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

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.

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 work was 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.

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 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 [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 [1996], Feinberg [1995, 1996], Morris [1994], and Samet [pear], who all showed that, in finite spaces, the CPA could be characterized by a disagreement in expectations, in a sense explained below. Feinberg [1996] 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 [1996] 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).

The rest of this paper is organized as follows. In Section 2, I define the language considered and its semantics. In Section 3, I consider the two types of characterizations. In addition to considering what happens with common knowledge in the language, I also show that without common knowledge, there are no new consequences of the CPA. This contrasts with a result of Lipman [1997], who 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. For interested readers, the axiom systems mentioned in the paper are described in the appendix. For reasons of space, proofs have been omitted.

2 Syntax and Semantics

I start with a brief review of the syntax and semantics of the language for reasoning about knowledge, common knowledge, and probability defined in [Fagin and Halpern 1994]. We start with a set Φ of primitive propositions (think of these as representing basic events such as "agent 1 went left on his last move") and n agents. We take $\mathcal{L}_n^{K,C,pr}$ to be the least set of formulas that includes Φ and is closed under the following construction rules: 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 φ"), $C\varphi$ ("φ 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 φ 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 φ") is an abbreviation for $K_1\varphi \wedge \ldots K_n\varphi$ and $E^{m+1}\varphi$ is an abbreviation for $E^1E^m\varphi$ ("everyone knows that everyone knows ... that everyone knows—$m+1$ times—φ"), 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 by using i-probability formulas, we can 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.

To assign truth values to formulas in $\mathcal{L}_n^{K,C,pr}$, we use 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 1996]); $\mathcal{K}_i(w)$ 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, is equivalent to using an equivalence relation.[1]

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. Formally, an interpretation π associates with each world w a truth assignment to the primitive propositions in Φ; 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 $M = (W, \mathcal{K}_1, \dots, \mathcal{K}_n, \mathcal{PR}_1, \dots, \mathcal{PR}_n, \pi)$, where $F = (W, \mathcal{K}_1, \dots, \mathcal{K}_n, \mathcal{PR}_1, \dots, \mathcal{PR}_n)$ is a frame and π 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.

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 φ 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 φ is true in structure M. We proceed by induction on the structure of φ, assuming that we have given the definition for all subformulas φ' of φ 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 $\pi(w)(p) = \mathbf{true}$

[1]Bonanno and Nehring [1996] assume only that the relation is serial, Euclidean, 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.

$(M, w) \models \varphi \wedge \varphi'$ 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 $k \geq 1$

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

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 [1996], Feinberg [1996], and Samet [pear] do not require the condition that the prior gives each component positive probability. However, this condition is necessary for Aumann's theorem to hold; if the common prior can give probability 0 to a component, then we can have two agents with a common prior that have common knowledge (in a component with prior probability 0) that they assign different probabilities to a particular event. 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.

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.[2]

A formula φ 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 φ 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 $\mathcal{F}$.

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 $\mathcal{F}$, (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; as I show in the full paper, this is necessary for the technical results to hold.

My results on frame distinguishability parallel those of Feinberg [1996]: 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 $a \neq b$.

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

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 [1996] 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. 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, 1996] and Samet [pear] showed that the converse also holds in finite spaces.[3] 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 $\mathcal{F}_2^{fin} - \mathcal{F}_2^{CP,fin}$.*

As Feinberg and Samet show, we can extend this characterization to $n > 2$ agents. Their characterization leads to an axiom CP$_n$ which allows us to distinguish $\mathcal{F}_n^{CP,fin}$ from $\mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$, for all $n > 0$. The details can be found in the full paper.

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 $\mathcal{F}_2 - \mathcal{F}_2^{CP}$. But, in fact, an even stronger result holds:

Theorem 3.3: *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 $\mathcal{F}_k - \mathcal{F}_k^{CP}$.*

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

[3]Essentially the same result is proved by Bonanno and Nehring [1996], but they were dealing with belief rather than knowledge, so rather than being equivalences, their $\mathcal{K}_i$ relations were serial, Euclidean, and transitive.

axioms and *inference rules*. 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.

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. We can 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.

In [Fagin and Halpern 1994], a complete axiomatization is provided for the language $\mathcal{L}_n^{K,pr}$ with respect $\mathcal{M}_n$. This axiom system, denoted $\mathrm{AX}_n^{K,pr}$, is described in the appendix. By adding a (well known) axiom and rule for reasoning about common knowledge (also given in the appendix), we get the system $\mathrm{AX}_n^{K,C,pr}$.

