Olivier Gossner

Game Theory

Lecture notes. **Modeling Information and Knowledge**

Information is a central concept in Game Theory. It is important when one wants to explore the foundations of equilibrium concepts (what do players need to know in order to play a Nash equilibrium, or a correlated equilibrium for instance?), or more simply when we are interested in agents who face some uncertainties about the strategic situation they face. Examples of such uncertainties include.

- Negotiation: objectives.
- Firms: cost functions
- Sale of a used car: the quality of the car
- Auctions: valuations of the bidders

1 States of nature and Knowledge

1.1 Possibility Correspondences

We introduce a way to model incomplete information. This is done through a space of "states of the world". Each state of the world describes a resolution of all uncertainties, including each

player's preferences, as well as each player's information. We start by describing a player's information at each state of the world. W is the (finite) set of states of the world.

I denotes the finite set of players, $i \in I$ denotes an individual player.

Definition 1.1. An *possibility correspondence* for player i a map $P_i \colon \mathsf{W} \mapsto P(\mathsf{W})$.

The interpretation of P_i is that, at w, $P_i(w)$ is the set of states that player i considers as possible.

Example 1.2. In the first example, there are two states of the world corresponding to different weather conditions, and the agent knows which state realizes.

$\mathsf{W} = \{\text{rain}, \text{shine}\}$, $P_i(\text{rain}) = \{\text{rain}\}$, $P_i(\text{shine}) = \{\text{shine}\}$.

In the second example, the agent doesn't know which state realizes.

$\mathsf{W} = \{\text{rain}, \text{shine}\}$, $P_i(\text{shine}) = P_i(\text{rain}) = \{\text{rain}, \text{shine}\}$.

In the third example, there are 4 states of the world, two in which the agent knows the weather, and two in which the agent doesn't know the weather.

$\mathsf{W} = \{\text{rain}_1, \text{shine}_1, \text{rain}_2, \text{shine}_2\}$. $P_i(\text{rain}_1) = \{\text{rain}_1\}$, $P_i(\text{shine}_1) = \{\text{shine}_1\}$, $P_i(\text{shine}_2) = P_i(\text{rain}_2) = \{\text{rain}_2, \text{shine}_2\}$.

The possibliity correspondences of Example 1.2 satisfy the two important properties below.

Non-delusion $w \in P_i(w)$.

Introspection If $w' \in P_i(w)$ then $P_i(w') = P_i(w)$

The interpretation of non-delusion is that the agent never excludes the truth. To understand introspection, imagine that the agent i knows W, and P_i. At state w, the agent knows the value of $P_i(w)$ (he knows which states he excludes and which he considers as possible). Then, the agent has the possibility to check all states w', and excludes the one for which $P_i(w')$ differs from $P_i(w)$.

A possibility correspondence P_i is *partitional* if there exists a partition of W such that, $P_i(w)$ is the element of the partition that contains w. In the previous example, all possibility correspondences are partitional.

The following Theorem draws the connection between non-delusion and introspection, and partitional possibility correspondences.

Theorem 1.3. *A possibility correspondence is partitional if and only if it satisfies Non-delusion and Introspection.*

Proof. First assume that P_i satisfies Non-delusion and Introspection. Let $\mathcal{P}$ be the partition of Ω such that ω, ω' are in the same element of $\mathcal{P}$ if and only if $P_i(\omega) = P_i(\omega')$, and let $\mathcal{P}(\omega) = \{\omega', P_i(\omega') = P_i(\omega)\}$ denote the element of $\mathcal{P}$ that contains ω. We prove that $P_i(\omega) = \mathcal{P}(\omega)$. According to Introspection, if $\omega' \in P_i(\omega)$ then $P_i(\omega') = P_i(\omega)$, hence $\omega' \in \mathcal{P}(\omega)$. If $\omega' \in \mathcal{P}(\omega)$, then $P_i(\omega) = P_i(\omega')$ and Non-delusion implies $\omega' \in P_i(\omega')$, hence $\omega' \in P_i(\omega)$.

