ECON 289, Lecture 2
Epistemic game theory
(or, How I learned to stop worrying and love the common prior)

Instructor: Ben Brooks

Spring, 2023

Modeling decision-making under uncertainty

- Consider a decision maker choosing some alternative $a \in A$
- The value a depends on some **state of the world** $\theta \in \Theta$
- The most widespread and enduring model of such an agent's preferences is the **subjective expected utility model** (Savage, 1954)
- The agent has
  1. A **belief** $\phi \in \Delta(\Theta)$;
  2. A **utility index** $u : A \times \Theta \to \mathbb{R}$
- Preferences are represented by:

$$\int_{\theta \in \Theta} u(a, \theta)\phi(d\theta)$$

Modeling strategic decision-making under uncertainty

- In strategic settings, players have preferences over everyone's actions
- Players $i = 1, \ldots, n$
- Player i has actions A_i, with $A = \prod_{i=1}^{n} A_i$
- Can still model preferences with a belief $\phi_i \in \Delta(\Theta)$ and preferences $u_i : A \times \Theta \to \mathbb{R}$, and expected utility:

$$\int_{\theta \in \Theta} u_i(a, \theta)\phi_i(d\theta)$$

Modeling strategic decision-making under uncertainty

- In strategic settings, players have preferences over everyone's actions
- Players $i = 1, \ldots, n$
- Player i has actions A_i, with $A = \prod_{i=1}^{n} A_i$
- Can still model preferences with a belief $\phi_i \in \Delta(\Theta)$ and preferences $u_i : A \times \Theta \to \mathbb{R}$, and expected utility:

$$\int_{\theta \in \Theta} u_i(a, \theta) \phi_i(d\theta)$$

- With multiple agents, a new complication:
  Can have $\phi_i \neq \phi_j$, and since u_i depends on a_j, player i "cares" about j's beliefs, since those determine a_j
- To close the model, need to model i's beliefs about ϕ_j

Modeling strategic decision-making under uncertainty

- In strategic settings, players have preferences over everyone's actions
- Players $i = 1, \ldots, n$
- Player i has actions A_i, with $A = \prod_{i=1}^{n} A_i$
- Can still model preferences with a belief $\phi_i \in \Delta(\Theta)$ and preferences $u_i : A \times \Theta \to \mathbb{R}$, and expected utility:

$$\int_{\theta \in \Theta} u_i(a, \theta)\phi_i(d\theta)$$

- With multiple agents, a new complication:
  Can have $\phi_i \neq \phi_j$, and since u_i depends on a_j, player i "cares" about j's beliefs, since those determine a_j
- To close the model, need to model i's beliefs about ϕ_j
- Moreover, a_j depends on what player j thinks about a_i, which depends on what j thinks about ϕ_i, etc

Modeling strategic decision-making under uncertainty

- In strategic settings, players have preferences over everyone's actions
- Players $i = 1, \ldots, n$
- Player i has actions A_i, with $A = \prod_{i=1}^{n} A_i$
- Can still model preferences with a belief $\phi_i \in \Delta(\Theta)$ and preferences $u_i : A \times \Theta \to \mathbb{R}$, and expected utility:

$$\int_{\theta \in \Theta} u_i(a, \theta)\phi_i(d\theta)$$

- With multiple agents, a new complication:
  Can have $\phi_i \neq \phi_j$, and since u_i depends on a_j, player i "cares" about j's beliefs, since those determine a_j
- To close the model, need to model i's beliefs about ϕ_j
- Moreover, a_j depends on what player j thinks about a_i, which depends on what j thinks about ϕ_i, etc
- Leads to an infinite regress in which preferences and strategic behavior could depend all **higher order beliefs**

Conceptual challenges

- One approach: explicitly model the belief hierarchy:
  1. Each player i's beliefs about θ;
  2. Each player i's beliefs about others' beliefs about θ;
  3. Beliefs about others' beliefs about others' beliefs, etc
- Would be very cumbersome

Conceptual challenges

- One approach: explicitly model the belief hierarchy:
  1. Each player i's beliefs about θ;
  2. Each player i's beliefs about others' beliefs about θ;
  3. Beliefs about others' beliefs about others' beliefs, etc
- Would be very cumbersome
- Also, it immediately leads to the following issue: Suppose that each player i is certain that
  1. $\theta = 1$;
  2. each player j is certain that $\theta = 0$

  What will happen?

Conceptual challenges

- One approach: explicitly model the belief hierarchy:
  1. Each player i's beliefs about θ;
  2. Each player i's beliefs about others' beliefs about θ;
  3. Beliefs about others' beliefs about others' beliefs, etc
- Would be very cumbersome
- Also, it immediately leads to the following issue: Suppose that each player i is certain that
  1. $\theta = 1$;
  2. each player j is certain that $\theta = 0$

  What will happen?
- The model is **incomplete** in the sense that the player i in our model is different from the player i that j thinks they are playing against
- In particular, the behavior that the analyst imputes for player i need not coincide with how j expects i to behave

Harsanyi's insight

- Harsanyi (1967) proposed a modeling device that addresses both of these issues:
  - Model hierarchies **implicitly** to avoid the cumbersome notation of higher order beliefs
  - **Complete** the model, by allowing for multiple **types** of each player, as an (implicit) description of every hierarchy of beliefs that players might impute to one another

Harsanyi's insight

- Harsanyi (1967) proposed a modeling device that addresses both of these issues:
  - Model hierarchies **implicitly** to avoid the cumbersome notation of higher order beliefs
  - **Complete** the model, by allowing for multiple **types** of each player, as an (implicit) description of every hierarchy of beliefs that players might impute to one another
- Formally, player i has a type t_i in a measurable set T_i
- Each type has a **belief** $\pi_i : T_i \to \Delta\left(\Theta \times T_{-i}\right)$ where $T_{-i} = \prod_{j \neq i} T_j$
- $\mathcal{T} = (\Theta, \{T_i, \pi_i\}_{i=1}^{n})$ is a **type space**
  (Also referred to as an **information structure**; types are sometimes referred to as **signals**, and Harsanyi called them "attribute vectors")

Completing the model

- Consider the example from two slides ago: Players are certain that $\theta = 0$ and that everyone else is certain that $\theta = 1$
- How do we model this with a Harsanyi type space?

Completing the model

- Consider the example from two slides ago: Players are certain that $\theta = 0$ and that everyone else is certain that $\theta = 1$
- How do we model this with a Harsanyi type space?
  - $T_i = \{0, 1\}$ for $i = 1, 2$
  - $\pi_i(\theta = 1, t_j = 0 | t_i = 1) = 1$
  - $\sum_{t_j} \pi_i(\theta = 0, t_j | t_i = 0) = 1$

Completing the model

- Consider the example from two slides ago: Players are certain that $\theta = 0$ and that everyone else is certain that $\theta = 1$
- How do we model this with a Harsanyi type space?
  - $T_i = \{0, 1\}$ for $i = 1, 2$
  - $\pi_i(\theta = 1, t_j = 0 | t_i = 1) = 1$
  - $\sum_{t_j} \pi_i(\theta = 0, t_j | t_i = 0) = 1$
- So, $t_i = 1$ corresponds to the types in our original description, and we have completed the model by adding $t_i = 0$

From types to beliefs

- NB by construction, the model is "closed": each type only assigns positive probability to types of other players that are part of the type space

From types to beliefs

- NB by construction, the model is "closed": each type only assigns positive probability to types of other players that are part of the type space
- Also, each type implies a particular belief hierarchy
- Why? Recall that $\pi_i(t_i) \in \Delta(\Theta \times T_{-i})$; the marginal on Θ gives the **first-order belief** $\phi_i^1(t_i)$:

$$\phi_i^1(Y|t_i) = \pi_i(Y \times T_{-i}|t_i) \ \forall Y \subseteq \Theta$$

From types to beliefs

- NB by construction, the model is "closed": each type only assigns positive probability to types of other players that are part of the type space
- Also, each type implies a particular belief hierarchy
- Why? Recall that $\pi_i(t_i) \in \Delta(\Theta \times T_{-i})$; the marginal on Θ gives the **first-order belief** $\phi_i^1(t_i)$:

$$\phi_i^1(Y|t_i) = \pi_i(Y \times T_{-i}|t_i)\ \forall Y \subseteq \Theta$$

- The **second-order belief** $\phi_i^2(t_i) \in \Delta(\Theta \times (\Delta(\Theta))^{n-1})$: For all $Y \subseteq \Theta \times (\Delta(\Theta)^{n-1})$,

$$\phi_i^2(Y|t_i) = \pi_i(\{(\theta, t_{-i})|(\theta, \phi_{-i}^1(t_{-i})) \in Y\}|t_i)$$

From types to beliefs

- NB by construction, the model is "closed": each type only assigns positive probability to types of other players that are part of the type space
- Also, each type implies a particular belief hierarchy
- Why? Recall that $\pi_i(t_i) \in \Delta(\Theta \times T_{-i})$; the marginal on Θ gives the **first-order belief** $\phi_i^1(t_i)$:

$$\phi_i^1(Y|t_i) = \pi_i(Y \times T_{-i}|t_i) \; \forall Y \subseteq \Theta$$

- The **second-order belief** $\phi_i^2(t_i) \in \Delta(\Theta \times (\Delta(\Theta))^{n-1})$: For all $Y \subseteq \Theta \times (\Delta(\Theta)^{n-1})$,

$$\phi_i^2(Y|t_i) = \pi_i(\{(\theta, t_{-i})|(\theta, \phi_{-i}^1(t_{-i})) \in Y\}|t_i)$$

