Approximately Common Priors[1]

Barton L. Lipman

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario N6A 5C2

email: blipman@julian.uwo.ca

October 1995

First Draft

---

[1] I wish to thank Jim Bergin, Eddie Dekel, and Steve Morris for helpful discussions, SSHRCC for financial support, and the University of Pennsylvania for a very enjoyable visit during which some of this research was carried out.

Abstract

I show that if one is only interested in local results — that is, theorems about what is true at a state — which only depend on finitely many orders of beliefs about beliefs, then the common prior assumption is essentially irrelevant. More precisely, given any model where priors have the same support and any finite N, there is another model with common priors which has the same n^{th} order beliefs for all $n \leq N$. In this sense, priors with the same support are (locally) approximately common.

1 Introduction

The common prior assumption — the assumption that all agents have the same prior over the relevant state space — is used in virtually all of game theory and information economics. It is a crucial assumption for many important results such as Aumann's agreeing–to–disagree result ([1]), the no–trade theorem (see, *e.g.*, Milgrom and Stokey [7]), and Aumann's [2] characterization of correlated equilibria. There has been a great deal of debate regarding the assumption — see, for example, Morris [9] for a summary of some of the arguments.

I provide a new result clarifying what the common prior assumption means as a statement about the agents' hierarchies of beliefs. My result, very loosely, is that the common prior assumption has only very weak implications for beliefs of finite order — that is, beliefs about the beliefs of others about the beliefs of . . . about the beliefs of others, where "beliefs about" is repeated only a finite number of times. Phrasing the result differently, most priors are locally approximately common. One implication of this is that dropping the common prior assumption is (almost) the same thing as replacing common knowledge with a high but finite level of mutual knowledge.

A precise statement of the result requires some background. Imagine we are given a complete description of the world by an outside observer. This description includes all true statements about some parameter which is unknown to the agents, the beliefs of the agents (their *first–order beliefs*) about this parameter, their beliefs about the beliefs of others (*second–order beliefs*), their beliefs about the others' beliefs about the others (*third–order beliefs*), etc. I will refer to a specification of the true parameter together with an infinite hierarchy of beliefs for each player as a *world*, reserving the term *state* for a related but different notion. As Mertens and Zamir [6] showed in their classic paper, such a description of the *actual* world generates a collection of *possible* worlds, one of which is the actual world. These other worlds can then be seen as fictitious constructs, used to clarify our understanding of the actual world.[1]

The actual world also identifies a prior over the set of possible worlds for each agent. The prior for a given agent is not unique since any prior yielding the same conditional probabilities on the appropriate events would serve the same purpose. However, the point is that the actual world identifies whether or not there is a single prior we could attribute to all the agents which generates the beliefs of each agent.

In other words, the actual world — the infinite hierarchy of beliefs about beliefs for each player — completely determines whether or not the beliefs are consistent with the common prior assumption. Hence even though we typically think of the common prior assumption as an *ex ante* statement — a statement regarding the agent's beliefs before getting information — we can view whether or not this assumption holds as something determined at the interim level

---

[1] See Dekel and Gul [4] for a similar perspective.

— that is, in the actual world or after receipt of information.

To state the result more precisely, then, suppose the agents' beliefs at a world ω are weakly consistent in the sense that it is common knowledge that all agents' beliefs at all levels give positive probability to the truth. Then for any finite N, there is another world, ω^N, which is consistent with the common prior assumption and where for all $n \leq N$, n^{th} order beliefs at ω^N are the same as those at world ω. In this sense, we can always approximate weakly consistent priors by common priors, at least for local results.

This result has several implications which are discussed in detail in Section 4. One is the implication mentioned above: dropping the common prior assumption is almost the same thing as replacing common knowledge with a high but finite degree of mutual knowledge. That is, given a world where a particular set of facts are common knowledge but priors are different, we can find a model with common priors in which nothing changes except that these facts are arbitrarily high mutual knowledge but *not* common knowledge. This has important methodological implications which are discussed in more detail in Section 4.2.

The theorem also tells us where the real "bite" of the common prior assumption lies. The theorem relies on an assumption of weak consistency — and the result is not true without this assumption. However, the theorem shows that weak consistency is the only implication the common prior assumption has if we are only concerned with finitely many orders of beliefs. In this sense, the common prior assumption is primarily an "infinite order" assumption since we know that the common prior assumption has a very significant effect at least as measured by the kinds of theorems that are true with it and false without it.

Finally, the result enables me to show that the set of worlds consistent with common priors is dense in the set of all worlds — that is, it is dense in the Mertens–Zamir universal belief space. In this sense, *all* belief hierarchies are approximately consistent with common priors. This implication is explained in Section 4.1. This is a very surprising result in light of the apparent generic failure of common priors when viewed differently. Suppose we fix a state set and choose priors randomly for the agents. Clearly, the probability that we choose the same priors for all agents is very small, suggesting that the common prior assumption, in some sense, fails generically.[2] My result shows that when we take the set of belief hierarchies as the relevant space, rather than the set of probability distributions on a fixed state space, we get a startlingly different view.

It is also worth mentioning that the result shows that the line between differences in information and differences in beliefs is more blurred than one might suppose. As advocates of the common prior assumption have long noted, the assumption is a formal statement of the intuitive idea that differences in beliefs should all stem from differences in information. My

---

[2] This description overlooks the fact that it is only conditional beliefs that are relevant. However, as Nyarko [10] shows, this statement can be formalized and demonstrated even taking this fact into account.

result indicates that if we only care about finitely many orders of belief, a “pure” difference of beliefs (noncommon priors) can always be translated into a difference in beliefs stemming from a difference in information. Section 2.2 gives an example showing how this translation works.

2 The Model and an Illustrative Example

2.1 Model

Tan and Werlang [11] and Brandenburger and Dekel [3] have shown that we can generate any point in the universal beliefs space by use of what I will term a *partitions model*. Given a *parameter space* S, and a finite set of players $\mathcal{I} = \{1, \ldots, I\}$, a partitions model consists of

1. a set of *states*, Ω
2. a function $f : \Omega \to S$ telling what the unknown parameter is in each state
3. for each player $i \in \mathcal{I}$, a partition Π_i of Ω
4. for each player $i \in \mathcal{I}$, a prior μ_i over Ω.