Now assume that P_i is partitional, *i.e.* $P_i(\omega) = \mathcal{P}(\omega)$ for some partition $\mathcal{P}$ of Ω. P_i is Non-delusional, since $\omega \in \mathcal{P}(\omega)$, it satisfies Introspection since $\mathcal{P}(\omega) = \{\omega', \mathcal{P}(\omega) = \mathcal{P}(\omega')\}$. □

Example 1.4. There are 100 states of nature, numbered 00 to 99. The agent can read the first digit, but not the second one. What is $P_i(\omega)$? Does it verify non-delusion? Introspection?

Example 1.5. There are 100 states of nature, numbered 00 to 99. The agent reads both digits, but the other way round. What is $P_i(\omega)$? Does it verify non-delusion? Introspection?

Example 1.6. There are two states of nature G and B, standing for "good news" or "bad news". The agent learns the state of nature, but forgets the bad news, and remembers only the good news. Does the possibility correspondence satisfy non-delusion, introspection?

Exercise 1.7. There are two states of nature G and B, standing for "good state" or "bad state". The agent always believes that the state is good. Does the possibility correspondence satisfy non-delusion, introspection?

1.2 Knowledge

We relate possibility correspondences to the notion of knowledge. An event E is a subset of Ω. The event E is "known at state ω" if $P_i(\omega) \subseteq E$. $K_i E$ is the event that E is known:

$$K_i E = \{\omega, P_i(\omega) \subseteq E\}$$

$K_i E$ is thus the set of states ω at which all possibilities considered as possible by the agent belong to E ; in all these possibilities E is true, therefore the agent "knows E".

Example 1.8. Consider the three possibility correspondences of Example 1.2. Let E be the event that represents the shiny weather.

In the first possibility correspondences, $E = \{\text{shine}\}$, and $K_i E = \{\text{shine}\}$.

In the second possibility correspondence, $E = \{\text{shine}\}$ and $K_i E = \emptyset$.

In the third possibility correspondence, $E = \{\text{shine}_1, \text{shine}_2\}$, and $K_i E = \{\text{shine}_1\}$.

The following properties seem "reasonable" properties of the knowledge operator associated to a rational agent. These properties can be understood as characteristics of the agent's reasoning abilities.

Deductive Closure $K_i(E \cap F) = K_iE \cap K_iF$: an agent knows E and F if and only if this agent knows E and knows F.

Truth $K_iE \subseteq E$: E must be true if E is known

Positive Introspection $K_iE \subseteq K_iK_iE$: If the agent knows E, he knows that he knows E

Negative Introspection $\neg K_iE \subseteq K_i\neg K_iE$: If E is not known, then it is known that E is not known.

Proposition 1.9. *Every knowledge operator that is derived from a partitional possibility correspondence satisfies deductive closure.*

Instead of modeling information with possibility correspondences, we could have started right away using knowledge operators K_i. Are the two approaches equivalent?

We already know how to associate a knowledge operator to a possibility correspondence, can we also associate a possibility correspondence to a knowledge operator? Suppose we were to observe the knowledge operator K_i of agent i. Could we explain this knowledge as coming from a possibility correspondence?

Say that K_i is derived from P_i if for every event E, $K_iE = \{w, P_i(w) \subseteq E\}$. What knowledge operators K_i are the ones derived from possibility correspondences P_i? We know how to characterize K_i from P_i. Is there a way we can characterize P_i from K_i?

The following proposition provides a simple answer. According to the first point K_i is associated to some P_i if and only if K_i satisfies deductive closure. If K_i satisfies deductive closure, then the corresponding P_i is given by point 2, which says that $P_i(w)$ is the intersection of all events known at w. By deductive closure, this is equivalent to saying that $P_i(w)$ is the *smallest* event known at w.

Proposition 1.10. *Let K_i be the agent's knowledge operator and P_i be a possibility correspondence. Then K_i is associated to P_i if and only if*

1. *K_i satisfies Deductive Closure*
2. *For every w:*

$$P_i(w) = \bigcap_{E \text{ s.t. } w \in K_iE} E$$

Proof. First take K_i and P_i such that $K_iE = \{w, P_i(w) \subseteq E\}$ for every E. For every E, F, $K_i(E \cap F) = \{w, P_i(w) \subseteq E \cap F\} = \{w, P_i(w) \subseteq E\} = \{w, P_i(w) \subseteq F\} = K_iE \cap K_iF$. Hence K_i satisfies Deductive Closure. We now prove $P_i(w) = \cap_{E \text{ s.t. } w \in K_iE} E$. To show $\subseteq$, if E is such that $w \in K_iE$, then $P_i(w) \subseteq E$. To show $\supseteq$, note that $w \in K_iP_i(w)$, hence $\bigcap_{E \text{ s.t. } w \in K_iE} E \subseteq P_i(w)$.

