Common Knowledge and Common Learning

Duarte Gonçalves

University College London

Topics in Economic Theory

3 people are prisoners on an island.

They cannot communicate with each other, cannot see their reflections, and all had green eyes.

The island is ruled by a despotic bear who imposed a peculiar rule:
a prisoner can ask to leave every night, but only prisoners with green eyes will be permitted to escape, whilst all others will be tossed in the volcano.

All the prisoners want to leave, but will never take action unless they are absolutely certain that they have green eyes.

You want to do something to help the prisoners.

The dictator allows you one thing only: to say a single sentence to the prisoners.

But there's a twist: you can't tell them anything that each didn't know.

What do you do?

‘There is at least one person with green eyes.’

Everyone knew that. On the first day, no one asks to leave. On the second day, neither. On the third day, everyone does.

If no one asks to leave on day 1, then it must be that everyone sees someone with green eyes
(otherwise they’d deduce they themselves have green eyes).

If no one asks to leave on day 2, then it must be that everyone sees two people with green eyes
(otherwise they’d deduce they themselves have green eyes).

On day 3, everyone is sure they have green eyes.

Common Knowledge

What happened? Saying something everyone knows makes it *commonly known*.

Who cares (other than for its own sake)?

- Highlights the role of public signals and announcements (monetary policy, auctions).
- Clarifies limits of coordination (distributed systems, protests, currency attacks).
- Provides epistemic foundations for solution concepts (backward induction, Bayesian Nash equilibrium).

This lecture: formalising knowledge and deriving implications.

Overview

1. Knowledge
2. An Aside: Syntactic Knowledge
3. Common Prior Assumption and Its Implications
4. Common Belief
5. Common Learning
6. Universal Type Space

Overview

1. Knowledge
   - Knowledge
   - Common Knowledge
2. An Aside: Syntactic Knowledge
3. Common Prior Assumption and Its Implications
4. Common Belief
5. Common Learning
6. Universal Type Space

Knowledge

$\omega \in \Omega$: state of the world. Finite.

Events $E, F \subseteq \Omega$.

Knowledge Function: $k : \Omega \to 2^{\Omega}$

When true state is ω, $k(\omega)$ represents what DM knows.

Or, $k(\omega)$ are the states DM cannot distinguish from ω.

Knowledge

Example

$\Omega = \{\omega, \omega'\}$, $k(\omega) = \{\omega, \omega'\}$, $k(\omega') = \{\omega'\}$.

ω= Skipped my stop; ω= Didn't skip my stop.

If I didn't skip my stop, I know I didn't. By if I did, I don't know that I did.

Issue: if introspect, I realise that if I didn't skip my stop I would know it, so if I don't know I must've skipped it!

If I knew k, introspection would rule this out.

Knowledge

$\omega \in \Omega$: state of the world. Finite.

Events $E, F \subseteq \Omega$.

Knowledge Function: $k : \Omega \to 2^{\Omega}$

When true state is ω, $k(\omega)$ represents what DM knows.

Or, $k(\omega)$ are the states DM cannot distinguish from ω.

Partitional Knowledge function: (i) $\forall \omega \in \Omega$, $\omega \in k(\omega)$ and (ii) $\omega' \in k(\omega) \implies k(\omega) = k(\omega')$.

Knowledge Operator: $K : 2^{\Omega} \to 2^{\Omega}$ s.t. $K(E) := \{\omega \in \Omega \mid k(\omega) \subseteq E\}$.

Note: $K(E) = \cup_{k(\omega) \subseteq E} k(\omega)$.

Knowledge

Lemma

If k is partitional, then:

1. $K(\Omega) = \Omega$. (Axiom of awareness)
2. $K(E) \cap K(F) = K(E \cap F)$.
3. $K(E) \subseteq E$. (Axiom of knowledge)
4. $K(E) = K(K(E))$. (axiom of transparency)
5. $\Omega \setminus K(E) = K(\Omega \setminus K(E))$. (Axiom of wisdom)
6. $F \subseteq E \implies K(F) \subseteq K(E)$. (Monotonicity)
7. $K(E) = \cup_{\omega \in K(E)} k(\omega)$.

Proof

1. $\forall \omega, \omega \in k(\omega) \subseteq \Omega \implies K(\Omega) = \Omega$.
2. $\omega \in K(E \cap F) \iff \exists \omega' : \omega \in k(\omega') = k(\omega) \subseteq E \cap F \iff \omega \in K(E) \cap K(F)$.
3. $\omega \in K(E) \implies \omega \in k(\omega) \subseteq E$.

Knowledge

Lemma

If k is partitional, then:

1. $K(\Omega) = \Omega$. (Axiom of awareness)
2. $K(E) \cap K(F) = K(E \cap F)$.
3. $K(E) \subseteq E$. (Axiom of knowledge)
4. $K(E) = K(K(E))$. (axiom of transparency)
5. $\Omega \setminus K(E) = K(\Omega \setminus K(E))$. (Axiom of wisdom)
6. $F \subseteq E \implies K(F) \subseteq K(E)$. (Monotonicity)
7. $K(E) = \cup_{\omega \in K(E)} k(\omega)$.

Proof

4. $K(K(E)) \subseteq K(E)$ from 3. Moreover, $\omega \in K(K(E)) \implies k(\omega) \subseteq K(E) \subseteq E \implies \omega \in K(E) \implies K(K(E)) \subseteq K(E)$ (Using 3.).

Knowledge

Lemma

If k is partitional, then:

1. $K(\Omega) = \Omega$. (Axiom of awareness)
2. $K(E) \cap K(F) = K(E \cap F)$.
3. $K(E) \subseteq E$. (Axiom of knowledge)
4. $K(E) = K(K(E))$. (axiom of transparency)
5. $\Omega \setminus K(E) = K(\Omega \setminus K(E))$. (Axiom of wisdom)
6. $F \subseteq E \implies K(F) \subseteq K(E)$. (Monotonicity)
7. $K(E) = \cup_{\omega \in K(E)} k(\omega)$.

Proof

5. Using 3., as $K(E) = K(K(E)) = \cup_{K(\{\omega\}) \subseteq K(E)} K(\{\omega\})$ $K(\Omega \setminus K(E)) \subseteq \Omega \setminus K(E)$ from 3. Moreover, as k partitional, $\omega \in \Omega \setminus K(E) \implies \omega \notin K(E) = K(K(E)) \implies \neg(k(\omega) \subseteq K(E)) \implies (\omega' \notin K(E) \forall \omega' \in k(\omega)) \implies k(\omega) \subseteq \Omega \setminus K(E) \implies \omega \in K(\Omega \setminus K(E))$.

Knowledge

Proposition

If k is partitional, then:

1. $K(\Omega) = \Omega$. (Axiom of awareness)
2. $K(E) \cap K(F) = K(E \cap F)$.
3. $K(E) \subseteq E$. (Axiom of knowledge)
4. $K(E) = K(K(E))$. (axiom of transparency)
5. $\Omega \setminus K(E) = K(\Omega \setminus K(E))$. (Axiom of wisdom)
6. $F \subseteq E \implies K(F) \subseteq K(E)$. (Monotonicity)
7. $K(E) = \cup_{\omega \in K(E)} k(\omega)$.

Proof

6. $F \subseteq E \implies K(F) = K(F \cap E) = K(F) \cap K(E) \subseteq K(E)$ (using 2).
7. $K(E) = \cup_{\omega \in K(E)} \{\omega\} \subseteq K(E) = \cup_{\omega \in K(E)} k(\omega)$. Moreover, $\forall \omega' \in \cup_{\omega \in K(E)} k(\omega)$, $\exists \omega'' \in K(E) : \omega' \in k(\omega'')$; and as $k(\omega') = k(\omega'') \subseteq E$, then $\omega' \in K(E)$.

Interactive Knowledge

Player $i \in \{1, ..., I\}$ with knowledge operator K_i.

Assume partitional k_i henceforth.

Definition

(i) There is **mutual knowledge** of $E \subseteq \Omega$ at ω if $\omega \in K^1(E) := \cap_i K_i(E)$.

(ii) Let $K^{n+1}(E) := K^1(K^n(E))$, for $n = 1, 2, ...$. There is **common knowledge** of $E \subseteq \Omega$ at ω if $\omega \in K^\infty(E) := \cap_n K^n(E)$.

Remark

If E is CK at ω, then $\forall F \supseteq E$, F is CK at ω.

Proof

By monotonicity of K_i, $\omega \in K^\infty(E) \subseteq K^\infty(F)$.

Can also consider CK for subset of players.

Properties of Mutual and Common Knowledge Operators

Proposition

If k_i is partitional $\forall i$, then, for any $n = 1, 2, ..., \infty$,

1. $K^n(\Omega) = \Omega$. (Axiom of awareness)
2. $K^n(E) \cap K^n(F) = K^n(E \cap F)$.
3. $K^n(E) \subseteq E$. (Axiom of knowledge)
4. $K^n(E) \supseteq K^n(K^n(E))$. (axiom of transparency)
5. $\Omega \setminus K^n(E) \supseteq K^n(\Omega \setminus K^n(E))$. (Axiom of wisdom)
6. $F \subseteq E \implies K^n(F) \subseteq K(E)$. (Monotonicity)

Properties of Mutual and Common Knowledge Operators

Proof

We prove for $n = 1$.

