

Lecture 2:

Epistemic foundation for Nash equilibrium

3 Epistemic Foundation for Nash Equilibrium

Most game theory courses are built around the concept of **Nash equilibrium**.

Possible reasons:

- Nash equilibrium has been crucial for development of game theory.
- Its mathematical definition is rather simple.

However, finding **epistemic conditions** that lead to Nash equilibrium is not easy.

3.1 Definition of Nash equilibrium

Consider a finite static game $\Gamma = (C_i, u_i)_{i \in I}$.

For every player i , let $\mu_i \in \Delta(C_i)$ be a probability distribution over i 's choices.

Let $\mu = (\mu_i)_{i \in I}$.

Interpretation from player i 's perspective:

For every opponent j , $\mu_j \in \Delta(C_j)$ is i 's **belief about j 's choice**.

Moreover, $\mu_i \in \Delta(C_i)$ is i 's **belief about the opponents' common belief about i 's choice**.

Let $\mu_{-i} := (\mu_j)_{j \neq i}$.

For every choice c_i , define

$$u_i(c_i, \mu_{-i}) := \sum_{c_{-i} \in C_{-i}} \left[\prod_{j \neq i} \mu_j(c_j) \right] u_i(c_i, c_{-i}).$$

This is i 's expected utility from choosing c_i while having belief μ_{-i} about opponents' choices.

Choice c_i is **optimal for player i under μ_{-i}** if

$$u_i(c_i, \mu_{-i}) \geq u_i(c'_i, \mu_{-i})$$

for all other choices c'_i .

The profile $\mu = (\mu_i)_{i \in I}$ is a **Nash equilibrium** (Nash, 1951) if for every player i , and every choice $c_i \in C_i$:

$\mu_i(c_i) > 0$ only if c_i is optimal for player i under μ_{-i} .

A choice c_i is a **Nash choice** (Perea, 2007) if there is a Nash equilibrium $\mu = (\mu_i)_{i \in I}$ such that c_i is optimal for player i under μ_{-i} .

Be careful:

Every choice c_i that receives positive probability in a Nash equilibrium is a Nash choice, but **not vice versa**:

There are Nash choices that do not receive positive probability in any Nash equilibrium.

Relation to common belief in rationality:

Every Nash choice can be chosen rationally under common belief in rationality.
(See below).

However, not every choice that can be made rationally under common belief in rationality is a Nash choice.

So: Common belief in rationality does not necessarily lead to Nash equilibrium!

3.2 Example: Teaching a lesson

Story: It is Friday, and teacher will give a surprise exam next week.

In order to pass, you must study last two days before exam.

In order to write a perfect exam, and gain a compliment by your father, you must study for six days.

Utilities for you: Pass: +5, Studying one day: -1, Compliment: +4

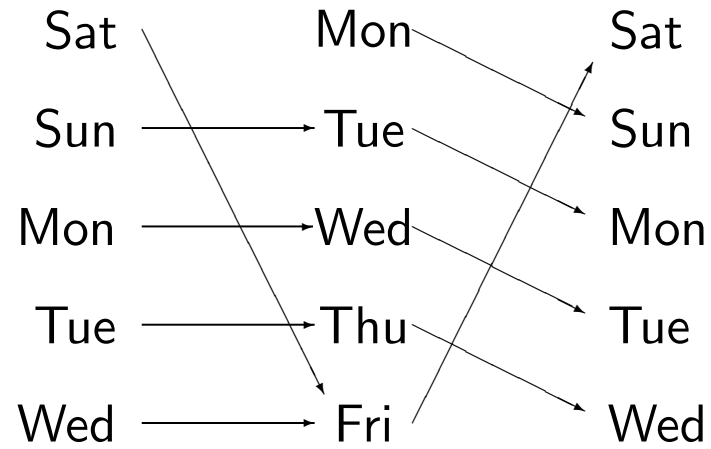
Utilities for teacher: Fail: +5, Studying one day: +1

