{K}_2, \mathcal{PR}_1, \mathcal{PR}_2)$ be the frame described in Figure 3: $W = \{w_1, w_2, \ldots\}$; $\mathcal{K}_1$ induces the partition $\{\{w_1\}, \{w_2, w_3\}, \{w_4, w_5\}, \ldots\}$ and $\mathcal{K}_2$ induces the partition $\{\{w_1, w_2\}, \{w_3, w_4\}, \ldots\}$; $\mathcal{PR}_1$ and $\mathcal{PR}_2$ are as described in the figure. As the figure shows, both agents think that all the worlds they consider possible at each

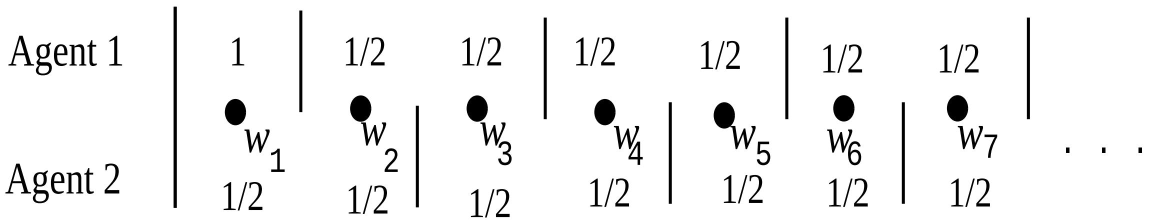


Figure 3: A frame that satisfies CP$_2$, but not the CPA.

world are equally likely (which means that they have probability 1/2 except in the case of agent 1 at worlds $w_1$).

It is easy to see that there is no common prior in $F^*$. For suppose that $\Pr_W$ is such a common prior. To get all the conditional probabilities to work out, we must have $\Pr_W(w_1) = \Pr_W(w_2) = \Pr_W(w_3) = \ldots$, and this is clearly impossible; there is no uniform distribution on a countable set.[8]

[8]There is a common improper prior on $W$, namely, the uniform measure, which assigns measure 1 to each world (and measure $\infty$ to every infinite set, including $W$). We might hope for a characterization of the CPA in infinite spaces using common improper priors. However, it is not hard to show that agents with a common improper (even uniform) prior can disagree about the expectation of a bounded random variable, so the obvious characterization does not work.

On the other hand, suppose that there exist mutually exclusive formulas $\varphi_1, \ldots, \varphi_m$ such that $C_1(a_1pr_1(\varphi_1) + \cdots + a_mpr_1(\varphi_m) > 0 \land a_1pr_2(\varphi_1) + \cdots + a_mpr_2(\varphi_m) < 0)$ is satisfied in some structure $M^*$ based on $\mathcal{F}$. This means that $(M^*, w_j) \models a_1pr_1(\varphi_1)+\cdots+ a_mpr_1(\varphi_m) > 0 \land a_1pr_2(\varphi_1)+\cdots+a_mpr_2(\varphi_m) < 0$ for all $j$. Since $\varphi_i$ and $\varphi_j$ are mutually exclusive, we must have that $[\![\varphi_i]\!]_{M^*} \cap [\![\varphi_j]\!]_{M^*} = \emptyset$ if $i \neq j$. Suppose, without loss of generality, that $\varphi_1, \ldots, \varphi_m$ are ordered so that $\min\{k : w_k \in [\![\varphi_i]\!]_{M^*}\} < \min\{k : w_k \in [\![\varphi_j]\!]_{M^*}\}$ if $i < j$. Note that this means that either $w_1 \notin [\![\varphi_j]\!]_{M^*}$ for all $j$ or $w_1 \in [\![\varphi_1]\!]_{M^*}$. Since $(M^*, w_1) \models a_1pr_1(\varphi_1) + \cdots + a_mpr_1(\varphi_m) > 0$, we must have that $w_1 \in [\![\varphi_1]\!]_{M^*}$ and that $a_1 > 0$. Since $(M^*, w_1) \models a_1pr_2(\varphi_1) + \cdots + a_mpr_2(\varphi_m) < 0$, we must have $w_2 \in [\![\varphi_2]\!]_{M^*}$ and $a_2 < -a_1$. An easy induction now shows that (a) $w_k \in [\![\varphi_k]\!]_{M^*}$, (b) $|a_k| > |a_{k-1}|$ for $k = 2, \ldots, m$, and (c) $a_j$ alternates in sign for $j = 1, \ldots, k$. Now suppose that $m$ is even (a similar argument works if $m$ is odd). In that case, $\mathcal{K}_1(w_m) = \{w_m, w_{m+1}\}$ and $a_m$ is negative. It follows that $(M^*, w_m) \models a_1pr_1(\varphi_1)+\cdots+a_mpr_1(\varphi_m) \leq 0$, contradicting our original assumption. Thus, every instance of $\mathrm{CP}_2$ holds in $M^*$. ∎

Example 3.4 shows that $\mathrm{CP}_2$ does not distinguish $\mathcal{F}_2^{CP}$ from $\mathcal{F}_2 - \mathcal{F}_2^{CP}$, since every instance of $\mathrm{CP}_2$ is valid in $F^* \in \mathcal{F}_2 - \mathcal{F}_2^{CP}$. We might hope to find a richer set of formulas that does allow us to distinguish $\mathcal{F}_2^{CP}$ from $\mathcal{F}_2 - \mathcal{F}_2^{CP}$; the following theorem shows that we cannot.

**Theorem 3.5:** *For all $k \geq 2$, there is no set $\mathcal{A}_k$ of formulas in $\mathcal{L}_k^{K,C,pr}$ that distinguishes $\mathcal{F}_k^{CP}$ from $\mathcal{F}_k - \mathcal{F}_k^{CP}$.*

**Proof:** See the appendix. ∎

The key step in the proof of Theorem 3.5 involves showing that every formula that is valid in $\mathcal{F}_2^{CP}$ is valid in the frame $F^*$ of Example 3.4. Proving this requires a characterization of the formulas that are valid in $\mathcal{F}_2^{CP}$; that is the subject of the next section.

## 3.2 A Sound and Complete Axiomatization of the CPA

The more standard approach to characterizing a notion like the CPA in the logic community is via a sound and complete axiomatization. An *axiom system* AX consists of a collection of *axioms* and *inference rules*. An axiom is a formula, and an inference rule has the form "from $\varphi_1, \ldots, \varphi_k$ infer $\psi$," where $\varphi_1, \ldots, \varphi_k, \psi$ are formulas. Typically (and, in particular, in this paper), the axioms are all instances of *axiom schemes*. Thus, for example, an axiom scheme such as $K_i\varphi \Rightarrow \varphi$ defines an infinite collection of axioms, one for each choice of $\varphi$. A *proof* in AX consists of a sequence of formulas, each of which is either an axiom in AX or follows by an application of an inference rule. A proof is said to be a *proof of the formula* $\varphi$ if the last formula in the proof is $\varphi$. We say $\varphi$ is *provable in AX*, and write $\mathrm{AX} \vdash \varphi$, if there is a proof of $\varphi$ in AX; similarly, we say that $\varphi$ is *consistent with AX* if $\neg\varphi$ is not provable in AX.

An axiom system AX is said to be *sound* for a language $\mathcal{L}$ with respect to a set $\mathcal{M}$ of structures if every formula in $\mathcal{L}$ provable in AX is valid with respect to every structure in $\mathcal{M}$. The system AX is *complete* for $\mathcal{L}$ with respect to $\mathcal{M}$ if every formula in $\mathcal{L}$ that is valid with respect to every structure in $\mathcal{M}$ is provable in AX. We think of AX as characterizing the class $\mathcal{M}$ if it provides a sound and complete axiomatization of that class. Soundness and completeness provide a connection between the *syntactic* notion of provability and the *semantic* notion of validity.[9]

In [Fagin and Halpern 1994], a complete axiomatization is provided for the language $\mathcal{L}_n^{K,pr}$ with respect to $\mathcal{M}_n$. The axiom system can be modularized into five components: axioms for propositional reasoning, axioms for reasoning about knowledge, axioms for reasoning about linear inequalities (since $i$-probability formulas are basically linear inequalities), axioms for reasoning about probability, and axioms for combined reasoning about knowledge and probability, forced by assumptions A1 and A2. Let $\mathrm{AX}_n^{K,pr}$ consist of the following axioms and inference rules, where $i \in \{1, \ldots, n\}$:

**I. Propositional Reasoning**

Prop. All instances of propositional tautologies.

R1. From $\varphi$ and $\varphi \Rightarrow \psi$ infer $\psi$.

**II. Reasoning About Knowledge**

K1. $(K_i\varphi \land K_i(\varphi \Rightarrow \psi)) \Rightarrow K_i\psi$.

K2. $K_i\varphi \Rightarrow \varphi$.

K3. $K_i\varphi \Rightarrow K_iK_i\varphi$.

K4. $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$.

RK. From $\varphi$ infer $K_i\varphi$.

**III. Axioms for reasoning about linear inequalities**

I1. $(a_1pr_i(\varphi_1)+\cdots+a_mpr_i(\varphi_m) \geq b) \Leftrightarrow (a_1pr_i(\varphi_1)+\cdots+a_mpr_i(\varphi_m)+0pr_i(\varphi_{k+1}) \geq b)$.

I2. $(a_1pr_i(\varphi_1) + \cdots + a_mpr_i(\varphi_m) \geq b) \Rightarrow (a_{j_1}pr_i(\varphi_{j_1}) + \cdots + a_{j_m}pr_i(\varphi_{j_m}) \geq b)$, if $j_1, \ldots, j_m$ is a permutation of $1, \ldots, m$.

[9] One could similarly define the notion of a sound and complete axiomatization with respect to a set of frames. Invariably, an axiom system is sound and complete with respect to a set of structures iff it is sound and complete with respect to the corresponding set of frames, since a formula is valid with respect to a frame iff it is valid with respect to all the structures based on it. Thus, for simplicity, I focus only on structures here.

I3. $(a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) \geq b) \wedge (a'_1 pr_i(\varphi_1) + \cdots + a'_m pr_i(\varphi_m) \geq b') \Rightarrow$
$(a_1 + a'_1) pr_i(\varphi_1) + \cdots + (a_m + a'_m) pr_i(\varphi_m) \geq (b + b')$.

I4. $(a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) \geq b) \Leftrightarrow (c_1 pr_i(\varphi_1) + \cdots + c_m pr_i(\varphi_m) \geq db)$ if $d > 0$.

I5. $(a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) \geq b) \vee (a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) \leq b)$.

I6. $(a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) \geq b) \Rightarrow (a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) > b')$ if $b > b'$.

**IV. Reasoning about probabilities**

P1. $pr_i(\varphi) \geq 0$.

P2. $pr_i(true) = 1$.

P3. $pr_i(\varphi \wedge \psi) + pr_i(\varphi \wedge \neg\psi) = pr_i(\varphi)$.

RP. From $\varphi \Leftrightarrow \psi$ infer $pr_i(\varphi) = pr_i(\psi)$.[10]

**V. Reasoning about knowledge and probabilities**

KP1. $K_i(\varphi) \Rightarrow pr_i(\varphi) = 1$.

KP2. $\varphi \Rightarrow K_i\varphi$, if $\varphi$ is an $i$-probability formula or the negation of an $i$-probability formula.

The axioms and rules for propositional reasoning and reasoning about knowledge together give the standard complete axiomatization for knowledge [Fagin, Halpern, Moses, and Vardi 1995]. The axioms and rules for reasoning about inequalities and reasoning about probability are taken from [Fagin, Halpern, and Megiddo 1990], where it is shown that, together with the the propositional component, they give a complete axiomatization for reasoning about probability. Note that P3 essentially captures finite additivity. Although our probability measures are countably additive, there is no axiom for countable additivity. This is essentially because the language is too weak to capture this inherently infinitary property.

What happens when we add common knowledge to the language? It is well known [Fagin, Halpern, Moses, and Vardi 1995; Halpern and Moses 1992] that adding the following to the axioms and rules for knowledge gives a complete axiomatization for the language of knowledge and common knowledge:[11]

**VI. Reasoning About Common Knowledge**

[10]In [Fagin and Halpern 1994], this inference rule is stated as the axiom $pr_i(\varphi) = pr_i(\psi)$ if $\varphi \Leftrightarrow \psi$ is a propositional tautology. We need the more general inference rule to prove, for example, that $pr_i(K_j\varphi) = pr_i(K_jK_j\varphi)$.

[11]In [Fagin, Halpern, Moses, and Vardi 1995; Halpern and Moses 1992] there is also an axiom that says $E\varphi \Leftrightarrow K_1\varphi \wedge \ldots \wedge K_n\varphi$. This axiom is unnecessary here because I have taken $E\varphi$ to be an abbreviation (whose definition is given by the axiom), rather than taking $E$ to be a primitive operator.