- Inductively, one can define kth order beliefs in this manner (and we will do so shortly)

Bayesian games and Bayes Nash equilibrium

- Type spaces are an invaluable modeling tool
- Allow us to model Bayesian games in **(agent) normal form**:
- Player i's strategy is a measurable mapping $\sigma_i : T_i \to \Delta(A_i)$
- Set of strategies is Σ_i, strategy profiles $\Sigma = \prod_i \Sigma_i$
- For $\sigma \in \Sigma$, t_i's utility is

$$U_i(\sigma|t_i) = \int_{\theta\in\Theta, t_{-i}\in T_{-i}} \int_{a\in A} u_i(a,\theta)\sigma(da|t)\pi_i(d(\theta, t_{-i})|t_i)$$

- σ is a **Bayes Nash equilibrium** if for all i, $t_i \in T_i$, $\sigma_i' \in A_i$

$$U_i(\sigma|t_i) \geq U_i(\sigma_i', \sigma_{-i}|t_i)$$

Questions about Harsanyi type spaces

- Are Harsanyi type spaces general? Do they implicitly restrict belief hierarchies?

Questions about Harsanyi type spaces

- Are Harsanyi type spaces general? Do they implicitly restrict belief hierarchies?
- Also, predictions depend on the particular type space
- What predictions are consistent with **some** type space? (And indeed, Harsanyi's original proposal was to study games for all type spaces)
- The union of type spaces is a type space, but is there some "canonical" largest type space that we could consider?

Questions about Harsanyi type spaces

- Are Harsanyi type spaces general? Do they implicitly restrict belief hierarchies?
- Also, predictions depend on the particular type space
- What predictions are consistent with **some** type space? (And indeed, Harsanyi's original proposal was to study games for all type spaces)
- The union of type spaces is a type space, but is there some "canonical" largest type space that we could consider?
- Mertens and Zamir (1985) show that Harsanyi type spaces are general in a sense, in that every belief hierarchy (subject to mild restrictions) corresponds to a type in some type space
- They also construct a **universal type space** that "contains" every type space
- Our exposition is closer to Brandenburger and Dekel (1993) (We use the Kolmogorov Extension Theorem)

Explicitly modeling beliefs

- First-order beliefs are probability measures on $X_0 = \Theta$, the measurable set of states
- Second-order beliefs are probability measures on

  $$X_1 = X_0 \times (\Delta(X_0))^{n-1}$$

  i.e., a belief over the state and others' beliefs about the state

Explicitly modeling beliefs

- First-order beliefs are probability measures on $X_0 = \Theta$, the measurable set of states
- Second-order beliefs are probability measures on

$$X_1 = X_0 \times (\Delta(X_0))^{n-1}$$

  i.e., a belief over the state and others' beliefs about the state
- Third-order beliefs are over

$$\begin{aligned} X_2 &= X_1 \times (\Delta(X_1))^{n-1} \\ &= X_0 \times (\Delta(X_0))^{n-1} \times \left(\Delta\left(X_0 \times (\Delta(X_0))^{n-1}\right)\right)^{n-1}, \end{aligned}$$

  i.e., a belief about
  1. The state;
  2. Others beliefs' about the state;
  3. Others' beliefs about (the state and others' beliefs about the state)

Explicitly modeling beliefs

- First-order beliefs are probability measures on $X_0 = \Theta$, the measurable set of states
- Second-order beliefs are probability measures on

$$X_1 = X_0 \times (\Delta(X_0))^{n-1}$$

  i.e., a belief over the state and others' beliefs about the state
- Third-order beliefs are over

$$\begin{aligned} X_2 &= X_1 \times (\Delta(X_1))^{n-1} \\ &= X_0 \times (\Delta(X_0))^{n-1} \times \left(\Delta\left(X_0 \times (\Delta(X_0))^{n-1}\right)\right)^{n-1}, \end{aligned}$$

  i.e., a belief about
  1. The state;
  2. Others beliefs' about the state;
  3. Others' beliefs about (the state and others' beliefs about the state)
- Inductively, the $k+1$th order belief is on

$$X_k = X_{k-1} \times (\Delta(X_{k-1}))^{n-1}$$

Belief hierarchies

- An element of $\Delta(X_{k-1})$ is called a k**th order belief**: a probability measure over the state and others' beliefs up to the $k-1$th order
- A **belief hierarchy** is an element

$$t = (t^1, t^2, \ldots) \in \prod_{k=0}^{\infty} \Delta(X_k) \equiv T^0$$

Belief hierarchies

- An element of $\Delta(X_{k-1})$ is called a k**th order belief**: a probability measure over the state and others' beliefs up to the $k-1$th order
- A **belief hierarchy** is an element

$$t = (t^1, t^2, \ldots) \in \prod_{k=0}^{\infty} \Delta(X_k) \equiv T^0$$

- The X_k's and T^0 inherit "nice" properties of Θ
- E.g., if Θ is complete/separable/metric/compact $\implies$ X_k's and T^0 are complete/separable/metric/compact

Interpretation as a stochastic process

- Consider the product

$$
\begin{aligned}
&\Theta \times (T_0)^{n-1} \\
&= \Theta \times (\Delta(X_0) \times \Delta(X_1) \times \dots)^{n-1} \\
&= \underbrace{\underbrace{\underbrace{X_0 \times \Delta(X_0)^{n-1}}_{\equiv X_1} \times \Delta(X_1)^{n-1}}_{\equiv X_2} \times \dots \times \Delta(X_{k-1})^{n-1}}_{\equiv X_k} \times \Delta(X_k)^{n-1} \times \dots
\end{aligned}
$$

- You can think of elements of this product as possible sample paths of a discrete time process, where
  - at period $k = 0$ we realize a state and
  - at $k > 0$ we realize a profile of kth order beliefs (of other players)
- The kth order belief in X_{k-1} is a distribution over the first k elements of this sequence

Consistency

- Thus, a kth order belief t^k also specifies lower order beliefs as well, by taking the marginal on X_{k-2}, X_{k-3}, etc
- Clearly, if a hierarchy $t = (t^1, t^2, \ldots)$ is derived from some type space, then it has to be that the belief at each level is consistent with the beliefs at lower levels
- In particular, we say that t is **consistent** if for every $k \geq 2$,

$$\operatorname*{marg}_{X_{k-2}} t^k = t^{k-1}$$

- Let $T^1 \subset T^0$ denote the consistent belief hierarchies

Beliefs over types from belief hierarchies

Proposition (Consistent beliefs)

There exists a homeomorphism $f : T^1 \to \Delta(\Theta \times (T^0)^{n-1})$.

Beliefs over types from belief hierarchies

Proposition (Consistent beliefs)

There exists a homeomorphism $f : T^1 \to \Delta(\Theta \times (T^0)^{n-1})$.

- In other words, f associates to each type $t \in T^1$ a belief about the state and others' types (in T^0), and this mapping is a continuous bijection and has a continuous inverse
- NB similar to a Harsanyi type space! Consistent hierarchies are identified with beliefs about the state and others' hierarchies

Proof

- Recall, for $t \in T^1$, each t^k is an element of

$$\Delta(X_{k-1}) = \Delta(X_0 \times \Delta(X_0)^{n-1} \times \Delta(X_1)^{n-1} \times \ldots \times \Delta(X_{k-2})^{n-1})$$

- This is a consistent family of distributions over the first k realizations of a discrete time stochastic process in $\Delta(X_{k-1})^{n-1}$
- In this case, the **Kolmogorov extension theorem** says that there is a unique probability measure over sample paths, in T^0, that has the specified marginals, which we denote by $f(t)$

Proof

- Recall, for $t \in T^1$, each t^k is an element of

$$\Delta(X_{k-1}) = \Delta(X_0 \times \Delta(X_0)^{n-1} \times \Delta(X_1)^{n-1} \times \ldots \times \Delta(X_{k-2})^{n-1})$$

- This is a consistent family of distributions over the first k realizations of a discrete time stochastic process in $\Delta(X_{k-1})^{n-1}$
- In this case, the **Kolmogorov extension theorem** says that there is a unique probability measure over sample paths, in T^0, that has the specified marginals, which we denote by $f(t)$
- NB this relies on the space Θ being complete and separable, so that X_k is complete and separable in the weak-* topology (Mertens and Zamir just use compactness; Heifetz and Samet (1998) give a different non-topological construction)

Proof, continued

- Now, we need to check that f is a homeomorphism
- It must be one-to-one, because $t^k = \mathrm{marg}_{X_{k-1}} f(t)$, so if $f(t) = f(\hat{t})$, then $t^k = \hat{t}^k$ for every k, so $t = \hat{t}$
- Onto follows from defining $f^{-1}(t)$ as these marginals
- Finally, we need to prove continuity: Fix a sequence $(t_l)_{l=0}^{\infty}$ and t; then

$$
\begin{aligned}
& t_l \to t \\
\iff & t_l^k \to t^k \\
\iff & t_l^k = f^k(t_l) \to f^k(t) = t^k \\
\iff & f(t_l) \to f(t)
\end{aligned}
$$