1. $\cap_i K_i(\Omega) = \cap_i \Omega = \Omega$.
2. $(\cap_i K_i(E)) \cap (\cap_i K_i(F)) = \cap_i (K_i(E) \cap K_i(F)) = \cap_i K_i(E \cap F)$.
3. $\cap_i K_i(E) \subseteq \cap_i E = E$.
4. Follows from 3.
5. Follows from 3.
6. $F \subseteq E \implies K_i(F) \subseteq K_i(E) \, \forall i \implies \cap_i K_i(F) \subseteq \cap_i K_i(E)$.

Iterate arguments to extend to $n > 1$.

Interactive Knowledge

Example

$\Omega = \{\omega_1, \omega_2, \omega_3\}$.

$k_1(\omega_1) = \{\omega_1\}, \quad k_1(\omega_2) = k_1(\omega_3) = \{\omega_2, \omega_3\}$.

$k_2(\omega_1) = k_2(\omega_2) = \{\omega_1, \omega_2\}, \quad k_2(\omega_3) = \{\omega_3\}$.

$E = \{\omega_2, \omega_3\} \implies K_1(E) = E, K_2(E) = \{\omega_3\} \implies K^1(E) = \{\omega_3\}$.

$K_1(K^1(E)) = \emptyset \implies K^n(E) = K^\infty(E) = \emptyset, \forall n \geq 2$.

Only Ω is CK.

How to get CK? By assumption or deriving CK from CK of something else.

Interactive Knowledge

Definition

Event E is evident if it is mutually known, $E \subseteq K^1(E)$.

If E happens, everyone knows E happens.

Remark

(i) E is evident $\implies E \subseteq K^1(E) \subseteq E \implies E = K^1(E)$.

(ii) E is evident $\iff k_i(\omega) \subseteq E, \forall \omega \in E$.

(iii) If E is evident, $E = K^{\infty}(E)$ and so E is CK at any $\omega \in E$.

Interactive Knowledge

Definition

Event E is evident if it is mutually known, $E \subseteq K^1(E)$.

Proposition (Monderer and Samet, 1989 GEB)

C is CK at ω if and only if there is an evident event E s.t. $\omega \in E$ and $E \subseteq K^1(C)$.

One could have just as well have written (...) "$\omega \in E$ and $E \subseteq C$"

Proof

If: $E \subseteq K^1(C) \implies \omega \in E = K^\infty(E) \subseteq K^\infty(C)$.

Only if: Let $E := K^\infty(C) := \cap_n K^n(C)$. Then, $K_i(E) = E \, \forall i \implies K^1(E) = E$, and E is evident.
Moreover, by transparency, $E = K^\infty(C) \subseteq K^1(C)$.

Overview

Syntactic Knowledge

In the words of Aumann (1999 IJGT, I)

> *One question that often arises is, what do the players know about the [semantic] formalism itself? Does each know the others' partitions? If so, from where does this knowledge derive? If not, how can the formalism indicate what each player knows about the others' knowledge?*
> *(...) More generally, the whole idea of "state of the world," and of a partition structure that accurately reflects the players' knowledge about other players' knowledge, is not transparent. What are the states? Can they be explicitly described? Where do they come from? Where do the information partitions come from?*

Syntactic Knowledge Model

Main Ingredients

Symbols: including Letters from an alphabet $\mathcal{X} := \{x, y, z, ...\}$ taken as fixed, and $\vee$, $\neg$, (), and κ_i.

Formula (or Propositions): finite string of symbols.

1. Every letter is a formula.
2. If f and g are formulas, so is $(f) \vee (g)$.
3. If f is a formula, so are $\neg(f)$ and $\kappa_i(f)$ for each i.

Interpretation

$\kappa_i f$: "i knows f".

$\vee$, $\neg$: 'or', 'it is not true that'.

Formula f: a finite concatenation of natural occurrences, using operators and connectives of propositional logic plus the knowledge operators.

Lists of Formulas

Lists

$f \implies g$ means $(\neg f) \vee g$.

$f \iff g$ means $f \implies g$ and $g \implies f$.

List is set of formulas.

Properties of List $\mathcal{L}$

- logically closed if ($f \in \mathcal{L}$ and $f \implies g \in \mathcal{L}$) implies $g \in \mathcal{L}$.
- epistemically closed if $f \in \mathcal{L}$ implies $\kappa_i f \in \mathcal{L}$.
- strongly closed if logically and epistemically closed.
  Strong closure of $\mathcal{L}$ is smalles strongly clost list that includes $\mathcal{L}$.
- coherent if $\neg f \in \mathcal{L}$ implies $f \notin \mathcal{L}$.
- complete if $f \notin \mathcal{L}$ implies $\neg f \in \mathcal{L}$.

Tautologies

Tautology is statement commonly believed by everyone. Formally:

Tautology: a formula in strong closure of the list of all formulas having one of the following forms:

(i) $(f \vee f) \implies f$.

(ii) $f \implies (f \vee g)$.

(iii) $(g \vee f) \implies (f \vee g)$.

(iv) $(f \implies g) \implies ((h \vee f) \implies (h \vee g))$.

(v) $\kappa_i f \implies f$.

(vi) $\kappa_i(f \implies g) \implies ((\kappa_i f) \implies (\kappa_i g))$.

(vii) $\neg\kappa_i f \implies \kappa_i \neg\kappa_i f$.

$\mathcal{T}$: list of all tautologies.

(Theorems are tautologies!)

g is consequence of f if $f \implies g$ is a tautology.

Syntax $\mathcal{S}$: set of all formulas with given population I and alphabet $\mathcal{X}$, countable.

Towards an Isomorphism for Syntactic-Semantic Knowledge

Canonical Semantic Knowledge System: For given syntax $\mathcal{S}$,

State ω:= a closed, coherent, complete list of formulas containing all tautologies.

Ω: set of all states.

Information function: $k_i : \Omega \to 2^{\Omega}$ s.t. $k_i(\omega)$ is set of formulas in ω starting with κ_i.

Events $E_f := \{\omega \in \Omega : f \in \omega\}$.

For any list $\mathcal{L}$, let $\mathcal{L}^*$ denote the strong closure of $\mathcal{L} \cup \mathcal{T}$.

$\mathcal{L}$ is consistent if $f \in \mathcal{L}^*$ implies $\neg f \notin \mathcal{L}^*$.

Proposition: A list is a state iff it is complete and consistent.

Syntactic and semantic approaches isomorphic (under conditions)

– Aumann (1999 IJGT I, §9).

So What?

Contractual complexity, intrincate financial derivatives, and failures of contigent reasoning: issues in dealing with logical deduction (costly, confusing).

Vague and complex contracts: e.g., Jakobsen (2020 AER), Piermont (2024 WP).

Expanding state space via introducing new concepts/possibilities.

(Fun fact: modern concept of concept originated mainly in Kant's work in 18th century.)

Overview

1. Knowledge
2. An Aside: Syntactic Knowledge
3. Common Prior Assumption and Its Implications
   - Adding Beliefs
   - Agreeing to Disagree
   - No-Trade Theorem
   - No-Trade Theorem
   - Charaterising CPA
4. Common Belief
5. Common Learning
6. Universal Type Space

Back to Semantics: Adding Beliefs to Partitional Model

Prior Belief $\mathbb{P}_i$ of player i over Ω. Assume full support, $\mathbb{P}_i > 0$.

Posterior Belief $\mathbb{P}_i(E|k_i(\omega)) = \mathbb{P}_i(E \cap k_i(\omega))/\mathbb{P}_i(k_i(\omega))$.

Common Prior $\mathbb{P} \in \Delta(\Omega)$ if $\mathbb{P}_i = \mathbb{P}\ \forall i$.

Harsanyi Doctrine: We are born equal; we have different views about the world because we receive different information.

Implications of common prior assumption (CPA)?

Agreeing to disagree and no-trade theorem.

Agreeing to Disagree

Aumann (1976 AMS): Two individuals with a common prior belief, even if they had very different information (attending different school, different upbring, etc.) cannot agree to disagree.

i.e., if differences are due to information and posteriors are common knowledge, then there can't actually be disagreement.

Theorem

Let there be a common prior $\mathbb{P}$. Suppose it is CK at ω^* that player 1's posterior beliefs on event E are m_1 whereas player 2's are m_2. Then, $m_1 = m_2$.

Agreeing to Disagree

Theorem

Let there be a common prior $\mathbb{P}$. Suppose it is CK at ω^* that player 1's posterior beliefs on event E are m_1 whereas player 2's are m_2. Then, $m_1 = m_2$.

Proof

Consider $D_i := \{\omega \mid \mathbb{P}(E \mid k_i(\omega)) = m_i\}$.