C1. $C\varphi \Leftrightarrow E(\varphi \wedge C\varphi)$.

RC. From $\varphi \Rightarrow E(\varphi \wedge \psi)$ infer $\varphi \Rightarrow C\psi$.

Let $\mathrm{AX}_n^{K,C,pr}$ be the system consisting of the axioms and rules of $\mathrm{AX}_n^{K,pr}$ together with C1 and RC.

**Theorem 3.6:** $\mathrm{AX}_n^{K,C,pr}$ *is a sound and complete axiomatization for* $\mathcal{L}_n^{K,C,pr}$ *with respect to both* $\mathcal{M}_n$ *and* $\mathcal{M}_n^{fin}$ *(and hence also with respect to both* $\mathcal{F}_n$ *and* $\mathcal{F}_n^{fin}$*).*

**Proof:** The proof is a straightforward (although lengthy and tedious) combination of the techniques of [Fagin and Halpern 1994] and [Halpern and Moses 1992]. The result is actually proved in the course of proving Theorem 3.8. ∎

It is worth noting that, although common knowledge is, in a sense, an infinitary notion (that is, $C$ can be defined in terms of an infinite conjunction of formulas involving the $K_i$'s), it can be characterized using a finitary axiom and inference rule—C1 and RC.

$\mathrm{AX}_n^{K,C,pr}$ is *not* a sound and complete axiomatization for $\mathcal{L}_n^{K,C,pr}$ with respect to $\mathcal{M}_n^{CP}$ and $\mathcal{M}_n^{CP,fin}$. If we restrict to structures that satisfy the CPA, we get new valid formulas. Indeed, as we have already seen, every instance of $\mathrm{CP}_n$ is valid in $\mathcal{M}_n^{CP}$ (and hence $\mathcal{M}_n^{CP,fin}$). In light of Theorem 3.3, we might hope that if we add $\mathrm{CP}_n$ to $\mathrm{AX}_n^{K,C,pr}$, this would give us a sound and complete axiomatization, at least for $\mathcal{M}_n^{CP,fin}$. Unfortunately, this is not the case.

To understand why, some background is helpful. Samet [1998] shows that, given a frame, the set of possible priors for agent $i$ (i.e., those that can generate the posteriors defined by $\Pr_{w,i}$) forms a closed convex set. If two agents do not have common prior, the corresponding sets of possible priors must be disjoint. He then makes use of a standard result of convex analysis [Rockafellar 1972] to conclude that these sets can be strictly separated by a hyperplane. The separating hyperplane gives the coefficients $a_1, \ldots, a_m$ in $\mathrm{CP}_2$. That is, strict separation by a hyperplane amounts to a disagreement in expectation.

If we consider the set of priors compatible with a given formula, it is no longer necessarily a closed set, so Samet's argument does not quite work. For example, let $\varphi_1$, $\varphi_2$, and $\varphi_3$ be the three mutually exclusive formulas $p \wedge q$, $p \wedge \neg q$, and $\neg p$, respectively. Let $\psi_1$ be $(pr_1(\varphi_1) > pr_1(\varphi_2)) \vee ((pr_1(\varphi_1) = pr_1(\varphi_2)) \wedge (pr_1(\varphi_3) > 1/2))$ and $\psi_2$ be $(pr_2(\varphi_1) < pr_2(\varphi_2)) \vee ((pr_2(\varphi_1) = pr_2(\varphi_2)) \wedge (pr_2(\varphi_3) \leq 1/2))$.[12]

Let $X^i$ consist of all prior probability distributions for agent $i$ that satisfy $\psi_i$, $i = 1, 2$. Then $X^1 = \{(x_1, x_2, x_3) : x_1 > x_2 \text{ or } x_1 = x_2, x_3 > 1/2\}$ (where $x_i$ is the probability of $\varphi_i$, $i = 1, 2, 3$) and $X^2 = \{(x_1, x_2, x_3) : x_1 < x_2 \text{ or } x_1 = x_2, x_3 \leq 1/2\}$. $X^1$ and $X^2$ are easily seen to be disjoint. Thus, there cannot be a common prior. However, although $X_1$ and $X_2$ are convex, they are not closed; it is easy to show that they cannot be strictly

[12] This example was suggested by Dov Samet.

separated by a hyperplane, and we do not have disagreement in expectation in the spirit of $\mathrm{CP}_2$. As a consequence, we get the following theorem.

**Theorem 3.7:** *The formula $\neg C(\psi_1 \wedge \psi_2)$ is valid in $\mathcal{M}_2^{CP}$, but is not provable in the system $\mathrm{AX}_2^{K,C,pr} + CP_2$.*

It follows from Theorem 3.7 that, if we are to obtain a completeness result, even in the case of two agents, we need something stronger than $\mathrm{CP}_2$. The key insight comes from examining the set $X^1$ and $X^2$ in this counterexample again. For all $(x_1, x_2, x_3) \in X^1$ and $(y_1, y_2, y_3) \in X^2$, we have

$$x_1 - x_2 \geq 0 \geq y_1 - y_2 \text{ and } x_1 - x_2 = y_1 - y_2 \Rightarrow (x_3 - x_1 - x_2) > 0 \geq (y_3 - y_1 - y_2).$$

This example generalizes. More precisely, any two disjoint convex (but not necessarily closed) sets $X_{00}$ and $X_{10}$ can be separated in expectation in the following more general sense. Let $\overline{X}_{00}$ and $\overline{X}_{10}$ denote the topological closure of $X_{00}$ and $X_{10}$, respectively . If $\overline{X}_{00}$ and $\overline{X}_{10}$ are disjoint, then they can be strictly separated by a hyperplane. If not, then they can be weakly separated by a hyperplane $H_1$. Let $X_{i1} = X_{i0} \cap H_1$, for $i = 0, 1$. Notice that $X_{i0}$ and $X_{i1}$ are disjoint, convex sets. Either $\overline{X}_{i0}$ and $\overline{X}_{i1}$ are disjoint, so they can be strictly separated by a hyperplane, or they are weakly separated by a hyperplane $H_2$. We can continue in this way to construct convex, disjoint sets $X_{ij}$, $i = 0, 1$ for $j = 0, 1, 2, \ldots$. For sufficiently large $j$, their closures must be disjoint, and hence strictly separable by a hyperplane. This is made precise in Lemma A.3 in the appendix, and generalized to more than two agents in Lemma A.4.

Essentially, this observation tells us that if the CPA holds, then two agents cannot disagree in expectation in this more general sense. As a consequence, the following axiom is valid.

$\mathrm{CP}'_2$. If $\varphi_1, \ldots, \varphi_m$ are mutually exclusive formulas and $i^* \in \{1, 2\}$, then

$$\begin{array}{rl}\neg C( & \sum_{j=1}^m a_{1j} pr_1(\varphi_j) \geq 0 \wedge \sum_{j=1}^m a_{1j} pr_2(\varphi_j) \leq 0 \wedge \\ & ((\sum_{j=1}^m a_{1j} pr_1(\varphi_j) = 0 \wedge \sum_{j=1}^m a_{1j} pr_2(\varphi_j) = 0) \Rightarrow \\ & \quad \ldots \wedge \\ & \quad (\sum_{j=1}^m a_{(h-1)j} pr_1(\varphi_j) \geq 0 \wedge \sum_{j=1}^m a_{(h-1)j} pr_2(\varphi_j) \leq 0 \wedge \\ & \quad ((\sum_{j=1}^m a_{(h-1)j} pr_1(\varphi_j) = 0 \wedge \sum_{j=1}^m a_{(h-1)j} pr_2(\varphi_j) = 0) \Rightarrow \\ & \qquad (\sum_{j=1}^m a_{hj} pr_{i^*}(\varphi_j) > 0 \wedge \sum_{j=1}^m a_{hj} pr_{2-i^*}(\varphi_j) \leq 0))) \ldots)).\end{array}$$

It is easy to see that the formula $\neg C(\psi_1 \wedge \psi_2)$ in Theorem 3.7 follows from $\mathrm{CP}'_2$. Indeed, Theorem 3.8 shows that (in the presence of the other axioms), all formulas valid in $\mathcal{M}_2^{CP}$ follow from $\mathrm{CP}'_2$.

Just as $\mathrm{CP}_2$ generalizes to $\mathrm{CP}_n$ with $n$ agents, so we get the following generalization of $\mathrm{CP}'_2$:

CP$'_n$. If $\varphi_1,\ldots,\varphi_m$ are mutually exclusive formulas, $a_{ikj}$, $i=1,\ldots,n$, $j=1,\ldots,m$, $k=1,\ldots,h$, are rational numbers such that $\sum_{i=1}^n a_{ikj}=0$, for $j=1,\ldots,m$, $k=1,\ldots,h$, and $i^*\in\{1,\ldots,n\}$, then

$$\begin{array}{l}\neg C(\quad \bigwedge_{i=1}^n(\sum_{j=1}^m a_{i1j}pr_i(\varphi_j)\geq 0)\wedge(\bigwedge_{i=1}^n(\sum_{j=1}^m a_{i1j}pr_i(\varphi_j)=0)\Rightarrow\\ \qquad\qquad \ldots\ \wedge\\ \qquad\qquad\quad (\bigwedge_{i=1}^n(\sum_{j=1}^m a_{i(h-1)j}pr_i(\varphi_j)\geq 0)\wedge(\bigwedge_{i=1}^n(\sum_{j=1}^m a_{i(h-1)j}pr_i(\varphi_j)=0)\Rightarrow\\ \qquad\qquad\qquad (\sum_{j=1}^m a_{i^*hj}pr_{i^*}(\varphi_j)>0)\wedge\bigwedge_{i\neq i^*}(\sum_{j=1}^m a_{ihj}pr_i(\varphi_j)\geq 0)))\ldots)).\end{array}$$

Although CP$'_n$ is not as elegant as we might hope, it does the job. Let AX$_n^{CP}$ consist of all the axioms and rules of AX$_n^{K,C,pr}$ together with CP$'_n$.

**Theorem 3.8:** AX$_n^{CP}$ *is a sound and complete axiomatization for* $\mathcal{L}_n^{K,C,pr}$ *with respect to both* $\mathcal{M}_n^{CP}$ *and* $\mathcal{M}_n^{CP,fin}$ *(and hence also with respect to both* $\mathcal{F}_n^{CP}$ *and* $\mathcal{F}_n^{CP,fin}$*).*

**Proof:** See the appendix. ∎

The fact that CP$'_n$ in addition to the other standard axioms suffices to characterize the CPA in finite structures may not be so surprising in light of Theorem 3.3. What may seem somewhat surprising that there is no difference between infinite structure and finite structures in Theorem 3.8. The contrast with Theorems 3.3 and 3.5 is striking; they show that there is a big distinction between finite and infinite frames when we try to characterize the CPA in terms of frame distinguishability. The key point is that, although this language is quite expressive in some ways, it is not expressive enough to distinguish finite structures from infinite ones. This is made precise in the following theorem, which shows that if a formula is satisfiable at all, it is satisfied in a finite structure. The result actually follows from the proof of Theorem 3.8, but I provide in the appendix an alternative proof, using a standard proof technique for proving such results from the modal logic literature known as *filtration.* Note that it follows from the result that finite frames cannot be distinguished from infinite frames (whether or not we assume the CPA) either using frame distinguishability or complete axiomatizations.

**Theorem 3.9:** *A formula in* $\mathcal{L}_n^{K,C,pr}$ *is valid with respect to* $\mathcal{M}_n^{CP}$ *(resp.,* $\mathcal{M}_n$*) iff it is valid with respect to* $\mathcal{M}_n^{CP,fin}$ *(resp.,* $\mathcal{M}_n^{fin}$*).*

**Proof:** See the appendix. ∎

## 3.3 Restricting the Language to $\mathcal{L}_n^{K,pr}$

What happens if we drop the common knowledge operator from the language? As I mentioned earlier, it is shown in [Fagin and Halpern 1994] that AX$_n^{K,pr}$ provides a sound and complete axiomatization for the language $\mathcal{L}_n^{K,pr}$ with respect to $\mathcal{M}_n$. Here, I show that it is also a complete axiomatization for the language $\mathcal{L}_n^{K,pr}$ with respect to $\mathcal{M}_n^{CP}$. That is, there are no new consequences in the languages $\mathcal{L}_n^{K,pr}$ that follow from CP. Moreover, restricting to finite structures does not change anything.