I will use $\mathcal{M}$ to denote a typical partitions model $(\Omega, f, \{(\Pi_i, \mu_i)\}_{i \in \mathcal{I}})$. Given $\mathcal{M}$ and a state $\omega \in \Omega$, we can define a point in the Mertens–Zamir universal beliefs space by “unravelling.” That is, player $\imath$’s first–order beliefs at ω are those induced by updating $\imath$’s prior given the event $\Pi_i(\omega)$ and using the function f to convert this to a belief on S. Formally, the first–order beliefs of i at state ω, say $\delta_i^1[\omega]$, are defined by

$$\delta_i^1[\omega](B) = \mu_i\left(\{\omega' \in \Omega \mid f(\omega') \in B\} \mid \Pi_i(\omega)\right)$$

for each measurable $B \subseteq S$. The second–order beliefs are calculated by performing the same updating but then using the partitions for players other than i to calculate i’s beliefs about these players’ beliefs, etc. That is, if we have two players, then player 1’s second–order beliefs at ω, say $\delta_1^2[\omega]$, are a probability distribution over $S \times \Delta(S)$, where $\Delta(S)$ is the set of first–order beliefs for player 2.[3] Formally, then,

$$\delta_1^2[\omega](B) = \mu_i\left(\left\{\omega' \in \Omega \mid (f(\omega), \delta_2^1(\omega)) \in B\right\} \mid \Pi_1(\omega)\right)$$

---

[3] We could equivalently define second–order beliefs to be a distribution over $S \times \Delta(S) \times \Delta(S)$ and then impose the condition that 1 knows his own beliefs. The approach in the text follows Brandenburger and Dekel [3], while the approach described in this footnote follows the original Mertens and Zamir formulation.

for each measurable $B \subseteq S \times \Delta(S)$. Etc.

This equivalence works in reverse as well: any point in the universal beliefs space can be constructed in this fashion from some partitions model.[4]

2.2 An Example

Suppose we have two players and two possible values of the unknown parameter, $S = \{s_1, s_2\}$. We also have two states, $\Omega = \{\omega_1, \omega_2\}$. The function f is given by $f(\omega_j) = s_j$, $j = 1, 2$. Both players' partitions are trivial: $\Pi_i = \{\Omega\}$ for $i = 1, 2$. Player 1's prior is $\mu_1(\omega_1) = 2/3$, while player 2's prior has $\mu_2(\omega_1) = 1/3$. Hence we have a "pure" difference in beliefs, not stemming from any difference in information. Let $\mathcal{M}$ denote this partitions model.

I now show that we can match beliefs at each state to an arbitrarily high order. I demonstrate this by constructing a sequence of new partitions models where each model has a state ω^n which matches the beliefs at ω in $\mathcal{M}$ up to order N. In each of these new models, the difference in beliefs is replaced by a difference in information. However, as we will see, this replacement is not without cost: we *cannot* avoid changing the beliefs at some very high orders.

So fix an integer N and construct the new model as follows. The new set of states, $\bar{\Omega}^N$, is taken to be

$$\bar{\Omega}^N = \Omega \times \{1, \ldots, 2^N\}.$$

Intuitively, think of state (ω, k) as the k^{th} copy of state ω from our original model. In line with this intuition, the new function relating states and the value of the unknown parameter, $\bar{f}^N$, is given by

$$\bar{f}^N(\omega, k) = f(\omega).$$

The common prior of both players is that all states in $\bar{\Omega}^N$ are equally likely.

To generate the appropriate beliefs for player 1, then, we will have to have his information reflect a greater likelihood of state ω_1 than ω_2. Hence his partition, $\bar{\Pi}_1^N$, includes all events of the form $\{(\omega_1, 2k-1), (\omega_1, 2k), (\omega_2, k)\}$ for $k = 1, \ldots, 2^{N-1}$. Since player 1 "uses up" his ω_1 copies faster than his ω_2 copies, we will also end up with an event $\{(\omega_2, 2^{N-1}+1), \ldots, (\omega_2, 2^N)\}$. Note that at *every* event except this last one, player 1's beliefs over S give probability 2/3 to s_1 and 1/3 to s_2.

The partition for player 2, $\bar{\Pi}_2^N$, is analogous. Since player 2 thinks ω_2 is more likely, we include every event of the form $\{(\omega_1, k), (\omega_2, 2k-1), (\omega_2, 2k)\}$ for $k = 1, \ldots, 2^{N-1}$. Analogously to the case of player 1, this leaves us with "extra" ω_1 copies, so player 2's partition also includes the event $\{(\omega_1, 2^{N-1}+1), \ldots, (\omega_1, 2^N)\}$. Again, at all events except this last one, player

---

[4] To be more precise, we may need σ–fields instead of partitions. See Brandenburger and Dekel [3].

2's beliefs over S match his beliefs from the original model. Let M^N denote this partitions model.

It is not hard to show that n^{th} order beliefs at $(\omega, 1)$ in M^N are the same as n^{th} order beliefs at ω in M for all $n \leq N$ for both $\omega = \omega_1$ and $\omega = \omega_2$. To see this, first note that we have obviously matched first–order beliefs. In state $(\omega, 1)$, player 1's first–order beliefs put probability 2/3 on s_1 and 1/3 on s_2 or, in the obvious vector notation, $(2/3, 1/3)$. This is precisely the same as player 1's first–order beliefs at ω_1 in the original partitions model. Similarly for player 2.

To get second–order beliefs for player 1, we must identify the events of 2's partition he considers possible. These events are

$$\{(\omega_1, 1), (\omega_2, 1), (\omega_2, 2)\}$$

and

$$\{(\omega_1, 2), (\omega_2, 3), (\omega_2, 4)\}.$$

The set of first–order beliefs for 2 at these events form the support of player 1's second–order beliefs. Note that 2's beliefs over S at each of these events are $(1/3, 2/3)$, so that player 1 puts probability 1 on these being player 2's first–order beliefs. This exactly matches player 1's second–order beliefs at ω_1 in the original partition model. Again, the analogous argument for player 2 works.