$D_1 \cap D_2$ CK at ω^*, $\exists$ evident event F s.t. $\omega^* \in F \subseteq D_1 \cap D_2$.

As F is evident, $F = K_i(F) = \{\omega \mid k_i(\omega) \subseteq F\}$. Hence, $F = \cup_{\omega \in F} k_i(\omega)$, where $\{k_i(\omega)\}_{\omega \in F}$ denotes a partition.

Since, for disjoint A, B, one has $\mathbb{P}(E|A) = \mathbb{P}(E|B) \implies \mathbb{P}(E|A) = \mathbb{P}(E|A \cup B)$, then $\mathbb{P}(E|F) = \mathbb{P}(E|k_i(\omega)) = m_i$.

Hence, $m_1 = m_2 = \mathbb{P}(E|F)$.

Theorem

Let there be a common prior $\mathbb{P}$. It cannot be CK at some ω^* that player 1's posterior beliefs on event E is strictly greater than player 2's.

Proof

Suppose not: $\exists\omega^*$ at which $D := \{\omega \mid \mathbb{P}(E \mid k_1(\omega)) > \mathbb{P}(E \mid k_2(\omega))\}$ is CK.

$\exists$ evident event F s.t. $\omega^* \in F \subseteq D$.

$\implies \forall\omega \in F, \mathbb{P}(\omega)\mathbb{P}(E|k_1(\omega)) > \mathbb{P}(\omega)\mathbb{P}(E|k_2(\omega))$.

$\implies P(F \cap E) = \sum_{\omega\in F} \mathbb{P}(\omega)P(E|k_1(\omega)) > \sum_{\omega\in F} \mathbb{P}(\omega)\mathbb{P}(E|k_2(\omega)) = \mathbb{P}(F \cap E)$, contradiction.

Agreeing to Disagree

It *can* be CK that two players have different beliefs at an event E.

Example

$\Omega = \{\omega_1, \omega_2\}$, $\mathbb{P}$ uniform common prior.

$\forall\omega$, $k_1(\omega) = \{\omega\}$ $k_2(\omega) = \Omega$.

$E = \Omega$.

"the two players have different posterior beliefs" at every state of the world, and this event is common knowledge: $K^{\infty}(\Omega) = \Omega$.

Agreeing to Disagree

Corollary

Let there be a common prior $\mathbb{P}$ and $X : \Omega \to \mathbb{R}$ a random variable. It cannot be CK at some ω^* that player 1's holds a higher expectation of X than player 2 does.

Proof

Suppose not: $\exists \omega^*$ at which $D := \{\omega \mid \mathbb{E}[X \mid k_1(\omega)] > \mathbb{E}[X \mid k_2(\omega)]\}$ is CK.

$\exists$ evident event F s.t. $\omega^* \in F \subseteq D$.

$\implies \forall \omega \in F, \frac{\mathbb{P}(\omega)}{\mathbb{P}(F)} \mathbb{E}[X|k_1(\omega)] > \frac{\mathbb{P}(\omega)}{\mathbb{P}(F)} \mathbb{E}[X|k_2(\omega)]$.

$\implies \mathbb{E}[X \mid F] = \sum_{\omega \in F} \frac{\mathbb{P}(\omega)}{\mathbb{P}(F)} \mathbb{E}[X|k_1(\omega)] > \sum_{\omega \in F} \frac{\mathbb{P}(\omega)}{\mathbb{P}(F)} \mathbb{E}[X|k_2(\omega)] = \mathbb{E}[X \mid F]$, contradiction.

No-Trade Theorem

Typical reason provided for trading: differences in information.

Milgrom and Stokey (1982 JET) show this is not exactly correct...

Definitions

- **Allocation:** $a : \Omega \to A$, a contract that associates each state with an allocation or transfer to all agents.
- **Payoffs:** Player i's state-dependent utility function $u_i(a(\omega), \omega)$.
- **Ex-ante Efficiency:** b is ex-ante efficient if $\nexists a$ s.t. $\forall i$ $\mathbb{E}[u_i(a(\omega), \omega)] \geq \mathbb{E}[u_i(b(\omega), \omega)]$ with a strict inequality for some i.
  i.e., $\sum_{\omega \in \Omega} \mathbb{P}(\omega) u_i(a(\omega), \omega) \geq \sum_{\omega \in \Omega} \mathbb{P}(\omega) u_i(b(\omega), \omega)$

No-Trade Theorem

Theorem (Milgrom and Stokey, 1982 JET)

Let there be a common prior $\mathbb{P}$. Suppose b is ex-ante efficient. If cannot be common knowledge that there is some allocation a that is weakly preferred to b by all players and strictly by at least one.

Then players cannot trade away from b even if new information k_i arrives.

Agreeing to disagree results have brutal implications for trading:

- with CPA, once we get to an ex-ante efficient allocation, there is no scope for purely information-based trade.

No-Trade Theorem

Theorem (Milgrom and Stokey, 1982 JET)

Let there be a common prior $\mathbb{P}$. Suppose b is ex-ante efficient. If cannot be common knowledge that there is some allocation a that is weakly preferred to b by all players and strictly by player 1.

Proof

Suppose $\exists a$ s.t. CK (with new information) that a is weakly preferred to b by everyone and strictly so by player 1.

Let F be evident event contained in

$\{\omega \in \Omega | \mathbb{E}[u_i(a(\omega), \omega) - u_i(b(\omega), \omega) \mid k_i(\omega)] \geq 0 \forall i \text{ and } \mathbb{E}[u_1(a(\omega), \omega) - u_1(b(\omega), \omega) \mid k_1(\omega)] > 0\}$.

Then, $\forall i$, $\mathbb{E}[u_i(a(\omega), \omega) - u_i(b(\omega), \omega) \mid F] \geq 0$, and $\mathbb{E}[u_1(a(\omega), \omega) - u_1(b(\omega), \omega) \mid F] > 0$.

Define contract c: $c(\omega) = a(\omega)$ if $\omega \in F$ and $c(\omega) = b(\omega)$ if otherwise. We get

$$\mathbb{E}[u_i(c(\omega), \omega) - u_i(b(\omega), \omega)] = \mathbb{P}(F)\mathbb{E}[u_i(a(\omega), \omega) - u_i(b(\omega), \omega)]$$

which is ≥ 0 for all i and > 0 for $i = 1$.

Contradicts b being ex-ante efficient.

No-Trade Theorem

Typical reason provided for trading: differences in information.

Milgrom and Stokey (1982 JET) show this is not exactly correct...

With CPA, once we get to an ex-ante efficient allocation, there is no scope for purely information-based trade.

Note! Prices do adjust to new information.
More: under weak conditions
(essentially strict risk-aversion, smooth EU, ex-ante efficiency
— see Milgrom and Stokey, 1982 JET, Theorem 3)
change in relative prices reveals new info available to traders and is independent of endowments, utility functions, prior beliefs, and initial allocation.

Charaterising CPA

CPA has important consequences, not just theoretical convenience.

If observe beliefs of different players, can we say whether there is a common prior?

Two fundamental contributions:

- Samet (1998 GEB), Common Priors and Separation of Convex Sets; and
- Samet (1998 GEB), Iterated Expectations and Common Priors.

Charaterising CPA

CPA has important consequences, not just theoretical convenience.

If observe beliefs of different players, can we say whether there is a common prior?

Samet (1998 GEB), Common Priors and Separation of Convex Sets

Each agent's set of priors = convex hull of agent's types.

Proposition: A common prior exists if and only if the intersection of these convex sets is nonempty.

Proof idea: Generalisation of the separation theorem: multiple convex, closed subsets of the simplex intersect $\iff$ no linear functional can simultaneously separate them.

Interpretation: Absence of common prior $\iff$ existence of a bet that everyone expects to win. (No-trade-theorem-like converse)

Charaterising CPA

CPA has important consequences, not just theoretical convenience.

If observe beliefs of different players, can we say whether there is a common prior?

Samet (1998 GEB), Iterated Expectations and Common Priors

Can we test for a common prior using only present beliefs?

Key idea: (1) Start with any random variable X. (2) Compute iterated expectations: Eve's expectation of X, Adam's expectation of Eve's expectation, Eve's expectation of Adam's expectation, ... These sequences always converge.

Proposition: Common prior exists if and only if for every X, all iterated expectation sequences converge to the same limit. Common limit is expectation under the common prior.

Proof idea: Represent type functions as Markov matrices. Common prior = invariant probability measure for all players' matrices.

Charaterising CPA

CPA has important consequences, not just theoretical convenience.

If observe beliefs of different players, can we say whether there is a common prior?

Two fundamental contributions:

- Samet (1998 GEB), Common Priors and Separation of Convex Sets; and
- Samet (1998 GEB), Iterated Expectations and Common Priors.

Also: Feinberg (2000 JET), Geanakoplos and Polemarchakis (1982 JET).

Related: Literature on merging of beliefs.
Epistemic foundations for solution concepts (Aumann 1987 Ecta; Aumann and Brandenburger 1995 Ecta).