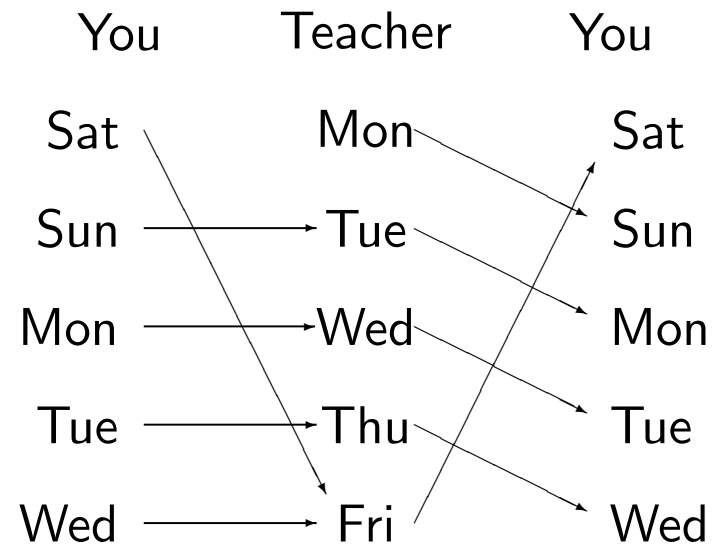
		Teacher				
		Mon	Tue	Wed	Thu	Fri
You	Sat	3, 2	2, 3	1, 4	0, 5	3, 6
	Sun	-1, 6	3, 2	2, 3	1, 4	0, 5
	Mon	0, 5	-1, 6	3, 2	2, 3	1, 4
	Tue	0, 5	0, 5	-1, 6	3, 2	2, 3
	Wed	0, 5	0, 5	0, 5	-1, 6	3, 2

When should you start studying?

	Mon	Tue	Wed	Thu	Fri
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	-1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	-1, 6	3, 2	2, 3	1, 4
Tue	0, 5	0, 5	-1, 6	3, 2	2, 3
Wed	0, 5	0, 5	0, 5	-1, 6	3, 2

You
Teacher
You





Every type expresses common belief in rationality.

So, under **common belief in rationality** you can start studying on **any day**.

	Mon	Tue	Wed	Thu	Fri
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	-1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	-1, 6	3, 2	2, 3	1, 4
Tue	0, 5	0, 5	-1, 6	3, 2	2, 3
Wed	0, 5	0, 5	0, 5	-1, 6	3, 2

Look for **Nash choices**.

So, first compute all Nash equilibria in the game.

We show that for every Nash equilibrium (μ_1, μ_2) , it is the case that $\mu_2(Fri) = 1$.

	Mon	Tue	Wed	Thu	Fri
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	-1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	-1, 6	3, 2	2, 3	1, 4
Tue	0, 5	0, 5	-1, 6	3, 2	2, 3
Wed	0, 5	0, 5	0, 5	-1, 6	3, 2

Suppose that $\mu_2(Thu) > 0$.

Then, $\mu_1(Wed) > 0$, since otherwise $u_2(Thu) < u_2(Fri)$.

Then, $\mu_2(Fri) = 1$, since otherwise $u_1(Wed) < u_1(Sat)$.

Contradiction.

	Mon	Tue	Wed	Thu	Fri
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	-1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	-1, 6	3, 2	2, 3	1, 4
Tue	0, 5	0, 5	-1, 6	3, 2	2, 3
Wed	0, 5	0, 5	0, 5	-1, 6	3, 2

Suppose that $\mu_2(Wed) > 0$.

Then, $\mu_1(Tue) > 0$, since otherwise $u_2(Wed) < u_2(Thu)$.

Then, $\mu_2(Thu) > 0$, since otherwise $u_1(Tue) < u_1(Sat)$.

Contradiction.

	Mon	Tue	Wed	Thu	Fri
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	-1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	-1, 6	3, 2	2, 3	1, 4
Tue	0, 5	0, 5	-1, 6	3, 2	2, 3
Wed	0, 5	0, 5	0, 5	-1, 6	3, 2

Suppose that $\mu_2(Tue) > 0$.

Then, $\mu_1(Mon) > 0$, since otherwise $u_2(Tue) < u_2(0.9(Wed) + 0.1(Thu))$.

Then, $\mu_2(Wed) > 0$ or $\mu_2(Thu) > 0$, since otherwise $u_1(Mon) < u_1(Sat)$.