More generally, we see that player 1's n^{th} order beliefs are completely determined by the events that are $n - 1$ "steps" from $(\omega, 1)$. More precisely, an event is $n - 1$ steps from $(\omega, 1)$ if we can construct a sequence of n events starting from $\{(\omega_1, 1), (\omega_1, 2), (\omega_2, 1)\}$ such that as we move from one event to the next, we move from an event in one player's partition to an intersecting event in the other player's partition. It is not hard to see that for $n \leq N$, all events reachable this way have the same first–order beliefs as in the original model. Hence $(\text{player 1 attaches probability 1 to player 2 attaching probability 1 to})^n$ to player 1's beliefs being $(2/3, 1/3)$ and player 2's beliefs being $(1/3, 2/3)$ as long as $n \leq N$.

Note that the new partitions model is "like" the old one in the sense that beliefs at certain states are the same as beliefs in the original model up to a high but finite order. On the other hand, the *ex ante* models are quite different. In the original model, there is zero probability that, say, player 1 puts probability 1 on s_2. Yet in the new model, there is a positive probability of this occurrence. Put differently, the notion of "closeness" here is an interim notion — focusing on the closeness of actual worlds, rather than the closeness of *ex ante* models.

A related point is that while the new model matches statements about mutual knowledge to a very high degree, the fact that it does so only up to a *finite* degree means that statements about common knowledge are not matched. In particular, in the original model, beliefs over S are common knowledge. In the new model, both players know that both know that ... that both know the beliefs over S for a very large number of "know that" 's. But the beliefs are *not* common knowledge.

A natural question to ask is whether this approach can be continued for all orders instead of only up to some finite upper bound. One answer is simply that extending the construction above does not work. If we replace K with ∞, we end up with a prior which is uniform on a countable set — an impossibility.

More generally, we know from Mertens–Zamir that no such construction is possible. As discussed in Section 1, Mertens–Zamir showed that if we specify beliefs about beliefs for all finite orders, this completely determines whether or not the beliefs of the players are consistent with the common prior assumption. Hence if they are not, there is no way to change the model to one with a common prior and simultaneously match *all* orders of beliefs.[5]

3 The Result

I will say that a partitions model $(S, \Omega, f, \{(\Pi_i, \mu_i)\}_{i \in I})$ is *finite* if Ω is finite and *countable* if Ω is finite or countably infinite. Given any probability distribution μ on Ω, let $\mathrm{supp}(\mu)$ denote the support of μ.

Definition 1 *A partitions model is* weakly consistent *if* $\mathrm{supp}(\mu_i) = \mathrm{supp}(\mu_j)$ *for all* $i, j \in I$.

In the introduction, I defined weak consistency as common knowledge that all players' beliefs at all levels give positive probability to the truth. Certainly, a common support seems to capture this idea to some extent. A more formal link is given in Section 4.1 where I show that weak consistency as defined here for the partitions model is equivalent to weak consistency as defined in the introduction for the equivalent belief hierarchies.

It is easy to see that weak consistency means that I may as well assume that every agent puts positive probability on every state.[6] Otherwise, I can restrict attention to the subset of Ω which is the common support for all agents' beliefs — and the fact that Ω is countable will mean that this support is countable also.

---

[5] There is one way that this might be possible. Heifetz [5] has shown that Mertens–Zamir rely on an assumption that we might term common knowledge of countable additivity. More specifically, he shows that while there is a unique countably additive belief which extends all finite orders of belief, there may be many finitely additive extensions. In light of this, it is conceivable that there are finitely additive extensions which are consistent with common priors even when the unique countably additive extension is not. While my construction does not seem to converge in any meaningful sense to such finitely additive extensions, it is possible that some other construction would yield such a "continuity" result.

[6] This overlooks the issue of how we would define beliefs on events with prior probability zero. If we simply require all events to have positive prior probability (and if we only care about conditional probabilities, the fact that Ω will be assumed to be countable means that we may as well assume this), then states all agents give zero prior probability to can be ignored.

Theorem 1 *Let $\mathcal{M}$ denote any weakly consistent countable partitions model. For any finite N, there exists a countable partitions model $\mathcal{M}^N$ in which*

1. *the common prior assumption is satisfied and*
2. *for every $\omega \in \mathcal{M}$, there is an $\omega^N \in \mathcal{M}^N$ such that n^{th} order beliefs at ω in $\mathcal{M}$ are the same as n^{th} order beliefs at ω^N in $\mathcal{M}^N$ for all $n \leq N$.*

The proof, which is contained in the Appendix, is tedious but not difficult. It provides a construction which generalizes the construction in the example of Section 2.2.

4 Implications

The result has a variety of implications which I now discuss in turn.

4.1 Implications for the Structure of the Universal Beliefs Space

The main result, as stated in the introduction, referred to infinite hierarchies of beliefs and used the assumption that it is common knowledge that all players give positive probability to the true world. But Theorem 1 was a result about partitions models and assumed that the players' beliefs have the same support. In this section, I first show that the statement in the introduction is, in fact, equivalent to Theorem 1. Once the theorem is restated in terms of belief hierarchies, it is not difficult to use it to show that every world in the Mertens–Zamir universal beliefs space is arbitrarily close to a world which satisfies in the common prior assumption. That is, the set of worlds satisfying the common prior assumption is dense in the universal beliefs space.

I will try to minimize the extra notation needed to make these statements precise and so will refer to Mertens–Zamir [6] for numerous technical details which are not directly relevant for my purposes. As before, we have a set of players $\{1, \ldots, I\}$ and a parameter space S. I assume S is compact. For any compact set X, let $\Delta(X)$ denote the set of probability measures on X endowed with the weak* topology. This set is compact as well. Then we can define an infinite sequence of sets recursively by

$$\begin{aligned} X_0 &= S \\ T_{k+1} &= \Delta(X_k) \\ X_{k+1} &= X_k \times [T_{k+1}]^I \end{aligned}$$

To understand this definition, note that $T_1 = \Delta(X_0) = \Delta(S)$ is the set of first–order beliefs for an agent. Hence X_1 is just S times the set of first–order beliefs for each agent. Hence this is the set over which second–order beliefs are defined, so $T_2 = \Delta(X_1)$ is the set of second–order beliefs, etc. Let $X = S \times \prod_{k=1}^{\infty} [T_k]^I$. This is a well–defined, compact space.