More recently Arieli, Babichenko, Sandomirskiy, and Tamuz (2021 JPE): use 'agree to disagree' to characterise "Feasible Joint Posterior Beliefs" and applications to information design.

Overview

Common Belief

$(\Omega, \Sigma, \mathbb{P})$ probability space, I finite set agents, k_i induces partition of Ω,
$\mathcal{H}_i := \sigma(\{k_i(\omega)\}_{\omega \in \Omega})$.

"i knows E at ω" = $\{k_i(\omega) \subseteq E\}$.

"i knows E" = $K_i(E) := \{\omega \in \Omega \mid k_i(\omega) \subseteq E\}$.

From "i knows E at ω" to "i believes E wp$\geq q$ at ω" $\equiv$ "i q-believes E at ω".

Definition

$B_i^q(E) := \{\omega \in \Omega \mid \mathbb{P}(E|k_i(\omega)) \geq q\} = \{\omega \in \Omega \mid \mathbb{P}(E|\mathcal{H}_i) \geq q\}$.

"i q-believes E at ω" = $\{\mathbb{P}(E|k_i(\omega)) \geq q\}$.

Event "i q-believes E" = $B_i^q(E)$.

Properties of q-Belief

Proposition

For $E, F \in \Sigma$, $q \in [0, 1]$, $i \in I$:

1. $\mathbb{P}(E | B_i^q(E)) \geq p$.
2. $B_i^q(E) \in \mathcal{H}_i$.
3. If $E \in \mathcal{H}_i$, then $B_i^q(E) = E$.
4. $B_i^q(B_i^q(E)) = B_i^q(E)$.
5. $E \subseteq F \implies B_i^q(E) \subseteq B_i^q(F)$.
6. If (E^n) is decreasing sequence of events, then $B_i^q(\cap_n E^n) = \cap_n B_i^q(E^n)$.

1-belief = knowledge with finite models; not in continuous models:

e.g., **1**-believe that uniform random draw from $[0, 1]$ is irrational, but we don't know it.

Common q-Belief

Definition

(i) There is **mutual q-belief** of E at ω if $\omega \in B^{q,1}(E) := \cap_i B_i^q(E)$.

(ii) Let $B^{q,n+1}(E) := B^{q,1}(B^{q,n}(E))$, for $n = 1, 2, \ldots$.
There is **common q-belief** of E at ω if $\omega \in B^{q,\infty}(E) := \cap_n B^{q,n}(E)$.

Common q-belief as 'almost CK'. Common q-belief vs CK:

C = currency attack starts at 9:00.

E=At 9:00, on a phone call, there is an announcement among traders that currency attack starts.

If everyone sees announcement, E is evident and C is CK.

But if there is a small chance not everyone is paying attention, E is not evident (may even not be q-evident for high q).

Common q-Belief

Definition

E is **evident q-believed** or **q-evident** if it is mutually q-believed, $E \subseteq B^q(E)$.

q-Evident event: whenever E occurs, everyone assigns probability $\geq q$ to its occurrence.

Proposition (Monderer and Samet, 1989 GEB)

C is common q-believed at ω if and only if there is a q-evident event E s.t. $\omega \in E$ and $E \subseteq B^{q,1}(C)$.

Proof

If: $E \subseteq B^q(C) \implies \omega \in E \subseteq B^q(E) \subseteq B^{q,\infty}(C)$.

Only if: Let $E := B^{q,\infty}(C)$. Then, $E = B^q(E) = B^{q,\infty}(C)$.
By transparency, $E = B^{q,\infty}(C) \subseteq B^{q,1}(C)$.

Overview

Common Learning: Attacking a Currency

Coordinating on a Currency Attack

Two traders coordinate on when to attack currency A or B.

Every day, each trader simultaneously decides to attack currency A, B, or wait.

They only stand to gain if *both* attack the weaker currency *and* at the same time.

Every day, each receives private signal about if it's best to attack A or B (the state).

Signals iid conditional on the state, but possibly correlated across traders.

Coordination Requirements

Coordination requires traders to be sufficiently convinced if state is A or B.

With fixed state, if traders will a.s. learn the state this is not an issue.

But... i also needs to be sufficiently convinced that j is sufficiently convinced state is the same. And i also needs to be sufficiently convinced that j is sufficiently convinced that i is sufficiently convinced ... etc.

Attacking A is optimal for trader i in some period t if and only if the trader assigns probability at least q to the joint event that (i) state is A and (ii) j attacks A too (which will depend on their beliefs about a symmetric event).

Is the attack ever carried out? When does individual learning imply common learning?

Trivial case: public signals (perfect correlation). Anything else?

Why Common *learning*?

Coordination needs higher-order belief convergence, not only individual learning.

Private signals may fail to generate common knowledge.

Question: when do private signals imply common learning?

Reference: Cripps, Ely, Mailath, and Samuelson (2008 Ecta).

Setup

Time $t = 0, 1, 2, \ldots$ Players $i, j \in \{1, 2\}$ (results extend to finite I).

Parameter $\theta \in \Theta$ finite.

Period-t signals $z_t = (z_{1t}, z_{2t}) \in Z_1 \times Z_2 =: Z$.

Every period, player i observes z_{it}. Conditional on θ, $\{z_t\}_{t \geq 0}$ iid over t.

Within-t correlation across players allowed unless stated.

States $\Omega = \Theta \times Z^{\infty}$.

Prior $\mathbb{P}$ (CPA); $\mathbb{P}^{\theta}(\cdot) = \mathbb{P}(\cdot \mid \theta)$; $\mathbb{E}^{\theta}[\cdot] = \mathbb{E}[\cdot \mid \theta]$.

Belief operators at time t

Agent i's Private history $h_{it} = (z_{i0}, ..., z_{it-1})$; induced filtration $\mathcal{H}_{it} := \sigma(z_{i0}, \ldots, z_{it})$.

Posterior $\mathbb{P}(E \mid h_{it}) = \mathbb{P}(E \mid \mathcal{H}_{it} = h_{it})$ for event $E \subseteq \Theta$.

q-belief: $B_{it}^q(E) := \{\omega \mid \mathbb{P}(E \mid \mathcal{H}_{it}) \geq q\}$.

Definition (Individual learning)

Agent i learns θ if $\forall q \in (0, 1)$, $\exists T$ s.t. $\forall t > T$, $\mathbb{P}^\theta(B_{it}^q(\theta)) > q$.
Agent i learns Θ if this holds for each $\theta \in \Theta$.

Note: Individual learning is equivalent to $\lim_{t \to \infty} \mathbb{P}^\theta(B_{it}^q(\theta)) = 1 \; \forall q \in (0, 1)$.

Belief operators at time t

Agent i's Private history $h_{it} = (z_{i0}, \ldots, z_{it-1})$; induced filtration $\mathcal{H}_{it} := \sigma(\{h_{it}\})$.

Posterior $\mathbb{P}(E \mid h_{it}) = \mathbb{P}(E \mid \mathcal{H}_{it} = h_{it})$ for event $E \subseteq \Theta$.

q-belief: $B_{it}^{q}(E) := \{\omega \mid \mathbb{P}(E \mid \mathcal{H}_{it}) \geq q\}$.

Definition (Individual learning)

Agent i learns θ if $\forall q \in (0,1)$, $\exists T$ s.t. $\forall t > T$, $\mathbb{P}^{\theta}(B_{it}^{q}(\theta)) > q$.
Agent i learns Θ if this holds for each $\theta \in \Theta$.

Mutual q-belief: $B_t^q(E) \equiv B_t^{q,1}(E) := \cap_i B_{it}^q(E)$. Iterate: $B_t^{q,n+1}(E) := B_t^{q,1}(B_t^{q,n}(E))$.

Common q-belief: $C_t^q(E) := \cap_{n \geq 1} B_t^{q,n}(E)$.

Definition (Common learning)

Players commonly learn θ if $\forall q \in (0,1)$, $\exists T$ s.t. $\forall t > T$, $\mathbb{P}^{\theta}(C_t^q(\theta)) > q$.
They commonly learn Θ if this holds for each $\theta \in \Theta$.

Note: Common learning is equivalent to $\lim_{t\to\infty} \mathbb{P}^{\theta}(C_t^q(\theta)) = 1 \; \forall q \in (0,1)$.

Characterisation via Time-t q-Evident Events

Proposition (adapt. Monderer Samet 1989)

C commonly q-believed at ω and time t if and only if $\exists$ q-evident event $E : \omega \in E \subseteq B_t^q(C)$.

Corollary

Agents commonly learn Θ if and only if $\forall \theta$ and $q \in (0, 1)$, there is events E_t and period T s.t. for all $t > T$

1. (High probability) $\mathbb{P}^{\theta}(E_t) > q$;
2. (q-belief of θ) θ is q-believed on E_t at time t;
3. (q-evidence) E_t is q-evident at time t.

Proof Strategy

Define E_t s.t.

Step 1 (High probability): Show E_t has high probability.