Now consider K_i that satisfies Deductive Closure, we and let $P_i(w) = \cap_{E \text{ s.t. } w \in K_iE} E$ for every w. We need to prove that for this particular P_i, $K_iE = \{w, P_i(w) \subseteq E\}$, hence we want to prove:

$$K_iE = \left\{ w, \bigcap_{E' \text{ s.t. } w \in K_iE'} E' \subseteq E \right\}$$

For $w' \in K_iE$, we have $\cap_{E' \text{ s.t. } w' \in K_iE'} E' \subseteq E$, hence $w' \in \{w, \cap_{E' \text{ s.t. } w \in K_iE'} E' \subseteq E\}$.

Now assume w is such that $\cap_{E' \text{ s.t. } w \in K_iE'} E' \subseteq E$. Deductive Closure shows that the intersection of all events known at w is also known at w: $w \in K_i(\cap_{E' \text{ s.t. } w \in K_iE'} E')$. Deductive Closure also implies that if $A \subseteq B$, $K_iA = K_iB \cap K_i(A \cap B) \subseteq K_iB$. Applying this to $A = \cap_{E' \text{ s.t. } w \in K_iE'} E'$, $B = E$, we deduce $w \in K_iA \subseteq K_iB = K_iE$.

□

Proposition 1.10 shows that every K_i that satisfies Deductive Closure can be derived from some P_i. Furthermore, this P_i is unique, and given by the formula in 2. We can thus either work with K_i that satisfies Deductive Closure, or, equivalently, using the corresponding possibility correspondence P_i.

Theorem 1.11. *Assume that K_i is associated to P_i. K_i verifies Truth if and only if P_i satisfies non-delusion. K_i verifies Positive Introspection and Negative Introspection if and only if P_i verifies Introspection.*

Proof. Truth: Assume P_i satisfies Non-Delusion, if $w \in K_iE$ then $w \in P_i(w) \subseteq E$, hence K_i satisfies Truth. If K_i satisfies Truth, w is in all events E such that $w \in K_iE$, hence in $P_i(w) = \cap_{E \text{ s.t. } w \in K_iE} E$, which is Non-Delusion.

Assume P_i verifies Introspection. To prove that K_i satisfies Positive Introspection, let $w \in K_iE$, *i.e.* $P_i(w) \subseteq E$. We need to show that $w \in K_iK_iE$, *i.e.* $P_i(w) \subseteq K_iE$. Let $w' \subseteq P_i(w)$, then $P_i(w') = P_i(w) \subseteq E$, which shows that $w' \in K_iE$, hence the result. To prove that K_i satisfies Negative Introspection, assume $w \in \neg K_iE$, *i.e.* $P_i(w) \not\subseteq E$. We want to prove $w \in K_i\neg K_iE$, *i.e.* $P_i(w) \subseteq \neg K_iE$. For every $w' \in P_i(w)$, $P_i(w') = P_i(w) \not\subseteq E$, which is the desired result.

Now assume that K_i satisfies Positive Introspection and Negative Introspection. According to Positive Introspection, $P_i(w) \subseteq E$ means that $w \in K_iE \subseteq K_iK_iE$, hence $P_i(w) \in K_iE$, which means that for every $w' \in P_i(w)$, $P_i(w') \subseteq E$. Applying this to $E = P_i(w)$ shows that for every $w' \in P_i(w)$, $P_i(w') \subseteq P_i(w)$. According to negative Introspection, $P_i(w) \not\subseteq E$, means that

$w \in \neg K_i E \subseteq K_i \neg K_i E$, thus for every $w' \in P_i(w)$, $P_i(w') \not\subseteq E$. We want to prove that for every $w' \in P_i(w)$, $P_i(w') \supseteq P_i(w)$. Assume by contradiction that $P_i(w') \not\supseteq P_i(w)$. The Negative Introspection property applied to $E = P_i(w')$ gives $P_i(w') \not\subseteq P_i(w')$, a contradiction. □

Exercise 1.12. This is taken from Conan Doyle's famous novel "Silver Blaze".