The universal beliefs space, which I denote $\mathcal{U}$, is the largest subset of X which, in the terminology of Brandenburger and Dekel [3], satisfies common knowledge of coherence. The exact definition of this term is irrelevant for my purposes. I will refer to a point in $\mathcal{U}$ as a *world*.

The main result of Mertens–Zamir is that there exists a set of *types*, $\mathcal{T}$, such that $\mathcal{U}$ is homeomorphic to $S \times \mathcal{T}^I$ and $\mathcal{T}$ is homeomorphic to $\Delta(S \times \mathcal{T}^{I-1})$. In other words, we can think of a world as specifying the true value of the unknown parameter and the type of each player where a type is a probability distribution on S and the types of the other players. Equivalently, a type for player i is a probability distribution on the set of worlds with the property that player i puts probability 1 on his own true type.

If some player i at world u has a belief which contains u' in its support, I will say that u' is *believed possible* by i at u. A set of worlds $U \subseteq \mathcal{U}$ is *belief–closed* if for every $u \in U$ and every u' believed possible by some player at u, we have $u' \in U$. In other words, every world believed possible by someone at a world in U is also in U.

Given a world, $u \in \mathcal{U}$, the *belief–closed subspace generated by u* is the smallest belief–closed set containing all worlds believed possible by any player at u.

My analysis makes use of four facts. First, as discussed in Section 2.1, any partitions model together with any state in that model uniquely identifies a particular world in the universal beliefs space by the unravelling procedure described earlier.

Second, as alluded to in Section 2.1, any belief–closed subspace U of $\mathcal{U}$ generates a partitions model. More specifically, if U is countable, we can find a partitions model with the property that the state set in the partitions model is one–to–one with U and where the infinite hierarchy of beliefs for any given player at a given state (defined by unravelling) is precisely the infinite hierarchy of that player at the corresponding world. (When U is uncountable, we need σ–fields in place of partitions, but this issue is irrelevant for my purposes.) When a partitions model $\mathcal{M}$ has this relationship to a belief–closed set U, I will say that $\mathcal{M}$ and U are *equivalent*. Similarly, I will say that a state ω in $\mathcal{M}$ is *equivalent* to a world $u \in U$ if the unraveling of ω generates u.

Third, as noted, $\mathcal{U}$ is a subset of X. In other words, it is a subset of an infinite product space. For this reason, it is natural to follow Mertens–Zamir in using for a topology on $\mathcal{U}$ the (relativized) product topology.

The final fact I require is a trivial implication of Mertens–Zamir's Theorem 3.1. The result is that for every $u \in \mathcal{U}$, there exists a sequence $\{u^N\}$ in $\mathcal{U}$ converging to u such that every u^N generates a finite belief–closed set. Less formally, any world $u \in \mathcal{U}$ is arbitrarily close to a world which generates a finite belief–closed subspace.

I now give a more precise definition of a weakly consistent world in preparation for restating Theorem 1 for belief hierarchies.

Definition 2 *Let $u \in \mathcal{U}$ generate belief–closed subspace U. Then u is weakly consistent if for all $u' \in U \cup \{u\}$, every player believes u' possible at u'.*

Intuitively, if every player believes the true world possible, this means that each player's first–order beliefs contain the true parameter in its support, each player's second–order beliefs contain the pair giving the true parameter value and true first–order beliefs in its support, etc. If this holds at every world in a belief–closed set, it means that every player knows that this holds, every player knows that every player knows it, etc. In other words, as stated in the introduction, weak consistency simply means that it is common knowledge that every player puts positive probability on the truth at every level of belief.

The following lemma uses the first two facts above and enables me to relate Theorem 1 to the version of the result stated in the introduction. The proof is in the Appendix.

Lemma 1 *If $u \in \mathcal{U}$ is weakly consistent and generates a countable belief–closed subspace U, then U is equivalent to a weakly consistent countable partitions model.*

One more definition is required.

Definition 3 *A world $u \in \mathcal{U}$ is* consistent with the common prior assumption *if the belief–closed subspace it generates is equivalent to a partitions model with common priors.*

Theorem 2 *If $u \in \mathcal{U}$ is weakly consistent and generates a countable belief–closed set U, then there is a sequence of worlds $\{u^N\}$ converging to u such that each u^N is consistent with the common prior assumption.*

Proof. Fix such a u and let U be the belief–closed set it generates. By Lemma 1, it has an equivalent partitions model, say $\mathcal{M}$, which is weakly consistent. Obviously, $\mathcal{M}$ is countable since Ω and U must be one–to–one. Let ω be the state in $\mathcal{M}$ which is equivalent to u. By Theorem 1, for any N, we can find a partitions model satisfying common priors and a state ω^N

in that model such that the n^{th} order beliefs at ω^N are the same as those at ω for all $n \leq N$. Let u^N be the world in U which is generated from ω^N by unraveling. By definition, u^N is consistent with the common prior assumption. Because u^N is the world generated by ω^N, u^N has the same parameter value as u and has the same n^{th} order beliefs for each player as u for all $n \leq N$. Hence u^N converges to u pointwise as $N \to \infty$. That is, $u^N \to u$. ▌

In other words, every weakly consistent world which generates a countable belief–closed set is arbitrarily close to a world with common priors.

In fact, every world is arbitrarily close to one which is weakly consistent and generates a finite belief–closed set. Hence *every* world is arbitrarily close to one consistent with common priors. The proof of the following lemma is in the Appendix.