- Hence, both f and its inverse are continuous $\square$

The homeomorphism is "natural"

- Note that $f(t)$ is a distribution over

$$\begin{aligned}\Theta \times (T^0)^{n-1} &= X_0 \times (\Delta(X_0) \times \Delta(X_1) \times \ldots)^{n-1} \\ &= X_0 \times \Delta(X_0)^{n-1} \times \Delta(X_1)^{n-1} \times \ldots\end{aligned}$$

The homeomorphism is "natural"

- Note that $f(t)$ is a distribution over

$$\begin{aligned}\Theta \times (T^0)^{n-1} &= X_0 \times \left(\Delta(X_0) \times \Delta(X_1) \times \ldots\right)^{n-1} \\ &= X_0 \times \Delta(X_0)^{n-1} \times \Delta(X_1)^{n-1} \times \ldots\end{aligned}$$

- Thus, for $k \geq 2$, $f(t)$ has a marginal over

$$X_{k-1} = X_0 \times \Delta(X_0)^{n-1} \times \ldots \times \Delta(X_{k-2})^{n-1}$$

The homeomorphism is "natural"

- Note that $f(t)$ is a distribution over

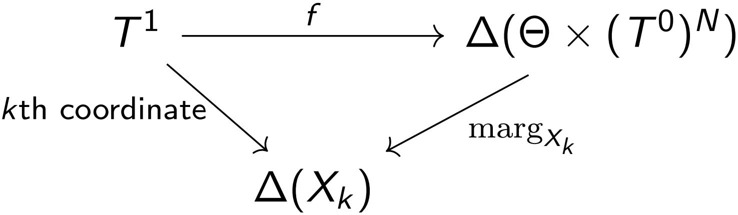
$$\begin{aligned}\Theta \times (T^0)^{n-1} &= X_0 \times (\Delta(X_0) \times \Delta(X_1) \times \dots)^{n-1} \\ &= X_0 \times \Delta(X_0)^{n-1} \times \Delta(X_1)^{n-1} \times \dots\end{aligned}$$

- Thus, for $k \geq 2$, $f(t)$ has a marginal over

$$X_{k-1} = X_0 \times \Delta(X_0)^{n-1} \times \dots \times \Delta(X_{k-2})^{n-1}$$

- This marginal is **exactly** t^k!

Commutative diagram



Closing the model

- Suppose we define $T_i = T^1$ and $\pi_i(\theta, t_{-i}|t_i) = f(t_i)(\theta, t_{-i})$; Is this a Harsanyi type space?

Closing the model

- Suppose we define $T_i = T^1$ and $\pi_i(\theta, t_{-i}|t_i) = f(t_i)(\theta, t_{-i})$; Is this a Harsanyi type space? No!
- Each $t \in T^1$ can be associated with beliefs $f(t)$ about Θ and others' hierarchies t_{-i}
- But only guaranteed that $t_{-i} \in (T^0)^{n-1}$ (outside the domain of the homeomorphism)
- We need to impose **common knowledge of consistency**

Closing the model

- Suppose we define $T_i = T^1$ and $\pi_i(\theta, t_{-i}|t_i) = f(t_i)(\theta, t_{-i})$; Is this a Harsanyi type space? No!
- Each $t \in T^1$ can be associated with beliefs $f(t)$ about Θ and others' hierarchies t_{-i}
- But only guaranteed that $t_{-i} \in (T^0)^{n-1}$ (outside the domain of the homeomorphism)
- We need to impose **common knowledge of consistency**
- For $k \geq 2$, let

$$T^k = \left\{ t \in T^{k-1} | f(t)(\Theta \times T^{k-1}) = 1 \right\}$$

- Let $T^* = \cap_{k \geq 2} T^k$, and let g denote the restriction of f to T^*
- The set T^* is the **universal type space**

Closing the model

- Suppose we define $T_i = T^1$ and $\pi_i(\theta, t_{-i}|t_i) = f(t_i)(\theta, t_{-i})$; Is this a Harsanyi type space? No!
- Each $t \in T^1$ can be associated with beliefs $f(t)$ about Θ and others' hierarchies t_{-i}
- But only guaranteed that $t_{-i} \in (T^0)^{n-1}$ (outside the domain of the homeomorphism)
- We need to impose **common knowledge of consistency**
- For $k \geq 2$, let

$$T^k = \left\{ t \in T^{k-1} | f(t)(\Theta \times T^{k-1}) = 1 \right\}$$

- Let $T^* = \cap_{k \geq 2} T^k$, and let g denote the restriction of f to T^*
- The set T^* is the **universal type space**

Proposition (Universal type space)

The function g is a homeomorphism from T^* to $\Delta(\Theta \times (T^*)^{n-1})$.

In what sense is the universal type space "universal"?

- Any type space $\mathcal{T} = (\Theta, \{T_i, \pi_i\}_{i=1}^n)$ can be "naturally" embedded in T^*
- Let $\phi_i^1 : T_i \to X^1 = \Delta(\Theta)$ denote player i's **first-order belief**: for any measurable set $Y \subseteq \Theta$

$$\phi_i^1(Y|t_i) = \pi_i(Y \times T_{-i}|t_i)$$

In what sense is the universal type space "universal"?

- Any type space $\mathcal{T} = (\Theta, \{T_i, \pi_i\}_{i=1}^n)$ can be "naturally" embedded in T^*
- Let $\phi_i^1 : T_i \to X^1 = \Delta(\Theta)$ denote player i's **first-order belief**: for any measurable set $Y \subseteq \Theta$

$$\phi_i^1(Y|t_i) = \pi_i(Y \times T_{-i}|t_i)$$

- Given embeddings ϕ^{k-1}, we can define the projection

$$\gamma_i^k(\theta, t_{-i}) = (\theta, \phi_{-i}^1(t_{-i}), \phi_{-i}^2(t_{-i}), \ldots, \phi_{-i}^{k-1}(t_{-i})) \in X_{k-1}$$

- kth order belief: for each $Y \subseteq X_{k-1}$, let

$$\phi_i^k(Y|t_i) = \pi_i((\gamma_i^k)^{-1}(Y)|t_i)$$

In what sense is the universal type space "universal"?

- Any type space $\mathcal{T} = (\Theta, \{T_i, \pi_i\}_{i=1}^n)$ can be "naturally" embedded in T^*
- Let $\phi_i^1 : T_i \to X^1 = \Delta(\Theta)$ denote player i's **first-order belief**: for any measurable set $Y \subseteq \Theta$

$$\phi_i^1(Y|t_i) = \pi_i(Y \times T_{-i}|t_i)$$

- Given embeddings ϕ^{k-1}, we can define the projection

$$\gamma_i^k(\theta, t_{-i}) = (\theta, \phi_{-i}^1(t_{-i}), \phi_{-i}^2(t_{-i}), \ldots, \phi_{-i}^{k-1}(t_{-i})) \in X_{k-1}$$

- kth order belief: for each $Y \subseteq X_{k-1}$, let

$$\phi_i^k(Y|t_i) = \pi_i((\gamma_i^k)^{-1}(Y)|t_i)$$

- Then, $\phi_i(t_i) = (\phi_i^1(t_i), \phi_i^2(t_i), \ldots)$ embeds T_i into T^*

The canonical embedding

- The measure $g(\phi_i(t_i))$ is a distribution over $\Theta \times (T^*)^{n-1}$

The canonical embedding

- The measure $g(\phi_i(t_i))$ is a distribution over $\Theta \times (T^*)^{n-1}$
- Given embeddings ϕ^{k-1}, we can define the mapping

$$\gamma(\theta, t_{-i}) = (\theta, \phi_{-i}(t_{-i}))$$

- $g(\phi_i(t_i))$ is exactly the pushforward measure defined by $\pi_i(t_i) \circ \gamma_i^{-1}$, i.e., for every measurable $Y \subseteq \Theta \times (T^*)^{n-1}$,

$$g(\phi_i(t_i))(Y) = \pi_i(\gamma_i^{-1}(Y)|t_i)$$

The canonical embedding

- The measure $g(\phi_i(t_i))$ is a distribution over $\Theta \times (T^*)^{n-1}$
- Given embeddings ϕ^{k-1}, we can define the mapping

$$\gamma(\theta, t_{-i}) = (\theta, \phi_{-i}(t_{-i}))$$

- $g(\phi_i(t_i))$ is exactly the pushforward measure defined by $\pi_i(t_i) \circ \gamma_i^{-1}$, i.e., for every measurable $Y \subseteq \Theta \times (T^*)^{n-1}$,

$$g(\phi_i(t_i))(Y) = \pi_i(\gamma_i^{-1}(Y)|t_i)$$

- In a sense, the belief relationships of types in $\mathcal{T}$ are preserved by their analogues in the universal type space
- Specifically, $\gamma_i^{-1}(Y)$ are the states and types in $\Theta \times T_{-i}$ such that the corresponding states and hierarchies are in Y
- Probabilities of such sets are pinned down by i's belief hierarchy

