Tutorial: Information in information theory and games

Nils Bertschinger, Eckehard Olbrich
MPI MIS Leipzig

Combining Information Theory and Game Theory
Santa Fe

13.08.2012
Introduction: Information in Games — Some Examples

Formalizations of information
  Information in game theory — information partitions
  Information in information theory — random variables and channels

Rate distortion theory and the value of Information

Example: Stackelberg game
Some topics that are covered in the reading list but will not be addressed in this tutorial:

- Cheap talk, signalling and screening
- Information in markets and prices
- Organization structures and information
- Information games: game theoretical characterizations of information theoretic concepts
Introduction: Information in Games — Some Examples
Example 1: Levine and Ponssard (1977)


1. Is information always valuable?
2. Is it better to acquire information secretly or in front of others?
3. Is private information better than public information?
The game

\[ p = \frac{1}{2} \]

### Nature

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>(a_{11}, a_{11})</td>
<td>(a_{21}, a_{12})</td>
</tr>
<tr>
<td>B</td>
<td>(a_{12}, a_{21})</td>
<td>(a_{22}, a_{22})</td>
</tr>
</tbody>
</table>

\[ A_1 \]

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>(a_{12}, a_{12})</td>
<td>(a_{22}, a_{11})</td>
</tr>
<tr>
<td>B</td>
<td>(a_{11}, a_{22})</td>
<td>(a_{21}, a_{21})</td>
</tr>
</tbody>
</table>

\[ A_2 \]

\[
\begin{align*}
    a_{11} &> a_{12} & \text{and} & & a_{21} &> a_{22} \\
    a_{11} + a_{12} &> a_{21} + a_{22} \\
    a_{12} + a_{21} &> a_{11} + a_{22}
\end{align*}
\]

- **Column player:** L is dominant in \(A_1\) and R in \(A_2\).
- **Row player:** T is dominant in \(A_1\) and B in \(A_2\).
Public, private and secret information

Public information: All players are informed and all players, e.g. by an public announcement and all players know this, thus establishing common knowledge.

Private information: In the case of private information Col player has information that is not available for the Row player, but the Row player knows that the Col player has specific information.

Secret information: In the case of secret information the Row player does not know that the Col player has additional information, i.e. the Row player is not aware of the possibility that the Col player could have additional information.
Example: Public information

\[ \begin{array}{c|cc}
 & L & R \\
\hline
T & a_{11}, a_{11} & a_{21}, a_{12} \\
B & a_{12}, a_{21} & a_{22}, a_{22} \\
\end{array} \quad \begin{array}{c|cc}
 & L & R \\
\hline
T & a_{12}, a_{12} & a_{22}, a_{11} \\
B & a_{11}, a_{22} & a_{21}, a_{21} \\
\end{array} \]

- \( a_{11} > a_{12} \) and \( a_{21} > a_{22} \)
- \( a_{11} + a_{12} > a_{21} + a_{22} \)
- \( a_{12} + a_{21} > a_{11} + a_{22} \)

- Players will play \((T, L)\) in \(A_1\) and \((B, R)\) in \(A_2\). Expected payoff is \(a_{11} + a_{21})/2\).
Example: Private information

<table>
<thead>
<tr>
<th></th>
<th>$A_1$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>$\frac{a_{11} + a_{12}}{2}$, $\frac{a_{11} + a_{12}}{2}$</td>
<td>$\frac{a_{11} + a_{22}}{2}$, $a_{11}$</td>
</tr>
<tr>
<td>B</td>
<td>$\frac{a_{11} + a_{12}}{2}$, $\frac{a_{21} + a_{22}}{2}$</td>
<td>$\frac{a_{12} + a_{21}}{2}$, $a_{21}$</td>
</tr>
</tbody>
</table>

\[ a_{11} > a_{12} \quad \text{and} \quad a_{21} > a_{22} \]
\[ a_{11} + a_{12} > a_{21} + a_{22} \]
\[ a_{12} + a_{21} > a_{11} + a_{22} \]