Lemma 2 *For any world $u \in U$, there exists a sequence of worlds $\{u^N\}$ converging to u such that each u^N is weakly consistent and generates a finite belief–closed set.*

Putting these results together gives

Theorem 3 *For any world $u \in U$, there exists a sequence of worlds $\{u^N\}$ converging to u such that each u^N is consistent with the common prior assumption. Hence the closure of the set of worlds consistent with common priors is U.*

Proof. Fix any world $u \in U$. By Lemma 2, we know that this world is arbitrarily close to a weakly consistent world generating a finite belief–closed subspace. By Theorem 2, each such world is arbitrarily close to a world which is consistent with common priors. ▌

4.2 Methodological Implications

In a sense, this result says that dropping the common prior assumption adds (weakly) fewer possibilities than dropping common knowledge assumptions. In other words, any prediction we could obtain from a model with common knowledge and noncommon priors could also be generated without common knowledge but with common priors, while the reverse may not be true. In this sense, if we wish to generalize from the usual common knowledge and common prior assumptions but to maintain as much predictive power as possible, it is better to drop the common prior assumption than the common knowledge assumptions. To see the point, suppose we knew that a particular outcome was possible in some state where certain facts were common knowledge but priors were not the same. If these priors satisfy weak consistency, then it would necessarily be true that this same outcome would be possible in a model with common priors but where the common knowledge facts are now only mutual knowledge of some high but finite order.

Another methodological implication: If we are willing to assume weak consistency and only care about beliefs up to some finite order, then there is no reason not to assume common priors. Put differently, any (local) result which only depends on finitely many orders of belief and which is proved with the common prior assumption is also true if we replace the common prior assumption by weak consistency.

By "local," I mean a result which is true about a state, not about the model as a whole. For example, Aumann's [1] agreeing to disagree result is a local result. The theorem hypothesizes that posterior beliefs are common knowledge at a state and asks what else must be true at that state. Similarly, no–trade theorems begin with the hypothesis that it is common knowledge at a state that a particular trade is mutually beneficial and characterize other things which must be true at that state. By contrast, Aumann's [2] result characterizing correlated equilibrium is a global result since it says what must be true in expectation over the entire set of states. My result says nothing about global results since the translation I do requires changing global properties of the model in order to preserve local properties.

Rephrasing this last methodological implication, we see that any result which requires common priors must depend on infinitely many orders of beliefs and hence on common knowledge assumptions. For example, the fact that the agreeing to disagree result and no–trade theorems don't hold without common priors[7] indicates that these results must also rely on their common knowledge assumptions, a fact which is well–known.

4.3 Interpretation of the Common Prior Assumption

Given the controversial nature of the common prior assumption, it is natural to ask whether my results are a defense or a criticism of the assumption. I believe one can find ammunition for either position.

The points which support the common prior assumption are obvious. First, if we are only interested in finitely many orders of beliefs about beliefs, then the common prior assumption is really no stronger than weak consistency. Second, there is a sense in which every world is arbitrarily close to one in which the common prior assumption holds.

On the other hand, we typically are *not* only interested in finitely many orders of beliefs about beliefs. If we restrict attention to finitely many orders, we cannot discuss common knowledge, making it obvious that this is quite a severe restriction. Similarly, the "closeness"

---

[7] See Morris [8] for a characterization how relaxations of the common prior assumption affect no–trade results.

of a world with common priors is misleading: the topology in which we obtain this result is not a topology in which our results are continuous. In other words, even though all possible beliefs hierarchies are “close” to satisfying the common prior assumption, this does not say that all behavior based on belief hierarchies is “close” to the behavior predicted by common priors.

One implication of the result can be taken as a criticism of the common prior assumption. The result clearly shows that the common prior assumption is, primarily, an “infinite order” restriction. That is, since it only imposes a minimal restriction on beliefs up to any finite order, the primary bite of the assumption is on the full infinite hierarchy of beliefs. Since we know that the common prior assumption has quite a lot of bite to it, the infinite order restrictions must be strong. This fact makes clear that a (nontrivial) characterization of what the common prior assumption means in terms of the hierarchies of beliefs at the actual world is likely to be extremely difficult and perhaps uninterpretable.

A Proof of Theorem 1

Fix an arbitrary integer N. As noted in the text, weak consistency means that we may as well assume $\mu_i(\omega) > 0$ for all $\omega \in \Omega$ and all $i \in I$. For each ω, define a permutation $\iota(\omega) = (\iota_1(\omega), \ldots, \iota_I(\omega))$ of $\{1, \ldots, I\}$ such that

$$\mu_{\iota_j(\omega)}(\omega) \leq \mu_{\iota_{j+1}(\omega)}(\omega), \quad j = 1, \ldots, I-1.$$

In other words, $\iota(\omega)$ renumbers the agents so that lower numbered agents give less prior probability to ω.

Define

$$\ell_1(\omega) = \mu_{\iota_1(\omega)}(\omega)$$

and for $j = 2, \ldots, I$,

$$\ell_j(\omega) = \mu_{\iota_j(\omega)}(\omega) - \mu_{\iota_{j-1}(\omega)}(\omega).$$

Let

$$\max(\omega) = \sum_{j=1}^{I} \ell_j(\omega) = \max\{\mu_1(\omega), \ldots, \mu_I(\omega)\}.$$

It is useful to define a function $j(i,\omega)$ giving the "rank" ι assigns to i for ω — that is, $j(i,\omega)$ is defined by

$$\iota_{j(i,\omega)}(\omega) = i.$$

Note for future use that

$$\sum_{j=1}^{j(i,\omega)} \ell_j(\omega) = \mu_i(\omega).$$

Also, let

$$\beta = 1 + \max_{\omega \in \Omega} \frac{\max(\omega) - \ell_1(\omega)}{\ell_1(\omega)}.$$

Let

$$\bar{\Omega}^N = \Omega \times \{1, \ldots, NI\}.$$

Naturally, let $\bar{f}^N(\omega, k) = f(\omega)$. The common prior $\bar{\mu}^N$ is given by

$$\bar{\mu}^N(\omega, k) = \alpha z_k(\omega)$$

where $\alpha > 0$ is a constant to be determined and $z_k(\omega)$ is defined as follows. For $k = 1, \ldots, I$,

$$z_k(\omega) = \ell_k(\omega)$$

Also,

$$z_{I+1}(\omega) = (\beta + 1)\ell_1(\omega) - \max(\omega).$$

Finally, for $k \geq I+2$,

$$z_k(\omega) = (\beta+1)z_{k-I}(\omega).$$

If all the z_k's are positive, then we can choose α so that the $\overline{\mu}^N$'s sum to 1 and ensure that this is a probability distribution. By construction, $\ell_k(\omega) \geq 0$ for all k, strictly so for $k=1$. Hence this is a legitimate probability distribution as long as

$$\beta\ell_1(\omega) \geq \max(\omega) - \ell_1(\omega)$$

which is guaranteed for all ω from the definition of β.