Commutative diagram

$$
\begin{array}{ccc}
T_i & \xrightarrow{\quad\quad\phi\quad\quad} & T^* \\
\Big\downarrow \pi_i & & \Big\downarrow g \\
\Delta(\Theta \times T_{-i}) & \xrightarrow{\gamma^{-1}} & \Delta(\Theta \times (T^*)^N)
\end{array}
$$

Redundant types

- Importantly, this does **not** mean that the likelihood of types t_{-i} only depends on their belief hierarchies
- The reason is that there might be more than one type that maps into the same belief hierarchy in the universal type space

Redundant types

- Importantly, this does **not** mean that the likelihood of types t_{-i} only depends on their belief hierarchies
- The reason is that there might be more than one type that maps into the same belief hierarchy in the universal type space
- As an example, consider a trivial case where $\Theta = 1$ and $T_1 = T_2 = \{0, 1\}$, and assume that $\pi_i(t_i|t_i) = 1$
- Then all types have the same higher order beliefs, but each type has different beliefs

Redundant types

- Importantly, this does **not** mean that the likelihood of types t_{-i} only depends on their belief hierarchies
- The reason is that there might be more than one type that maps into the same belief hierarchy in the universal type space
- As an example, consider a trivial case where $\Theta = 1$ and $T_1 = T_2 = \{0, 1\}$, and assume that $\pi_i(t_i|t_i) = 1$
- Then all types have the same higher order beliefs, but each type has different beliefs
- Such "redudant" types could be important for characterizing behavior under certain solution concepts, e.g., Bayes Nash equilibrium, since they act as correlation devices

Example with redundant types

State L, prob $1/2$

	A	B
A	$(1,1)$	$(0,0)$
B	$(0,0)$	$(1,1)$

State R, prob $1/2$

	A	B
A	$(0,0)$	$(1,1)$
B	$(1,1)$	$(0,0)$

Example with redundant types

State L, prob $1/2$

	A	B
A	$(1,1)$	$(0,0)$
B	$(0,0)$	$(1,1)$

State R, prob $1/2$

	A	B
A	$(0,0)$	$(1,1)$
B	$(1,1)$	$(0,0)$

- $\mathcal{T}^1$:
  - $T_i = \{t_i\}$, $\pi_i(\theta, t_j | t_i) = 1/2$
  - Common knowledge that both players think both states are equally likely, and equilibrium payoffs are $1/2$

Example with redundant types

State L, prob $1/2$

	A	B
A	$(1,1)$	$(0,0)$
B	$(0,0)$	$(1,1)$

State R, prob $1/2$

	A	B
A	$(0,0)$	$(1,1)$
B	$(1,1)$	$(0,0)$

- $\mathcal{T}^1$:
  - $T_i = \{t_i\}$, $\pi_i(\theta, t_j | t_i) = 1/2$
  - Common knowledge that both players think both states are equally likely, and equilibrium payoffs are $1/2$
- $\mathcal{T}^2$:
  - $T_i = \{t_i, s_i\}$, $\pi_i(R, s_j | t_i) = \pi_i(R, t_j | s_i) = \pi_i(L, s_j | s_i) = \pi_i(L, t_j | t_i) = 1/2$
  - Again, common knowledge of no information
  - But the types are correlated with the state (they match if $\theta = L$ and mismatch if $\theta = R$
  - $\sigma_i(A | t_i) = \sigma_i(B | s_i) = 1$ is an equilibrium, with payoff of 1!

Richer type spaces

- Bottom line: Higher order beliefs are not the only strategically relevant feature of the type space... we will return to this next week when we study robust predictions

Richer type spaces

- Bottom line: Higher order beliefs are not the only strategically relevant feature of the type space... we will return to this next week when we study robust predictions
- Ely and Pęski (2006) in fact construct a richer universal type space, where each type is identified with a hierarchy of higher order **conditional** beliefs, i.e., the beliefs that one would have if they learned others' types; they showed that this expanded type is rich enough to pin down rationalizable behavior

Richer type spaces

- Bottom line: Higher order beliefs are not the only strategically relevant feature of the type space... we will return to this next week when we study robust predictions
- Ely and Pęski (2006) in fact construct a richer universal type space, where each type is identified with a hierarchy of higher order **conditional** beliefs, i.e., the beliefs that one would have if they learned others' types; they showed that this expanded type is rich enough to pin down rationalizable behavior
- But... this only works for two players, and even conditional belief hierarchies don't characterize Nash equilibrium outcomes, which may depend on the presence of "pure" correlation devices that are independent of the state

The meaning of the product topology

- Throughout, we have been using the weak-$*$ and product topologies
- Even though T^* is an infinite dimensional vector space, it is very well-behaved under this topology

The meaning of the product topology

- Throughout, we have been using the weak-* and product topologies
- Even though T^* is an infinite dimensional vector space, it is very well-behaved under this topology
- It is well-known that if Z is complete/separable/metrizable/compact/etc, then so is $\Delta(Z)$
- Moreover, these properties also carry over to the countable product $\Delta(Z_1) \times \Delta(Z_2) \times \cdots$ under the product topology

The meaning of the product topology

- Throughout, we have been using the weak-* and product topologies
- Even though T^* is an infinite dimensional vector space, it is very well-behaved under this topology
- It is well-known that if Z is complete/separable/metrizable/compact/etc, then so is $\Delta(Z)$
- Moreover, these properties also carry over to the countable product $\Delta(Z_1) \times \Delta(Z_2) \times \cdots$ under the product topology
- Thus, if Θ is "nice", then T^* will be nice as well
- More than that, the product topology has a natural interpretation: convergence of belief hierarchies is equivalent to convergence of beliefs at every level

Is this the right topology?

Is this the right topology? Probably not

- Rubinstein's "electronic mail game" shows that rationalizable actions are not lower-hemicontinuous in the product topology

Is this the right topology? Probably not

- Rubinstein's "electronic mail game" shows that rationalizable actions are not lower-hemicontinuous in the product topology
- Two players, "generals" deciding whether to attack
- Binary state of the world $\theta \in \{B, G\}$ (good or bad)
- $\theta = G$ with probability $p < 1/2$ (common prior)

Is this the right topology? Probably not

- Rubinstein's "electronic mail game" shows that rationalizable actions are not lower-hemicontinuous in the product topology
- Two players, "generals" deciding whether to attack
- Binary state of the world $\theta \in \{B, G\}$ (good or bad)
- $\theta = G$ with probability $p < 1/2$ (common prior)
- Each player has two actions $\{R, A\}$ (retreat or attack), with payoffs being

State B, prob $1 - p$

	R	A
R	(M, M)	$(M, -L)$
A	$(-L, M)$	$(0, 0)$

State G, prob p

	R	A
R	$(0, 0)$	$(0, -L)$
A	$(-L, 0)$	(M, M)

Is this the right topology? Probably not

- Rubinstein's "electronic mail game" shows that rationalizable actions are not lower-hemicontinuous in the product topology
- Two players, "generals" deciding whether to attack
- Binary state of the world $\theta \in \{B, G\}$ (good or bad)
- $\theta = G$ with probability $p < 1/2$ (common prior)
- Each player has two actions $\{R, A\}$ (retreat or attack), with payoffs being

State B, prob $1-p$

	R	A
R	(M, M)	$(M, -L)$
A	$(-L, M)$	$(0, 0)$

State G, prob p

	R	A
R	$(0, 0)$	$(0, -L)$
A	$(-L, 0)$	(M, M)

- Players want to attack iff the state is good, but don't like attacking when the other player retreats
- Assume $L > M > 0$

Noisy communication

- When the true state is common knowledge, there's an equilibrium in which the players fight iff the state is good and get a payoff M

Noisy communication

- When the true state is common knowledge, there's an equilibrium in which the players fight iff the state is good and get a payoff M
- Alternative info: player 1 knows the state for sure

Noisy communication

- When the true state is common knowledge, there's an equilibrium in which the players fight iff the state is good and get a payoff M
- Alternative info: player 1 knows the state for sure
- Players communicate back and forth via a "noisy channel," parametrized by $\epsilon \in (0, 1)$:

Noisy communication

- When the true state is common knowledge, there's an equilibrium in which the players fight iff the state is good and get a payoff M
- Alternative info: player 1 knows the state for sure
- Players communicate back and forth via a "noisy channel," parametrized by $\epsilon \in (0, 1)$:
  - When $\theta = B$, both players receive no messages
  - When $\theta = G$, player 1 gets a message that the state is G

Noisy communication

- When the true state is common knowledge, there's an equilibrium in which the players fight iff the state is good and get a payoff M
- Alternative info: player 1 knows the state for sure
- Players communicate back and forth via a "noisy channel," parametrized by $\epsilon \in (0, 1)$:
  - When $\theta = B$, both players receive no messages
  - When $\theta = G$, player 1 gets a message that the state is G
  - Player 1 then sends a message to player 2 that is lost with probability ϵ

Noisy communication

- When the true state is common knowledge, there's an equilibrium in which the players fight iff the state is good and get a payoff M
- Alternative info: player 1 knows the state for sure
- Players communicate back and forth via a "noisy channel," parametrized by $\epsilon \in (0, 1)$:
  - When $\theta = B$, both players receive no messages
  - When $\theta = G$, player 1 gets a message that the state is G
  - Player 1 then sends a message to player 2 that is lost with probability ϵ
  - Conditional on receiving a message, player 2 sends back a "confirmation," that gets lost with prob ϵ

Noisy communication

- When the true state is common knowledge, there's an equilibrium in which the players fight iff the state is good and get a payoff M
- Alternative info: player 1 knows the state for sure
- Players communicate back and forth via a "noisy channel," parametrized by $\epsilon \in (0, 1)$:
  - When $\theta = B$, both players receive no messages
  - When $\theta = G$, player 1 gets a message that the state is G
  - Player 1 then sends a message to player 2 that is lost with probability ϵ
  - Conditional on receiving a message, player 2 sends back a "confirmation," that gets lost with prob ϵ
  - Player 1's confirmation of the confirmation is lost with prob ϵ, etc.