Step 2 (q-belief of θ): Show that, on E_t, θ is q-believed at time t.

Step 3 (q-evidence): Show that E_t is q-evident at time t.

Independent Signals

Theorem

Suppose (i) each player individually learns Θ; and (ii) signals are independent across players. Then players commonly learn Θ.

Proof

Fix θ, $q \in (0,1)$ and define $E_t := \{\theta\} \cap B_t^{\sqrt{q}}(\theta)$.

Step 1 (High probability):

Learning + independence $\implies \exists T : \forall t > T, \mathbb{P}^{\theta}(E_t) = \mathbb{P}^{\theta}(B_{1t}^{\sqrt{q}}(\theta))\, \mathbb{P}^{\theta}(B_{2t}^{\sqrt{q}}(\theta)) > \sqrt{q}$.

Step 2 (q-belief of θ):

On E_t, $E_t \subseteq B_{it}^{\sqrt{q}}(\theta) \implies \mathbb{P}(\theta \mid \mathcal{H}_{it}) \geq \sqrt{q}$.

Step 3 (q-evidence): WTS $E_t \subseteq B_{it}^{q}(E_t)$, $i = 1, 2$. Focus on $i = 1$; symmetric for player 2.

Independent Signals

Theorem

Suppose (i) each player individually learns Θ; and (ii) signals are independent across players. Then players commonly learn Θ.

Proof

Fix θ, $q \in (0,1)$ and define $E_t := \{\theta\} \cap B_t^{\sqrt{q}}(\theta)$.

Step 3 (q-evidence): WTS $E_t \subseteq B_{it}^{q}(E_t)$, $i = 1, 2$. Focus on $i = 1$; symmetric for player 2.

$$\mathbb{P}(E_t|\mathcal{H}_{1t}) = \mathbb{P}(\{\theta\} \cap B_t^{\sqrt{q}}(\theta)|\mathcal{H}_{1t}) = \mathbb{P}(\{\theta\} \cap B_{1t}^{\sqrt{q}}(\theta) \cap B_{2t}^{\sqrt{q}}(\theta)|\mathcal{H}_{1t}).$$

As $B_{1t}^{\sqrt{q}}(\theta) \in \mathcal{H}_{1t}$ and from independence of signals:

$$\mathbb{P}(E_t|\mathcal{H}_{1t}) = \mathbb{P}(\{\theta\} \cap B_{1t}^{\sqrt{q}}(\theta) \cap B_{2t}^{\sqrt{q}}(\theta)|\mathcal{H}_{1t}) = \mathbf{1}_{\{B_{1t}^{\sqrt{q}}(\theta)\}}\mathbb{P}(\{\theta\} \cap B_{2t}^{\sqrt{q}}(\theta)|\mathcal{H}_{1t})$$

$$= \mathbf{1}_{\{B_{1t}^{\sqrt{q}}(\theta)\}}\mathbb{P}(\theta|\mathcal{H}_{1t})\mathbb{P}^{\theta}(B_{2t}^{\sqrt{q}}(\theta)).$$

Independent Signals

Theorem

Suppose (i) each player individually learns Θ; and (ii) signals are independent across players. Then players commonly learn Θ.

Proof

Fix θ, $q \in (0,1)$ and define $E_t := \{\theta\} \cap B_t^{\sqrt{q}}(\theta)$.

Step 3 (q-evidence): WTS $E_t \subseteq B_{it}^q(E_t)$, $i = 1, 2$. Focus on $i = 1$; symmetric for player 2.

$$\mathbb{P}(E_t|\mathcal{H}_{1t}) = \mathbf{1}_{\{B_{1t}^{\sqrt{q}}(\theta)\}}\mathbb{P}(\theta|\mathcal{H}_{1t})\mathbb{P}^{\theta}(B_{2t}^{\sqrt{q}}(\theta)).$$

$$\omega' \in E_t \implies \omega' \in B_{1t}^{\sqrt{q}}(\theta) \implies \text{on } \omega',\ \mathbf{1}_{\{B_{1t}^{\sqrt{q}}(\theta)\}}\mathbb{P}(\theta|\mathcal{H}_{1t}) \geq \sqrt{q}.$$

As for large t, $\mathbb{P}^{\theta}(B_{2t}^{\sqrt{q}}(\theta)) > \sqrt{q}$,

$$\omega' \in E_t \implies \omega' \in \{\omega \mid \mathbb{P}(E_t|\mathcal{H}_{1t}) = \mathbf{1}_{\{B_{1t}^{\sqrt{q}}(\theta)\}}\mathbb{P}(\theta|\mathcal{H}_{1t})\mathbb{P}^{\theta}(B_{2t}^{\sqrt{q}}(\theta)) > q\} = B_{1t}^q(E_t).$$

i.e., $E_t \subseteq B_{1t}^q(E_t)$. Done.

Beyond Independence and Perfect Correlation

Common learning holds trivially with perfect correlation.

Common learning also holds with independence.

Failing indepedence but less than perfect correlation: no more uniform bound on 1's beliefs about 2's beliefs used for q-evidence.

- Issue: Agent 1 may observe signals typical of θ but which lead 1 to believe 2 has seen signals less typical of θ.
- Can happen $P^{\theta}(B_{jt}^{q}(\theta) \mid \mathcal{H}_{it}) < P^{\theta}(B_{jt}^{q}(\theta))$
- i.e., conditioning complicates proof.

Setup: Finite Signal Sets

Time $t = 0, 1, 2, \ldots$ Players $i, j \in \{1, 2\}$ (results extend to finite I).

Parameter $\theta \in \Theta$ finite.

Period-t signals $z_t = (z_{1t}, z_{2t}) \in Z_1 \times Z_2 =: Z$.

Every period, player i observes z_{it}. Conditional on θ, $\{z_t\}_{t \geq 0}$ iid over t.

Within-t correlation across players allowed unless stated.

States $\Omega = \Theta \times Z^\infty$.

Prior $\mathbb{P}$ (CPA); $\mathbb{P}^\theta(\cdot) = \mathbb{P}(\cdot \mid \theta)$; $\mathbb{E}^\theta[\cdot] = \mathbb{E}[\cdot \mid \theta]$.

Theorem

If Z_1, Z_2 are finite and each player individually learns Θ, then players commonly learn Θ.

Maintained simplifying assumption: $\mathbb{P}^\theta(z) > 0$ for all z and θ.

Not necessary, but significantly lightens notational burden and proof complexity.

Finite Signal Sets

Marginals: $Q_i^\theta := (\mathbb{P}^\theta(z_i))_{z_i}$. Individual learning requires: $\theta \neq \theta' \implies Q_i^\theta \neq Q_i^{\theta'}$.

Markov kernels: $M_i^\theta := [\mathbb{P}^\theta(z_j|z_i)]_{z_i,z_j}$. ($i$'s belief about j's signal at t given each z_{it}).

Expected frequency: $Q_i^\theta M_i^\theta = Q_j^\theta$. Note: $\|M_i^\theta\|_1 = 1$

2-step kernels: $M_{ij}^\theta = M_i^\theta M_j^\theta = [\sum_{z_j} \mathbb{P}^\theta(z_i|z_j)\mathbb{P}^\theta(z_j|z_i')]_{z_i,z_i'}$.

Stationarity: $Q_i^\theta M_{ij}^\theta = Q_i^\theta$.

Empirical frequencies (up to t): $\hat{Q}_{it}$.

$\hat{Q}_{it}$: empirical frequency of i's signals.

Expected empirical frequencies of other: $\hat{Q}_{it} M_i^\theta$.

i's expectation of 2's expectation of i's empirical frequencies: $\hat{Q}_i M_{ij}^\theta$.

Theorem

If Z_1, Z_2 are finite and each player individually learns Θ, then players commonly learn Θ.

Key Lemma A: Own Frequencies Concentrate

Lemma A

For each θ, there are $\delta_t \downarrow 0$ and T s.t. $\forall t > T$, $\mathbb{P}^{\theta}(\|\hat{Q}_{it} - Q_{it}^{\theta}\| \mid \mathcal{H}_{it}) \geq 1 - \delta_t$.

Proof

$\hat{Q}_{it}$ is an avg of conditionally iid rv with $\mathbb{E}^{\theta}[\hat{Q}_{it}] = Q_{it}^{\theta} \in \Delta(\Theta)$.

Result follows from WLLN.

Key Lemma B: Concentrated Own Frequencies $\Longrightarrow$ High Posterior

Lemma B

Fix θ. There is T and $\beta_t \downarrow 0$ s.t. for all $t > T$, on the event of Lemma A ($\{||\hat{Q}_{it} - Q_i^{\theta}|| < \delta_t\}$), $P(\theta|\mathcal{H}_{it}) \geq 1 - \beta_t$.