Theorem 3.4: $\mathrm{AX}_n^{K,C,pr}$ *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* $\mathcal{F}_n^{fin}$*).*

$\mathrm{AX}_n^{K,C,pr}$ 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 CP_n is valid in $\mathcal{M}_n^{CP}$ (and hence $\mathcal{M}_n^{CP,fin}$). We might hope that if we add 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 [pear] shows that, given a frame, the set of possible priors for agent i (i.e., those that can generate the posteriors defined by $\mathrm{Pr}_{w,i}$) is closed and convex. 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 CP_2. That is, strict separation by a hyperplane amounts to a disagreement in expectation.

Unfortunately, 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 φ_1, φ_2, and φ_3 be the three mutually exclusive formulas $p \wedge q$, $p \wedge \neg q$, and $\neg p$, respectively. Let ψ_1 be $(pr_1(\varphi_1) > pr_1(\varphi_2)) \vee ((pr_1(\varphi_1) \geq pr_1(\varphi_2)) \wedge (pr_1(\varphi_3) > 1/2))$ and let ψ_2 be $(pr_2(\varphi_1) < pr_2(\varphi_2)) \vee ((pr_2(\varphi_1) \leq pr_2(\varphi_2)) \wedge (pr_2(\varphi_3) \leq 1/2))$.[4] Let X^i consist of all prior probability distributions for agent i that satisfy ψ_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 φ_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;

[4]This example was suggested by Dov Samet.

it is easy to show that they cannot be strictly separated by a hyperplane, and we do not have disagreement in expectation in the spirit of CP_2. As a consequence, we get:

Theorem 3.5: *The formula* $\neg C(\psi_1 \wedge \psi_2)$ *is valid in* $\mathcal{M}_2^{CP}$, *but is not provable from* $AX_2^{K,C,pr} + CP_2$.

It follows from Theorem 3.5 that, if we are to obtain a completeness result, even in the case of two agents, we need something stronger than CP_2. The key insight comes from examining the sets 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

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

As is shown in the full paper, this example generalizes. Any two disjoint convex (but not necessarily closed) sets can be separated in expectation in this more general sense.

This observation suggests the following axiom in the case of two agents.

CP_2'. If $\varphi_1, \ldots, \varphi_m$ are mutually exclusive formulas and $i^* \in \{1, 2\}$, then

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

It is easy to see that the formula $\neg C(\psi_1 \wedge \psi_2)$ in Theorem 3.5 follows from CP_2'.

Just as CP_2 generalizes to CP_n with n agents, we can generalize CP_2' to CP_n' (see the appendix for a description of CP_n'). Let AX_n^{CP} consist of all the axioms and rules of $AX_n^{K,C,pr}$ together with CP_n'.

Theorem 3.6: 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* $\mathcal{F}_n^{CP,fin}$*).*

It may seem somewhat surprising that there is no difference between infinite structure and finite structures in Theorem 3.6. The contrast with Theorems 3.2 and 3.3 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. In fact, standard techniques of modal logic can be used to show that if a formula is satisfiable at all, it is satisfied in a finite structure.

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 $\mathrm{AX}_n^{K,pr}$ 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 3.7: $\mathrm{AX}_n^{K,pr}$ *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* $\mathcal{F}_n^{CP,fin}$*).*

We do no better with frame distinguishability.

Theorem 3.8: *For all* n*, no set* $\mathcal{A}$ *of formulas in* $\mathcal{L}_n^{K,pr}$ *distinguishes* $\mathcal{F}_n^{CP,fin}$ *from* $\mathcal{F}_n^{fin} - \mathcal{F}_n^{CP,fin}$ *(or* $\mathcal{F}_n^{CP}$ *from* $\mathcal{F}_n^{CP} - \mathcal{F}_n^{CP}$*).*

These results stand in distinction to those proved by Lipman [1997]. Lipman showed that there are consequences of the CPA (given his formalization of it) even without common knowledge in the language. That is because, in his formalization of the CPA, all information sets are required to have positive prior probability. Lipman's (slightly stronger) version of the CPA can be formalized as follows:

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

Although CP^s seems only slighty stronger that CP, it has an impact on all the results of this paper. As Lipman shows, the formula $pr_i(\varphi \wedge pr_j(\varphi) = 0) = 0$ (which is in $\mathcal{L}_n^{K,pr}$) is valid in structures satisfying CP^s, so Theorem 3.7 does not hold. In addition, as I show in the full paper, $\mathrm{AX}_n^{K,C,pr}$ is no longer a complete axiomatization for the language $\mathcal{L}_n^{K,C,pr}$, and CP_n no longer distinguishes finite frames that satisfy CP^s from ones that do not (so that Feinberg's result really depends on the fact that he uses CP rather than CP^s). I discuss these issues in more detail in the full paper.