The partitions are as follows. First,

$$\overline{\Pi}_i^N(\omega', 1) = \left\{(\omega, k) \mid \omega \in \Pi_i(\omega') \text{ and } k \leq j(\omega, i)\right\}.$$

For $n = 1, \ldots, N-1$,

$$\overline{\Pi}_i^N(\omega', (n-1)I + j(i,\omega') + 1) = \left\{(\omega, k) \;\middle|\; \begin{array}{l} \omega \in \Pi_i(\omega') \text{ and} \\ k \in \{(n-1)I + j(i,\omega)+1, \ldots, nI + j(i,\omega)\} \end{array} \right\}.$$

Finally, for ω' such that $j(i,\omega') < I$,

$$\overline{\Pi}_i^N(\omega', (N-1)I + j(i,\omega') + 1) = \{(\omega, k) \mid j(i,\omega) < I \text{ and } k \in \{(N-1)I + j(i,\omega)+1, \ldots, NI\}\}.$$

It is not difficult to verify that the events so defined form a partition of $\overline{\Omega}^N$. Let M^N denote the partitions model so constructed.

The key property of this construction is that certain conditional beliefs are preserved. The following definition specifies the events for which these conditional beliefs are appropriately preserved.

Definition 4 *An event $E \subseteq \Omega^N$ is* Ω–measurable *if there exists $B \subseteq \Omega$ such that*

$$E = \{(\omega, k) \in \overline{\Omega}^N \mid \omega \in B\}.$$

In this case, call B the Ω projection *of E. An event $E \subseteq \Omega^N$ is* conditionally Ω–measurable for i at (ω, k) *if there exists an Ω–measurable event F such that*

$$E \cap \overline{\Pi}_i^N(\omega, k) = F \cap \overline{\Pi}_i^N(\omega, k).$$

In this case, define the Ω projection of E to be the Ω projection of F.

Intuitively, if an event E in $\overline{\Omega}^N$ is of the form $\{(\omega, k) \mid \omega \in B\}$, then it naturally corresponds to the event B in Ω — hence the name Ω–measurable. Even if E is not Ω–measurable, it may be true that conditional on $\imath$'s information, E may as well be in the sense that the conditional probability $\imath$ attaches to E is the same as what he would give to some event which is Ω–measurable.

The construction given above guarantees that conditional probabilities attached to the conditionally Ω–measurable events correspond appropriately to the conditional probabilities given the analogous events in Ω. The following lemma states this more precisely.

Lemma 3 *For all* $\omega, \omega' \in \Omega$*, all* i*, and all* $k' \leq (N-1)I + j(i,\omega') + 1$*,*

$$\sum_k \bar{\mu}^N[(\omega,k) \mid \bar{\Pi}_i^N(\omega',k')] = \mu_i[\omega \mid \Pi_i(\omega')].$$

Hence for any i*, any* ω'*, any* $k' \leq (N-1)I + j(i,\omega') + 1$*, and any event* E *which is conditionally* Ω*–measurable by* i *at* (ω',k')*, the probability* i *gives to* E *at* (ω',k') *in* M^N *is the same as the probability* i *gives to the* Ω *projection of* E *at* ω' *at* M*.*

Proof of Lemma. First, note that if $\omega \notin \Pi_i(\omega')$, then the statement of the lemma holds trivially since both sides of the equation are zero. So suppose $\omega \in \Pi_i(\omega')$. Then we wish to show that

$$\sum_{k \in R_i(\omega',k')} \frac{\bar{\mu}^N(\omega,k)}{\sum_{(\omega'',k'') \in \bar{\Pi}_i^N(\omega',k')} \bar{\mu}^N(\omega'',k'')} = \frac{\mu_i(\omega)}{\sum_{\omega'' \in \Pi_i(\omega')} \mu_i(\omega'')} \tag{1}$$

where $R_i(\omega',k')$ is the set of k such that $(\omega,k) \in \bar{\Pi}_i^N(\omega',k')$. (Note that this is independent of ω as long as $\omega \in \Pi_i(\omega')$.) To show this, first consider $k' \leq j(i,\omega')$. For k' in this range, the left–hand side of (1) is

$$\begin{aligned}
\sum_{k=1}^{2N} \bar{\mu}^N[(\omega,k) \mid \bar{\Pi}_1^N(\omega',1)] &= \frac{\sum_{k=1}^{j(i,\omega)} \alpha z_k(\omega)}{\sum_{\omega'' \in \Pi_i(\omega')} \sum_{k=1}^{j(i,\omega'')} \alpha z_k(\omega'')} \\
&= \frac{\mu_i(\omega)}{\sum_{\omega'' \in \Pi_i(\omega')} \mu_i(\omega'')},
\end{aligned}$$

as was to be shown.