- Column player has private information: Column player plays $L$ in $A_1$ and $R$ in $A_2$.
- Row player plays $B$
- Expected payoffs are $(a_{12} + a_{21})/2$ and $a_{21}$.
<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>((a_{11} + a_{12})/2, (a_{11} + a_{12})/2)</td>
<td>((a_{21} + a_{22})/2, (a_{11} + a_{12})/2)</td>
</tr>
<tr>
<td>B</td>
<td>((a_{11} + a_{12})/2, (a_{21} + a_{22})/2)</td>
<td>((a_{21} + a_{22})/2, (a_{21} + a_{22})/2)</td>
</tr>
</tbody>
</table>

\(a_{11} > a_{12}\)  and  \(a_{21} > a_{22}\)

\(a_{11} + a_{12} > a_{21} + a_{22}\)

\(a_{12} + a_{21} > a_{11} + a_{22}\)
The game becomes a coordination game. Levine and Ponsarrd assume that the players will coordinate at \((T, L)\). Expected payoff is \((a_{11} + a_{12})/2\).
Example: Secret information

- Row player assume the situation of no information:

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>((a_{11} + a_{12})/2, (a_{11} + a_{12})/2)</td>
<td>((a_{21} + a_{22})/2, (a_{11} + a_{12})/2)</td>
</tr>
<tr>
<td>B</td>
<td>((a_{11} + a_{12})/2, (a_{21} + a_{22})/2)</td>
<td>((a_{21} + a_{22})/2, (a_{21} + a_{22})/2)</td>
</tr>
</tbody>
</table>
Example: Secret information

- Row player assume the situation of no information:

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>((a_{11} + a_{12})/2, (a_{11} + a_{12})/2)</td>
<td>((a_{21} + a_{22})/2, (a_{11} + a_{12})/2)</td>
</tr>
<tr>
<td>B</td>
<td>((a_{11} + a_{12})/2, (a_{21} + a_{22})/2)</td>
<td>((a_{21} + a_{22})/2, (a_{21} + a_{22})/2)</td>
</tr>
</tbody>
</table>
Example: Secret information

- Row player assume the situation of no information:

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>( \frac{a_{11} + a_{12}}{2}, \frac{a_{11} + a_{12}}{2} )</td>
<td>( \frac{a_{21} + a_{22}}{2}, \frac{a_{11} + a_{12}}{2} )</td>
</tr>
<tr>
<td>B</td>
<td>( \frac{a_{11} + a_{12}}{2}, \frac{a_{21} + a_{22}}{2} )</td>
<td>( \frac{a_{21} + a_{22}}{2}, \frac{a_{21} + a_{22}}{2} )</td>
</tr>
</tbody>
</table>

- Column player has secret information

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>( a_{11}, a_{11} )</td>
<td>( a_{21}, a_{12} )</td>
</tr>
<tr>
<td>B</td>
<td>( a_{12}, a_{21} )</td>
<td>( a_{22}, a_{22} )</td>
</tr>
</tbody>
</table>

\[ A_1 \]

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>( a_{12}, a_{12} )</td>
<td>( a_{22}, a_{11} )</td>
</tr>
<tr>
<td>B</td>
<td>( a_{11}, a_{22} )</td>
<td>( a_{21}, a_{21} )</td>
</tr>
</tbody>
</table>

\[ A_2 \]
Example: Secret information

- Row player assume the situation of no information:

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>((a_{11} + a_{12})/2),((a_{11} + a_{12})/2)</td>
<td>((a_{21} + a_{22})/2),((a_{11} + a_{12})/2)</td>
</tr>
<tr>
<td>B</td>
<td>((a_{11} + a_{12})/2),((a_{21} + a_{22})/2)</td>
<td>((a_{21} + a_{22})/2),((a_{21} + a_{22})/2)</td>
</tr>
</tbody>
</table>

- Column player has secret information

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>(a_{11},a_{11})</td>
<td>(a_{21},a_{12})</td>
</tr>
<tr>
<td>B</td>
<td>(a_{12},a_{21})</td>
<td>(a_{22},a_{22})</td>
</tr>
</tbody>
</table>

\(A_1\)

because \(a_{11} > a_{12}\).