**Theorem 3.10:** $\mathrm{AX}_n^{K,pr}$ *is a sound and complete axiomatization for* $\mathcal{L}_n^{K,pr}$ *with respect to both* $\mathcal{M}_n^{CP}$ *and* $\mathcal{M}_n^{CP,fin}$ *(and hence also with respect to both* $\mathcal{F}_n^{CP}$ *and* $\mathcal{F}_n^{CP,fin}$*).*

**Proof:** See the appendix. ∎

We do no better with frame distinguishability. Of course, we already know from Theorem 3.5 that formulas in $\mathcal{L}_n^{K,C,pr}$ cannot distinguish arbitrary (infinite) frames satisfying the CPA from ones that do not. But Theorem 3.3 tells us that we can distinguish finite frames satisfying the CPA from ones that do not, using formulas that involve common knowledge. It is almost immediate from Theorem 3.10 that this use of common knowledge is necessary. The real point here is that, since we do not have infinite conjunctions in the language, common knowledge is not definable in terms of knowledge. Moreover, finite conjunctions of formulas involving knowledge and probability do not suffice for characterizing the CPA; infinite conjunctions (particularly, the infinite conjunctions defined by the $C$ operator) are necessary.

**Theorem 3.11:** *For all* $n$*, no set* $\mathcal{A}$ *of formulas in* $\mathcal{L}_n^{K,pr}$ *distinguishes* $\mathcal{F}_n^{CP,fin}$ *from* $\mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$.

**Proof:** See the appendix. ∎

These results are qualitatively similar to those proved by Lipman [1997], although there are nontrivial technical differences. Lipman shows that given a structure $M$ not satisfying (his formalization of) the CPA, a world $w$ in $M$, and $N \geq 0$, there is a structure $M_N$ that satisfies the CPA and world $w_N$ in $M_N$ such that $w$ and $w_N$ agree on all formulas of depth at most $N$ (where the *depth* of a formula is the depth of nesting of the modal operators in the language; thus, for example, $K_1p$ has depth 1, $K_1K_2p$ and $K_i(pr_1(p) < 1)$ have depth 2, and $pr_1(pr_2(K_1p) < 1) > 1/2$ has depth 3). On the other hand, in Lipman's framework, there are consequences of the CPA even without common knowledge in the language. In particular, Lipman shows that agents' belief must be *weakly consistent* in the sense that it is impossible for agents to have false beliefs. For example, given his formalization of the CPA, it is impossible for agent 1 to ascribe positive probability to the event that $p$ is true but agent 2 ascribing probability 0 to it. That is, $pr_1(p \wedge pr_2(p) = 0) > 0$ is inconsistent.

Note that this formula is consistent in $\mathcal{M}_2^{CP}$. Consider the structure described in Figure 4. There are two worlds, $w_1$ and $w_2$. Agent 2 cannot distinguish them while agent 1 can (so agent 2's partition has one equivalence class—$\{w_1, w_2\}$—while agent 1's has two—$\{w_1\}$ and $\{w_2\}$). Agent 2 ascribes probability 1 to $w_2$ and probability 0 to $w_1$. Obviously, agent 1's probability at $w_1$ and $w_2$ is determined. If $p$ is true at $w_1$ and

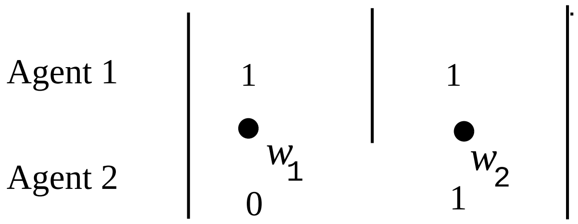


Figure 4: A frame satisfying CP but not CP$^s$.

false at $w_2$, then clearly $pr_1(p \wedge pr_2(p) = 0) = 1$ is true at $w_1$. Moreover, this structure satisfies CP; we can take the prior to agree with agent 2's probability measure. Notice that the common prior assigns $\mathcal{K}_1(w_1)$ probability 0. This is precisely what is disallowed by Lipman.

Lipman's (slightly stronger) version of the CPA can be formalized as follows:

CP$^s$. There exists a probability space $(W, \mathcal{X}_W, \Pr_W)$ such that, for all $i$, $w$, if $\mathcal{PR}_i(w) = (\mathcal{K}_i(w), \mathcal{X}_{w,i}, \Pr_{w,i})$, then $\mathcal{X}_{w,i} \subseteq \mathcal{X}_W$, $\Pr_W(\mathcal{K}_i(w)) > 0$, and $\Pr_{w,i}(U) = \Pr_W(U|\mathcal{K}_i(w))$ for all $U \in \mathcal{X}_{w,i}$.

Let $\mathcal{F}_n^{CP^s}$, $\mathcal{F}_n^{CP^s,fin}$, $\mathcal{M}_n^{CP^s}$, and $\mathcal{M}_n^{CP^s,fin}$ denote the sets of frames (resp., finite frames, structures, finite structures) for $n$ agents that satisfy CP$^s$.

Lipman's results characterizing the consequences of CP$^s$ in the language $\mathcal{L}_n^{K,pr}$ can be viewed, in terms of the framework here, as a combination of results regarding axiomatizations and frame distinguishability. I briefly review his results here (translated to this framework).

Lipman first shows that the language $\mathcal{L}_n^{K,pr}$ cannot distinguish structures satisfying CP$^s$ from those satisfying the weaker *common support* assumption. A structure $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi)$ satisfies the common support assumption if it satisfies the following condition:

CS. For all worlds $w \in W$, agents $i, j$, and events $E \subseteq \mathcal{K}_i(w) \cap \mathcal{K}_j(w)$, if $\Pr_{w,i}(E) = 0$ then $\Pr_{w,j}(E) = 0$.

Let $\mathcal{F}_n^{CS}$ (resp., $\mathcal{F}_n^{CS,fin}$) consist of those frames in $\mathcal{F}_n$ (resp., $\mathcal{F}_n^{fin}$) that satisfy CS. CS is clearly weaker than CP$^s$ (i.e., $\mathcal{F}_n^{CP^s} \subseteq \mathcal{F}_n^{CS}$). Intuitively, it holds as long as $i$ and $j$'s priors assign probability 0 to the same events, and does not require that they assign the same probability to all events. However, Lipman shows that the same formulas in $\mathcal{L}_n^{K,pr}$ are valid in both sets of structures.

**Theorem 3.12:** [Lipman 1997] *For all $\varphi \in \mathcal{L}_n^{K,pr}$, we have $\mathcal{F}_n^{CS} \models \varphi$ iff $\mathcal{F}_n^{CP^s} \models \varphi$.*[13]

[13]Actually, Lipman proves this result only for countable frames. Since, by using the techniques of Theorem 3.9, we can show that a formula is satisfied in $\mathcal{F}_n^{CS}$ (resp., $\mathcal{F}_n^{CP^s}$) iff it is satisfied in $\mathcal{F}_n^{CS,fin}$ (resp., $\mathcal{F}_n^{CP^s,fin}$), the result holds for arbitrary frames as well.

Lipman further shows that weak consistency distinguishes frames satisfying CS from those that do not. More precisely, consider the following axiom:

WC. $pr_i(\varphi \wedge pr_j(\varphi) = 0) = 0$.

**Theorem 3.13:** [Lipman 1997] *WC distinguishes $\mathcal{F}_n^{CS}$ from $\mathcal{F}_n - \mathcal{F}_n^{CS}$.*[14]

We might at first think that it follows from Theorems 3.12 and 3.13 that WC distinguishes frames satisfying CP$^s$ from those that do not, but it is easy to see that this is not true. It is trivial to construct a 2-world frame that satisfies WC but does not satisfy CP$^s$. In fact, I conjecture that there are no formulas in $\mathcal{L}_n^{K,pr}$ that can distinguish frames satisfying CP$^s$ from ones that do not, although I have not proved this.

What happens when we add common knowledge to the language again? I have not examined this situation in detail, although I conjecture that analogues to Theorems 3.3, 3.5, and 3.8 hold. Note, however, that we need something stronger than CP$_n$ together with WC to distinguish finite frames satisfying CP$^s$ from those that do not, as the following example shows.

**Example 3.14:** Consider the structure $M$ described in Figure 5. There are four worlds, $\{w_1, w_2, w_3, w_4\}$. Agent 1's partition is $\{w_1, w_2, w_3\}$ and $\{w_4\}$, while agent 2's is $\{w_1, w_2\}$. $M$ clearly satisfies CS, since both agents agree that $w_3$ gets probability 0. It also satisfies

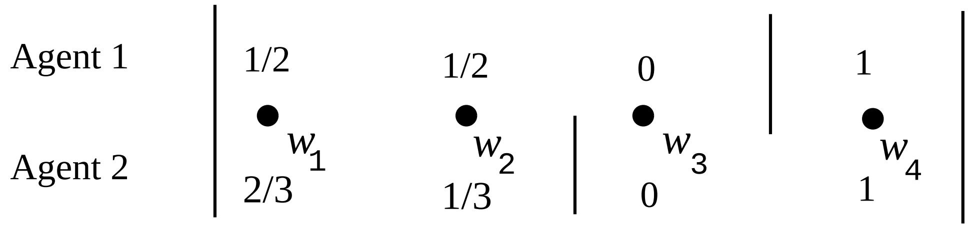


Figure 5: A frame satisfying CS and CP, but not CP$^s$.

CP, since there is a common prior which gives $w_4$ probability 1. However, it does not satisfy CP$^s$: There can be no common prior that gives $\{w_1, w_2\}$ positive probability. Since $M$ satisfies CS and CP, it satisfies all instances of CP$_2$ (in fact, CP$'_2$) and WC. Thus, these formulas cannot distinguish even finite frames satisfying CP$^s$ from ones that do not. ∎

The following strengthening of CP$_2$ is valid in $\mathcal{F}_n^{CP^s}$ (and not in the frame of Example 3.14):

[14]Lipman actually considers structures rather than frames and imposes an additional condition he calls *nonredudancy*, which, roughly speaking, says that any two worlds are distinguishable by some formula. By working at the level of frames, we avoid the need for the nonredundancy condition.

$\mathrm{CP}_2^s$. If $\varphi_1, \ldots, \varphi_m$ are mutually exclusive formulas, then

$$\neg C(a_1 pr_1(\varphi_1) + \cdots + a_m pr_1(\varphi_m) \geq 0 \ \land \ a_1 pr_2(\varphi_1) + \cdots + a_m pr_2(\varphi_m) \leq 0 \\ \land \neg C \neg (a_1 pr_1(\varphi_1) + \cdots + a_m pr_1(\varphi_m) > 0)).$$

Whether this formula (and its obvious generalization to $n$ agents) suffices to distinguish finite frames satisfying $\mathrm{CP}^s$ from those that do not remains open, as does the problem of providing a sound and complete axiomatization in the language $\mathcal{L}_n^{K,C,pr}$ for frames satisfying $\mathrm{CP}^s$.

# 4 Discussion

In this paper, I have considered two different ways of characterizing the CPA—by frame distinguishability and by complete axiomatizations. The notion of frame distinguishability is closer to the notions typically used in the economics community. If $\mathcal{F}$ can be distinguished from $\mathcal{F}'$, that amounts to saying that we have a *test* that can distinguish frames in $\mathcal{F}$ from those in $\mathcal{F}'$. That is analogous to saying that we have a test that distinguishes gold from bronze. Clearly, whether or not we have a distinguishing test depends on how sharp our tools are. In this context, "sharpness of tools" amounts to the expressive power of the language.

Having a test that distinguishes gold from bronze does not mean we have a complete characterization of the properties of gold. But what is a "complete characterization" of gold? Does it suffice to talk about its molecular structure, or do we also have to mention its color and the fact that it glitters in the sun? It should be clear that the notion of "complete characterization" is language dependent. We have a complete characterization of gold in a given language $\mathcal{L}$ if we can describe everything that can be said about gold in $\mathcal{L}$. In general, having a complete characterization in one language tells us nothing about getting a characterization in a richer language. For example, if we have a weak language, it may be easy to find a complete characterization, because there are not many interesting properties of gold in that language. That does not give us any hint of what would constitute a complete characterization in a richer language. (By way of contrast, if we have a distinguishing test in one language, the same test works for any more powerful language.)

We observed this phenomenon with the CPA: in the language $\mathcal{L}_n^{K,pr}$, there is nothing interesting that we can say about the CPA. There are no new axioms over and above the axioms for reasoning about knowledge and probability in all structures (Theorem 3.10). Once we add common knowledge to the language, there are a great many more interesting things that can be said about (structures satisfying) the CPA.