A dialog takes place between the famous detective Sherlock Holmes and the Scottland Yard detective Gregory:

> Gregory: *Is there any other point to which you would wish to draw my attention?*
>
> Holmes: *To the curious incident of the dog in the night-time.*
>
> Gregory: *The dog did nothing in the night-time.*
>
> Holmes: *That was the curious incident.*

From noticing that the dog did not bark in the night-time, Holmes infers that no one intruded in the house. On the other hand, Gregory, who was able to make the same observations as Holmes did, failed to reach this conclusion. Had the dog barked, Watson would certainly have inferred that someone intruded.

Model Holmes and Watson's knowledge. Do they satisfy non-delusion, positive introspection, negative introspection?

2 Interactive knowledge

The set of states at which E is *mutually known* (known by all players) is

$$KE = \cap_i K_i E$$

Now consider the common knowledge operator K. What can we say about the poroperties that it satisfies, based on the properties satisfied by each of the K_i?

Proposition 2.1. *If each K_i satisfies Deductive Closure, then K also satisfies deductive closure. If at least one K_i satisfies Truth, then K satisfies Truth.*

Proof left as an exercise.

The first part of the Theorem is quite useful. If each K_i can be derived from a possibility correspondence P_i, then K can also be derived from a possibility correspondence, call it P. In many examples, it is in fact easier to describe P than K, as P requires the description of an subset of W for each element of W, but K requires the description of a subset of W for each subset of W.

On the other hand, even if each K does not inherit properties of Positive and Negative introspection from the individual K_i's. Consider the following example:

Exercise 2.2. $\mathsf{W} = \{1,2,3,4,5\}$. $P_1 = \{\{1\},\{2,3\},\{4,5\}\}$, $P_2 = \{\{1\},\{2\},\{3,4\},\{5\}\}$.

1. Let $E = \{3,4,5\}$. At which states is E mutually known? At which states is it mutually known of order 2? At which states is it commonly known?
2. Same questions for $F = \{2,3,4,5\}$.
3. What events are mutually known at in $w = 2$?
4. What events are mutually known of order 2 at $w = 2$?
5. What events are commonly known at $w = 2$?

It can be seen in the example that $P(w) = \cup_i P_i(w)$. This is in fact logical "if one person in the group cannot exclude a state w', then we cannot say that the group commonly knows that the state is not w'".

We now move from mutual knowledge to mutual knowledge of higher orders. The set of states at which all players know that all players know that they know ... that they know (k times) E is

$$K^k E = K \dots K E$$

It can be easily seen that K^k satisfies deductive closure if each K_i does. What can be said about the possibility correspondence P^k associated to K^k? It is left as an exercise to show the following:

$$P^k(w) = \{w', \exists w_1, \dots, w_k, w_1 = w, w_k = w', \forall 1 \leq i \leq k w_{i+1} \in P(w_i\}$$

This shows that $P^k(w)$ is the set of states w' which can be reached from w using a chain of lengths k, where the chain is obtained by following a possibility correspondence P_i at each iteration.

The set of states at which E is *common knowledge* is

$$CKE = \cap_k K^k E$$

Again, Common Knowledge satisfies deductive closure if each K_i does. Hence in this case it can be derived from a possibility correspondence, call it CP. CP can be derived as follows:

$$CP(w) = \{w', \exists k, w_1, \ldots, w_k, w_1 = w, w_k = w', \forall 1 \leq i \leq k w_{i+1} \in P(w_i\}$$

So, $CP(w)$ is now the set of elements w' that can be reached from w using a chain, where the chain can be of arbitrary length. In particular two consecutive chains can be attached to create a new chain, which implies a form of transitivity: $w' \in CP(w)$ and $w'' \in CP(w')$ together imply $w'' \in CP(w)$.

The following result is left as an exercise.

Theorem 2.3. *Assume each P_i is partitional, and CK represents common knowledge. Then, CK is derived from a* partitional *possibility correspondence CP.*

When each P_i is partitional, CP is the finest partition that is as coarse as each of the P_i's. This is also known as the "meet" of the partitions P_i's.