- Expected payoffs are \((a_{11} + a_{22})/2\), \(a_{11}\) for raw and column player respectively.
No information: \((T, L)\) is dominating any other outcome. Expected payoff for C is \((a_{11} + a_{12})/2\).
No information: \((T, L)\) is dominating any other outcome. Expected payoff for C is \((a_{11} + a_{12})/2\).

Public information: \((T, L)\) in \(A_1\) and \((B, R)\) in \(A_2\). Expected payoff for C is \((a_{11} + a_{21})/2\).
No information: \((T, L)\) is dominating any other outcome. Expected payoff for C is \((a_{11} + a_{12})/2\).

Public information: \((T, L)\) in \(A_1\) and \((B, R)\) in \(A_2\). Expected payoff for C is \((a_{11} + a_{21})/2\).

Column player has private information: Expected payoff for C is \(a_{21}\).
No information: \((T, L)\) is dominating any other outcome. Expected payoff for C is \((a_{11} + a_{12})/2\).

Public information: \((T, L)\) in \(A_1\) and \((B, R)\) in \(A_2\). Expected payoff for C is \((a_{11} + a_{21})/2\).

Column player has private information: Expected payoff for C is \(a_{21}\).

Column player has secret information: Expected payoff for C is \(a_{11}\).
Payoffs of column player

\[ A_1 \]

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>4,4</td>
<td>(a_{21}, 3)</td>
</tr>
<tr>
<td>B</td>
<td>3,(a_{21})</td>
<td>1,1</td>
</tr>
</tbody>
</table>

\[ A_2 \]

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>3,3</td>
<td>1,4</td>
</tr>
<tr>
<td>B</td>
<td>4,1</td>
<td>(a_{21}, a_{21})</td>
</tr>
</tbody>
</table>
Example 2: Leader’s advantage

Abstracted from a duopoly market entry game. Two players - leader (L) and follower (F).

Players move simultaneously:

- Nash equilibrium: Stackelberg outcome \((S, S')\)
- Nash equilibrium: Cournot outcome \((C, C)\)
- “Leader’s advantage” \(5 > 4\).
- Follower prefers \((C, C)\) — having information about the leaders move “hurts” the follower ⇒ negative value of information
• Kyle Bagwell, *Commitment and Observability in Games*, GEB,8,271-280 (1995): An arbitrary small amount of noise destroys the leader’s advantage under the assumption of full rationality.
Follower has to specify its moves for both channel outcomes.

<table>
<thead>
<tr>
<th>L</th>
<th>F</th>
<th>$S'$ $C'$</th>
<th>$S'$ $C'$</th>
<th>$S'$ $C'$</th>
<th>$S'$ $C'$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>(5, 2)</td>
<td>(5 - 2\epsilon, 2 - \epsilon)</td>
<td>(3 + 2\epsilon, 1 + \epsilon)</td>
<td>(3, 1)</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>(6, 3)</td>
<td>(4 + 2\epsilon, 4 - \epsilon)</td>
<td>(6 - 2\epsilon, 3 + \epsilon)</td>
<td>(4, 4)</td>
</tr>
</tbody>
</table>

Pure strategy Nash equilibria:

- $\epsilon = 0$: (S,SC) (and (C,CC) that is not subgame perfect).
- $\epsilon > 0$: Only (C,CC).
Follower has to specify its moves for both channel outcomes.