For similar reasons, we may be able to completely characterize a notion without being able to distinguish frames that satisfy it from ones that do not. Again, we saw this phenomenon with the CPA. We can completely characterize the CPA in the language

$\mathcal{L}_n^{K,pr}$ (in a not particularly interesting way, as Theorem 3.10 shows), although $\mathcal{L}_n^{K,pr}$ is of no help in providing tests to distinguish frames satisfying the CPA from ones that do not (Theorem 3.11). If we add common knowledge to the language, then we can distinguish finite frames satisfying the CPA from ones that do not (Theorem 3.3—this is essentially the result proved by Feinberg, Samet, and Bonanno and Nehring), but cannot distinguish infinite frames satisfying the CPA from those that do not (Theorem 3.5); nevertheless, we can completely characterize the properties of (finite or infinite) frames satisfying the CPA (Theorem 3.8).

As I observed in the introduction, the fact that a language not rich enough to provide a distinguishing test can still completely characterize all the properties of a notion of interest is a standard phenomenon in logic. This leads to an obvious open question: is there a natural language that is sufficiently rich to distinguish infinite frames satisfying the CPA from ones that do not (given only their posterior information). Note, however, that such a sufficiently rich language may not be axiomatizable.

In general, the relative merits of one language relative to another is an issue that needs to be debated. For example, I have considered a language with common knowledge here, whereas Feinberg did not consider a language with common knowledge. Is it reasonable to add common knowledge to the language? In general, there is a tradeoff between the expressive power of a language and its complexity. Enriching a language may make it easier to express some notions, but in general makes it harder to decide whether a formula is valid. For example, although there is an algorithm for deciding if a formula is valid whether or not the language includes common knowledge, without common knowledge in the language, the problem is *polynomial-space* complete; with common knowledge, it becomes *exponential-time* complete. (See [Fagin, Halpern, Moses, and Vardi 1995, Chapter 3] for further discussion of these issues.) Another issue to be considered is that of axiomatizations. It may be more difficult to axiomatize a richer language.[15] It is typical in the economics literature to define the common knowledge operator in terms of the knowledge operator, leaving it out of the language. The economics literature is thus implicitly taking infinite conjunctions (actually, infinite intersections, since in economics there are events, not formulas) for granted. However, infinite conjunctions are not expressible in the language $\mathcal{L}_n^{K,pr}$, which allows only finite conjunctions. Logicians have typically avoided infinitary languages; they typically require infinitary axioms and rules of inference and are difficult to deal with computationally. Following tradition, I have used an explicit $C$ operator rather than introducing infinite conjunctions. In the end, perhaps the best argument for including common knowledge here is that the results are so much more elegant with it than without it. Having said that, it should be clear that the decision of what to include in the language is, in general, not one to be taken

[15]Although this is not necessarily the case. For example, it is easier to give a complete axiomatization of the logic of probability if linear combinations of probabilities are allowed than if only comparisons of the form $pr(\varphi) > \alpha$ are allowed. The axiom P3, which captures the fact that the probability of the union of two disjoint sets is the sum of the individual probabilities, cannot be expressed in a logic that does not allow linear combinations.

lightly.

To sum up, I have tried to clarify here two distinct notions of "characterization". As I tried to indicate in the introduction, both have their uses. Frame distinguishability is perhaps the more appropriate notion when a frame is given; axiomatizations are more useful to a modeler who is only give some facts about the frame, rather than a complete description of the frame. In any case, it is important to to be clear about the differences between the notions.

# A Appendix: Proofs

The order of proofs here is different from the order in which the results are stated in the main text, since some of the earlier theorems (particularly Theorem 3.5) depend on some of the later results. The statements are repeated for the convenience of the reader.

**Theorem 3.2:** *$CP_2$ distinguishes $\mathcal{F}_2^{CP,fin}$ from $\mathcal{F}_2^{fin} - \mathcal{F}_2^{CP,fin}$.*

**Proof:** It is easy to see that every instance of $\mathrm{CP}_2$ is valid in every frame of $\mathcal{F}_2^{CP,fin}$; this is essentially Aumann's [1976] argument. I repeat his proof here to make the paper self-contained, since essentially the same idea is used in a number of other proofs.

Suppose $F \in \mathcal{F}_2^{CP,fin}$, $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi)$ is a structure based on $F$, $w \in W$, and $\varphi_1, \ldots, \varphi_m$ are mutually exclusive. Suppose by way of contradiction that

$$(M, w) \models C(a_1 pr_1(\varphi_1) + \cdots + a_m pr_1(\varphi_m) > 0 \wedge a_1 pr_2(\varphi_1) + \cdots + a_m pr_2(\varphi_m) < 0).$$

Sets of the form $\mathcal{K}_1(w')$ partition $\mathcal{C}(w)$. Let $U_1, \ldots, U_k$ be a partition of $\mathcal{C}(w)$ into sets of this form. Since $F \in \mathcal{F}_2^{CP,fin}$, there is a common prior $\Pr_W$ on $W$ as required by CP. Since $(M, w) \models C(a_1 pr_1(\varphi_1) + \cdots + a_m pr_1(\varphi_m)) > 0$, it follows that as long as $\Pr_W(U_j) > 0$, we have that $a_1 \Pr_W([\![\varphi_1]\!]_M \cap U_j) + \cdots + a_m \Pr_W([\![\varphi_m]\!]_M \cap U_j) > 0$, for $j = 1, \ldots, k$. (Of course, $a_1 \Pr_W([\![\varphi_1]\!]_M \cap U_j) + \cdots + a_m \Pr_W([\![\varphi_m]\!]_M \cap U_j) = 0$ if $\Pr_W(U_j) = 0$.) Moreover, we have $\Pr_W([\![\varphi_i]\!]_M \cap \mathcal{C}(w)) = \sum_{j=1}^k \Pr_W([\![\varphi_i]\!]_M \cap U_j)$, for $i = 1, \ldots, m$. Since $\Pr_W(\mathcal{C}(w)) > 0$, we must have $\Pr_W(U_j) > 0$ for some $j$, so $a_1 \Pr_W([\![\varphi_1]\!]_M \cap \mathcal{C}(w)) + \cdots + a_m \Pr_W([\![\varphi_m]\!]_M \cap \mathcal{C}(w)) > 0$. On the other hand, a similar argument using the fact that $(M, w) \models C(a_1 pr_2(\varphi_1) + \cdots + a_m pr_2(\varphi_m)) < 0$, shows that $a_1 \Pr_W([\![\varphi_1]\!]_M \cap \mathcal{C}(w)) + \cdots + a_m \Pr_W([\![\varphi_m]\!]_M \cap \mathcal{C}(w)) < 0$. This gives us the desired contradiction.

For the converse, suppose that $F = (W, \mathcal{K}_1, \mathcal{K}_2, \mathcal{PR}_1, \mathcal{PR}_2) \in \mathcal{F}_2^{fin} - \mathcal{F}_2^{CP,fin}$. Feinberg and Samet show that there is a random variable $X$ such that for each world $w \in W$, agent 1's expectation of $X$ is positive and agent 2's is negative. That is, if $\Pr_{w,i}$ is agent $i$'s probability distribution at world $w$, for $i = 1, 2$, then we have $\sum_{w' \in \mathcal{K}_1(w)} X(w') \Pr_{w,1}(w') > 0$ and $\sum_{w' \in \mathcal{K}_1(w)} X(w') \Pr_{w,2}(w') < 0$ for each world $w \in W$. Without loss of generality, we can assume that $X(w)$ is rational for each world $w \in W$. Suppose $W = \{w_1, \ldots, w_N\}$; let $K = \lceil \log_2(N) \rceil$. Then we can easily write $N$ mutually exclusive propositional formulas

$\varphi_1, \ldots, \varphi_N$ using the primitive propositions $p_1, \ldots, p_K$; these all have the form $q_1 \wedge \ldots \wedge q_K$, where each $q_i$ is either $p_i$ or $\neg p_i$. We can then define a structure $M$ based on $F$ with an interpretation $\pi$ such that $[\![\varphi_j]\!]_M = \{w_j\}$, $j = 1, \ldots, N$. Taking $a_j = X(w_j)$, $j = 1, \ldots, m$, then $C(a_1 pr_1(\varphi_1) + \cdots + a_N pr_1(\varphi_1) > 0 \wedge a_1 pr_2(\varphi_1) + \cdots + a_N pr_2(\varphi_1) < 0)$ is satisfied (in fact, valid) in $M$. ∎

Note that this proof crucially depended on being able to define an interpretation $\pi$ appropriately. This is why frames rather than structures are used in Definition 3.1.

As I said earlier, I defer the proof of Theorem 3.5 until after that of Theorem 3.8, continuing instead with the proof of Theorem 3.7.

**Theorem 3.7:** *The formula $\neg C(\psi_1 \wedge \psi_2)$ is valid in $\mathcal{M}_2^{CP}$, but is not provable in the system $\mathrm{AX}_2^{K,C,pr} + CP_2$.*

**Proof:** First I show that $\neg C(\psi_1 \wedge \psi_2)$ is valid in $\mathcal{M}_2^{CP}$. Suppose that $(M, w) \models C(\psi_1 \wedge \psi_2)$ for some $M \in \mathcal{M}_2^{CP}$. Let $\Pr_W$ be the common prior in $M$, let $W_1$ be the set of worlds in $\mathcal{C}(w)$ where $pr_1(\varphi_1) > pr_1(\varphi_2)$ is satisfied, let $W_2$ be the set of worlds in $\mathcal{C}(w)$ where $pr_2(\varphi_1) < pr_2(\varphi_2)$ is satisfied, and let $W_3$ be the set of worlds in $\mathcal{C}(w)$ where $pr_1(\varphi_1) = pr_1(\varphi_2) \wedge pr_2(\varphi_1) = pr_2(\varphi_2)$ is satisfied. As in the proof of Theorem 3.2, let $U_1, \ldots, U_m$ be a partition of $\mathcal{C}(w)$ into sets the form $\mathcal{K}_i(w')$. Thus, $\Pr_W([\![\varphi_1]\!]_M \cap \mathcal{C}(w)) = \sum_{j=1}^m \Pr_W([\![\varphi_1]\!]_M \cap U_j)$. Since $\psi_1$ is common knowledge at $w$, it follows that $\Pr_W([\![\varphi_1]\!]_M \cap U_j) \geq \Pr_W([\![\varphi_2]\!]_M \cap U_j)$ for $j = 1, \ldots, m$. Thus, $\Pr_W([\![\varphi_1]\!]_M) \geq \Pr_W([\![\varphi_2]\!]_M \cap \mathcal{C}(w))$. Moreover, if $\Pr_W(W_1) > 0$, then $\Pr_W([\![\varphi_1]\!]_M \cap \mathcal{C}(w)) > \Pr_W([\![\varphi_2]\!]_M \cap \mathcal{C}(w))$.

Similarly, since $\psi_2$ is common knowledge at $w$, it follows that $\Pr_W([\![\varphi_1]\!]_M \cap \mathcal{C}(w)) \leq \Pr_W([\![\varphi_2]\!]_M \cap \mathcal{C}(w))$. Moreover, if $\Pr_W(W_2) > 0$, then $\Pr_W([\![\varphi_1]\!]_M \cap \mathcal{C}(w)) < \Pr_W([\![\varphi_2]\!]_M \cap \mathcal{C}(w))$. Thus, we must have $\Pr_W(W_1) = \Pr_W(W_2) = 0$. It follows that $\Pr_W(W_3) = \Pr_W(\mathcal{C}(w)) > 0$. (Recall that CP requires $\Pr_W$ to give every component positive measure.) But it follows from $C\psi_1$ that $\Pr_W(W_3 \cap [\![\varphi_3]\!]_M) > 1/2 \Pr_W(W_3)$ and from $C\psi_2$ that $\Pr_W(W_3 \cap [\![\varphi_3]\!]_M) \leq 1/2 \Pr_W(W_3)$. This gives us the desired contradiction.

I next show that $\neg C(\psi_1 \wedge \psi_2)$ is not provable $\mathrm{AX}_2^{K,C,pr} + \mathrm{CP}_2$. As usual, let $\Delta^m$ denote the $(m-1)$-dimensional simplex, that is $\{(x_1, \ldots, x_m) \in \mathbb{R}^m : x_1 + \cdots + x_m = 1,\ x_i \geq 0, i = 1, \ldots, m\}$. Let $W = \{w_1, w_2, w_3\}$ and, for $(x_1, x_2, x_3) \in \Delta^3$, define the probability measure $\Pr^{(x_1,x_2,x_3)}$ on $W$ by taking $\Pr^{(x_1,x_2,x_3)}(w_i) = x_i$ for $i = 1, 2, 3$. Consider structures of the form $M^{(x_1,x_2,x_3)} = (W, \mathcal{K}_1, \mathcal{K}_2, \mathcal{PR}_1^{(x_1,x_2,x_3)}, \mathcal{PR}_2, \pi)$, where $\mathcal{K}_1$ and $\mathcal{K}_2$ are the universal relations on $W$ (that is, both agents have only one cell in their partition, consisting of all of $W$), $\Pr_{w,1}^{(x_1,x_2,x_3)} = \Pr^{(x_1,x_2,x_3)}$ and $\Pr_{w,2} = \Pr^{(1/4,1/4,1/2)}$ for all $w \in W$, and $\pi$ is such that $p$ is true at $w_1$ and $w_2$ and false at $w_3$, while $q$ is true at $w_1$ and $w_3$ and false at $w_2$. Let $\mathcal{M} = \{M^{\vec{x}} : M^{\vec{x}} \models C(\psi_1 \wedge \psi_2\}$. Thus, $\mathcal{M} = \{M^{(x_1,x_2,x_3)} : x_1 > x_2 \text{ or } x_1 = x_2 < 1/4\}$.