Contradiction.

	Mon	Tue	Wed	Thu	Fri
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	-1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	-1, 6	3, 2	2, 3	1, 4
Tue	0, 5	0, 5	-1, 6	3, 2	2, 3
Wed	0, 5	0, 5	0, 5	-1, 6	3, 2

Suppose that $\mu_2(Mon) > 0$.

Then, $\mu_1(Sun) > 0$, since otherwise

$$u_2(Mon) < u_2(0.9(Tue) + 0.09(Wed) + 0.01(Thu)).$$

Then, $\mu_2(Tue) > 0$, since otherwise $u_1(Sun) < u_1(Mon)$.

Contradiction.

	Mon	Tue	Wed	Thu	Fri
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	-1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	-1, 6	3, 2	2, 3	1, 4
Tue	0, 5	0, 5	-1, 6	3, 2	2, 3
Wed	0, 5	0, 5	0, 5	-1, 6	3, 2

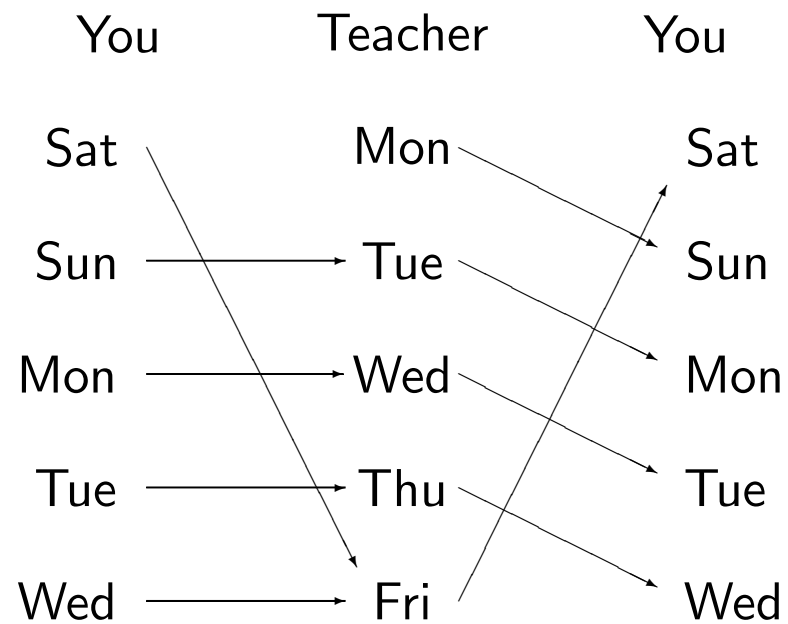
So, in every Nash equilibrium (μ_1, μ_2) , we must have that $\mu_2(Fri) = 1$.

Hence, *Sat* and *Wed* are the only Nash choices for you.

Question:

Which epistemic conditions separate **common belief in rationality** from **Nash equilibrium**?

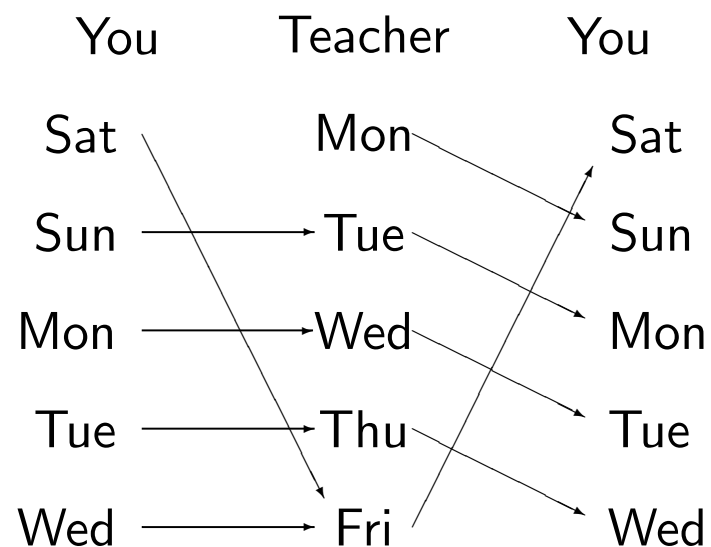
So, which additional conditions should we impose **beyond** common belief in rationality to arrive at Nash equilibrium?