\[
\begin{array}{c|cc|cc|cc}
L & \text{F} & S' & C' & S' & C' & S' & C' \\
\hline
S & SS & (5, 2) & \text{SC} & (5 - 2\epsilon, 2 - \epsilon) & \text{CS} & (3 + 2\epsilon, 1 + \epsilon) & \text{CC} & (3, 1) \\
C & \text{SS} & (6, 3) & \text{SC} & (4 + 2\epsilon, 4 - \epsilon) & \text{CS} & (6 - 2\epsilon, 3 + \epsilon) & \text{CC} & (4, 4) \\
\end{array}
\]

Pure strategy Nash equilibria:

\( \epsilon = 0 \): (S,SC) (and (C,CC) that is not subgame perfect).

\( \epsilon > 0 \): Only (C,CC).

Leader’s advantage “survives” as a mixed strategy (van Damme and Hurkens, 1997).
What could be gained by using information theory?

- Allows to quantify information transmission
- Distinguishing between the capacity of the channel, i.e. potential information, and the actually transmitted information
- Differential value of information
Example 3: Quantifying information

Example 3: Quantifying information

- Experimental study of the value of information in financial markets
Example 3: Quantifying information

- Experimental study of the value of information in financial markets
- Result: “There is a wide range of information levels (from zero information to above average information levels) where additional information does not yield higher returns.”
- Average profit per information level:

![Average profit per information level in treatment T1](image)
How is information measured?

- The asset’s intrinsic value is encoded in a randomly drawn binary string of length 10. The number of “1”s in the string is the value of the asset.
- Information level “k” means that the trader has seen the first $k$ bits of the string.
How is information measured?

- The asset’s intrinsic value is encoded in a randomly drawn binary string of length 10. The number of “1”s in the string is the value of the asset.
- Information level “k” means that the trader has seen the first $k$ bits of the string.

Information theoretic analysis:
- The different bits of the string add different amounts of information about the value of the asset.

Mutual information between number of revealed bits $k$ and the value:

![Graph showing mutual information](image)
How is information measured?

![Average profit per information level in treatment T1](image1)

![Length of substring vs. information level in bits](image2)
Formalizations of information
In the examples, an uninformed player could not distinguish between different states of nature or moves of the leader.

Formally, information in games is described by an information structure consisting of:

- A probability space \((\Omega, \Sigma, \mu)\) with set of outcomes \(\Omega\), Sigma-algebra \(\Sigma\) and measure \(\mu\).
- A measurable partition \(\Pi_i\) for each player \(i \in I\) which partitions \(\Omega\) into a (finite) set of elements \(E_{i,k} \subseteq \Omega\).

A player cannot distinguish between different states \(\omega, \omega' \in \Omega\) whenever they belong to the same element \(E_{i,k}\) of his information partition \(\Pi_i\).
An information structure encodes beliefs of the players:

- When player $i$ is in element $E_{i,k}$ of his information partition, he believes an event $A \subset \Omega$ with probability

$$\mu(A|E_{i,k}) = \frac{\mu(A \cap E_{i,k})}{\mu(E_{i,k})}$$
An information structure encodes beliefs of the players:

- When player $i$ is in element $E_{i,k}$ of his information partition, he believes an event $A \subset \Omega$ with probability
  \[ \mu(A|E_{i,k}) = \frac{\mu(A \cap E_{i,k})}{\mu(E_{i,k})} \]

- The set of all states $\omega$ where player $i$ believes an event $A$ with probability at least $q$, is itself an event:
  \[ B^q_i(A) = \{ \omega | \mu(A|\Pi_i(\omega)) \geq q \} = \bigcup_{k: \mu(A|E_{i,k}) \geq q} E_{i,k} \]
  where $\Pi_i(\omega)$ denotes the element of $\Pi_i$ containing $\omega$. 
An information structure encodes beliefs of the players:

- When player $i$ is in element $E_{i,k}$ of his information partition, he believes an event $A \subset \Omega$ with probability

$$
\mu(A|E_{i,k}) = \frac{\mu(A \cap E_{i,k})}{\mu(E_{i,k})}
$$

- The set of all states $\omega$ where player $i$ believes an event $A$ with probability at least $q$, is itself an event:

$$
B^q_i(A) = \{\omega|\mu(A|\Pi_i(\omega)) \geq q\} = \bigcup_{k: \mu(A|E_{i,k}) \geq q} E_{i,k}
$$

where $\Pi_i(\omega)$ denotes the element of $\Pi_i$ containing $\omega$.