2.1 Beliefs and consensus

When agents take decisions under uncertainty, not only the set of possible states of nature matters to them, but also the relative likelihood of these states. We introduce probabilistic beliefs the agent holds on possible states of nature.

Before receiving any information, agent i holds a belief p_i on W. Thus, $p_i(w) \geq 0$ is the probability that the agent assigns to the stage $w \in \mathsf{W}$, and $\mathring{a}_w\, p_i(w) = 1$. The probability distribution p_i is called agent i's prior belief on W. For an event E, the prior belief that E realizes is given by $p_i(E) = \mathring{a}_{w \in E}\, p_i(w)$.

Once receiving information, the agent does not consider all states in W as being possible, but only those in $p_i(w)$. Given this information, the agent assigns probabilities to states $w' \in \mathsf{W}$ according to Bayes's rule:

$$p_i(w'|P_i(w)) = \frac{p(\{w'\} \cap P_i(w))}{p(P_i(w))}$$

Note that if $w' \notin P_i(w)$, then $p_i(w'|P_i(w)) = 0$. For an event $E \subseteq \mathsf{W}$,

$$p_i(E|P_i(w)) = \frac{p(E \cap P_i(w))}{p(P_i(w))}$$

The probability distribution $p_i(\cdot|P_i(w))$ is called agent i's *posterior* probability distribution at state w. The term *prior* thus refers to *before* receiving information, while *posterior* means *after* this information is received.

Now consider different agents. In principle, agents i and j could hold different prior probabilities p_i and p_j on W. We often make the assumption that before receiving any information, agents hold the same probabilistic beliefs. This assumption, called the *common prior assumption*, means that there exists a probability distribution p such that every agent has prior p.

As shown by the next example, two agents can have same priors on W, but different posterior probabilities.

Example 2.4. $\mathsf{W} = \{1,2,3,4,5\}$. $P_1 = \{\{1\},\{2,3\},\{4,5\}\}$, $P_2 = \{\{1\},\{2\},\{3,4\},\{5\}\}$.

Let p be uniform on W, and $E = \{3,4,5\}$.

- $p(E|P_1(w))$ takes the values $(0, \frac{1}{2}, \frac{1}{2}, 1, 1)$ in the different states $1 \ldots 5$,
- $p(E|P_2(w))$ takes the values $(0,0,1,1,1)$.

We like to ask whether it is possible that agent's posterior are known to each other, while they differ. The answer is yes, as shown by our next example.

Example 2.5. Let $\mathsf{W} = \{a_1,a_2,a_3,b_1,b_2,b_3,c_1,c_2,c_3\}$ and p is uniform on W.
$P_1 = \{\{a_1,a_2,a_3\},\{b_1,b_2,b_3,c_1,c_2,c_3\}\}$, $P_2 = \{\{a_1,a_2,a_3,b_1,b_2,b_3\},\{c_1,c_2,c_3\}\}$.
Consider the event $E = \{a_1,a_2,b_1,c_1,c_2,c_3\}$. $p(E|P_i(w))$ takes the values

$$(\frac{2}{3},\frac{2}{3},\frac{2}{3},\frac{2}{3},\frac{2}{3},\frac{2}{3},\frac{2}{3},\frac{2}{3},\frac{2}{3})$$

for player 1, and

$$(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},1,1,1)$$

for player 2.

In states $\{a_1,a_2,a_3\}$, player 1 knows that player 2's posterior on E is $\frac{1}{2}$, player 2 knows that player 1's posterior is $\frac{2}{3}$. Nevertheless, player's beliefs on E disagree at these states.

Note that in that example, in states where player's posterior probabilities disagree and are mutual knowledge, these posterior probabilities are not common knowledge. In the next example, when posterior probabilities are common knowledge, they are the same.

Example 2.6. $\Omega = \{1,2,3,4,5,6,7\}$. $P_1 = \{\{1,2,3\},\{4,5\},\{6,7\}\}, P_2 = \{\{1\},\{2\},\{3\},\{4,5,6,7\}\}$.

Let p be uniform on Ω, $E = \{3,5,6\}$. What are $p(E|P_1(\omega))$ and $p(E|P_2(\omega))$ in all states ω? When is $p(E|P_1(\omega))$ common knowledge? When is $p(E|P_2(\omega))$ common knowledge?