Proof

Step 1: Define Log-likelihood ratio. $\lambda_{it}^{\theta\theta'} := \ln\left(\frac{\mathbb{P}(\theta|h_{it})}{\mathbb{P}(\theta'|h_{it})}\right)$.

$$\lambda_{it}^{\theta\theta'} = \lambda_{it-1}^{\theta\theta'} + \ln\left(\frac{Q_i^{\theta}(z_{it-1})}{Q_i^{\theta'}(z_{it-1})}\right) = \lambda_{i0}^{\theta\theta'} + \sum_{s=0}^{t-1} \ln\left(\frac{Q_i^{\theta}(z_{is})}{Q_i^{\theta'}(z_{is})}\right) = \lambda_{i0}^{\theta\theta'} + t\sum_{z_i} \hat{Q}_{it}(z_i) \ln\left(\frac{Q_i^{\theta}(z_i)}{Q_i^{\theta'}(z_i)}\right).$$

Lemma B

Fix θ. There is T and $\beta_t \downarrow 0$ s.t. for all $t > T$, on the event of Lemma A ($\{||\hat{Q}_{it} - Q_i^\theta|| < \delta_t\}$), $P(\theta|\mathcal{H}_{it}) \geq 1 - \beta_t$.

Proof

Step 2: Uniform bound on Log-likelihood ratio.

Recall Kulback-Leibler divergence $D_{KL}(p||q) := \sum_x p(x)\ln(p(x)/q(x))$.

Let $\tilde{\delta} < \min_{\theta',z_i} Q^{\theta'}(z_i)$; $\quad \hat{\delta} := \min_{\theta' \neq \theta''} D_{KL}(Q_i^{\theta''}||Q_i^{\theta'})/2$; $\quad b := \max_{\theta,\theta',z_i} Q_i^\theta(z_i)/Q_i^{\theta'}(z_i)$.

Note also that when $||\hat{Q}_{it} - Q_{it}^\theta|| < \tilde{\delta}$ then we must have $\hat{Q}_{it}^\theta > 0$.

$$\left|\lambda_{it}^{\theta\theta'} - tD_{KL}(Q_i^\theta||Q_i^{\theta'})\right| = t\left|\sum_{z_i}(\hat{Q}_{it}(z_i) - Q_i^\theta(z_i))\ln\left(\frac{Q_i^\theta(z_i)}{Q_i^{\theta'}(z_i)}\right)\right|$$

$$\leq t||\hat{Q}_{it} - Q_i^\theta|| \max_{\theta,\theta',z_i} \ln(Q_i^\theta(z_i)/Q_i^{\theta'}(z_i)) = t||\hat{Q}_{it} - Q_i^\theta|| \ln b.$$

Event from Lemma A: Take T s.t. for $t > T$, $||\hat{Q}_{it} - Q_i^\theta|| < \delta_t < \min\{\tilde{\delta},\hat{\delta}\}/(\ln b)$.

Then, $\lambda_{it}^{\theta\theta'} \geq \lambda_{i0}^{\theta\theta'} + t(D_{KL}(Q_i^\theta||Q_i^{\theta'}) - ||\hat{Q}_{it} - Q_i^\theta|| \ln b) \geq \lambda_{i0}^{\theta\theta'} + t\hat{\delta}$.

Key Lemma B: Concentrated Own Frequencies $\implies$ High Posterior

Lemma B

Fix θ. There is T and $\beta_t \downarrow 0$ s.t. for all $t > T$, on the event of Lemma A $(\{||\hat{Q}_{it} - Q_i^\theta|| < \delta_t\})$, $\mathbb{P}(\theta|\mathcal{H}_{it}) \geq 1 - \beta_t$.

Proof

Step 3: Obtaining β_t.

$$\forall \theta' \neq \theta,\ \lambda_{it}^{\theta\theta'} \geq \lambda_{i0}^{\theta\theta'} + t\hat{\delta} \iff \forall \theta' \neq \theta,\ \frac{\mathbb{P}(\theta')}{\mathbb{P}(\theta)} \geq \frac{\mathbb{P}(\theta'|h_{it})}{\mathbb{P}(\theta|h_{it})} e^{t\hat{\delta}}$$

$$\implies \frac{1-\mathbb{P}(\theta)}{\mathbb{P}(\theta)} \geq \frac{1-\mathbb{P}(\theta|h_{it})}{\mathbb{P}(\theta|h_{it})} e^{t\hat{\delta}} \iff \mathbb{P}(\theta|h_{it}) \geq \left(1 + \frac{1-\mathbb{P}(\theta)}{\mathbb{P}(\theta)} e^{-t\hat{\delta}}\right)^{-1} =: 1 - \beta_t.$$

Key Lemma C: Beliefs about Others' Frequencies Concentrate

Lemma C

Fix θ. $\forall \exists T$ and $\gamma_t \downarrow 0$ s.t. for all $t > T$, $\mathbb{P}^{\theta}(\|\hat{Q}_{it} M_i^{\theta} - \hat{Q}_{jt}\| < \gamma_t \mid \mathcal{H}_{it}) \geq 1 - \gamma_t$.

Proof

Step 1: Boole's inequality. Define $\bar{Q}_{jt}^{\theta} := \hat{Q}_{it} M_i^{\theta}$.

Noting that $\{\sum_{\ell=1}^{n} X_\ell > c\} \subseteq \cup_{\ell=1}^{n} \{X_\ell > c/n\}$, by Boole's inequality,

$$\mathbb{P}^{\theta}(\|\bar{Q}_{jt}^{\theta} - \hat{Q}_{jt}\| \geq \gamma_t \mid h_{it}) \leq \sum_{z_j} \mathbb{P}^{\theta}\left(\left|\bar{Q}_{jt}^{\theta}(z_j) - \hat{Q}_{jt}(z_j)\right| \geq \gamma_t / |Z_j| \mid h_{it}\right).$$

Reduces problem to bounding deviation of each component of the frequency vector.

Step 2: Hoeffding's Inequality. For fixed z_j, empirical frequency $\hat{Q}_{jt}(z_j) = \frac{1}{t} \sum_{s=0}^{t-1} \mathbf{1}_{\{z_{js} = z_j\}}$ is avg of independent, bounded random variables.

Conditional on h_{it} (and θ), not iid, but Hoeffding's inequality still applies.

Note conditional mean is $\mathbb{E}^{\theta}[\hat{Q}_{jt}(z_j) \mid h_{it}] = \bar{Q}_{jt}^{\theta}(z_j)$.

Hoeffding's inequality gives, for any $\varepsilon > 0$:

$$\mathbb{P}^{\theta}(|\hat{Q}_{jt}(z_j) - \bar{Q}_{jt}^{\theta}(z_j)| \geq \varepsilon \mid h_{it}) = \mathbb{P}^{\theta}(|\hat{Q}_{jt}(z_j) - \mathbb{E}^{\theta}[\hat{Q}_{jt}(z_j) \mid h_{1t}]| \geq \varepsilon \mid h_{it}) \leq e^{-2t\varepsilon^2}.$$

Key Lemma C: Beliefs about Others' Frequencies Concentrate

Lemma C

Fix θ. $\forall \exists T$ and $\gamma_t \downarrow 0$ s.t. for all $t > T$, $\mathbb{P}^{\theta}(||\hat{Q}_{it} M_i^{\theta} - \hat{Q}_{jt}|| < \gamma_t \mid \mathcal{H}_{it}) \geq 1 - \gamma_t$.

Proof

Step 1: Boole's inequality. Define $\bar{Q}_{jt}^{\theta} := \hat{Q}_{it} M_i^{\theta}$.

$\mathbb{P}^{\theta}(||\bar{Q}_{jt}^{\theta} - \hat{Q}_{jt}|| \geq \gamma_t \mid h_{it}) \leq \sum_{z_j} \mathbb{P}^{\theta}\left(\left|\bar{Q}_{jt}^{\theta}(z_j) - \hat{Q}_{jt}(z_j)\right| \geq \gamma_t/|Z_j| \mid h_{it}\right)$.

Step 2: Hoeffding's Inequality. $\mathbb{P}^{\theta}(|\bar{Q}_{jt}^{\theta}(z_j) - \hat{Q}_{jt}(z_j)| \geq \varepsilon \mid h_{it}) \leq 2e^{-2t\varepsilon^2}$.

Step 3: Combine the bounds. Let $\varepsilon = \gamma_t/|Z_j|$. Substituting into the sum from Step 1:

$$\mathbb{P}^{\theta}(||\bar{Q}_{jt}^{\theta} - \hat{Q}_{jt}|| \geq \gamma_t \mid h_{it}) \leq 2|Z_j| e^{-2t\gamma_t^2/|Z_j|^2}.$$

Step 4: Choose γ_t. Let $\gamma_t = t^{-1/3}$. For large enough t, $2|Z_j| e^{-2t^{1/3}/|Z_j|^2} \leq t^{-1/3} = \gamma_t$.

Hence, $\mathbb{P}^{\theta}(||\bar{Q}_{jt}^{\theta} - \hat{Q}_{jt}|| < \gamma_t \mid h_{it}) = 1 - \mathbb{P}^{\theta}(||\bar{Q}_{jt}^{\theta} - \hat{Q}_{jt}|| \geq \gamma_t \mid h_{it}) \geq 1 - \gamma_t$.