Remember: Your only Nash choices are *Sat* and *Wed*.

These choices are supported by your type t_1^{Sat} , which believes that the teacher chooses *Fri*.

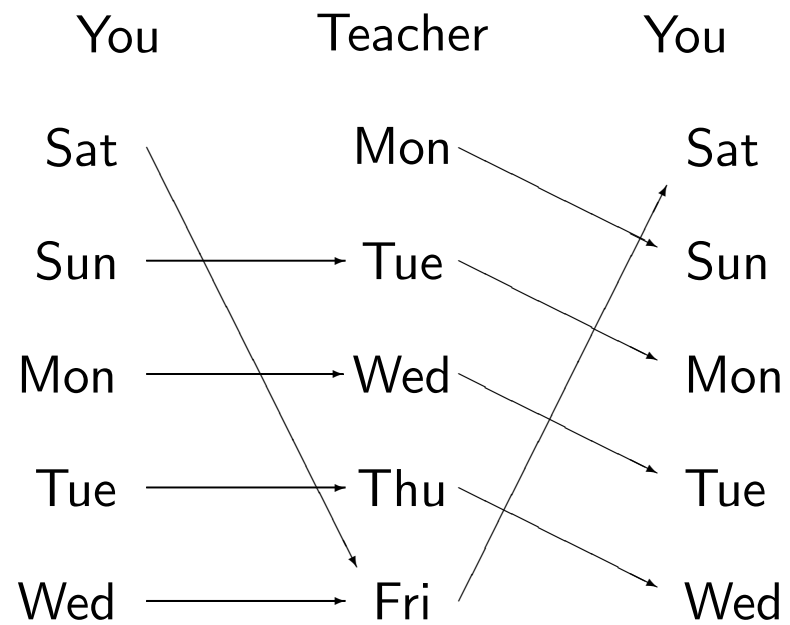
What is special about t_1^{Sat} ?



At type t_1^{Sat} , you believe that teacher is correct about your belief.

We say: Type t_1^{Sat} has **self-referential beliefs** (Perea, 2007).

At type t_1^{Sat} , you **believe that teacher has self-referential beliefs.**



Your other types do not have self-referential beliefs.

3.3 Self-referential beliefs

Let $\Gamma = (C_i, u_i)_{i \in I}$ be a finite static game.

Let $\mathbf{M} = (T_i, b_i)_{i \in I}$ be a finite epistemic model for Γ where

- T_i is the finite set of types for player i , and
- b_i is a function that assigns to every type $t_i \in T_i$ a probabilistic belief $b_i(t_i) \in \Delta(C_{-i} \times T_{-i})$.

Type t_i has **self-referential beliefs** (Perea, 2007) if,

whenever t_i believes an event $E \subseteq C_{-i} \times T_{-i}$ with probability p ,

then he believes that every opponent j believes that he believes E with probability p .

Consequence:

If type t_i has self-referential beliefs, then he believes that every opponent believes that his type is t_i .

3.4 Epistemic foundation for two players

For games with two players, one can show that:

- believing in the opponents' rationality
- believing that the opponent believes in your rationality,
- having self-referential beliefs, and
- believing that the opponent has self-referential beliefs,

leads to Nash equilibrium.

Theorem 3.1. (Perea, 2007)

Consider a finite static game with two players, i and j .

Let (T_i, T_j, b_i, b_j) be a finite epistemic model, and $t_i \in T_i$ a type such that

- t_i believes in j 's rationality, and believes that j believes in i 's rationality,
- t_i has self-referential beliefs, and believes that j has self-referential beliefs.

Then, every rational choice for t_i is a **Nash choice**.

In fact, if

- t_i believes in j 's rationality, and believes that j believes in his i 's rationality,
- t_i has self-referential beliefs, and believes that j has self-referential beliefs,

then t_i expresses **common** belief in rationality.

Moreover, none of the four epistemic conditions above can be dropped.