Noisy communication

- When the true state is common knowledge, there's an equilibrium in which the players fight iff the state is good and get a payoff M
- Alternative info: player 1 knows the state for sure
- Players communicate back and forth via a "noisy channel," parametrized by $\epsilon \in (0, 1)$:
  - When $\theta = B$, both players receive no messages
  - When $\theta = G$, player 1 gets a message that the state is G
  - Player 1 then sends a message to player 2 that is lost with probability ϵ
  - Conditional on receiving a message, player 2 sends back a "confirmation," that gets lost with prob ϵ
  - Player 1's confirmation of the confirmation is lost with prob ϵ, etc.
  - Communication ends when a message gets lost

Type distribution

Θ	t_1	t_2	Prob
B	0	0	$1-p$
G	1	0	$p\epsilon$
G	1	1	$p(1-\epsilon)\epsilon$
G	2	1	$p(1-\epsilon)^2\epsilon$
G	2	2	$p(1-\epsilon)^3\epsilon$
G	3	2	$p(1-\epsilon)^4\epsilon$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Notation for types

- A player's type is the number of messages they receive
- If you have received $t_i > 0$ messages, then
  - You are sure that $\theta = G$;
  - You are sure that the other player is sure that $\theta = G$
  - You are sure that the other player is sure that you are sure that $\theta = G$
  - $\vdots$
  - $\times t_i$
- As $t_i \to \infty$, beliefs converge pointwise to common knowledge that $\theta = G$

Unraveling

Proposition

R is the unique rationalizable action for all types.

- Thus, no matter how "close" beliefs get to common knowledge of the state in the product topology, they can never rationalize attacking
- More precisely, this is a failure of lower hemicontinuity of the rationalizable action correspondence

Proof: Base steps

- First, when $\theta = B$, player 1 knows the state is B and has a strictly dominant strategy to play R

Proof: Base steps

- First, when $\theta = B$, player 1 knows the state is B and has a strictly dominant strategy to play R
- Next, suppose player 2 does not receive **any** messages
- This could be because the state is B, which occurs with probability $1 - p > 1/2$, or because the state is G and the message did not get through, which occurs with probability $p\epsilon < 1/2$

Proof: Base steps

- First, when $\theta = B$, player 1 knows the state is B and has a strictly dominant strategy to play R
- Next, suppose player 2 does not receive **any** messages
- This could be because the state is B, which occurs with probability $1 - p > 1/2$, or because the state is G and the message did not get through, which occurs with probability $p\epsilon < 1/2$
- In ex ante probability units, the payoff from R is $(1 - p)M$, and the payoff from A is at most

$$(1 - p)(-L) + p\epsilon M < p\epsilon M$$

(assuming player 2 attacks whenever they get a message)

Proof: Base steps

- First, when $\theta = B$, player 1 knows the state is B and has a strictly dominant strategy to play R
- Next, suppose player 2 does not receive **any** messages
- This could be because the state is B, which occurs with probability $1 - p > 1/2$, or because the state is G and the message did not get through, which occurs with probability $p\epsilon < 1/2$
- In ex ante probability units, the payoff from R is $(1 - p)M$, and the payoff from A is at most

$$(1 - p)(-L) + p\epsilon M < p\epsilon M$$

  (assuming player 2 attacks whenever they get a message)
- Thus, regardless of ϵ, the unique best response is R

Proof: Inductive steps

- Now, suppose player i has received a given positive number of messages
- There are two possibilities:
  - Player j has k messages (received a confirmation for his last message but didn't send one back)
  - Player j has $k - 1$ messages (didn't receive a confirmation)
- The likelihoods are proportional to $\epsilon(1 - \epsilon)$ and ϵ, respectively

Proof: Inductive steps

- Now, suppose player i has received a given positive number of messages
- There are two possibilities:
  - Player j has k messages (received a confirmation for his last message but didn't send one back)
  - Player j has $k-1$ messages (didn't receive a confirmation)
- The likelihoods are proportional to $\epsilon(1-\epsilon)$ and ϵ, respectively
- Suppose inductively that player j retreats when they have $k-1$ messages

Proof: Inductive steps

- Now, suppose player i has received a given positive number of messages
- There are two possibilities:
  - Player j has k messages (received a confirmation for his last message but didn't send one back)
  - Player j has $k-1$ messages (didn't receive a confirmation)
- The likelihoods are proportional to $\epsilon(1-\epsilon)$ and ϵ, respectively
- Suppose inductively that player j retreats when they have $k-1$ messages
- Player i's payoff from R is 0, but the payoff from A is at most what they get if player j attacks for sure after k messages:

$$(-L)\epsilon + M\epsilon(1-\epsilon) < 0$$

(since $L > M$) □

What is the right topology?

- This is an ongoing debate; there is broad agreement that nearby types should have similar behavior under natural solution concepts
- Dekel, Fudenberg, and Morris (DFM) (2006) define a "strategic topology", which formalizes this idea
- Chen et. al (2010) define what they call a "uniform weak topology" that is slightly coarser than the DFM topology, but does not make explicit reference to games
- These guys also have a more recent paper that evidently closes the gap with DFM's topology, again without explicit reference to games
- Related to a "box topology" on belief hierarchies

The common prior assumption

- In the early days of von Neumann and Morgenstern (1944) and Luce and Raiffa (1957), they analyzed games with uncertainty using an extensive form, where chance moves at the beginning of the game according to commonly agreed probabilities, and private information is captured by players' information sets
- Importantly, we model a **prior stage** before the players have their private information, and where all the players agree on the distribution over states that will determine payoffs and types
- Card games, e.g., poker, fall into this class, where the players agree about the distribution of orders of cards in the deck, before it has been shuffled and the cards have been dealt
- Games like this have the feature that **differences in beliefs are only due to differences in information**
- We would now say that information that is derived in this manner satisfies the **common prior assumption**

Common priors

- Formally, fix a type space $\mathcal{T} = (\Theta, \{T_i, \pi_i\}_{i=1}^n)$ such that T_i is finite for all i
- $\mathcal{T}$ satisfies the **common prior assumption** (CPA) if there exist $\pi \in \Delta(\Theta \times T)$ and $\lambda_i : T_i \to \mathbb{R}_{++}$ for all i such that

$$\pi(\theta, t) = \lambda_i(t_i)\pi_i(t_{-i}, \theta | t_i)$$

- NB Harsanyi referred to common prior types as "consistent", which is how we have referred to belief hierarchies in T_0 (following Mertens and Zamir; Brandenburger and Dekel call types in T_0 "coherent")

Aside on terminology

- In the language of von Neumann and Morgenstern (1944) and Luce and Raiffa (1957), "complete information" seems to refer to any game theoretic model in which the description of players' information is complete, as in poker, whereas "incomplete information" refers to a model which is not fully specified, as when the players' belief hierarchies are not belief-closed
- vNM and LR separately make the distinction between **perfect and imperfect information** in extensive forms, where the former simply means that each information set consists of a single history; in such games, players are fully informed of the history of play, and information is trivial
- Harsanyi changed this terminology so that incomplete information refers to any setting in which players have higher order uncertainty, and "complete information" is synonymous with the CPA
- Today, incomplete vs complete usually refers to whether or not there is non-trivial higher order uncertainty, and incomplete information includes common prior and non-common prior beliefs

Implications of the CPA: Correlated equilibrium

- Let's explore the implications of the CPA regarding predictions for behavior
- Aumann (1987): foundations of correlated equilibrium
- He uses a formalism that is distinct from the type space, that is sometimes referred to as an **epistemic model**
- Let Ω be a (finite) set of states, and let $\mathbb{P}_i$ be a partition of Ω for each player i
- We interpret $\mathbb{P}_i$ as the "events" that player i can distinguish (replaces the type t_i in the Harsanyi model)
- Agent i has a **subjective belief** $\rho_i \in \Delta(\Omega)$
- There is a **common prior** if $\rho_1 = \rho_2 = \cdots = \rho_n$