The bound is uniform over all histories h_{it}.

Key Lemma D: A Contraction Mapping

Lemma D

For each θ, the 2-step kernel $M_{ij}^{\theta} = M_i^{\theta} M_j^{\theta}$ is a contraction mapping on $\Delta(Z_i)$ with contraction modulus $(1 - r)$, where $r = \sum_{z_i'} \min_{z_i} \sum_{z_j} \mathbb{P}^{\theta}(z_i \mid z_j)\, \mathbb{P}^{\theta}(z_j \mid z_i') > 0$.

Proof

$M_{ij}^{\theta} > 0$ and M_{ij}^{θ} is a product of two stochastic matrices, hence also a stochastic matrix. The result follows from Stokey, Lucas, and Prescott (1989, Lemma 11.3).

Intuition: My expectation of your expectation of my frequencies ($\hat{Q}_{it} M_{ij}^{\theta}$) is "more accurate" (closer to the truth Q_i^{θ}) than my initial frequencies $\hat{Q}_{it}$.

$\|\hat{Q}_{it} M_{ij}^{\theta} - Q_i^{\theta}\| = \|\hat{Q}_{it} M_{ij}^{\theta} - Q_i^{\theta} M_{ij}^{\theta}\| \leq (1 - r)\|\hat{Q}_{it} - Q_i^{\theta}\|$.

This property stabilises the entire belief hierarchy.

Proof of the Theorem

Proof

Proof Strategy:

Use the Monderer-Samet corollary. For any $q < 1$, we must find an event E_t that, for large t, has high probability, implies q-belief in θ, and is q-evident.

Let E_t be the event where all players' empirical frequencies are very close to the true ones (condition on the fixed parameter):

$$E_t := \{\theta\} \cap \bigcap_{k \in \{i,j\}} \{\omega \mid \|\hat{Q}_{kt} - Q_k^{\theta}\| < \delta_t\}$$

where $\delta_t \to 0$ is chosen from Lemma A.

Step 1 (High probability): By Lemma A (LLN), $\mathbb{P}^{\theta}(E_t) \to 1$. So for large t, $\mathbb{P}^{\theta}(E_t) > q$.

Step 2 (q-belief of θ): By Lemma B, for any $\omega \in E_t$, players' posteriors on θ converge to 1. So for large t, $E_t \subseteq B_t^q(\theta)$.

Proof of the Theorem

Proof

Step 3 (q-evidence): For any $q < 1$, WTS $E_t \subseteq B_t^q(E_t)$ for large t.

Focus on player i. For any $\omega \in E_t$, we need to show $\mathbb{P}(E_t \mid \mathcal{H}_{it}) \geq q$.

Since $\omega \in E_t \implies \omega \in E_{it}$ (which is $\mathcal{H}_{it}$-measurable) and $\mathbb{P}(\theta \mid \mathcal{H}_{it}) \to 1$ (by Lemma B), this task reduces to showing that for any $\omega \in E_{it}$:

$$\mathbb{P}^\theta(E_{jt} \mid \mathcal{H}_{it}) = \mathbb{P}^\theta(||\hat{Q}_{jt} - Q_j^\theta|| < \delta_t \mid \mathcal{H}_{it}) \quad \text{is high.}$$

Step 3.1: Triangle Inequality. Recall player i's expectation of j's frequencies, $\bar{Q}_{jt}^\theta := \hat{Q}_{it} M_i^\theta$. For any realisation of frequencies: $||\hat{Q}_{jt} - Q_j^\theta|| \leq \underbrace{||\hat{Q}_{jt} - \bar{Q}_{jt}^\theta||}_{\text{Term 1}} + \underbrace{||\bar{Q}_{jt}^\theta - Q_j^\theta||}_{\text{Term 2}}$.

Step 3.2: Bounding the Terms. Choose $\delta_t : \gamma_t / r < \delta_t$ (we can always choose δ_t converging slower to zero).

Term 1: By Lemma C, conditional on any h_{it} and θ, the event $\{||\hat{Q}_{jt} - \bar{Q}_{jt}^\theta|| < \gamma_t\} \supseteq \{||\hat{Q}_{jt} - \bar{Q}_{jt}^\theta|| < r\delta_t\}$ occurs with probability at least $1 - \gamma_t$.

Term 2: As $\omega \in E_{it} \implies ||\hat{Q}_{it} - Q_i^\theta|| < \delta_t$, by Lemma D:

$$||\bar{Q}_{jt}^\theta - Q_j^\theta|| = ||\hat{Q}_{it} M_i^\theta - Q_i^\theta M_i^\theta|| \leq (1 - r)||\hat{Q}_{it} - Q_i^\theta|| < (1 - r)\delta_t.$$

Proof of the Theorem

Proof

Step 3 (q-evidence): For any $q < 1$, WTS $E_t \subseteq B_t^q(E_t)$ for large t.

Focus on player i. For any $\omega \in E_t$, we need to show $\mathbb{P}(E_t \mid \mathcal{H}_{it}) \geq q$.

Since $\omega \in E_t \implies \omega \in E_{it}$ (which is $\mathcal{H}_{it}$-measurable) and $\mathbb{P}(\theta \mid \mathcal{H}_{it}) \to 1$ (by Lemma B), this task reduces to showing that for any $\omega \in E_{it}$:

$$\mathbb{P}^{\theta}(E_{jt} \mid \mathcal{H}_{it}) = \mathbb{P}^{\theta}(\|\hat{Q}_{jt} - Q_j^{\theta}\| < \delta_t \mid \mathcal{H}_{it}) \quad \text{is high.}$$

Step 3.3: Conclusion. On event E_{it}, player i knows $\|\hat{Q}_{jt} - Q_j^{\theta}\| < \delta_t$ with prob. $\geq 1 - \gamma_t$.

Therefore, for large enough t and $\omega \in E_t$:

$$\mathbb{P}(E_t \mid \mathcal{H}_{it}) = 1_{\{E_{it}\}} \cap \mathbb{P}(E_{jt} \mid \mathcal{H}_{it}) = \mathbb{P}(\theta \mid \mathcal{H}_{it})\mathbb{P}^{\theta}(\|\hat{Q}_{jt} - Q_j^{\theta}\| < \delta_t \mid \mathcal{H}_{it}) > (1-\beta_t)(1-\gamma_t) \geq q.$$

We conclude $E_t \subseteq B_{it}^q(E_t)$ for $i = 1, 2$. Done.

Recap and Extensions

Common learning: $C_t^q(\theta)$ holds with probability $\to 1$.

Characterisation via high-probability q-evident events F_t.

Theorems: independence suffices; finite signals suffice.

Extensions

- Don't need assumptions on $\mathbb{P}^\theta(z)$ (just made our already long proof a bit easier).
- Extends to finitely many players.
- Extends to uncertainty and differences of belief about conditional signal-generating distributions: e.g., signals iid conditional on (θ, ρ^θ).

Limits to Common Learning

Remark

With countably infinite signal sets, players may individually learn $\boldsymbol{\Theta}$ yet fail to commonly learn: $\exists q \in (\mathbf{0}, \mathbf{1})$ s.t. $\forall t$, $\mathbb{P}^{\boldsymbol{\theta}}(C_t^q(\boldsymbol{\theta})) < q$.

Construction idea

State-specific decisive signals with probabilities decreasing in t.

Each player's posterior $\to \mathbf{1}$ almost surely (individual learning).

But each remains unsure whether the other already saw a decisive signal.

No q-evident event forms for q near $\mathbf{1}$ at any finite t.

Common Learning: Back to Attacking a Currency

Coordinating on a Currency Attack

Two traders coordinate on when to attack currency A or B.

Every day, each trader simultaneously decides to attack currency A, B, or wait.

They only stand to gain if *both* attack the weaker currency *and* at the same time.

Every day, each receives private signal about if it's best to attack A or B (the state).

Signals iid conditional on the state, but possibly correlated across traders.

Coordination Requirements

Coordination requires traders to be sufficiently convinced if state is A or B.

With fixed state, if traders will a.s. learn the state this is not an issue.

But... i also needs to be sufficiently convinced that j is sufficiently convinced state is the same. And i also needs to be sufficiently convinced that j is sufficiently convinced that i is sufficiently convinced ... etc.

Attacking A is optimal for trader i in some period t if and only if the trader assigns probability at least q to the joint event that (i) state is A and (ii) j attacks A too (which will depend on their beliefs about a symmetric event).

Is the attack ever carried out? Yes, a.s. with independent or finite signals.

Overview

1. Knowledge
2. An Aside: Syntactic Knowledge
3. Common Prior Assumption and Its Implications
4. Common Belief
5. Common Learning
6. Universal Type Space

And Now for Something Completely Different

Incomplete Information in Games:

- Players are uncertain about what actions other players can take, and their payoffs.
- Players are uncertain about their opponents' uncertainty, etc.
- Higher-order interactive uncertainty.