Note that $M^{(1/4,1/4,1/2)} \in \mathcal{M}_2^{CP,fin}$. Thus, $M^{(1/4,1/4,1/2)}$ satisfies every instance of $\mathrm{CP}_2$. Clearly $M^{(1/4,1/4,1/2)} \notin \mathcal{M}$. However, it is easy to see that for each $\epsilon > 0$, there exists a tuple $\vec{x} \in \Delta^3$ such that $|\vec{x} - (1/4, 1/4, 1/2)| < \epsilon$ and $M^{\vec{x}} \in \mathcal{M}$, where $|\vec{x} - \vec{y}| =$

$\max_{i\in\{1,2,3\}} |x_i - y_i|$ for $\vec{x}, \vec{y} \in \Delta^3$. Thus, $M^{(1/4,1/4,1/2)}$ is in the closure of $\mathcal{M}$, in an appropriate topology.

**Claim A.1:** For every instance $\sigma$ of $\mathrm{CP}_2$, there exists $\epsilon_\sigma > 0$ such that $M^{\vec{x}} \models \sigma$ for all $\vec{x}$ such that $|\vec{x} - (1/4, 1/4, 1/2)| < \epsilon_\sigma$.

I shall prove Claim A.1 shortly; first I show why it suffices to prove the theorem. Suppose there is a proof of $\neg C(\psi_1 \wedge \psi_2)$ in the system $\mathrm{AX}^{K,C,pr} + \mathrm{CP}_2$. By definition, this means there is a sequence of formulas $\varphi_1, \ldots, \varphi_m$, each of which is an axiom of $\mathrm{AX}^{K,C,pr} + \mathrm{CP}_2$ or follows from previous steps by an application of an inference rule, such that $\varphi_m = \neg C(\psi_1 \wedge \psi_2)$. Let $\epsilon$ be the minimum of $\epsilon_\sigma$ for all instances $\sigma$ of $\mathrm{CP}_2$ that arise in $\varphi_1, \ldots, \varphi_m$. It is easy to see that each formula $\varphi_i$, $i = 1, \ldots, m$ is valid in $M^{\vec{x}}$ if $|\vec{x} - (1/4, 1/4, 1/2)| < \epsilon$: Each formula $\varphi_j$ that is an instance of an axiom other than $\mathrm{CP}_2$ is valid in every structure; if $\varphi_j$ is an instance of $\mathrm{CP}_2$, this follows from Claim A.1 and the choice of $\epsilon$; and if $\varphi_j$ follows from previous formulas by application of an inference rule, this follows since inference rules preserve validity if the formulas they are being applied to are valid. In particular, $\varphi_m = \neg C(\psi_1 \wedge \psi_2)$ is valid in every structure $M^{\vec{x}}$ such that $|\vec{x} - (1/4, 1/4, 1/2)| < \epsilon$. But this contradicts the fact that $C(\psi_1 \wedge \psi_2)$ is valid in every structure in $\mathcal{M}$, by choice of $\mathcal{M}$.

Thus, it remains to prove Claim A.1. This claim may seem obvious. Consider any instance $\sigma = \neg C(a_1 pr_1(\sigma_1) + \cdots + a_m pr_1(\sigma_m) > 0 \wedge a_1 pr_2(\sigma_1) + \cdots + a_m pr_2(\sigma_m) < 0)$ of $\mathrm{CP}_2$. Since $\sigma$ is valid in $M^{(1/4,1/4,1/2)}$, it seems clear that making only slight changes to agent 1's probability shouldn't affect the validity of $\sigma$. This intuition is in fact true; however, there is one subtlety involved in proving it: We must show that the event corresponding to $\sigma_j$ does not change as a result of small changes in the probability. This is the content of the next claim, which says that we can partition the set of structure $M^{\vec{x}}$ into convex regions over which the event corresponding to a given formula is constant.

**Claim A.2:** For all formulas $\varphi \in \mathcal{L}_2^{K,C,pr}$, there is a partition $\Pi_\varphi$ of $\Delta^3$ into a finite number of convex sets (defined by linear inequalities) such that for all $D \in \Pi_\varphi$ and all subformulas $\psi$ of $\varphi$, there is a subset $W_\psi^D$ of $W$ such that $[\![\psi]\!]_{M^{\vec{x}}} = W_\psi^D$ for all $\vec{x} \in D$; that is, for each $D \in \Pi_\varphi$, the set of worlds where $\psi$ is true in $M^{\vec{x}}$ is the same for all $\vec{x} \in D$.

**Proof:** This result follows easily by induction on the structure of $\varphi$. If $\varphi$ is a primitive proposition, then we can take $\Pi_\varphi = \{\Delta^3\}$; for example, $[\![p]\!]_{M^{\vec{x}}} = \{w_1, w_2\}$ for all $\vec{x} \in \Delta^3$. We can take $\Pi_{\neg\varphi} = \Pi_{K_i\varphi} = \Pi_{C\varphi} = \Pi_\varphi$ and take $\Pi_{\varphi\wedge\psi} = \{A \cap B : A \in \Pi_\varphi, B \in \Pi_\psi\}$. Finally, suppose $\varphi$ has the form $a_1 pr_i(\sigma_1) + \cdots + a_m pr_i(\sigma_m) > b$. If $i = 2$, it is easy to see that we can take $\Pi_\varphi = \Pi_{\sigma_1 \wedge \ldots \wedge \sigma_m}$. If $i = 1$, consider a cell $D$ in $\Pi_{\sigma_1 \wedge \ldots \wedge \sigma_m}$. Let $W_j^D = [\![\sigma_j]\!]_{M^{\vec{x}}}$ for $\vec{x} \in D$, $j = 1, \ldots, m$, let $D^{>} = \{\vec{x} \in D : a_1 \Pr^{\vec{x}}(W_1^D) + \cdots + a_m \Pr^{\vec{x}}(W_m^D) > b\}$, and let $\Pi_\varphi = \{D^{>}, D - D^{>} : D \in \Pi_{\sigma_1 \wedge \ldots \wedge \sigma_m}\}$. It is easy to check that this partition has the desired properties. ∎

We can now prove Claim A.1. Suppose, by way of contradiction, that it does not hold for some instance $\sigma = \neg C(a_1 pr_1(\sigma_1) + \cdots + a_m pr_1(\sigma_m) > 0 \wedge a_1 pr_2(\sigma_1) + \cdots + a_m pr_2(\sigma_m) < 0)$ of $\mathrm{CP}_2$. Since $M^{(1/4,1/4,1/2)}$ satisfies every instance of $\mathrm{CP}_2$, it must be the case that $M^{(1/4,1/4,1/2)} \models \sigma$. Since a formula of the form $C\psi$ is true at either all worlds in $W$ or none of them, the set $W_\sigma^D$ must be either $\emptyset$ or $W$ for each $D \in \Pi_\sigma$. Since Claim A.1 is assumed to fail for $\sigma$, there must be some set $D \in \Pi_\sigma$ such that $(1/4, 1/4, 1/2) \in \overline{D}$ and $[\![\sigma]\!]_{M^{\vec{x}}} = \emptyset$ for all $\vec{x} \in D$ (i.e., $W_\sigma^D = \emptyset$). Since $M^{\vec{x}} \models a_1 pr_1(\sigma_1) + \cdots + a_m pr_1(\sigma_m) > 0$ for all $\vec{x} \in D$, we have $a_1 \Pr^{\vec{x}}(W_{\sigma_1}^D) + \cdots + a_m \Pr^{\vec{x}}(W_{\sigma_m}^D) > 0$ for all $\vec{x} \in D$. On the other hand, since $M^{\vec{x}} \models a_1 pr_2(\sigma_1) + \cdots + a_m pr_2(\sigma_m) < 0$ for $\vec{x} \in D$, we have $a_1 \Pr^{(1/4,1/4,1/2)}(W_{\sigma_1}^D) + \cdots + a_m \Pr^{(1/4,1/4,1/2)}(W_{\sigma_m}^D) < 0$. Since $(1/4, 1/4, 1/2) \in \overline{D}$, this gives us the desired contradiction, proving Claim A.1 and the theorem. ∎

Before proving Theorem 3.8, we need a technical lemma regarding separation of convex sets. It is well known that two closed convex subsets of $\Delta^m$ can be separated by a hyperplane. (See Rockafellar [1972] for this and all the other standard facts and definitions from convex analysis used below.) As Samet [1998] observes, we can take the separating point to be 0. That is, if $X$ and $Y$ are two closed convex subsets of $\Delta^m$, there exists a vector $\vec{a} \in I\!R^m$ such that for all $\vec{x} \in X$ and $\vec{y} \in Y$, we have $\vec{a} \cdot \vec{x} > 0 > \vec{a} \cdot \vec{y}$, where $\cdot$ denotes inner product. The following lemma generalizes this result to the case where $X$ and $Y$ are not necessarily closed. Roughly speaking, it says that either two convex subsets of $\Delta^m$ can be separated by a hyperplane $H_1$, or they can be weakly separated by $H_1$ (where weak separation here means that both sets may intersect $H_1$) and, if we consider the intersection of the two sets with $H_1$, these sets can be separated by a hyperplane $H_2$, or they can be weakly separated by $H_2$ and, if we consider the intersection the intersection of the sets with $H_2$, ...; moreover, this process stops after a finite number of sets in such a way that the resulting sets can be (strongly) separated by a hyperplane.

**Lemma A.3:** *Suppose that $X^1$ and $X^2$ are disjoint, convex (but not necessarily closed) subsets of $\Delta^m$. Then, for some $i^* \in \{1, 2\}$, $h \leq m-1$, and vectors $\vec{a}_1, \ldots, \vec{a}_h$, for all $\vec{y}^1 \in X^{i^*}$ and $\vec{y}^2 \in X^{2-i^*}$, we have*

$$\begin{array}{l} \vec{a}_1 \cdot \vec{y}^1 \geq 0 \wedge \vec{a}_1 \cdot \vec{y}^2 \leq 0 \wedge (\vec{a}_1 \cdot \vec{y}^1 = 0 \wedge \vec{a}_1 \cdot \vec{y}^2 = 0 \Rightarrow \\ \quad \ldots \wedge \\ \quad (\vec{a}_{h-1} \cdot \vec{y}^1 \geq 0 \wedge \vec{a}_{h-1} \cdot \vec{y}^2 \leq 0 \ \wedge (\vec{a}_{h-1} \cdot \vec{y}^1 = 0 \wedge \vec{a}_{h-1} \cdot \vec{y}^2 = 0 \Rightarrow \\ \qquad \vec{a}_h \cdot \vec{y}^1 > 0 \wedge \vec{a}_h \cdot \vec{y}^2 \leq 0)) \ldots). \end{array} \tag{1}$$

*Moreover, if $X^1$ and $X^2$ are defined by a finite collection of linear equations and inequalities with rational coefficients, the vectors $\vec{a}_1, \ldots, \vec{a}_h$ can all be taken to be rational.*

**Proof:** The proof proceeds by induction on the maximum dimension of $X^1$ and $X^2$. If it is 1, then both $X^1$ and $X^2$ are lines. It is well known that in this case there exists a vector $\vec{a}$, $i^* \in \{1, 2\}$, and constant $c$ such that $\vec{a} \cdot \vec{y}^1 > c \geq \vec{a} \cdot \vec{y}^2$ for all $\vec{y}^1 \in X^{i^*}$ and $\vec{y}^2 \in X^{2-i^*}$. Moreover, if $X^1$ and $X^2$ are defined by linear equations and inequalities with rational coefficients, $c$ and all the coordinates of $\vec{a}$ can be taken to be rational. Finally, as Samet observes, since $\vec{y}^1, \vec{y}^2 \in \Delta^m$, if we take $\vec{a}'$ to be the result of subtracting $c$ from all the coordinates of $\vec{a}$, we have $\vec{a}' \cdot \vec{y}^1 > 0 \geq \vec{a}' \cdot \vec{y}^2$.