- Iterating this construction leads to higher-order beliefs, e.g.

$$
B^q_i(B^q_j(A)), B^q_i(B^q_i(B^q_j(A))), \ldots
$$

In a game $G$ with action spaces $(A_i)_{i \in I}$, payoff functions $(u_i : \times_{i \in I} A_i \to \mathbb{R})_{i \in I}$ and information structure $(\Omega, \Sigma, \mu, (\Pi_i)_{i \in I})$

1. A state $\omega \in \Omega$ is drawn according to $\mu$.
2. Each player is informed about the corresponding element of his information partition.
3. Based on this information, each player $i$ then chooses a (mixed) move $s_i \in \Delta(A_i)$. Here, $\Delta(A)$ denotes the set of all probability distributions on $A$.
4. The payoff to each player $i$ depends on the chosen moves of all players, i.e. $u_i((s_j)_{j \in I})$.

Thus, in a game with information structure the strategy space of each player consists of functions $\Pi_i \to \Delta(A_i)$ mapping elements of the information partition to possible moves.
Thus, in a game with information structure the strategy space of each player consists of functions $\Pi_i \to \Delta(A_i)$ mapping elements of the information partition to possible moves.

An information partition $\Pi = \{E_1, \ldots, E_k\}$ is finer than $\Pi' = \{E'_1, \ldots, E'_{k'}\}$ if for every $E_{j'} \in \Pi'$ there exists $E_j \in \Pi$ such that $E_j \subseteq E_{j'}$. 
Thus, in a game with information structure the strategy space of each player consists of functions $\Pi_i \rightarrow \Delta(A_i)$ mapping elements of the information partition to possible moves.

An information partition $\Pi = \{E_1, \ldots, E_k\}$ is finer than $\Pi' = \{E'_1, \ldots, E'_{k'}\}$ if for every $E_{j'} \in \Pi'$ there exists $E_j \in \Pi$ such that $E_j \subseteq E_{j'}$.

From this we observe that

- Finer information partition $\implies$ more available strategies!
- The converse implication also holds, but the proof is rather technical and requires infinite partitions.
Thus, in a game with information structure the strategy space of each player consists of functions $\Pi_i \rightarrow \Delta(A_i)$ mapping elements of the information partition to possible moves.

An information partition $\Pi = \{E_1, \ldots, E_k\}$ is finer than $\Pi' = \{E'_1, \ldots, E'_{k'}\}$ if for every $E_{j'} \in \Pi'$ there exists $E_j \in \Pi$ such that $E_j \subseteq E_{j'}$.

From this we observe that

- Finer information partition $\Rightarrow$ more available strategies!
- The converse implication also holds, but the proof is rather technical and requires infinite partitions.

Refining an information partition corresponds to a qualitative change of the information structure. What about quantitative changes, e.g. perturbing the measure $\mu$?

Random variables

Another way to think about information is in terms of random variables.

Formally, a random variables is a measurable function $X : \Omega \rightarrow \mathcal{X}$ from a probability space $(\Omega, \Sigma, \mu)$ to a (finite) set of outcomes $\mathcal{X}$.

In the following, we will denote random variables by capital letters $X, Y, \ldots$ and individual outcomes by lower case letters $x, y, \ldots$.