Harsanyi's Proposal: model uncertainty via $(S \times T_1 \times \cdots \times T_I, (\pi_i))$

- $\pi_i : T_i \to \Delta(S \times T_{-i})$: if player i is of type t_i, believes { true game is s and opponents' type is t_{-i} } with probability $\pi_i(t_i)$.
- For each t_{-i} in support of $\pi_i(t_i)$, player $-i$ believes player i's type (plus game s) is distributed according to $\pi_{-i}(t_{-i})$.
- Introduces higher-order interactive uncertainty over S.
- If S is a singleton, this reduces to the partition model with state space $\Omega = \times_i T_i$ and $k_i = \{t_i\} \times T_{-i}$.

Can this simple type construction capture any possible belief hierarchy? Or are we missing something with Harsanyi's model?

Question first resolved by Mertens and Zamir (1985 IJGT).

- Brandenburger and Dekel (1993 JET) provide a simpler treatment.

Universal Type Space

Setup:

Two players for notational simplicity.

$X_0 = S$: set of basic uncertainties, complete, separable, metric (Polish) space.

Belief Hierarchies

$\mu_1 \in \Delta(X_0)$ is player's belief over basic uncertainties (first-order).

$X_1 := X_0 \times \Delta(X_0)$: enriched space of uncertainties.

$\mu_2 \in \Delta(X_1)$ is player's belief over (S, opponent's first-order belief) (second-order).

Inductively: $X_n = X_{n-1} \times \Delta(X_{n-1})$.

$\mu_{n+1} \in \Delta(X_n)$ is player's $(n+1)$th-order belief.

Hierarchy of beliefs: $\mu = (\mu_1, \mu_2, ...) \in \times_{n=0}^{\infty} \Delta(X_n) =: T^0$.

Describes all possible incomplete information situations.

Universal Type Space

Coherent beliefs

Example: $\mu_2 \in \Delta(X_1) = \Delta(X_0 \times \Delta(X_0))$. Marginal on X_0 must be $\mu_1 \in \Delta(X_0)$.

Let $\text{marg}_{X_{n-2}} \mu_n$ denote the marginal distribution of μ_n on X_{n-2}.

Belief hierarchy μ is **coherent** if $\text{marg}_{X_{n-2}} \mu_n = \mu_{n-1}$ for all $n \geq 2$.

$T^1 \subseteq T^0$: set of all coherent belief hierarchies.

Universal Type Space

Theorem

There is a homeomorphism (continuous, one-to-one, onto, inverse function) $\pi : T^1 \to \Delta(S \times T^0)$.

Proof: Step 1 – Construction

Let $Z_0 := X_0$ and $Z_n := \Delta(X_{n-1})$ for $n \geq 1$.

Then $X_n = Z_0 \times \cdots \times Z_n$, and $S \times T^0 = \times_{n=0}^{\infty} Z_n$.

For any $\mu \in T^0$, each level is $\mu_{n+1} \in \Delta(Z_0 \times \cdots \times Z_n)$.

If $\mu \in T^1$, coherence requires $\text{marg}_{X_{n-2}} \mu_n = \mu_{n-1}$ for $n \geq 2$.

By Kolmogorov's Extension Theorem, there exists a unique measure $\mu' \in \Delta(S \times T^0)$ with marginals (μ_{n+1}).

Define $\pi(\mu) := \mu'$.

Universal Type Space

Proof: Step 2 – Properties of π

Well-defined: Each $\mu \in T^1$ yields a unique $\mu' \in \Delta(S \times T^0)$.

Injective: If $\pi(\mu) = \pi(\tilde{\mu})$, then all finite marginals coincide $\implies \mu = \tilde{\mu}$.

Onto: Given $\nu \in \Delta(S \times T^0)$:

- Define μ_{n+1} as the marginal of ν on $Z_0 \times \cdots \times Z_n$.
- Marginals are automatically consistent, hence define coherent hierarchy $\mu \in T^1$.
- Then $\pi(\mu) = \nu$.

Continuous: T^1 has product weak* topology; $\Delta(S \times T^0)$ has weak* topology.

- Weak* convergence in $\Delta(S \times T^0)$ determined by convergence on bounded continuous functions depending on finitely many coordinates.
- By construction, $\text{marg}_{Z_0 \times \cdots \times Z_n} \pi(\mu) = \mu_{n+1}$.
- If $\mu^k \to \mu$ in T^1, then $\mu^k_{n+1} \to \mu_{n+1}$ for each n.
- Hence $\pi(\mu^k) \to \pi(\mu)$ weak*, since all finite marginals converge.
- Same reasoning shows π^{-1} is continuous.

Universal Type Space

Proof: Step 3 – Conclusion

π is a bijection between T^1 and $\Delta(S \times T^0)$.

Both π and π^{-1} are continuous.

Therefore $\pi : T^1 \to \Delta(S \times T^0)$ is a homeomorphism.

Two equivalent representations of types:

- as infinite coherent belief hierarchies, or
- as single Harsanyi-type probability measures.

The Point of the Universal Type Space Theorem

Theorem

There is a homeomorphism $\pi : T^1 \to \Delta(S \times T^0)$.

Establishes an equivalence between two ways of describing beliefs.

View 1: Belief Hierarchies (Explicit but Complex)

A type is an infinite sequence: belief about S, belief about (S, others' beliefs), etc.

T^1 is the set of all coherent infinite hierarchies.

View 2: Harsanyi Types (Implicit but Simple)

A type is a single probability distribution over S and opponents' types.

$\Delta(S \times T^0)$ collects these "compressed" descriptions.

"So What?"

π is a *lossless compression*: every coherent hierarchy $\iff$ a unique measure.

Being homeomorphism is important. Implies mapping is well-behaved: onto and continuous, preserving the topological structure.

Validates Harsanyi's idea: all incomplete-information situations can be modelled via types.

The Universal Type Space

The theorem maps $\mu \in T^1$ to a belief over (S, all possible hierarchies).

But coherent players should (perhaps) only consider coherent opponents.

Enforce this iteratively:

T^2: types in T^1 assigning prob. only to T^1.

T^k: types in T^{k-1} assigning prob. only to T^{k-1}.

Universal type space is the limit: $T = \cap_{k \geq 1} T^k$.

A type $\mu \in T$ represents a coherent hierarchy, and it is CK that all players' hierarchies are coherent.

Theorem

(S, T, π) is the universal type space, with $\pi : T \to \Delta(S \times T)$ a homeomorphism.

"Universal" $\implies$ any other type space can be embedded in it.

Canonical model of incomplete information.

Epistemic Game Theory

Literature dedicated to arriving at equilibrium via introspection.

- Epistemic foundations underlying correlated equilibrium (Aumann 1987 Ecta), Nash equilibrium (Aumann and Brandenburger 1993 Ecta; Polak 1999 Ecta), level-*k* (Brandenburger, Friedenberg, and Kneeland 2025 WP).
- Epistemics underlie selection argument in global games and coordination games via higher-order uncertainty and contamination arguments (Morris, Rob, and Shin 1995 Ecta; Morris, Shin, and Yildiz 2016 JET). Laplacian selection (best-response to uniform distribution) in global games (Morris and Yang 2022 REStud).
- Miscellanea:
  - Epistemics on networks (with applications to macro and finance): Golub and Morris (2017 WP, 2018 WP).
  - Undated communication can break down common learning via infection arguments (Steiner and Stewart 2011 JET)

Ending with a teaser

Morris (2014) “Coordination, timing and common knowledge”

There is a continuum of individuals whose clocks are not perfectly synchronized. In particular, they are slow by an amount between 0 and 4 min relative to the “true time”, with the delay uniformly distributed in the population.

Each individual does not know how slow his clock is and has a uniform belief about the delay.

At what time does it become common knowledge that the true time is, say, 8:00 a.m. or later?

The answer is never.

Only when the true time reaches 8:04 does everyone know that the true time has reached 8:00.

Only at 8:08 does everyone know that everyone knows that it is past 8:00, and so on.

Thus it never becomes common knowledge.

Morris (2014) "Coordination, timing and common knowledge"

But the paradox gets worse. When does it become common 3/4-belief (...) that the true time has reached 8:00?

An individual only assigns probability 3/4 to the true time being after 8:00 when his own clock reaches 7:59.

At this point, the true time is after 8:00 as long as his clock is delayed by at least 1 min, a probability 3/4 event.

It is not until a true time of 8:02 that proportion 3/4 of individuals observe a time after 7:59: this is because at true time 8:02, individual clock times are uniformly distributed between 7:58 and 8:02.

Thus only at 8:02 is it 3/4-believed – i.e., 3/4 of the population assign probability at least 3/4 – that the true time is after 8:00.

It is only at 8:04 that it is 3/4-believed that it is 3/4-believed that the time is after 8:00, and so on. So it is also never common 3/4-belief that the time is after 8:00.

We will formalize and generalize this argument and verify that for any $p > 1/2$ it is never common p-belief that the true time is after 8:00.

But, for any $p \leq 1/2$, there is common p-belief that the true time is after 8:00 from 8:00 on.