Now suppose by induction the result holds for sets of maximum dimension $k$, and suppose that in fact the maximum dimension of $X^1$ and $X^2$ is $k+1$. Again, by standard results, we know that there exists a vector $\vec{a}_1$ such that $\vec{a}_1 \cdot \vec{x}^1 \geq c \geq \vec{a}_1 \cdot \vec{y}$ for all $\vec{x}^1 \in X^1$ and $\vec{x}^2 \in X$. As above, we can assume without loss of generality that $c = 0$ and, if $X^1$ and $X^2$ are defined by linear equations and inequalities with rational coefficients, that the coordinates of $\vec{a}_1$ are rational. If at least one of the inequalities above is strict, we are done (replacing $\vec{a}_1$ by $-\vec{a}_1$ if necessary). If not, let $Y^1 = \{\vec{x}^1 \in X^1 : \vec{a}_1 \cdot \vec{x}^1 = 0\}$ and let $Y^2 = \{\vec{x}^2 \in X^2 : \vec{a}_1 \cdot \vec{x}^2 = 0\}$. $Y^1$ and $Y^2$ are disjoint convex sets of dimension at most $k$. Moreover, if $X^1$ and $X^2$ are defined by a finite number of linear equations with rational coefficients, then so are $Y^1$ and $Y^2$. The result now follows from the induction hypothesis. ∎

The expression in (1) is actually an expression in a formal language for reasoning about linear inequalities introduced in [Fagin, Halpern, and Megiddo 1990]. Since this will come up again later, it is worth making it a little more precise now. Suppose that we start with a fixed infinite set of variables. A *basic inequality formula* is one of the form $a_1x_1 + \cdots + a_kx_k > b$, where $a_1, \ldots, a_k, b$ are rational numbers and $x_1, \ldots, x_k$ are variables. For example, $2x_1 - x_2 > 3$ is a basic inequality formula. An *inequality formula* is a Boolean combination of basic inequality formulas. An *assignment (to variables)* is a function $A$ that assigns a real number to every variable. We define

$$A \models a_1x_1 + \cdots + a_kx_k > b \text{ iff } a_1A(x_1) + \cdots + a_kA(x_k) > b.$$

We then extend $\models$ to arbitrary inequality formulas, which are just Boolean combinations of basic inequality formulas, in the obvious way, namely

$$\begin{array}{lll} A \models \neg f & \text{iff} & A \not\models f \\ A \models f \wedge g & \text{iff} & A \models f \text{ and } A \models g. \end{array}$$

As usual we say an inequality formula $f$ is *valid* if $A \models f$ for all $A$ that are assignments to variables. If $f$ is a valid inequality and we obtain a formula $\sigma$ in $\mathcal{L}_n^{K,pr}$ by replacing the variables in $f$ by probability terms of the form $pr_i(\varphi)$ (replacing each occurrence of a variable $x_j$ by the same probability term), then the resulting formula is clearly also valid in $\mathcal{M}_n$. Moreover, as shown in [Fagin, Halpern, and Megiddo 1990], it is provable using just the axioms I1–I6 for reasoning about linear inequalities and propositional reasoning (Prop and R1). This fact will be used in the proof of Theorem 3.8.

Continuing with the main line of our proof, Samet [1998] shows how to generalize the special case of Lemma A.3 where $X^1$ and $X^2$ are closed convex sets to the case of $n$ sets. The following result is the analogous generalization here. I omit the proof, since it proceeds in much the same spirit as Samet's, using the ideas of Lemma A.3.

**Lemma A.4:** *Suppose that $X^1, \ldots, X^n$ are convex (but not necessarily closed) subsets of $\Delta^m$ such that $\cap_{i=1}^n X^i = \emptyset$. Then, for some $h \le m-1$, $i^* \in \{1,\ldots,n\}$, and vectors $\vec{a}_{ik}$, $i = 1,\ldots,n$, $k = 1,\ldots,h$, such that $\sum_{i=1}^n \vec{a}_{ik} = \vec{0}$, for $k = 1,\ldots,h$, for all $\vec{x}^i \in X^i$, $i = 1,\ldots,n$, we have*

$$\begin{array}{l} \bigwedge_{i=1}^n \vec{a}_{i1} \cdot \vec{x}^i \ge 0 \wedge (\bigwedge_{i=1}^n \vec{a}_{i1} \cdot \vec{x}^i = 0 \Rightarrow \\ \quad \ldots \; \wedge \\ \quad (\bigwedge_{i=1}^n \vec{a}_{i(h-1)} \cdot \vec{x}^i \ge 0 \wedge (\bigwedge_{i=1}^n (\vec{a}_{i(h-1)} \cdot \vec{x}^i = 0 \Rightarrow \\ \qquad (\vec{a}_{i^*h} \cdot \vec{x}^{i^*} > 0 \wedge \bigwedge_{i \ne i^*} \vec{a}_{i(h-1)} \cdot \vec{x}^i \ge 0))) \ldots). \end{array} \tag{2}$$

*Moreover, if the sets $X^i$, $i = 1,\ldots,n$, are each defined by a finite collection of linear equations and inequalities with rational coefficients, the coordinates of $\vec{a}_{ik}$ can all be taken to be rational.*

We are now ready to prove Theorem 3.8.

**Theorem 3.8:** $\mathrm{AX}_n^{CP}$ *is a sound and complete axiomatization for $\mathcal{L}_n^{K,C,pr}$ with respect to both $\mathcal{M}_n^{CP}$ and $\mathcal{M}_n^{CP,fin}$ (and hence also both $\mathcal{F}_n^{CP}$ and $\mathcal{F}_n^{CP,fin}$).*

**Proof:** The completeness proof follows closely along the lines of the completeness proof given in [Fagin and Halpern 1994] (which in turn uses a combination of techniques from [Fagin, Halpern, and Megiddo 1990; Halpern and Moses 1992; Makinson 1966]), which shows that $\mathrm{AX}_n^{K,pr}$ is a sound and complete axiomatization for $\mathcal{L}_n^{K,pr}$ with respect to $\mathcal{M}_n$. The added complications in this proof are dealing with the fact we have common knowledge in the language and with CP. The techniques for dealing with common knowledge are well known [Fagin, Halpern, Moses, and Vardi 1995; Halpern and Moses 1992], so I focus here on dealing with CP.

We want to show that if $\varphi \in \mathcal{L}_n^{K,C,pr}$ is valid with respect to $\mathcal{M}_n^{CP,fin}$, then it is provable in $\mathrm{AX}_n^{CP}$. Equivalently, we must show if $\varphi$ is consistent with $\mathrm{AX}_n^{CP}$, then $\varphi$ is satisfied in some structure in $\mathcal{M}_n^{CP,fin}$. The proof actually shows how to construct such a structure.

Let $Sub(\varphi)$ be the set of all subformulas of $\varphi$ and let $Sub^+(\varphi)$ be the set of subformulas of $\varphi$ and their negations.

If $w$ is a finite set of formulas, let $\varphi_w$ be the conjunction of the formulas in $w$. The set $w$ is a *maximal* consistent subset of $Sub^+(\varphi)$ if $w \subseteq Sub^+(\varphi)$, $\varphi_w$ is consistent with $\mathcal{M}_n^{CP,fin}$, and for every subformula $\psi$ of $\varphi$, either $\psi$ or $\neg\psi$ is in $w$. (Note that $w$ cannot include both $\psi$ and $\neg\psi$, for then $\varphi_w$ would not be consistent.) Following Makinson [1966] (see also [Fagin, Halpern, Moses, and Vardi 1995; Halpern and Moses 1992]), we first construct a Kripke structure for knowledge (but not probability) $(W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi)$ as follows: we take $W$, the set of worlds, to consist of all maximal consistent subsets of $Sub^+(\varphi)$. If $u$ and $v$ are worlds, then $(u,v) \in \mathcal{K}_i$ precisely if $u$ and $v$ contain the same formulas of the form $C\psi$, $K_i\psi$, and $pr_i(\psi_1) + \cdots + pr_i(\psi_k) > b$. We define $\pi$ so that for a primitive proposition $p$, we have $\pi(s)(p) = \mathbf{true}$ iff $p$ is one of the formulas in the set $s$. Our goal is to define a probability assignments $\mathcal{PR}_1, \ldots, \mathcal{PR}_n$ such that $M = (S, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi) \in \mathcal{M}_n^{CP}$ (in fact, it will be in $\mathcal{M}_n^{CP,fin}$, since $W$ is clearly finite) and, moreover, for every world $w \in W$ and every formula $\psi \in Sub^+(\varphi)$, we have

$$(M, w) \models \psi \text{ iff } \psi \in w. \qquad (*)$$

Since $\varphi$ is consistent, we must have $\varphi \in w$ for some $w \in W$. Hence, once we show that there exist $\mathcal{PR}_1, \ldots, \mathcal{PR}_n$ such that $M$ satisfies (*), we are done.

It is easy to see that the formulas $\varphi_w$ are mutually exclusive for $w \in W$. Moreover, we can show that $\mathrm{AX}_n^{CP} \vdash \psi \Leftrightarrow \vee_{\{v \in W \mid \psi \in v\}} \varphi_v$, for all $\psi \in Sub^+(\varphi)$. Using these observations, we can show, using P1–P3 and RP (and propositional reasoning, i.e., Prop and R1) that $\mathrm{AX}_n^{CP} \vdash pr_i(\psi) = \sum_{\{v \in W \mid \psi \in v\}} pr_i(\varphi_v)$ (cf. [Fagin, Halpern, and Megiddo 1990, Lemma 2.3]). Using this fact together with I1 and I3, we can show that an $i$-probability formula $\psi \in Sub^+(\varphi)$ is provably equivalent to a formula of the form $\sum_{v \in W} c_v \mu_i(\varphi_v) > b$, for some appropriate coefficients $c_v$.

For each world $u$ and agent $i$, we associate a set $L_{ui}$ of linear equalities and inequalities over variables of the form $x_{iv}$, for $v \in \mathcal{K}_i(u)$. We can think of $x_{iv}$ as representing $\mathrm{Pr}_{u,i}(v)$, i.e., the probability of world $v$ under agent $i$'s probability distribution at world $u$. We have one inequality in $L_{ui}$ corresponding to every $i$-probability formula $\psi$ in $Sub^+(\varphi)$. Assume that $\psi$ is equivalent to $\sum_{v \in W} c_v pr_i(\varphi_v) > b$. If $\psi \in u$, then the corresponding inequality is

$$\sum_{v \in \mathcal{K}_i(u)} c_v x_{iv} > b.$$

(Note that there are no terms with coefficient $x_{iv}$ for $v \notin \mathcal{K}_i(u)$. Intuitively, this is because $\mathrm{Pr}_{u,i}$ is a probability measure on $\mathcal{K}_i(u)$, so we can treat $\mathrm{Pr}_{u,i}(v)$ as 0 for $v \notin \mathcal{K}_i(u)$.) Similarly, if $\neg\psi \in u$, then the corresponding inequality is

$$\sum_{v \in \mathcal{K}_i(u)} c_v x_{iv} \le b.$$

Finally, we add to $L_{ui}$ the equality

$$\sum_{v \in \mathcal{K}_i(u)} x_{iv} = 1.$$

Note that if $u' \in \mathcal{K}_i(u)$, then $L_{u'i} = L_{ui}$, since the set of $i$-probability formulas in $u$ and $u'$ is the same.

As shown in [Fagin, Halpern, and Megiddo 1990, Theorem 2.2], since $\varphi_u$ is consistent, there exists a probability measure $\mathrm{Pr}^*_{u,i}$ satisfying (the equation and inequalities in) $L_{ui}$ (taking $x_{iv} = \mathrm{Pr}^*_{u,i}(v)$). If we were not concerned with CP, then we could just define $\mathcal{PR}_i$ so that $\mathrm{Pr}_{u,i} = \mathrm{Pr}^*_{u,i}$. Since $L_{u'i} = L_{ui}$ for $u' \in \mathcal{K}_i(u)$, we could also assume without loss of generality that $\mathrm{Pr}_{u,i} = \mathrm{Pr}_{u',i}$ for $u' \in \mathcal{K}_i(u)$. The techniques of [Fagin and Halpern 1994] (and of [Halpern and Moses 1992] in the case of common knowledge) then show that (*) holds for the resulting Kripke structure. This suffices to prove Theorem 3.6. However, we must work harder to complete the proof of Theorem 3.8, since the probability assignments do not necessarily satisfy CP.