The game form

- In addition, we assume that the players take actions that are part of the description of the world
- Let A_i be a (finite) set of actions for each player, with $A = \prod_{i=1}^{n} A_i$
- There is a mapping $\alpha_i : \Omega \to A_i$ which says which action is taken in which state, and is measurable with respect to $\mathbb{P}_i$ (so players know their own actions)
- $u_i : A \to \mathbb{R}$ is player i's payoff function
- We say that an agent is **rational at** ω if for all $a_i \in A_i$

$$\mathbb{E}[u_i(\alpha(\omega'))|\mathbb{P}_i](\omega) \geq \mathbb{E}[u_i(a_i, \alpha_{-i}(\omega'))|\mathbb{P}_i](\omega)$$

Interpretation: $\alpha_i(\omega)$ is a best response to player i's beliefs about $\alpha_{-i}(\omega)$ at ω

Connection to Harsanyi's model

- For some purposes, it is equivalent to the Harsanyi model, if we take $\Omega = \Theta \times T$ and $\mathbb{P}_i = \{\Theta \times \{t_i\} \times T_{-i} | t_i \in T_i\}$
- Can always pick $\rho_i(\omega)$ for each agent so that $\pi_i(\theta, t_{-i}|t_i)$ is the posterior given t_i (any weighted average of the π_i's)
- A key difference with the Harsanyi model is that the state ω determines everything, including players information, and also "endogenous" outcomes like actions

Reduced form descriptions of behavior

- The epistemic model contains lots of information, but we might more simply be interested in a reduced form description of rational behavior
- A natural question is what are the distributions of $\alpha(\omega)$ that could arise from some epistemic model where players are rational?
- But this immediately begs the question: distribution from whose perspective?
- If there is a common prior, however, then the players all agree about the distribution of ω, and we can assess the action distribution from this vantage point
- In particular, an epistemic model with a common prior **induces** a distribution $\mu \in \Delta(A)$, which is just the distribution of $\alpha(\omega)$:

$$\mu(a) = \sum_{\{\omega | \alpha(\omega) = a\}} \rho(\omega)$$

Necessary conditions from rationality

- Can any μ be induced by some common prior epistemic model in which players are rational?

Necessary conditions from rationality

- Can any μ be induced by some common prior epistemic model in which players are rational? No! Obviously players cannot take strictly dominated actions at any state of the world

Necessary conditions from rationality

- Can any μ be induced by some common prior epistemic model in which players are rational? No! Obviously players cannot take strictly dominated actions at any state of the world
- Indeed, we can manipulate the expression for rationality to get a simple necessary condition on μ
- Fix an action a_i
- By measurability, $\alpha_i^{-1}(a_i)$ is a union of elements of $\mathbb{P}_i$
- For any $P_i \in \mathbb{P}_i$ with $\subseteq \alpha_i^{-1}(a_i)$, rationality says that at any $\omega \in P_i$, the following is maximized at $a_i' = a_i$:

$$\mathbb{E}[u_i(a_i', \alpha_{-i}(\omega'))|\mathbb{P}_i](\omega) = \frac{1}{\rho(P_i)} \sum_{\omega' \in P_i} \rho(\omega') u_i(a_i', \alpha_{-i}(\omega'))$$

Necessary conditions from rationality

- Can any μ be induced by some common prior epistemic model in which players are rational? No! Obviously players cannot take strictly dominated actions at any state of the world
- Indeed, we can manipulate the expression for rationality to get a simple necessary condition on μ
- Fix an action a_i
- By measurability, $\alpha_i^{-1}(a_i)$ is a union of elements of $\mathbb{P}_i$
- For any $P_i \in \mathbb{P}_i$ with $\subseteq \alpha_i^{-1}(a_i)$, rationality says that at any $\omega \in P_i$, the following is maximized at $a_i' = a_i$:

$$\mathbb{E}[u_i(a_i', \alpha_{-i}(\omega'))|\mathbb{P}_i](\omega) = \frac{1}{\rho(P_i)} \sum_{\omega' \in P_i} \rho(\omega') u_i(a_i', \alpha_{-i}(\omega'))$$

- Summing over all $P_i \in \alpha_i^{-1}(a_i)$, with weights $\rho(P_i)$, we find that $a_i' = a_i$ maximizes

$$\begin{aligned}\sum_{\{\omega' | \alpha_i(\omega') = a_i\}} \rho(\omega') u_i(a_i', \alpha_{-i}(\omega')) &= \sum_{a_{-i}} u_i(a_i', , a_{-i}) \sum_{\{\omega' | \alpha(\omega') = (a_i, a_{-i})\}} \rho(\omega') \\ &= \sum_{a_{-i}} u_i(a_i', , a_{-i}) \mu(a_i, a_{-i})\end{aligned}$$

Correlated equilibrium

- More broadly, we say that $\mu \in \Delta(A)$ is a **correlated equilibrium** if for all i, a_i, and a_i',

$$\sum_{a_{-i}} \mu(a_i, a_{-i})(u_i(a_i, a_{-i}) - u_i(a_i', a_{-i})) \geq 0$$

- Interpretation: a mediator secretly recommends actions to players, according to μ
- The "obedience constraints" represent players' willingness to obey their recommendations

Correlated equilibrium

- More broadly, we say that $\mu \in \Delta(A)$ is a **correlated equilibrium** if for all i, a_i, and a_i',

$$\sum_{a_{-i}} \mu(a_i, a_{-i})(u_i(a_i, a_{-i}) - u_i(a_i', a_{-i})) \geq 0$$

- Interpretation: a mediator secretly recommends actions to players, according to μ
- The "obedience constraints" represent players' willingness to obey their recommendations
- NB Includes all of the Nash equilibria
- NB The joint distribution of actions is allowed to be correlated, in contrast to mixed-strategy Nash equilibrium, under which the joint distribution must be independent
- NB Slightly different from Aumann's definition

Aumann's theorem

Theorem

If there is a common prior and if every player is rational at every state of the world, then the distribution of $\alpha(\omega)$ is a correlated equilibrium.

- Proof: We just did it! $\square$
- Thus, correlated equilibrium is a consequence of having a common prior over states (where "state" is broad enough to include the actions that players are using) and common knowledge of rationality

Strengthening to if and only if

- In fact, we could strengthen to an only if, in the sense that if μ is a correlated equilibrium, then there is an epistemic model in which players have a common prior and players are rational at every state of the world
- How?

Strengthening to if and only if

- In fact, we could strengthen to an only if, in the sense that if μ is a correlated equilibrium, then there is an epistemic model in which players have a common prior and players are rational at every state of the world
- How? Just take $\Omega = A$, $\mathbb{P}_i = \{\{a_i\} \times A_{-i} | a_i \in A_i\}$, and $\rho_i = \mu$!

Correlated equilibrium: example

- Consider the game BoS:

	B	S
B	$(3, 1)$	$(0, 0)$
S	$(0, 0)$	$(1, 3)$

- Three Nash equilibria: (B, B), (S, S), and $(3/4B + 1/4S, 1/4B + 3/4S)$

Correlated equilibrium: example

- Consider the game BoS:

	B	S
B	$(3,1)$	$(0,0)$
S	$(0,0)$	$(1,3)$

- Three Nash equilibria: (B,B), (S,S), and $(3/4B + 1/4S, 1/4B + 3/4S)$
- Correlated equilibrium consists of $\mu(B,B)$, $\mu(B,S)$, $\mu(S,B)$, $\mu(S,S) \in [0,1]$ that sum to 1
- Obedience constraints:
  - B to S for player 1: $3\mu(B,B) \geq \mu(B,S)$
  - S to B for player 1: $\mu(S,S) \geq 3\mu(S,B)$
  - B to S for player 2: $\mu(B,B) \geq 3\mu(S,B)$
  - S to B for player 2: $3\mu(S,S) \geq \mu(B,S)$

How often could players miscoordinate?

- As a little numerical exercise, let us compute the maximum probability of miscoordination across all correlated equilibria

How often could players miscoordinate?

- As a little numerical exercise, let us compute the maximum probability of miscoordination across all correlated equilibria
- This is a linear program!

$$
\begin{aligned}
&\max_{\mu \geq 0} \mu(B,S) + \mu(S,B) \\
\text{s.t. } & \mu(B,S) \leq 3\mu(B,B) \\
& \mu(S,B) \leq \mu(S,S)/3 \\
& \mu(S,B) \leq \mu(B,B)/3 \\
& \mu(B,S) \leq 3\mu(S,S) \\
& \mu(B,B) + \mu(B,S) + \mu(S,B) + \mu(S,S) = 1
\end{aligned}
$$

How often could players miscoordinate?

- As a little numerical exercise, let us compute the maximum probability of miscoordination across all correlated equilibria
- This is a linear program!

$$\begin{aligned} \max_{\mu \geq 0} \; & \mu(B,S) + \mu(S,B) \\ \text{s.t.} \; & \mu(B,S) \leq 3\mu(B,B) \\ & \mu(S,B) \leq \mu(S,S)/3 \\ & \mu(S,B) \leq \mu(B,B)/3 \\ & \mu(B,S) \leq 3\mu(S,S) \\ & \mu(B,B) + \mu(B,S) + \mu(S,B) + \mu(S,S) = 1 \end{aligned}$$

- How do we solve it?