For k' strictly between $j(i,\omega')$ and $(N-1)I + j(i,\omega') + 1$, the event $\bar{\Pi}_i^N(\omega',k')$ takes the form $\bar{\Pi}_i^N(\omega',(n-1)I + j(i,\omega') + 1)$ for some n in $\{1,\ldots,N-1\}$. Analogously to the above, for $\omega \in \Pi_i(\omega')$,

$$\sum_{k=1}^{2N} \bar{\mu}^N[(\omega,k) \mid \bar{\Pi}_1^N(\omega',(n-1)I + j(i,\omega') + 1)] = \frac{\sum_{k=(n-1)I+j(i,\omega)+1}^{nI+j(i,\omega)} \alpha z_k(\omega)}{\sum_{\omega'' \in \Pi_i(\omega')} \sum_{k=(n-1)I+j(i,\omega'')+1}^{nI+j(i,\omega'')} \alpha z_k(\omega'')}$$

After cancelling α from numerator and denominator, the numerator is

$$(\beta+1)^{n-1}\sum_{k=j(i,\omega)+1}^{j(i,\omega)+I} z_k(\omega) = (\beta+1)^{n-1}\left\{\sum_{k=j(i,\omega)+1}^{I}\ell_k(\omega)+z_{I+1}(\omega)+(\beta+1)\sum_{k=2}^{j(i,\omega)}\ell_k(\omega)\right\}$$

where, to cover the case of $j(i,\omega)=1$, we treat $\sum_{k=2}^{1}x_k=0$ for any x_k sequence. (The denominator is analogous, where ω'' replaces ω and we sum over ω'' in $\Pi_i(\omega')$.) Substituting for $z_{I+1}(\omega)$ on the right–hand side and rearranging gives

$$\begin{aligned}
&(\beta+1)^{n-1}\left\{\sum_{k=j(i,\omega)+1}^{I}\ell_k(\omega)+(\beta+1)\sum_{k=1}^{j(i,\omega)}\ell_k(\omega)-\max(\omega)\right\}\\
&\qquad = (\beta+1)^{n-1}\left\{\sum_{k=1}^{I}\ell_k(\omega)+\beta\sum_{k=1}^{j(i,\omega)}\ell_k(\omega)-\max(\omega)\right\}\\
&\qquad = (\beta+1)^{n-1}\beta\sum_{k=1}^{j(i,\omega)}\ell_k(\omega)\\
&\qquad = (\beta+1)^{n-1}\beta\mu_i(\omega).
\end{aligned}$$

Hence

$$\sum_{k=1}^{2N}\bar{\mu}^N[(\omega,k)\mid\bar{\Pi}_1^N(\omega',(n-1)I+j(i,\omega')+1)]=\frac{\mu_i(\omega)}{\sum_{\omega''\in\Pi(\omega')}\mu_i(\omega'')},$$

as was to be shown. ▮

I now use this lemma to show that for any $\omega'\in\Omega$, n^{th} order beliefs at state $(\omega',1)$ in M^N are the same as n^{th} order beliefs at ω' in M for all $n\le N$. I show this by demonstrating the stronger statement that for all $n\le N$, all i, all $\omega'\in\Omega$, and all $k'\le(N-n)I+j(i,\omega')$, $\imath$'s n^{th} order beliefs at ω' in M are the same as his n^{th} order beliefs at (ω',k') in M^N. This is shown by simply establishing that all relevant events in $\bar{\Omega}^N$ are conditionally Ω–measurable and then appealing to Lemma 3.

The proof is by induction on n. So first consider $n=1$. Note that the value of s at a state (ω,k) is independent of k. That is, for any value of s, the event that the parameter takes on this value is Ω–measurable. Hence by Lemma 3, for all i, all ω', and all $k'\le(N-1)I+j(i,\omega')+1$, $\imath$'s first–order beliefs at (ω',k') must be the same as his first–order beliefs at ω'.

So suppose we have established that for all i, all $\omega'\in\Omega$, and all $k'\le(N-n)I+j(i,\omega')$, $\imath$'s n^{th} order beliefs at (ω',k') in M^N are the same as his n^{th} order beliefs at ω' in M for $n=1,\ldots,\bar{n}-1$ where $\bar{n}-1<N$. I now show that this implies that the same is true for $n=\bar{n}$.

So fix any i, any ω', and any $k'\le(N-\bar{n})I+j(i,\omega')$. Recall that $\imath$'s $\bar{n}^{\text{th}}$ order beliefs are a probability distribution on tuples consisting of s and the n^{th} order beliefs of the other players for all $n < \bar{n}$. Fix any such tuple and let E denote the event where the parameter and lower order beliefs take on this value. Suppose $(\omega, k) \in E \cap \overline{\Pi}_i^N(\omega', k')$. Clearly, then, at every (ω, k''), we have the same value of the parameter since its value is independent of k. Note from the construction of partitions, that we must have $k \leq (N - \bar{n})I + j(i, \omega')$. Since $j(i, \omega') \leq I$ for all i, we have

$$k \leq (N - \bar{n})I + j(i, \omega') \leq (N - \bar{n})I + I < (N - (\bar{n} - 1))I + j(j, \omega').$$

By the induction hypothesis, then, for any $j \neq i$, ■'s lower order beliefs at (ω, k) must be the same as his lower order beliefs at (ω, k'') for all $k'' \leq (N - \bar{n} + 1)I + j(j, \omega)$ since both must match his lower order beliefs at ω in the original model. Since, as shown above,

$$(N - \bar{n})I + j(i, \omega) < (N - \bar{n} + 1)I + j(j, \omega),$$

this implies that E is conditionally Ω–measurable for i at (ω', k'). Hence by Lemma 3, $\imath$'s $\bar{n}^{\text{th}}$ order beliefs at (ω', k') in M^N must equal his n^{th} order beliefs at ω' in M, completing the proof. ▮

B Proof of Lemma 1

Suppose u^* is weakly consistent and generates a countable belief–closed subspace U. I define a partitions model which is equivalent to it and is consistent. So let $\Omega = U$. Define f by setting $f(u)$ equal to the projection of u onto S for each $u \in U$. Define player $\imath$'s partition by setting $\Pi_i(u)$ equal to the set of $u' \in U$ such that the projection of u' onto the type space for i is the same as the projection of u. In other words, $u' \in \Pi_i(u)$ if and only if $\imath$'s type at u' is the same as his type at u. Obviously, this gives a partition for each player.