Note we can identify a probability measure on $W$ with an element of $\Delta^{|W|}$. We can thus use the tuple $\langle x_v : v \in W\rangle$ to denote a generic probability measure on $W$. We say that a probability measure $\Pr = \langle x_v : v \in W\rangle$ is *compatible* with $L_{ui}$ if $\Pr(\cdot|\mathcal{K}_i(u))$ satisfies $L_{ui}$ as long as $\Pr(\mathcal{K}_i(u)) \neq 0$. (More precisely, as long as the tuple $\langle x_{iv} : v \in \mathcal{K}_i(v)\rangle$ satisfies $L_{ui}$, where $x_{iv} = x_v / \Pr(\mathcal{K}_i(u))$.) Let $X^i$ consist of all the probabilities measures on $W$ compatible with $L_{ui}$ for all $u \in W$. If $\cap_{i=1}^n X^i \neq \emptyset$, then we are done: Choose $\Pr \in \cap_{i=1}^n X^i$, and define $\mathcal{PR}_i$ so that $\Pr_{u,i} = \Pr(\cdot|\mathcal{K}_i(u))$ if $\Pr(\mathcal{K}_i(u)) \neq 0$ and $\Pr_{u,i}$ is some arbitrary probability measure satisfying $L_{ui}$ if $\Pr(\mathcal{K}_i(u)) = 0$. As I mentioned above, with this choice of $\mathcal{PR}_i$, (*) holds.

Now suppose, by way of contradiction, that $\cap_{i=1}^n X^i = \emptyset$. Since $X^1, \ldots, X^n$ are defined by linear equations and inequalities with rational coefficients, by Lemma A.4, there exist $h \leq |W|$, $i^* \in \{1, \ldots, n\}$, and vectors $\vec{a}_{ik}$, $i = 1, \ldots, n$, $k = 1, \ldots, h$, satisfying (2) (from Lemma A.4) such that the coordinates of $\vec{a}_{ik}$ are all rational. Denote by $f^*$ the inequality formula obtained by using these particular vectors $\vec{a}_{ik}$ in (2), and taking the vectors $\vec{x_i}$ to be $\langle x_{iv} : v \in W\rangle$. Let $L^*_{ui}$ consist of all the equations and inequalities in $L_{ui}$ together with the equations $x_{iv} = 0$ for all $v \notin \mathcal{K}_i(u)$. Note that if $\langle x_{iv} : v \in W\rangle$ satisfies $L^*_{ui}$ for all $u \in W$, then it is in $X^i$. Let $\wedge_{i=1}^n L^*_{ui}$ denote the inequality formula that is the conjunction of the linear inequalities in $L^*_{ui}$, $i = 1, \ldots, n$. By Lemma A.4, $\wedge_{i=1}^n L^*_{ui} \Rightarrow f^*$ is a valid inequality formula, for each $u \in W$.

Let $\sigma_u$ be the formula in $\mathcal{L}_n^{K,pr}$ obtained by replacing each occurrence of $x_{iv}$ in $\wedge_{i=1}^n L^*_{ui}$ by $pr_i(\varphi_v)$; similarly, let $\sigma^*$ be the formula obtained by replacing each occurrence of $x_{iv}$ in $f^*$ by $pr_i(\varphi_v)$. As I mentioned earlier, by results of [Fagin, Halpern, and Megiddo 1990], the formula $\sigma_u \Rightarrow \sigma^*$ is provable using I1–I6, Prop, and R1, and hence provable in $\mathrm{AX}_n^{CP}$. Let $\sigma_W$ be $\vee_{u\in W}\sigma_u$. By straightforward propositional reasoning, we have

$$\mathrm{AX}_n^{CP} \vdash \sigma_W \Rightarrow (\sigma_W \wedge \sigma^*). \tag{3}$$

As shown in [Halpern and Moses 1992, pp. 344–345], we have

$$\mathrm{AX}_n^{CP} \vdash \sigma_W \Rightarrow E(\sigma_W). \tag{4}$$

(In fact, all we need for this proof are the axioms for reasoning knowledge and common knowledge; the axioms for probability and inequalities play no role.) Moreover, using (3), Prop, K1, R1, and R2, it is straightforward to show that

$$\mathrm{AX}_n^{CP} \vdash E(\sigma_W) \Rightarrow E(\sigma_W \wedge \sigma^*). \tag{5}$$

From (4), (5), and propositional reasoning, we get that

$$\mathrm{AX}_n^{CP} \vdash \sigma_W \Rightarrow E(\sigma_W \wedge \sigma^*). \tag{6}$$

Thus, from (6) and RC, we have that

$$\mathrm{AX}_n^{CP} \vdash \sigma_W \Rightarrow C(\sigma^*). \tag{7}$$

Propositional reasoning and (7) tells us that

$$\mathrm{AX}_n^{CP} \vdash \sigma_w \Rightarrow C(\sigma^*), \tag{8}$$

for all $w \in W$. But note that $C(\sigma^*)$ is the negation of an instance of $\mathrm{CP}'_n$. This says that $\sigma_w$ is inconsistent, for each $w \in W$. But this contradicts the assumption that $w$ is a (maximal) consistent set.

This contradiction completes the proof, since it shows that $\cap_{i=1}^n X^i \neq \emptyset$. ∎

Using Theorem 3.8, we can now prove Theorem 3.5.

**Theorem 3.5:** *For all $k \geq 2$, there is no set $\mathcal{A}_k$ of formulas in $\mathcal{L}_k^{K,C,pr}$ that distinguishes $\mathcal{F}_k^{CP}$ from $\mathcal{F}_k - \mathcal{F}_k^{CP}$.*

**Proof:** First suppose $k = 2$ and, by way of contradiction, that there is some set $\mathcal{A}_2$ of formulas that distinguishes $\mathcal{F}_2^{CP}$ from $\mathcal{F}_2 - \mathcal{F}_2^{CP}$. By part (a) of Definition 3.1, $\mathcal{A}_2$ must be a subset of the set of formulas valid in $\mathcal{F}_2^{CP}$. Now consider the frame $F^*$ of Example 3.4. Since $F^* \in \mathcal{F}_2 - \mathcal{F}_2^{CP}$, there must be a formula in $\mathcal{A}_2$ that is not valid in $F^*$. Thus, to get a contradiction, it suffices to show that every formula valid in $\mathcal{F}_2^{CP}$ is also valid in $F^*$. By Theorem 3.8, it suffices to show that every instance of an axiom of $\mathrm{AX}_2^{CP}$ is valid in $F^*$. By Theorem 3.6, it is immediate that every axiom other than $\mathrm{CP}'_2$ is valid in $F^*$. The proof that $\mathrm{CP}'_2$ is valid in $F^*$ proceeds along the same lines as the proof that $\mathrm{CP}_2$ is valid in $F^*$, so I omit details here.

Finally, in the case that $k > 2$, define $F_k^* = (W, \mathcal{K}_1, \ldots, \mathcal{K}_k, \mathcal{PR}_1, \ldots, \mathcal{PR}_k)$, where $W$, $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{PR}_1$, and $\mathcal{PR}_2$ are as in $F^*$, and $\mathcal{K}_2 = \ldots = \mathcal{K}_k$, $\mathcal{PR}_2 = \ldots = \mathcal{PR}_k$. Again, it is straightforward to show that every instance of $\mathrm{CP}'_k$ is valid in $F_k^*$. This suffices for the proof, just as in the case $k = 2$. ∎

**Theorem 3.9:** *A formula in $\mathcal{L}^{K,C,pr}$ is valid with respect to $\mathcal{M}_n^{CP}$ (resp., $\mathcal{M}_n$) iff it is valid with respect to $\mathcal{M}_n^{CP,fin}$ (resp., $\mathcal{M}_n^{fin}$).*

**Proof:** I start by considering the case of $\mathcal{M}_n^{CP}$ and $\mathcal{M}_n^{CP,fin}$. Clearly if $\varphi$ is valid with respect to $\mathcal{M}_n^{CP}$, it is also valid with respect to $\mathcal{M}_n^{CP,fin}$. For the converse, it suffices to show that if $\varphi$ is satisfied in $\mathcal{M}_n^{CP}$, then it is satisfied in $\mathcal{M}_n^{CP,fin}$; that is, if $\varphi$ is satisfiable at all, it is satisfied in a finite structure. This follows from Theorem 3.8 and its proof. If $\varphi$ is satisfied in $\mathcal{M}_n^{CP}$ then, by Theorem 3.8, $\varphi$ must be consistent with $\mathrm{AX}_n^{CP}$. The proof of Theorem 3.8 then shows how to construct a structure in $\mathcal{M}_n^{CP,fin}$ satisfying $\varphi$. (In fact, the structure has at most $2^{|\varphi|}$ worlds, where $|\varphi|$ is the length of $\varphi$, viewed as a string of symbols, since it is not hard to show by induction on $|\varphi|$ that $|Sub(\varphi)| \leq |\varphi|$). I provide an alternate proof of this result here, since it gives further insight into what is going on.

Suppose that $\varphi$ is satisfied in some structure $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi) \in \mathcal{M}_n^{CP}$. Since $M \in \mathcal{M}_n^{CP}$, there is some probability $\Pr_W$ on $W$ as required by CP. Define an equivalence relation $\sim$ on the worlds in $M$ by taking $w \sim w'$ if $w$ and $w'$ agree on all formulas in $Sub(\varphi)$. That is, if $(M, w) \models \psi$ iff $(M, w') \models \psi$ for all $\psi \in Sub(\varphi)$. Let $[w]$ be the equivalence class of $w$ according to $\sim$; that is, $[w] = \{w' : w \sim w'\}$. Note that there are at most $2^{|Sub(\varphi)|}$ ($\leq 2^{|\varphi|}$) equivalence classes.

Define a structure $M' = (W', \mathcal{K}'_1, \ldots, \mathcal{K}'_n, \mathcal{PR}'_1, \ldots, \mathcal{PR}'_n, \pi')$ as follows:

- $W' = \{[w] : w \in W\}$,
- $\mathcal{K}'_i([w]) = \{[w'] : [w]$ and $[w']$ agree on all formulas in $Sub(\varphi)$ of the form $K_i\psi$, $C\psi$, and $a_1 pr_i(\varphi_1) + \cdots + a_m pr_i(\varphi_m) > b\}$,
- $\mathcal{PR}'_i([w]) = (\mathcal{K}_i([w]), 2^{\mathcal{K}_i([w])}, \Pr_{[w],i})$, where $\Pr_{[w],i} = \Pr_W(\cdot|\mathcal{K}_i([w]))$ if $\Pr_W(\mathcal{K}_i([w]) > 0$, while if $\Pr_W(\mathcal{K}_i([w]) = 0$, then $\Pr_{[w],i}$ is a probability measure on $\mathcal{K}_i([w])$ that satisfies all the constraints in $L_{wi}$,
- $\pi'([w])(p) = \mathbf{true}$ iff $p \in Sub(\varphi)$ and $\pi(w)(p) = \mathbf{true}$.

Now a straightforward proof by induction on the structure of shows that $(M', [w]) \models \psi$ iff $(M, w) \models \psi$, for all $w \in W$ and $\psi \in Sub(\varphi)$. The ideas are standard (see, for example, the completeness proofs in [Halpern and Moses 1992]), so I leave details to the reader. Thus, if $\varphi$ is satisfied at some world in $M$, say $w_0$, then $(M', [w_0]) \models \varphi$. Moreover, $\Pr_W$ defines a common prior on $W'$. Hence, $M' \in \mathcal{M}_n^{CP,fin}$. This completes the proof.

The argument in the case of $\mathcal{M}_n$ and $\mathcal{M}_n^{fin}$ is almost identical, and is also left to the reader. ∎

**Theorem 3.10:** $\mathrm{AX}_n^{K,pr}$ *is a sound and complete axiomatization for* $\mathcal{L}_n^{K,pr}$ *with respect to both* $\mathcal{M}_n^{CP}$ *and* $\mathcal{M}_n^{CP,fin}$ *(and hence also both* $\mathcal{F}_n^{CP}$ *and* $\mathcal{F}_n^{CP,fin}$*).*

**Proof:** Clearly $\mathrm{AX}_n^{K,pr}$ is sound with respect to $\mathcal{M}_n^{CP}$ and $\mathcal{M}_n^{CP,fin}$, since it is already sound with respect to $\mathcal{M}_n$. We want to show that every formula in $\mathcal{L}_n^{K,pr}$ that is valid in $\mathcal{M}_n^{CP}$ (resp., $\mathcal{M}_n^{CP,fin}$) is provable in $\mathrm{AX}_n^{K,pr}$. As in the proof of Theorem 3.8, it suffices to show that every formula in $\mathcal{L}_n^{K,pr}$ consistent with $\mathrm{AX}_n^{K,pr}$ is satisfied in some structure in $\mathcal{M}_n^{CP,fin}$. Since $\mathrm{AX}_n^{K,pr}$ is complete with respect to $\mathcal{M}_n$, we know that every formula consistent with $\mathrm{AX}_n^{K,pr}$ is satisfied in some structure in $\mathcal{M}_n$. By Theorem 3.9, we can assume without loss of generality it is satisfied in a structure in $\mathcal{M}_n^{fin}$. Thus, it suffices to show that every formula that is satisfied in some structure in $\mathcal{M}_n^{fin}$ is also satisfied in some structure in $\mathcal{M}_n^{CP,fin}$.