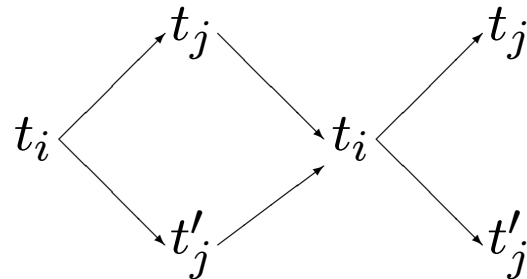
We prove the theorem by number of intermediate results which are interesting on their own.

Step 1. Let t_i be a type that

- has self-referential beliefs, and
- believes that j has self-referential beliefs.

Then, t_i assigns probability 1 to a single type t_j for player j .

Proof of step 1. Suppose that t_i would assign positive probability to two different types t_j and t'_j for opponent j .



Then, t_j believes, with prob.1, that i believes, with positive prob., that j 's type is t'_j .

So, t_j would not have self-referential beliefs. This is a contradiction.

Step 2. Let t_i be a type that

- believes in j 's rationality, and believes that j believes in i 's rationality,
- has self-referential beliefs, and believes that j has self-referential beliefs.

Then, t_i 's belief about j 's choice, together with t_i 's belief about j 's belief about i 's choice, form a **Nash equilibrium**.

Proof of step 2. By Step 1, the situation from t_i 's perspective is as follows:

$$t_i \xrightarrow{\mu_j} t_j \xrightarrow{\mu_i} t_i$$

Let $\mu_j \in \Delta(C_j)$ be t_i 's belief about j 's choice, and $\mu_i \in \Delta(C_i)$ be t_j 's belief about i 's choice.

As t_i **believes that j chooses rationally,**

$$\mu_j(c_j) > 0 \text{ only if } c_j \text{ optimal under } \mu_i.$$

As t_i **believes that j believes that i chooses rationally,**

$$\mu_i(c_i) > 0 \text{ only if } c_i \text{ optimal under } \mu_j.$$

Hence, (μ_j, μ_i) is a **Nash equilibrium.**

Step 3. Let t_i be a type that

- believes in j 's rationality, and believes that j believes in i 's rationality,
- has self-referential beliefs, and believes that j has self-referential beliefs.

Then, every choice that is rational for t_i is a **Nash choice**.

Proof of step 3. By Step 2, t_i 's belief μ_j about j 's choice is part of a Nash equilibrium (μ_j, μ_i) .

Hence, every choice that is rational for t_i must be a **Nash choice**.

In fact, the opposite of Theorem 3.1 is also true.

Theorem 3.2. (Perea, 2007)

Consider a finite static game with two players, i and j . Let c_i be a **Nash choice**. Then, there is a finite epistemic model and a type $t_i \in T_i$ such that

- t_i believes in j 's rationality, and believes that j believes in i 's rationality,
 - t_i has self-referential beliefs, and believes that j has self-referential beliefs, and
 - c_i is **rational for** t_i .
-

Consequence: Every Nash choice can be chosen rationally under common belief in rationality.

Nash (1951) has shown that a Nash equilibrium always exist for games with finitely many choices.

So, a Nash choice always exists.

In particular, the four epistemic conditions in Theorems 3.1 and 3.2 never lead to logical contradictions in such games.

3.5 Example: The lazy professor

Story: Within three days you have an exam.

Exam is about Chapters 1, 2, 3 and 4, and consists of one question only.

Question is about one of the chapters.