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 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 3.7). 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 $\mathcal{L}_n^{K,pr}$ (Theorem 3.7), 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 3.8). If we add common knowledge to the language, then we can distinguish finite frames satisfying the CPA from ones that do not (Theorem 3.2 and its extension to n agents—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 3.3); nevertheless, we can completely characterize the properties of (finite or infinite) frames satisfying the CPA (Theorem 3.6).

This leads to both a technical question and a pragmatic one. The technical question is whether, in a sufficiently rich language, the agents can distinguish infinite frames satisfying the CPA from ones that do not (given only their posterior information). I leave this question open. The pragmatic question is which of the two notions I have considered is more appropriate. That, of course, depends on the application. If we are interested in knowing if we can test whether or not the CPA holds in a given structure, this is a question essentially about frame distinguishability. On the other hand, if we are interested in knowing what properties hold in a given situation, given a finite collection Σ of facts about the agents' knowledge and beliefs and about the true situation (all expressed as formulas in $\mathcal{L}_n^{K,C,pr}$) and that the CPA holds, this is a question that can be answered using a complete axiomatization—frame distinguishability is of no help.

Although both notions are useful, it is helpful to be clear about the differences between them. This paper attempts to provide some clarification.

Appendix: Axiom Systems

This appendix describes all the axiom systems mentioned in the paper, for the interested reader.

The axiom system $AX_n^{K,pr}$ 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. I describe each component below:

I. Propositional Reasoning

Prop. All instances of propositional tautologies

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

II. Reasoning About Knowledge

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

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 φ 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 $1, \ldots, m$

I3. $(a_1pr_i(\varphi_1) + \cdots + a_mpr_i(\varphi_m) \geq b) \wedge (a'_1pr_i(\varphi_1) + \cdots + a'_mpr_i(\varphi_m) \geq b') \Rightarrow$
$(a_1 + a'_1)pr_i(\varphi_1) + \cdots + (a_m + a'_m)pr_i(\varphi_m) \geq (b + b')$

I4. $(a_1pr_i(\varphi_1) + \cdots + a_mpr_i(\varphi_m) \geq b) \Leftrightarrow (c_1pr_i(\varphi_1) + \cdots + c_mpr_i(\varphi_m) \geq db)$ if $d > 0$

I5. $(a_1pr_i(\varphi_1) + \cdots + a_mpr_i(\varphi_m) \geq b) \vee (a_1pr_i(\varphi_1) + \cdots + a_mpr_i(\varphi_m) \leq b)$

I6. $(a_1pr_i(\varphi_1) + \cdots + a_mpr_i(\varphi_m) \geq b) \Rightarrow (a_1pr_i(\varphi_1) + \cdots + a_mpr_i(\varphi_m) > b')$ if $b > b'$

IV. Reasoning about probabilities

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

P2. $pr_i(\mathit{true}) = 1$

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

RP. From $\varphi \Leftrightarrow \psi$ infer $pr_i(\varphi) = pr_i(\psi)$ [5]

V. Reasoning about knowledge and probabilities

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

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

$\mathrm{AX}_n^{K,C,pr}$ consists of all the axioms and rules of $\mathrm{AX}_n^{K,pr}$, together with the following axiom and rule for common knowledge:[6]

VI. Reasoning About Common Knowledge

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

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

Finally, here is the axiom CP'_n used in $\mathrm{AX}_n^{K,C,pr}$, which generalizes 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

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

Acknowledgements:

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.

[5]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 $pr_i(K_j\varphi) = pr_i(K_jK_j\varphi)$.

[6]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.

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.

Bonanno, G. and K. Nehring (1996). How to make sense of the common prior assumption under incomplete information. Unpublished manuscript.

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, 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. (1996). Characterizing common priors in the form of posteriors. Unpublished manuscript.

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.

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

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, to appear). Common priors as separation of convex sets. *Games and Economic Behavior*.