The probability $\mathbb{P}\{X \in A\}$ is given by

$$\mathbb{P}\{X \in A\} = \mu(X^{-1}(A))$$

and for finite outcome spaces we will use $p(x)$ as a shorthand for $\mathbb{P}\{X = x\}$. 
Information structures

Consider a finite information partition $\Pi = \{E_1, \ldots, E_K\}$ with elements $E_k \subset \Omega$. From this we can define a random variable $X : \Omega \to \{1, \ldots, K\}$ by

$$X(\omega) = k \text{ iff } \omega \in E_k$$
Consider a finite information partition $\Pi = \{E_1, \ldots, E_K\}$ with elements $E_k \subset \Omega$. From this we can define a random variable $X : \Omega \rightarrow \{1, \ldots, K\}$ by

$$X(\omega) = k \text{ iff } \omega \in E_k$$

Thus, an information structure $(\Omega, \Sigma, \mu, (\Pi_i)_{i \in I})$ can be considered as a collection of random variables $(X_i)_{i \in I}$ over $(\Omega, \Sigma, \mu)$ describing signals that are available to the individual agents.

These signals have a joint distribution $p((x_i)_{i \in I})$ on $\times_{i \in I} \mathcal{X}_i$ and each agent $i$ can only observe his outcome $x_i$. 
Consider a finite information partition $\Pi = \{E_1, \ldots, E_K\}$ with elements $E_k \subset \Omega$. From this we can define a random variable $X : \Omega \to \{1, \ldots, K\}$ by

$$X(\omega) = k \text{ iff } \omega \in E_k$$

Thus, an information structure $(\Omega, \Sigma, \mu, (\Pi_i)_{i \in I})$ can be considered as a collection of random variables $(X_i)_{i \in I}$ over $(\Omega, \Sigma, \mu)$ describing signals that are available to the individual agents.

These signals have a joint distribution $p((x_i)_{i \in I})$ on $\times_{i \in I} \mathcal{X}_i$ and each agent $i$ can only observe his outcome $x_i$.

At present, it is not clear how the construction of higher-order beliefs can be expressed in terms of random variables.
Why random variables? Information theory

What do we gain by describing information structures in terms of random variables?

Methods and tools from information theory:

• Unified description of uncertainty in decision situations
• Quantitative notion of uncertainty and information
• Fundamental limits of information processing
• New viewpoints: Differential value of information
Information theory provides tools to quantify the information of random variables:

- The uncertainty of a random variable $X$ is given by its entropy
  \[
  H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)
  \]

- The mutual information between two random variables $X$ and $Y$ is defined as the reduction of uncertainty:
  \[
  I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y)
  \]
Mutual information provides tools to quantify the information of random variables:

- The uncertainty of a random variable $X$ is given by its *entropy*

\[
    H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x)
\]

- The mutual information between two random variables $X$ and $Y$ is defined as as reduction of uncertainty:

\[
    I(X;Y) = H(X) - H(X|Y)
    = H(X) + H(Y) - H(X,Y)
\]

Mutual information can also be considered as a measure of statistical dependency:

\[
    I(X;Y) = 0 \text{ iff } X \text{ and } Y \text{ are independent}
\]
A conditional distribution $p(y|x)$ can be interpreted as a noisy channel from $X$ to $Y$.

The mutual information then quantifies how much information from the source with distribution $p(x)$ is transferred via the channel.

The maximal possible amount of information that can be transferred via the channel is called the channel capacity $C$ and defined as follows:

$$ C = \max_{p(x)} I(X; Y) $$
Consider a set of random variables $X_1, \ldots, X_K$ with joint distribution $p(x_1, \ldots, x_K)$. The joint distribution can be factored as follows:

$$p(x_1, \ldots, x_K) = p(x_1) \cdot \prod_{i=2}^{K} p(x_i|x_1, \ldots, x_{i-1})$$
Consider a set of random variables $X_1, \ldots, X_K$ with joint distribution $p(x_1, \ldots, x_K)$.