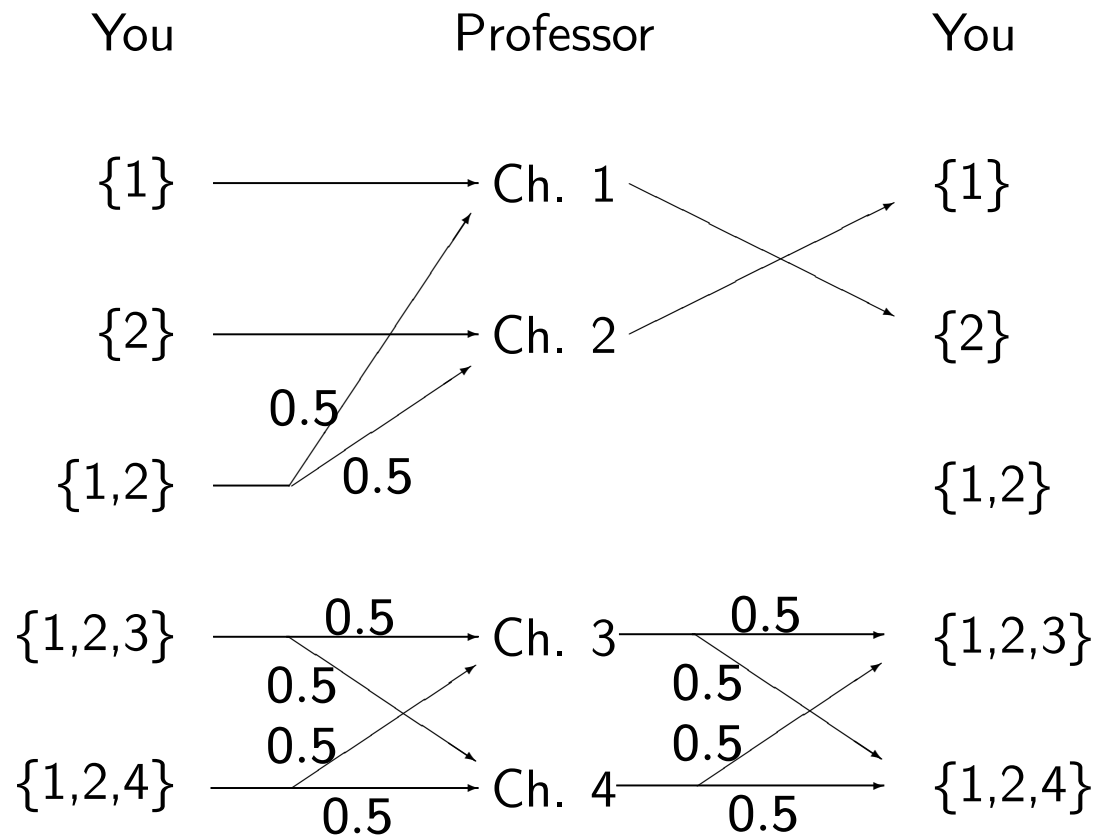
Studying one chapter takes one day.

For understanding Chapter 3 or 4, you must first have studied Chapters 1 and 2.

Utilities for you: Pass: +5, Studying one day: -1.

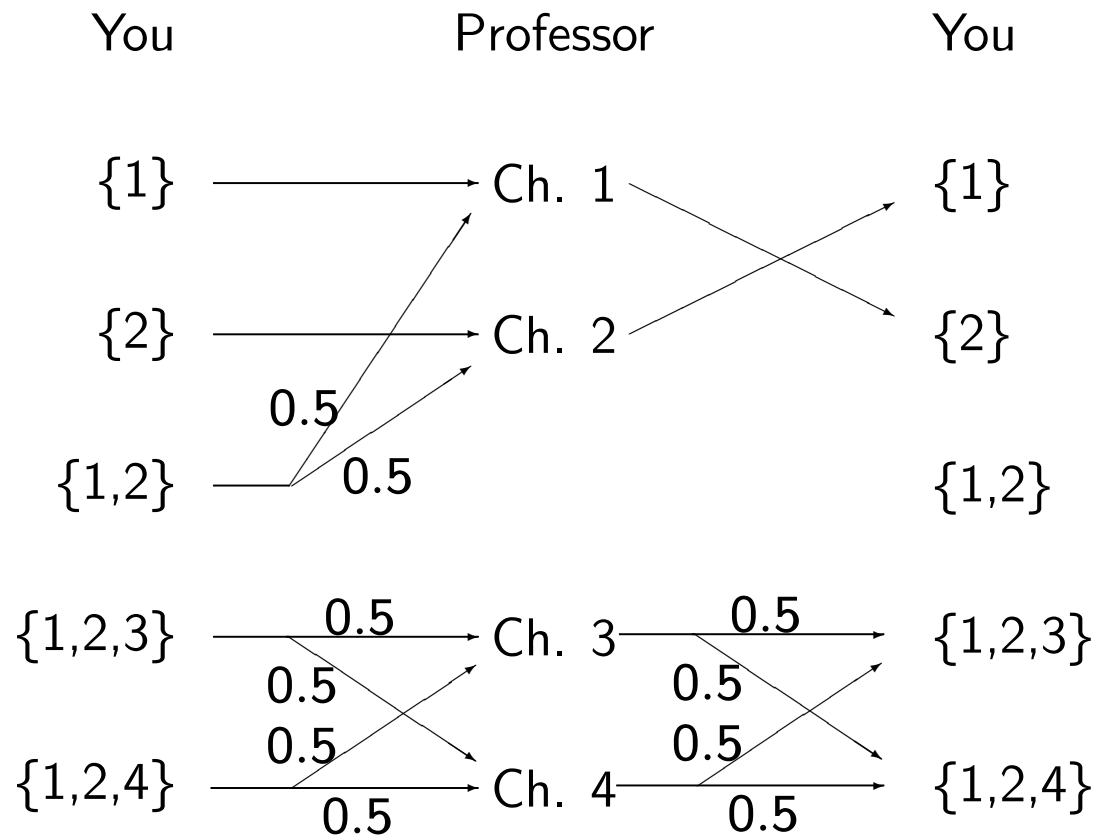
Utilities for professor: Fail: +5, Preparing a question about Ch. 1 or 2: -1, Preparing a question about Ch. 3 or 4: -2.