Define the *depth* of a formula $\varphi$ in $\mathcal{L}_n^{K,pr}$, denoted $d(\varphi)$, as follows:

- $d(p) = 0$ for a primitive proposition $p$,
- $d(\neg\psi) = d(\psi)$,
- $d(\psi \wedge \psi') = \max(d(\psi), d(\psi'))$,

- $d(K_i\psi) = 1 + d(\psi)$,
- $d(a_1 pr_i(\psi_1) + \cdots + a_k pr_i(\psi_k) > b) = 1 + \max(d(\psi_1), \ldots, d(\psi_k))$.

Let a *situation* be a pair $(M, w)$ consisting of a structure $M$ and a world $w$ in $M$. Two situations $(M, w)$ and $(M', w')$ are *equivalent up to depth* $k$, denoted $(M, w) \equiv^k (M', w')$, if, whenever $\varphi$ is a formula with $d(\varphi) \leq k$, then $(M, w) \models \varphi$ iff $(M', w') \models \varphi$.

The proof depends on two key observations, which I state informally here and then make more precise.

1. If a formula $\varphi_0 \in \mathcal{L}_n^{K,pr}$ is satisfiable at all, it is satisfied at the root of a "treelike" structure of height at most $d(\varphi_0)$.
2. Adding worlds to the leaves of this treelike structure that are "distance" greater than $d(\varphi_0)$ away from the root does not affect the truth of $\varphi_0$ at the root.

To make this precise, I use a standard idea from modal logic of "unwinding" a structure to a tree. Given a structure $M = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{PR}_1, \ldots, \mathcal{PR}_n, \pi) \in \mathcal{M}_n^{fin}$, we define a "treelike" structure $T^*_{M,w,k}$, for each world $w \in W$ and $k \geq 0$, such that $(T^*_{M,w,k}, r) \equiv^k (M, w)$, where $r$ is the "root" of $T^*_{M,w,k}$, as follows: The first step is to define (rooted, labeled, directed) trees $T_{M,w,k}$, by induction on $k$. The tree $T_{M,w,0}$ just consists of a single node $r$, labeled $w$. (In general, nodes will be labeled by worlds in $W$, but more than one node may be labeled with the same world, and edges will be labeled by agents.) $T_{M,w,k+1}$ consists of a root node $r$ labeled by $w$ and, for each agent $i$ and world $w' \neq w$ such $w' \in \mathcal{K}_i(w)$, a directed edges labeled by $i$ leading from $r$ to the root $r'$ of $T^i_{M,w,k}$, where $T^i_{M,w,k}$ is the result of removing all the $i$-successors of $r'$ in $T_{M,w',k}$ (and all the nodes reachable from these $i$-successors). We can easily show by induction on $k$ that this construction guarantees that there is no path in $T_{M,w,k+1}$ that contains two consecutive $i$-edges, for any agent $i$.

Now let the structure $T^*_{M,w,k} = (W^{M,w,k}, \mathcal{K}_1^{M,w,k}, \ldots, \mathcal{K}_n^{M,w,k}, \mathcal{PR}_1^{M,w,k}, \ldots, \mathcal{PR}_n^{M,w,k}, \pi^{M,w,k})$ be defined as follows:

- $W^{M,w,k}$ consists of the nodes in $T_{M,w,k}$.
- $\mathcal{K}_i^{M,w,k}$ is the smallest equivalence relation such that if $n'$ is an $i$-successor of $n$ in $T_{M,w,k}$, then $(n, n') \in \mathcal{K}_i^{M,w,k}$.
- $\mathcal{PR}_i^{M,w,k}(n) = (\mathcal{K}_i^{M,w,k}(n'), 2^{\mathcal{K}_i^{M,w,k}(n)}, \Pr_{n,i})$, where if $n$ is not a leaf in $T_{M,w,k}$ or if $n$ is a leaf in $T_{M,w,k}$ and is the $i$-successor of some (non-leaf) node, then for all $n' \in \mathcal{K}_i^{M,w,k}(n)$, we have $\Pr_{n,i}(n') = \Pr_{f(n),i}(f(n'))$, where $f$ is the function that associates with each node in $T_{M,w,k}$ the world that labels it; if $n$ is a leaf in $T_{M,w,k}$ and is not the $i$-successor of some non-leaf node, then $\Pr_{n,i}(n) = 1$. (In this case, it is easy to see that $\mathcal{K}_i^{M,w,k}(n) = \{n\}$.)

- $\pi^{M,w,k}(n)(p) = \pi(f(n))(p)$.

It is easy to check that $\Pr_{n,i}$ is indeed a probability measure on $\mathcal{K}_i^{M,w,k}(n)$ for each agent $i$ and $n \in W^{M,w,k}$.

**Lemma A.5:** *$(M,w) \equiv^k (T^*_{M,w,k}, r)$, where $r$ is the root of $T_{M,w,k}$.*

**Proof:** For each node $n \in W^{M,w,k}$, let $dist(r,n)$ be the distance from $r$ to $n$ in $T_{M,w,k}$. A straightforward induction on $d(\psi)$, which I leave to the reader, can be used to show that if $d(\psi) + dist(r,n) \leq k$, then $(T^*_{M,w,k}, n) \models \psi$ iff $(M, f(n)) \models \psi$. Since $d(r,r) = 0$ and $f(r) = w$, this gives us the desired result. ∎

Lemma A.5 actually shows proves both of the informal observations above. It shows that if a formula $\varphi_0$ is satisfiable at all, it is satisfied in a treelike structure of height at most $d(\varphi_0)$, since if $\varphi_0$ is satisfied at the situation $(M,w)$, then $T^*_{M,w,k}$ is the required treelike structure. Moreover, it shows that making changes in this treelike structure by adding worlds to leaves does not affect the truth of $\varphi_0$, since if $M'$ is the resulting structure, we will still have $T^*_{M,w,k} = T^*_{M',w,k}$. The remainder of the argument uses this second point (and makes it more precise).

Let $n$ be a leaf of $T_{M,w,k}$ and suppose that $n$ is the $i$-successor of some node $n'$. We construct a structure $M^*$ that is almost identical to $T^*_{M,w,k}$. Informally, we add a new world $n^*$ which is the $i'$-successor of $n$ for some $i' \neq i$, and assume that all agents assign $n^*$ probability 1. More precisely, let $M^* = (W^*, \mathcal{K}_1^*, \ldots, \mathcal{K}_n^*, \mathcal{PR}_1^*, \ldots, \mathcal{PR}_m^*, \pi^*)$, where

- $W^* = W^{M,w,k} \cup \{n^*\}$,
- $\mathcal{K}_j^* = \mathcal{K}_j^{M,w,k} \cup \{(n^*,n^*)\}$ for $j \neq i'$; $\mathcal{K}_{i'} = \mathcal{K}_i^{M,w,k} \cup \{(n,n^*),(n^*,n),(n^*,n^*)\}$,
- $\mathcal{PR}_j^*(n') = (\mathcal{K}_j^*(n'), 2^{\mathcal{K}_j^*(n')}, \Pr^*_{n',j})$, where $\Pr^*_{n',j} = \Pr_{n',j}$ if $n' \neq n^*$ and $(n',j) \neq (n,i')$, and $\Pr_{n',j}$ is the unique probability measure such that $\Pr_{n',j}(n^*) = 1$ if $n' = n^*$ or $(n',j) = (n,i')$,
- $\pi^*(n') = \pi(n')$ if $n' \neq n^*$ (the definition of $\pi^*(n^*)$ is irrelevant).

Our construction guarantees that (a) $T^*_{M,w,k} = T^*_{M^*,w,k}$ (since the way we changed $T^*_{M,w,k+1}$ to get $M^*$ involved only the addition of a node $k+1$ away from the root) and (b) $M^* \in \mathcal{M}_n^{CP,fin}$. To see (b), note that there is a common prior that gives probability 1 to $n^*$.

We can now easily complete the proof of Theorem 3.10. Suppose $\varphi_0$ is a formula satisfied in some situation $(M_0,w_0)$, where $M \in \mathcal{M}_n^{fin}$ and $d(\varphi_0) = k$. Using the construction above, we get a structure $M^* \in \mathcal{M}_n^{CP,fin}$ such that $T^*_{M,w,k} = T^*_{M^*,w,k}$. Thus, if $r$ is the root of $M^*$, we have $(M^*, r) \models \varphi_0$. ∎

**Theorem 3.11:** *For all $n$, no set $\mathcal{A}$ of formulas in $\mathcal{L}_n^{K,pr}$ distinguishes $\mathcal{F}_n^{CP,fin}$ from $\mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$.*

**Proof:** Suppose $\mathcal{A}$ distinguishes $\mathcal{F}_n^{CP,fin}$ from $\mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$. Let $F \in \mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$. By part (b) of Definition 3.1, there must be some formula $\varphi \in \mathcal{A}$ that is not valid in $F$. But by part (a) of Definition 3.1, $\mathcal{A}$ must be a subset of the set of formulas valid in $\mathcal{F}_n^{CP,fin}$. By Theorem 3.10, it follows that the formulas in $\mathcal{A}$ are also valid in $\mathcal{F}_n^{fin}$. Thus, $\varphi$ must also be valid in $F$, giving us a contradiction. ∎


**Acknowledgments:**

I would like to thank Bart Lipman for useful discussions regarding the differences between CP and CP$^s$ and Dov Samet for his examples regarding separation of convex sets.


# References


Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics 4*(6), 1236–1239.

Aumann, R. J. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica 55*, 1–18.

Aumann, R. J. (1989). Notes on interactive epistemology. Cowles Foundation for Research in Economics working paper.

Aumann, R. J. (1998). Common priors: a reply to Gul. *Econometrica 66*(4), 929–938.

Bonanno, G. and K. Nehring (1999). How to make sense of the common prior assumption under incomplete information. *International Journal of Game Theory 28*(3), 409–434.

Cormen, T. H., C. E. Leiserson, and R. L. Rivest (1990). *Introduction to Algorithms.* Cambridge, MA/New York: MIT Press/McGraw Hill.

Davis, M. (1977). *Applied Nonstandard Analysis.* New York: Wiley.

Fagin, R. and J. Y. Halpern (1994). Reasoning about knowledge and probability. *Journal of the ACM 41*(2), 340–367.

Fagin, R., J. Y. Halpern, and N. Megiddo (1990). A logic for reasoning about probabilities. *Information and Computation 87*(1/2), 78–128.

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

Feinberg, Y. (1995). A converse to the Agreement Theorem. Technical Report Discussion Paper #83, Center for Rationality and Interactive Decision Theory.

Feinberg, Y. (2000). Characterizing common priors in the form of posteriors. *Journal of Economic Theory 91*(2), 127–179.

Gul, F. (1998). A comment on Aumann's Bayesian view. *Econometrica 66*(4), 923–927.

Halpern, J. Y. and Y. Moses (1992). A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence 54*, 319–379.

Harsanyi, J. (1968). Games with incomplete information played by 'Bayesian' players, parts I-III. *Management Science 14*, 159–182, 320–334, 486–502.

Kaplan, D. (1966). Review of "A semantical analysis of modal logic I: normal modal propositional calculi". *Journal of Symbolic Logic 31*, 120–122.

Lemmon, E. J. (1977). *The "Lemmon Notes": An Introduction to Modal Logic*. Oxford, U.K.: Basil Blackwell. Written in collaboration with Dana Scott; edited by Krister Segerberg. American Philosophical Quarterly Monograph Series. Monograph No. 11.

Lipman, B. L. (1997). Finite order implications of common priors. Unpublished manuscript.

Makinson, D. (1966). On some completeness theorems in modal logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 12*, 379–384.

Morris, S. (1994). Trade with heterogeneous prior beliefs and asymmetric information. *Econometrica 62*, 1327–1348.

Morris, S. (1995). The common prior assumption in economic theory. *Economics and Philosophy 11*, 227–253.

Rockafellar, R. T. (1972). *On Measures of Information and Their Characterizations, second printing*. Princeton, N.J.: Princeton University Press.

Samet, D. (1998). Common priors and separation of convex sets. *Games and Economic Behavior 24*, 172–174.

Tarski, A. (1951). *A Decision Method for Elementary Algebra and Geometry* (2nd ed.). Univ. of California Press.