The joint distribution can be factored as follows:

$$p(x_1, \ldots, x_K) = p(x_1) \cdot \prod_{i=2}^{K} p(x_i|x_1, \ldots, x_{i-1})$$

Quite often, conditional independencies can be exploited to drop some of the conditioning variables, i.e.

$$p(x_i|x_1, \ldots, x_{i-1}) = p(x_i|x_{Pa}(X_i))$$

where $Pa(X_i) \subset \{x_1, \ldots, x_{i-1}\}$
A *Bayes net* is a graphical representation - in terms of a directed acyclic graph $G = (V, E)$ - of such a factorization:

- Each random variable $X_i$ corresponds to a vertex $v_i \in V$ of $G$
- There is a directed edge $v_j \rightarrow v_i$ in $G$ iff $X_j \in Pa(X_i)$

For example,

\[
\begin{array}{ccc}
X_1 & \rightarrow & X_2 \\
\rightarrow & & \rightarrow \\
& & X_3 \\
\end{array}
\]

represents the factorization

\[
p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)
\]
Multi-agent influence digrams (MAID) are an extension of Bayes nets to represent games:

- Moves of nature as well as of the players are represented as random variables.
- The conditional probabilities $p(x_i|x_{Pa}(X_i))$ are fixed for nature nodes. All nodes representing moves of player $i$ can be set by him, corresponding to his behavioral strategies.
- Utility nodes illustrate the dependencies of the utility function, e.g. the utility $u_i$ of player $i$ might not depend on the moves of all players, but only a subset $Pa(u_i)$ of player nodes.

The Levine and Ponssard games can be represented by the following MAID:

The secret information case cannot be handled in this formalism, since the structure of the game is not common knowledge in this case. The row player is unaware of the existence of the information channel.
The Levine and Ponssard games can be represented by the following MAID:

- **No information**
- **Public information**
- **Private information**

- Game theory: Qualitative change of game structure
- Information theory: Quantitative change of channel capacity
The Leader’s advantage game can be represented by the following MAIDs:

Here, \( L \) denotes the leader and \( F \) is the follower. The noisy channel is represented by the nature node \( \epsilon \).

The game is sequential, since the followers conditional distribution can use the noisy channel as input which in turn depends on the leaders move.

The utility \( U \) of both players depends on both of their moves.
Rate distortion theory and the value of Information
In games against nature, i.e. decision problems only one agent tries to maximize its utility, while the other, nature, simply draws its move from a probability distribution.

Adding information about the moves or states of nature creates a larger strategy space of the agent, because moves of the agent can be conditioned using finer distinctions. Thus the agent might find better strategies.

Adding information $\Rightarrow$ higher expected utility $\Rightarrow$ positive value of information
A game against nature is very similar to the Leader’s advantage game:

- The leader is not a player, but replaced by nature $X$
- Nature’s move can only be observed via a noisy channel $\epsilon$
- The player $A$ chooses his move based on the noisy observation $\epsilon$ of the state of nature.
We can abstract away the channel:

Now, we have to constrain the choice of the player:

- The player cannot choose any conditional distribution \( p(a|x) \), but is only allowed to use a maximum amount of information \( R \), i.e. we demand that

\[
I(A; X) \leq R
\]

\[\implies\] Rate distortion theory
Rate distortion theory

- Distortion measure \( d(\hat{x}, x) \) is a measure of the costs of representing state \( x \) by state \( \hat{x} \).
- Source/nature emits \( x \) with \( p(x) \)
- Rate distortion function

\[
R(D) = \min_{p(\hat{x}|x); \mathbb{E}[d(\hat{x}, x)] \leq D} I(\hat{X}; X)
\]

determines the minimal channel capacity needed for the reproduction of \( x \) at a distortion level \( D \).
- Distortion rate function

\[
D(R) = \min_{p(\hat{x}|x); I(\hat{X}; X) \leq R} \mathbb{E}[d(\hat{x}, x)]
\]

determines the minimal distortion that can be achieved by using information \( R \).
Rate distortion function for normally distributed Gaussian source with squared error distortion