		Professor			
		Ch.1	Ch.2	Ch.3	Ch.4
You	{1}	4, -1	-1, 4	-1, 3	-1, 3
	{2}	-1, 4	4, -1	-1, 3	-1, 3
	{1, 2}	3, -1	3, -1	-2, 3	-2, 3
	{1, 2, 3}	2, -1	2, -1	2, -2	-3, 3
	{1, 2, 4}	2, -1	2, -1	-3, 3	2, -2



You can rationally make **any choice** under common belief in rationality.

However, types $t_1^{\{1\}}$, $t_1^{\{2\}}$ and $t_1^{\{1,2\}}$ do **not** have **self-referential beliefs**.



Which study plans can you rationally choose if you hold **self-referential beliefs**, and **believe that professor holds self-referential beliefs** ?

		Professor			
		Ch.1	Ch.2	Ch.3	Ch.4
You	{1}	4, -1	-1, 4	-1, 3	-1, 3
	{2}	-1, 4	4, -1	-1, 3	-1, 3
	{1, 2}	3, -1	3, -1	-2, 3	-2, 3
	{1, 2, 3}	2, -1	2, -1	2, -2	-3, 3
	{1, 2, 4}	2, -1	2, -1	-3, 3	2, -2

There is a unique Nash equilibrium:

$$\left(\frac{1}{2}\{1, 2, 3\} + \frac{1}{2}\{1, 2, 4\}, \frac{1}{2}Ch.3 + \frac{1}{2}Ch.4\right).$$

So, the only Nash choices are to study $\{1, 2, 3\}$ and $\{1, 2, 4\}$.

	Ch.1	Ch.2	Ch.3	Ch.4
{1}	4, -1	-1, 4	-1, 3	-1, 3
{2}	-1, 4	4, -1	-1, 3	-1, 3
{1, 2}	3, -1	3, -1	-2, 3	-2, 3
{1, 2, 3}	2, -1	2, -1	2, -2	-3, 3
{1, 2, 4}	2, -1	2, -1	-3, 3	2, -2

Hence, if you hold **self-referential beliefs**, and **believe that the professor holds self-referential beliefs**, then you must study Chapters 1,2,3 or Chapters 1,2,4.

3.6 Extension to more than two players

If there are more than two players, the epistemic conditions in Theorem 3.1 are no longer enough to imply Nash equilibrium.

Problem 1: Player 1's belief about 2's choice may not be independent from his belief about 3's choice.

So, player 1's belief about opponents' choices may not be representable by $\mu_2 \in \Delta(C_2)$ and $\mu_3 \in \Delta(C_3)$.

Extra condition:

Type t_i must have **conditionally independent beliefs**.