An educated guess

- Obedience can be rewritten

$$\mu(S,B) + \mu(B,S) \leq (10/3)\min\{\mu(B,B), \mu(S,S)\}$$

- We can guess that this constraint should bind; if not, then we can decrease $\mu(B,B)$ and $\mu(S,S)$ and increase $\mu(S,B) + \mu(B,S)$, while maintaining the total probability
- Moreover, we can guess that $\mu(B,B) = \mu(S,S)$; if not, decrease the larger and increase the smaller to make a little slack in the constraint
- This suggests that the optimal solution satisfies

$$(10/3)\mu(B,B) + \mu(B,B) + \mu(B,B) = 1$$
$$\iff \mu(B,B) = \mu(S,S) = 3/16, \mu(S,B) + \mu(B/S) = 10/16$$

- But this can only work if all obedience constraints bind, meaning that $\mu(S,B) = 1/16$ and $\mu(B,S) = 9/16$
- This is the outcome in the mixed strategy Nash!

Constructing a saddle point to confirm

- The variables in the dual are
  - Non-negative multipliers on obedience constraints $\alpha_1(B,S)$, $\alpha_1(S,B)$, $\alpha_2(B,S)$, $\alpha_2(S,B)$
  - A free multiplier λ on the probability constraint
- There are four constraints, one for each primal variable
- The dual is:

$$\min_{\alpha \geq 0, \lambda} \lambda$$

$$\begin{aligned}
\text{s.t. } & \lambda \geq 3\alpha_1(B,S) + \alpha_2(S,B)/3 \quad [\mu(B,B)] \\
& \lambda \geq \alpha_1(S,B)/3 + 3\alpha_2(B,S) \quad [\mu(S,S)] \\
& \lambda \geq 1 - \alpha_1(B,S) - \alpha_2(S,B) \quad [\mu(S,B)] \\
& \lambda \geq 1 - \alpha_1(S,B) - \alpha_2(B,S) \quad [\mu(B,S)]
\end{aligned}$$

- Just need to find a feasible solution such that $\lambda = 10/16$, and all constraints bind
- Obviously works with $\alpha_i(a_i, a_i') = 3/16$ for all i, a_i, a_i'!

Correlated equilibrium and linear programming

- Incidentally, this example illustrates the computational tractability of correlated equilibria, being the intersection of finitely many linear inequalities
- In particular, the optimization of a linear objective over correlated equilibria is a linear program
- On your pset, I will ask you to expand on this exercise and compute the set of equilibrium payoffs across all correlated equilibria

Interpretations of the CPA: Prior stage

- What does the CPA mean?
- As we discussed, one view is that, initially, the agents do not have private information, and they all share the same belief over states of the world
- Differences in beliefs ex-post are generated by differences in information, i.e., privately observed types or partition elements
- Importantly, at the prior stage, the distribution of ω and information partitions are common knowledge

Interpretations of the CPA: Prior stage

- What does the CPA mean?
- As we discussed, one view is that, initially, the agents do not have private information, and they all share the same belief over states of the world
- Differences in beliefs ex-post are generated by differences in information, i.e., privately observed types or partition elements
- Importantly, at the prior stage, the distribution of ω and information partitions are common knowledge
- This makes perfect sense when there really is a prior stage and there is some physical reason why all players should agree about probabilities at that stage (e.g., poker)
- But often there isn't a prior stage or a compelling argument for probabilities to agree, and then the CPA is more controversial

Problems with the prior stage

- Gul (1998) articulated some critiques along these lines
- Our axiomatic foundations of beliefs only give us "personal probabilities"
- At the prior stage, why should agents' personal probabilities over Ω necessarily agree? Even in the case of poker, players could disagree about whether the deck is stacked!
- This seems even more problematic when Ω includes a description of endogenous features of the world that are only privately known, e.g., actions taken at each state
- Much stronger than just assuming a common prior over information!
- Moreover, there often is no actual prior stage, and we only observe agents once they have their interim beliefs

Interpretations of CPA interim beliefs

- Samet's (1998) characterization of the CPA:
- For each $E \subseteq \Omega$
  - Agent i has a belief p_i about the likelihood of E
  - Agent j has an expectation p_{ji} of p_i
  - Agent k has an expectation p_{kji} of p_{ji}
- The type space satisfies the CPA if and only if, for all E and for all such sequences of iterated expectations, the sequence of probabilities converges to the same value
- The limit is the prior probability of E!
- Pretty cool, but it seems no clearer whether this property should hold

The CPA and no trade

- An alternative characterization of the CPA is interms of whether risk-neutral agents would strictly prefer to trade
- Different priors would imply that agents would accept bets on the events about whose probability they disagree
- We do see such betting in prediction markets
- But the fact that such bets are not omnipresent suggests that violations of the CPA are modest, at least relative to agents' risk aversion
- We can develop this formally in the risk neutral benchmark, following Morris (1994)
- (Cf Milgrom and Stokey 1982, who only show sufficiency, in a world with risk aversion)

Trades

- A **trade** is a function $\gamma : T \times \Theta \to \mathbb{R}^n$
- It is **feasible** if for all (t, θ)

$$\sum_{i=1}^{n} \gamma_i(t, \theta) \leq 0$$

- It is **acceptable** if for all i and t_i,

$$\sum_{t_{-i},\theta} \pi_i(t_{-i}, \theta | t_i)\gamma_i(t_i, t_{-i}, \theta) \geq 0,$$

  with strict inequality for some i and t_i
- There is **no trade** if there does not exist a feasible and acceptable trade

No trade and common priors

Theorem

There is no trade if and only if $\mathcal{T}$ satisfies the CPA.

No trade and common priors

Theorem

There is no trade if and only if $\mathcal{T}$ satisfies the CPA.

- We will prove with some helper lemmas from convex duality, starting with:

Lemma (Farkas' lemma)

Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, exactly one of the following is true:

(i) *There exists $x \in \mathbb{R}^n$ such that $Ax = b$ and $x \geq 0$*

(ii) *There exists $y \in \mathbb{R}^m$ such that $yA \geq 0$ and $yb < 0$*

Variant I

Lemma

Given $A \in \mathbb{R}^{m \times n}$ and a column j, exactly one of the following is true:

(i) *There exists $x \in \mathbb{R}^n$ such that $x \geq 0$, $Ax = 0$, and $x_j > 0$*

(ii) *There exists $y \in \mathbb{R}^m$ such that $yA \geq 0$, with strict inequality for column j*

Variant I

Lemma

Given $A \in \mathbb{R}^{m \times n}$ and a column j, exactly one of the following is true:

(i) *There exists $x \in \mathbb{R}^n$ such that $x \geq 0$, $Ax = 0$, and $x_j > 0$*

(ii) *There exists $y \in \mathbb{R}^m$ such that $yA \geq 0$, with strict inequality for column j*

- Pf. Take an instance of the Farkas alternative where

$$A' = \begin{bmatrix} A \\ v \end{bmatrix}, \; v_j = 1, v_{j'} = 0, \; b = (0, \ldots, 0, 1)$$

Variant I

Lemma

Given $A \in \mathbb{R}^{m\times n}$ and a column j, exactly one of the following is true:

(i) *There exists $x \in \mathbb{R}^n$ such that $x \geq 0$, $Ax = 0$, and $x_j > 0$*

(ii) *There exists $y \in \mathbb{R}^m$ such that $yA \geq 0$, with strict inequality for column j*

- Pf. Take an instance of the Farkas alternative where

$$A' = \begin{bmatrix} A \\ v \end{bmatrix}, \; v_j = 1, v_{j'} = 0, \; b = (0, \ldots, 0, 1)$$

- Farkas $\implies$ exactly one of the following is true:

Variant I

Lemma

Given $A \in \mathbb{R}^{m \times n}$ and a column j, exactly one of the following is true:

(i) *There exists $x \in \mathbb{R}^n$ such that $x \geq 0$, $Ax = 0$, and $x_j > 0$*

(ii) *There exists $y \in \mathbb{R}^m$ such that $yA \geq 0$, with strict inequality for column j*

- Pf. Take an instance of the Farkas alternative where

$$A' = \begin{bmatrix} A \\ v \end{bmatrix}, \ v_j = 1, v_{j'} = 0, \ b = (0, \ldots, 0, 1)$$

- Farkas $\implies$ exactly one of the following is true:
  (i′) There exists $x \in \mathbb{R}^n$ with $A'x = b$ and $x \geq 0$, in which case $Ax = 0$ and $x_j = 1$, i.e., (i) holds

Variant I

Lemma

Given $A \in \mathbb{R}^{m \times n}$ and a column j, exactly one of the following is true:

(i) *There exists $x \in \mathbb{R}^n$ such that $x \geq 0$, $Ax = 0$, and $x_j > 0$*

(ii) *There exists $y \in \mathbb{R}^m$ such that $yA \geq 0$, with strict inequality for column j*

- Pf. Take an instance of the Farkas alternative where

$$A' = \begin{bmatrix} A \\ v \end{bmatrix}, \; v_j = 1, v_{j'} = 0, \; b = (0, \ldots, 0, 1)$$