As stated by the next (important) theorem, if players have common priors, and if their posterior probabilities are *common knowledge*, then these posterior probabilities are identical. The common interpretation of that result is that players can't agree to have different opinions, they can't agree to disagree.

Theorem 2.7 (We can't agree to disagree)**.** *Assume each player's information is partitional, and let E be an event. Assume that, at some state ω, each agent's belief about E is commonly known among all players. Then, all these posterior beliefs are equal.*

Proof. Let $q_i = p(E|P_i(\omega))$ be agent i's posterior belief on E at ω. Let $M(\omega)$ be the element of the common knowledge partition that contains ω. Let $E_i = \{\omega', p(E|P_i(\omega') = q_i)\}$. Since E_i is common knowledge at ω, $\omega \in E_i$, $M(\omega) \subseteq E_i$: agent i's belief is constant on $M(\omega)$.

$$P(E|P_i(\omega')) = q_i \quad \text{for } \omega' \in M(\omega)$$

Decompose $M(\omega)$ as a union of elements of the partition of player i: $M(\omega) = P_{i,1} \cup \ldots P_{i,k}$. Then,

$$P(E|M(\omega)) = \sum_{1 \le k' \le k} P(P_{i,k'}|M(\omega))P(E|P_{i,k'}) = \sum_{1 \le k' \le k} P(P_{i,k'}|M(\omega))q_i = q_i.$$

Since this is true for every player all player's posterior beliefs at ω are the same. □

The next example shows that it can be commonly known that two players have different posterior beliefs about an event, as long as the value of these beliefs are not commonly known.

Example 2.8. Let $\Omega = \{a,b\}$ with the uniform probability p. Player's information partitions are $P_1 = \{\{a,\},\{b\}\}$ and $P_2 = \{\{a,b\}\}$. For the event $E = \{a\}$, $p(E|P_1(a)) = 1$ and $p(E|P_1(b)) = 0$, whereas $p(E|P_2(\omega)) = \frac{1}{2}$ for every ω.

Exercise 2.9. Show that it cannot be commonly known that one player as a posterior belief strictly larger than another player.

We have already show in an example that the theorem fails if we replace "commonly known" by "mutually known". Similarly, it can be shown that the theorem fails if "common knowledge" is replaced by "mutually known at order k", for any value of k.

3 Examples and riddles

Example 3.1. Cheryl has two friends, Albert and Bernard. She gives them a list of 10 possible dates for her birthday;

- May 15th 16th 19th,
- June 17th 18th,
- July 14th 16th,
- and August 14th 15th 17th.

Then she tells the month to Albert, and the day of the month to Bernard.

Albert then says: “I don’t know when Cheryl’s birthday is, but I know that Bernard doesn’t know either.”

Bernard replies: “In the beginning, I didn’t know when Cheryl’s birthday was, but now I do.”

Finally Albert concludes: “Then I know her birthday too.”.

Can you guess Cheryl’s birthday too?

Example 3.2. 100 monks live together in a monastery in very harsh conditions. They are not allowed to talk or communicate in any form. They are not allowed to touch their own body, or look at themselves in a mirror. They pray all day long, except for supper when they all meet in the communal dining room for their only meal of the day.

One day, their leader exceptionally breaks the silence to make an announcement. Some monks are sick, and their disease creates a red spot on their forehead. The disease is not contagious, but sick members are requested to leave the order. The leader does not designate the sick members, they have to understand on their own that they are sick and must leave.

7 days later, after the meal, without any other remark, comment or announcement by their leader, all sick monks silently leave the community.

How many sick members where there and how did they know they were sick?

Example 3.3. Three students stand one behind the other. A professor comes to the room with 2 white hats and 3 black hats. The professor puts one hat on the head of each student, without telling them the colour.

Each student sees the hats of other students in front of him. The professor says that a student who can guess the colour of his hat gets the maximal grade.

The last student, who sees the hats of both others, says: ”I don’t know the colour of my hat.”

Then, the student in the middle, who sees the hat of the one in front of him, says: ”Neither do I.”.

Finally, the student in front of the line says: ”I know the colour of my hat.”.

Can you guess the colour of each of the student’s hats?