- The rate distortion function is a nonincreasing convex function of $D$.
- The distortion rate function is the inverse of the rate distortion function

$$D(R) = R^{-1}(D)$$

Therefore also $D(R)$ is nonincreasing and convex.
• Instead of minimizing the distortion function, maximize the utility function $U(a, x)$, with $a$ being the action of the agent and $x$ the move of nature, i.e. $D(a, x) \equiv -U(a, x)$

• Utility rate function

$$U(R) = \max_{p(a|x); I(A;X) \leq R} \mathbb{E}[U(a, x)]$$

provides the maximal utility that can be reached by using information $R$.

• Monoticity of $D(R) \Rightarrow$ monoticity $U(R) \Rightarrow$ positive value of information

• $D(R)$ convex $\Rightarrow U(R)$ concave $\Rightarrow$ decreasing marginal utility of information

• Differential value of information $\frac{dU}{dR}$
What happens in game theory?

Game against nature

Stackelberg game

\[ X \rightarrow e \rightarrow A \]

\[ L \rightarrow e \rightarrow F \]

\[ U \]

\[ U \]
Example: Stackelberg game
Example: Stackelberg game - Leaders advantage

Abstracted from a duopoly market entry game. Two players - leader (L) and follower (F).

Nash Equilibria:

- Leader moves first: Stackelberg outcome \((S, S)\)
- Players move simultaneously \(\equiv\) both players do not know the move of the other player: Cournot outcome \((C, C)\)
Noisy channel

\[ S', 1 - \epsilon \]
\[ C', \epsilon \]
\[ S', \epsilon \]
\[ C', 1 - \epsilon \]

\[ 5,2 \]
\[ 3,1 \]
\[ 5,2 \]
\[ 3,1 \]
\[ 6,3 \]
\[ 4,4 \]
\[ 6,3 \]
\[ 4,4 \]
Follower has to specify its moves for both channel outcomes.

<table>
<thead>
<tr>
<th>L</th>
<th>F</th>
<th>S′ C′</th>
<th>S′ C′</th>
<th>S′ C′</th>
<th>S′ C′</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SS</td>
<td>SC</td>
<td>CS</td>
<td>CC</td>
</tr>
<tr>
<td>S</td>
<td>(5, 2)</td>
<td>(5 − 2ε, 2 − ε)</td>
<td>(3 + 2ε, 1 + ε)</td>
<td>(3, 1)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>(6, 3)</td>
<td>(4 + 2ε, 4 − ε)</td>
<td>(6 − 2ε, 3 + ε)</td>
<td>(4, 4)</td>
<td></td>
</tr>
</tbody>
</table>

Pure strategy Nash equilibria:

ε = 0: (S,SC) (and (C,CC) that is not subgame perfect).

ε > 0: Only (C,CC).

In addition there are two mixed equilibria.
Payoffs and mutual informations in the mixed equilibria

For $\epsilon \to 0$ the blue equilibrium converges to the Cournot solution and the red one corresponds to the Stackelberg solution.

In equilibrium higher channel capacity does not necessarily imply higher mutual information!
Differential value of information

Dependence of payoffs on the channel capacity and the mutual information between leader and follower respectively.

![Graph showing payoffs and mutual information](image-url)
Differential value of information

Dependence of payoffs on the channel capacity and the mutual information between leader and the signal respectively, i.e. the information available to the follower.

An extension of rate distortion theory for games would need to constrain the channel capacity and not the actually used information!
Examples for positive and negative value of information in games.

Information theory allows to quantify information.

Information structures can be formulated in terms of random variables. MAIDs express the structure of games in terms of dependent random variables.

The value of information in games against nature is always positive. Rate distortion theory investigates the trade-off between information and utility.

Mixed equilibria change smoothly when varying the amount of information: Differential value of information.