To define priors, first fix a probability distribution ϕ_i on Π_i for each i where we require $\phi_i(\pi) > 0$ for all $\pi \in \Pi_i$. Then define $\imath$'s prior probability on u to be $\phi_i(\Pi_i(u))$ times the probability his type at u puts on u. It is easy to see that this definition gives a partitions model which is equivalent to U. Also, by weak consistency, we know that every player i at every world u puts strictly positive probability on u. Hence $\mu_i(u) > 0$ for all i and all $u \in \Omega$. Hence the partitions model is weakly consistent. ▮

C Proof of Lemma 2

In light of Mertens–Zamir's Theorem 3.1, we only need to show that given any world generating a finite belief–closed subspace, there is another arbitrarily nearby which is weakly consistent. Let u^* denote a world which generates a finite belief–closed subspace and let U denote this subspace. Note that the finiteness of U for each player i, every $u \in U$, and every k, $\imath$'s k^{th} order beliefs at u have a finite support.

So fix a sequence $\{\varepsilon_n\}$ converging to zero from above. For each $u \in U$, I construct a sequence of worlds $\{u^n(u)\}$ as follows. Let $s(u)$ be the parameter value at u (the projection of u onto S). Let $s(u)$ also be the parameter value at $u^n(u)$ for all n and all u. For any player i, let $\imath$'s first–order beliefs at $u^n(u)$ be a convex combination of his first–order beliefs at u and a degenerate distribution with probability 1 on $s(u)$ with weight $1-\varepsilon_n$ on the former and ε_n on the latter.

Second–order beliefs are slightly more complex. Player $\imath$'s second–order beliefs at $u^n(u)$ give the pair consisting of $s(u)$ and the true first–order beliefs at $u^n(u)$ probability equal to ε_n plus $1-\varepsilon_n$ times the probability $\imath$'s second–order beliefs at u gave to the pair $s(u)$ and the true first–order beliefs at u. For the pair consisting of $s(u')$ and the first–order beliefs at $u^n(u')$, player $\imath$ gives probability $1-\varepsilon_n$ times the probability his second–order beliefs at u gave to $s(u')$ and the true first–order beliefs at u'. Probability 0 is given to any other pair.

We proceed analogously to this for higher levels. That is, player $\imath$'s k^{th} order beliefs at $u^n(u)$ give the tuple consisting of $s(u)$ and the true lower order beliefs at $u^n(u)$ probability equal to ε_n plus $1-\varepsilon_n$ times the probability his beliefs at u gave to the pair consisting of $s(u)$ and the true lower order beliefs at u. His beliefs give the tuple consisting of $s(u')$ and the true lower order beliefs at $u^n(u')$ probability equal to $1-\varepsilon$ times the probability his beliefs at u gave $s(u')$ and the true lower order beliefs at u'. Probability 0 is given to any other tuple.

It is easy to show that the following construction on the types space is equivalent. We simply define $u^n(u)$ to be the world with parameter $s(u)$ and where the type of player i is the one with beliefs equal to a convex combination of probability 1 on $u^n(u)$ and a distribution, say δ_i^n, described shortly, with weight ε_n on the former and $1-\varepsilon_n$ on the latter. The distribution δ_i^n simply gives $u^n(u')$ the same probability that $\imath$'s type gave u' at world u.

By this construction, we see that these beliefs do indeed satisfy common knowledge of coherence in the sense of Brandenburger and Dekel [3], so $u^n(u) \in U$ for all n and all $u \in U$. The construction shows that each $u^n(u)$ has the property that every player believes $u^n(u)$ possible at $u^n(u)$ and that for each n, $U^n = \{u^n(u) \mid u \in U\}$ is belief–closed. Hence for every n and $u \in U$, $u^n(u)$ is weakly consistent.

It is not hard to show by induction that $u^n(u) \to u$ as $n \to \infty$. To see this, note that for any player i and any u, $\imath$'s first–order beliefs along the sequence $\{u^n(u)\}$ always have a finite support. For any point, say s', in the support of $\imath$'s first–order beliefs at u, there is a sequence of points, say $\{s^n\}$ in the supports of his beliefs at $\{u^n(u)\}$ converging to s' and the sequence of probabilities given to s^n converges to the probability given s' at u. This is, of course, sufficient for convergence in the weak* topology. Hence first–order beliefs for every player converge. But then this establishes precisely the analogous property for second–order beliefs, so that for every player i, *ı*’s second–order beliefs along the sequence $\{u^n(u)\}$ converge to his second–order beliefs at u, etc. Hence $u^n(u)$ converges pointwise — and therefore in the product topology — to u. ▮

References

[1] Aumann, R., “Agreeing to Disagree,” *Annals of Statistics*, **4**, 1976, pp. 1236–1239.

[2] Aumann, R., “Correlated Equilibrium as an Expression of Bayesian Rationality,” *Econometrica*, **55**, January 1987, pp. 1–18.

[3] Brandenburger, A., and E. Dekel, “Hierarchies of Beliefs and Common Knowledge,” *Journal of Economic Theory*, **59**, February 1993, pp. 189–198.

[4] Dekel, E., and F. Gul, “Rationality and Knowledge in Game Theory,” working paper, 1995.

[5] Heifetz, A., “Non–Wellfounded Type Spaces,” working paper, 1995.

[6] Mertens, J.-F., and S. Zamir, “Formalization of Bayesian Analysis for Games with Incomplete Information,” *International Journal of Game Theory*, **14**, 1985, pp. 1–29.

[7] Milgrom, P., and N. Stokey, “Information, Trade, and Common Knowledge,” *Journal of Economic Theory*, **26**, 1982, pp. 17–27.

[8] Morris, S., “Trade with Heterogeneous Prior Beliefs and Asymmetric Information,” *Econometrica*, **62**, November 1994, pp. 1327–1347.

[9] Morris, S., “The Common Prior Assumption in Economic Theory,” *Economics and Philosophy*, forthcoming.

[10] Nyarko, Y., “Most Games Violate the Harsanyi Doctrine,” New York University working paper, 1991.

[11] Tan, T., and S. Werlang, “On Aumann’s Notion of Common Knowledge — An Alternative Approach,” working paper, 1988.