Problem 2: Player 1's belief about 3's choice may be different from his belief about 2's belief about 3's choice.

So, $\mu_3 \in \Delta(C_3)$ may have an ambiguous interpretation.

Extra condition:

Type t_i must have **projective beliefs**.

Suppose that type t_i assigns positive probability to the opponents' profile of types $t_{-i} \in T_{-i}$.

By $b_i(t_i)(c_{-i}|t_{-i})$ we denote the **conditional probability** that $b_i(t_i)$ assigns to the profile of choices $c_{-i} \in C_{-i}$, **conditional on the event that the opponents' types are t_{-i} .**

By $b_i(t_i)(c_j|t_j)$ we denote the **conditional probability** that $b_i(t_i)$ assigns to j 's choice $c_j \in C_j$, **conditional on the event that j 's type is t_j .**

Type t_i has **conditionally independent beliefs** (Brandenburger and Friedenberg, 2007) if

$$b_i(t_i)(c_{-i}|t_{-i}) = \prod_{j \neq i} b_i(t_i)(c_j|t_j)$$

for all t_{-i} to which t_i assigns positive probability.

So, **conditional on the opponents' types being** t_{-i} , belief about j 's choice and belief about k 's choice are uncorrelated.

Let k be an opponent for player i .

An **event about player k** is a set $E_k \subseteq C_k \times T_k$.

Type t_i has **projective beliefs** (Perea, 2007) if for every two opponents j and k , and for every event E_k about player k :

if t_i believes E_k with probability k , then t_i believes, with probability 1, that j believes E_k with probability k .

So, t_i projects his belief about player k on player j .

Theorem 3.3. (Perea, 2007)

Consider a finite static game with at least three players. Let t_i be a type that:

- believes in the opponents' rationality, believes that his opponents believe in their opponents' rationality, has self-referential beliefs, believes that his opponents have self-referential beliefs,
- has conditionally independent beliefs, believes that every opponent has conditionally independent beliefs, and believes that every opponent has projective beliefs.

Then, every rational choice for t_i is a **Nash choice**.

None of the seven epistemic conditions in Theorem 3.3 can be dropped.

Why is it not imposed that t_i itself has **projective beliefs**?

It follows from the other conditions:

If t_i has self-referential beliefs, and believes that every opponent has projective beliefs, that t_i itself must have projective beliefs.

The opposite of Theorem 3.3 is also valid.

Theorem 3.4. (Perea, 2007)

Consider a finite static game with at least three players. Let c_i be a **Nash choice** for player i .

Then, there is a finite epistemic model, and a type t_i in this model, such that

- t_i satisfies the seven conditions in Theorem 3.3, and
 - c_i is **rational** for t_i .
-

Since a Nash choice always exists, one can always construct an epistemic model and a type t_i in this model that satisfies the seven epistemic conditions in Theorem 3.3.

So, the seven conditions in Theorem 3.3 do not lead to logical contradictions.

3.7 Related Models

Other epistemic foundations for Nash equilibrium have been provided by:

Brandenburger and Dekel (1987)

Aumann and Brandenburger (1995)

Asheim (2006)

These papers impose conditions on the beliefs of **all** players, whereas Perea (2007) only imposes conditions on the belief of a **single** player.

3.8 References

John Nash (1951): “Non-cooperative games”, *Annals of Mathematics* 54, 286-295.

Andrés Perea (2007): “A one-person doxastic characterization of Nash strategies”, *Synthese* 158, 251-271. (*Knowledge, Rationality and Action* 341-361).

Adam Brandenburger and Amanda Friedenberg (2007): “Intrinsic correlation in games”, Forthcoming in *Journal of Economic Theory*, Downloadable from <http://pages.stern.nyu.edu/~abranden/>

Adam Brandenburger and Eddie Dekel (1987): “Rationalizability and correlated equilibria”, *Econometrica* 55, 1391-1402.

Robert Aumann and Adam Brandenburger (1995): “Epistemic conditions for Nash equilibrium”, *Econometrica* 63, 1161-1180.

Geir Asheim (2006): “The consistent preferences approach to deductive reasoning in games”, *Theory and Decision Library*, Springer, Dordrecht, The Netherlands.