- Farkas $\implies$ exactly one of the following is true:
  - (i′) There exists $x \in \mathbb{R}^n$ with $A'x = b$ and $x \geq 0$, in which case $Ax = 0$ and $x_j = 1$, i.e., (i) holds
  - (ii′) There exists $y \in \mathbb{R}^m$ and $z \in \mathbb{R}$ such that $(y, z)b = z < 0$, $yA + zv \geq 0$, i.e., $yA \geq 0$ and strictly for column j, i.e., (ii) holds □

Variant II

Lemma

Given $A \in \mathbb{R}^{m \times n}$ and a set S of columns, exactly one of the following is true:

(i) *There exists an $x \in \mathbb{R}^n$ such that $x \geq 0$ and $Ax = 0$ and $x_j > 0$ for all $j \in S$*

(ii) *There exists $j \in S$ and $y \in \mathbb{R}^m$ such that $yA \geq 0$ for all j, with strict inequality for column j*

Variant II

Lemma

Given $A \in \mathbb{R}^{m \times n}$ and a set S of columns, exactly one of the following is true:

(i) *There exists an $x \in \mathbb{R}^n$ such that $x \geq 0$ and $Ax = 0$ and $x_j > 0$ for all $j \in S$*

(ii) *There exists $j \in S$ and $y \in \mathbb{R}^m$ such that $yA \geq 0$ for all j, with strict inequality for column j*

- Pf. From Variant I, (ii) implies $\neg$(i)

Variant II

Lemma

Given $A \in \mathbb{R}^{m \times n}$ and a set S of columns, exactly one of the following is true:

(i) *There exists an $x \in \mathbb{R}^n$ such that $x \geq 0$ and $Ax = 0$ and $x_j > 0$ for all $j \in S$*

(ii) *There exists $j \in S$ and $y \in \mathbb{R}^m$ such that $yA \geq 0$ for all j, with strict inequality for column j*

- Pf. From Variant I, (ii) implies $\neg$(i)
- Moreover, from Variant I, if (ii) doesn't hold, then for each $j \in S$, there exists $x^j \in \mathbb{R}^n$ such that $x^j \geq 0$, $Ax^j = 0$, and $x^j_j > 0$

Variant II

Lemma

Given $A \in \mathbb{R}^{m \times n}$ and a set S of columns, exactly one of the following is true:

(i) *There exists an $x \in \mathbb{R}^n$ such that $x \geq 0$ and $Ax = 0$ and $x_j > 0$ for all $j \in S$*

(ii) *There exists $j \in S$ and $y \in \mathbb{R}^m$ such that $yA \geq 0$ for all j, with strict inequality for column j*

- Pf. From Variant I, (ii) implies $\neg$(i)
- Moreover, from Variant I, if (ii) doesn't hold, then for each $j \in S$, there exists $x^j \in \mathbb{R}^n$ such that $x^j \geq 0$, $Ax^j = 0$, and $x^j_j > 0$
- So take $x = \sum_{j \in S} x^j$, which satisfies (i) $\square$

Proof of the theorem

- Existence of a trade means that there exist $\gamma_i : T \times \Theta \to \mathbb{R}$ for each $i = 1, \ldots, n$ such that

$$-\sum_{i=1}^{n} \gamma_i(t, \theta) \geq 0 \ \forall (t, \theta);$$

$$\sum_{t_{-i}, \theta} \pi_i(t_{-i}, \theta | t_i) \gamma_i(t_i, t_{-i}, \theta) \geq 0 \ \forall i, t_i$$

  and strict for some t_i

Proof of the theorem

- Existence of a trade means that there exist $\gamma_i : T \times \Theta \to \mathbb{R}$ for each $i = 1, \ldots, n$ such that

$$-\sum_{i=1}^{n} \gamma_i(t, \theta) \geq 0 \ \forall (t, \theta);$$

$$\sum_{t_{-i}, \theta} \pi_i(t_{-i}, \theta | t_i) \gamma_i(t_i, t_{-i}, \theta) \geq 0 \ \forall i, t_i$$

  and strict for some t_i

- This maps into Variant II case (ii) with rows indexed by $T \times \Theta \times \{1, \ldots, n\}$, columns indexed by $(T \times \Theta) \sqcup (\sqcup_{i=1}^{n} T_i)$, and

$$S = \sqcup_i T_i$$

  (where γ is the y)

Proof of the theorem

- Existence of a trade means that there exist $\gamma_i : T \times \Theta \to \mathbb{R}$ for each $i = 1, \ldots, n$ such that

$$-\sum_{i=1}^{n} \gamma_i(t, \theta) \geq 0 \ \forall (t, \theta);$$

$$\sum_{t_{-i}, \theta} \pi_i(t_{-i}, \theta | t_i) \gamma_i(t_i, t_{-i}, \theta) \geq 0 \ \forall i, t_i$$

  and strict for some t_i
- This maps into Variant II case (ii) with rows indexed by $T \times \Theta \times \{1, \ldots, n\}$, columns indexed by $(T \times \Theta) \sqcup (\sqcup_{i=1}^{n} T_i)$, and

$$S = \sqcup_i T_i$$

  (where γ is the y)
- So, no trade $\iff$ Variant II case (i) holds, i.e., there exist $\tilde{\phi} : T \times \Theta \to \mathbb{R}_+$ and $\tilde{\lambda}_i : T_i \to \mathbb{R}_{++}$ such that for all (i, t, θ),

$$\tilde{\lambda}_i(t_i) \pi_i(t_{-i}, \theta | t_i) - \tilde{\phi}(t, \theta) = 0$$

Proof of the theorem

- Existence of a trade means that there exist $\gamma_i : T \times \Theta \to \mathbb{R}$ for each $i = 1, \ldots, n$ such that

$$-\sum_{i=1}^{n} \gamma_i(t, \theta) \geq 0 \ \forall (t, \theta);$$

$$\sum_{t_{-i}, \theta} \pi_i(t_{-i}, \theta | t_i) \gamma_i(t_i, t_{-i}, \theta) \geq 0 \ \forall i, t_i$$

  and strict for some t_i
- This maps into Variant II case (ii) with rows indexed by $T \times \Theta \times \{1, \ldots, n\}$, columns indexed by $(T \times \Theta) \sqcup (\sqcup_{i=1}^{n} T_i)$, and

$$S = \sqcup_i T_i$$

  (where γ is the y)
- So, no trade $\iff$ Variant II case (i) holds, i.e., there exist $\tilde{\phi} : T \times \Theta \to \mathbb{R}_+$ and $\tilde{\lambda}_i : T_i \to \mathbb{R}_{++}$ such that for all (i, t, θ),

$$\tilde{\lambda}_i(t_i) \pi_i(t_{-i}, \theta | t_i) - \tilde{\phi}(t, \theta) = 0$$

- Then $\tilde{\phi} / \sum_{(t,\theta)} \tilde{\phi}(t, \theta)$ is a common prior $\square$

Interpreting the no trade theorem

- No trade relies on (at least) three key assumptions:
  1. Common priors
  2. Risk neutrality
  3. All trades between agents are feasible
- Relaxing any one of these three will break the equivalence
- So, in particular, agents may not want to trade because they are risk averse or because a trade is infeasible
- But, to the extent that agents are "close" to being risk neutral, then agents should have common priors over the events based on which they can trade

But do we use the CPA just at the interim stage?

- The CPA is often much more than an assumption about interim beliefs:
  - It's the "neutral" perspective with which we evaluate Bayesian welfare criteria, e.g., revenue in auctions
    Harsanyi refers to such an evaluator as a **properly informed observer**
  - It's also identified with the empirical distribution that is measured by an econometrician
- So, the CPA is often paired with one or both of the following even stronger assumptions:
  - the analyst is an extra agent in the model with a single type
  - the agents have **rational expectations**: the common prior coincides with the empirical distribution of types and states
- In a stationary world, one could imagine that players converge to a common prior through learning about the long run distribution of (θ, t), but to my knowledge this has not been formalized

Other arguments for the CPA

- Following Aumann (1998): the CPA is equivalent to the hypothetical statement that "if agents had the same information, they would have the same beliefs"
- Follows from considering counterfactual information $\mathbb{P}'_i = \{\Omega\}$, in which the above statement implies that $\rho_i = \rho_j$
- If agents acquire more information, then their beliefs are updated from that common prior
- Aumann argues that even if we do not observe the prior stage, it can still a useful abstraction for understanding how beliefs are formed and a natural benchmark that could lead to interesting insights about behavior

A more pragmatic view

- There is little to no formal distinction between priors and preferences; without restricting prior, we could explain any behavior by assuming that each player puts probability one on a state in which that behavior is optimal
- As Myerson (2004) writes:

  > *If there is no common prior, then we can only say that these difference in people's beliefs are just a fundamental assumption of our model. But then we must face the question: If we can assume any arbitrary characteristics for the individuals in our model, then why could we not explain the surprising behavior even more simply by assuming that each individual has a payoff function that is maximized by this behavior? Thus, to avoid such trivialization of the economic problem, applied theorists have generally limited themselves to models that satisfy Harsanyi's consistency assumption.*

- Put differently, the profession has coordinated on the (basically untestable) hypothesis of common priors as a way of disciplining our models, so that they generate new insights into human behavior
- A general aspiration, though, is to understand better when a particular insight about behavior truly relies on the CPA, or would weaker common knowledge assumptions suffice