Evidence Games:
Right to Remain Silent,
Left to Disclose

Sergiu Hart

April 2015
Evidence Games: Right to Remain Silent, Left to Disclose
Joint work with

Ilan Kremer
Motty Perry

Hebrew University of Jerusalem
University of Warwick
Disclose
Evidence Games

Left

Disclose

Right

Remain Silent
Evidence Games

Left

Disclose

Right

Remain Silent
Evidence Games

Left
Disclose

Right
Remain Silent
Evidence Games

- Left (Disclose)
- Right (Remain Silent)
Q: "Do you deserve a pay raise?"
Q: "Do you deserve a pay raise?"
A: "Of course."
Q: "Do you deserve a pay raise?"
A: "Of course."

Q: "Are you guilty and deserve punishment?"
Q: "Do you deserve a pay raise?"
A: "Of course."

Q: "Are you guilty and deserve punishment?"
A: "Of course not."
Q: "Do you deserve a pay raise?"
A: "Of course."

Q: "Are you guilty and deserve punishment?"
A: "Of course not."

How can one obtain reliable information?
Q: "Do you deserve a pay raise?"
A: "Of course."

Q: "Are you guilty and deserve punishment?"
A: "Of course not."

How can one obtain reliable information?
How can one determine the "right" reward, or punishment?
Q: "Do you deserve a pay raise?"
A: "Of course."

Q: "Are you guilty and deserve punishment?"
A: "Of course not."

- How can one obtain reliable information?
- How can one determine the "right" reward, or punishment?
- How can one "separate" and avoid "unraveling" (Akerlof 70)?
Basic Setup

- Agent who is informed
**Basic Setup**

- **Agent** who is informed
- **Principal** who takes decision but is uninformed
Basic Setup

- **Agent** who is informed
- **Principal** who takes decision but is uninformed
- Agent *transmits* information to Principal: explicitly (message) or implicitly (action)
Two Setups
Two Setups

 SETUP 1: Principal decides \textit{after} observing Agent’s move
Two Setups

- SETUP 1: Principal decides *after* observing Agent’s move
- SETUP 2: Principal chooses a *policy before* Agent’s move
Two Setups

- **SETUP 1**: Principal decides *after* observing Agent’s move
- **SETUP 2**: Principal chooses a *policy* *before* Agent’s move

"policy": a function that assigns a decision of Principal to each move of Agent
Two Setups

**SETUP 1**: Principal decides *after* observing Agent’s move

**SETUP 2**: Principal chooses a *policy before* Agent’s move

"*policy*": a function that assigns a decision of Principal to each move of Agent (Agent knows the policy when making his move)
Two Setups

- **SETUP 1**: Principal decides *after* observing Agent’s move
- **SETUP 2**: Principal chooses a *policy* *before* Agent’s move

*"policy":* a function that assigns a decision of Principal to each move of Agent (Agent knows the policy when making his move)

Principal is *committed* to the policy
Two Setups

- **GAME**: Principal decides *after* observing Agent’s move

- **MECHANISM**: Principal chooses a *policy* before Agent’s move

  "*policy*": a function that assigns a decision of Principal to each move of Agent (Agent knows the policy when making his move)

- Principal is *committed* to the policy
Literature
A’s payoff depends on A’s type
on A’s move

<table>
<thead>
<tr>
<th>GAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>MECHANISM</td>
</tr>
</tbody>
</table>
### Literature

<table>
<thead>
<tr>
<th>A’s payoff depends</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>on A’s type</td>
<td>yes</td>
</tr>
<tr>
<td>on A’s move</td>
<td>yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GAME</th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MECHANISM</td>
<td></td>
</tr>
</tbody>
</table>

(1) signaling: Spence 73  
handicap principle: Zahavi 75
### Literature

<table>
<thead>
<tr>
<th>A’s payoff depends</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>on A’s type</td>
<td>yes</td>
</tr>
<tr>
<td>on A’s move</td>
<td>yes</td>
</tr>
</tbody>
</table>

| GAME               | (1)    |
| MECHANISM          | (2)    |

(1) signaling: Spence 73
handicap principle: Zahavi 75
(2) screening: Rothschild–Stiglitz 76
### A’s payoff depends

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>on A’s type</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>on A’s move</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

### GAME

<table>
<thead>
<tr>
<th>MECHANISM</th>
<th>(1)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>signaling: Spence 73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>handicap principle: Zahavi 75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>screening: Rothschild–Stiglitz 76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cheap talk: Crawford–Sobel 82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krishna–Morgan 07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A’s payoff depends</td>
<td>on A’s type</td>
<td>yes</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>on A’s move</td>
<td>yes</td>
</tr>
<tr>
<td>GAME</td>
<td>(1)</td>
<td>(3)</td>
</tr>
<tr>
<td>MECHANISM</td>
<td>(2)</td>
<td>(4)</td>
</tr>
</tbody>
</table>

(1) signaling: Spence 73  
   handicap principle: Zahavi 75  
(2) screening: Rothschild–Stiglitz 76  
(3) cheap talk: Crawford–Sobel 82  
(4) Krishna–Morgan 07
### Literature

<table>
<thead>
<tr>
<th>A’s payoff depends</th>
<th>on A’s type</th>
<th>yes</th>
<th>yes</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>on A’s move</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>GAME</td>
<td>(1)</td>
<td>(3)</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>MECHANISM</td>
<td>(2)</td>
<td>(4)</td>
<td>✗</td>
</tr>
</tbody>
</table>

(1) signaling: Spence 73  
handicap principle: Zahavi 75  
(2) screening: Rothschild–Stiglitz 76  
(3) cheap talk: Crawford–Sobel 82  
(4) Krishna–Morgan 07
### A’s payoff depends

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>yes</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>on A’s type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>on A’s move</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GAME</strong></td>
<td>(1)</td>
<td>(3)</td>
<td>X</td>
</tr>
<tr>
<td><strong>MECHANISM</strong></td>
<td>(2)</td>
<td>(4)</td>
<td>X</td>
</tr>
</tbody>
</table>
### Literature

#### A’s payoff depends

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>yes</th>
<th>no</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>on A’s type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on A’s move</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>restricted</td>
</tr>
</tbody>
</table>

#### GAME

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(3)</th>
<th>×</th>
</tr>
</thead>
</table>

#### MECHANISM

<table>
<thead>
<tr>
<th></th>
<th>(2)</th>
<th>(4)</th>
<th>×</th>
</tr>
</thead>
</table>
### Literature

A’s payoff depends

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>yes</th>
<th>no</th>
<th>no</th>
<th>restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>on A’s type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on A’s move</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**GAME**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(3)</th>
<th>×</th>
<th>(5)</th>
</tr>
</thead>
</table>

**MECHANISM**

<table>
<thead>
<tr>
<th></th>
<th>(2)</th>
<th>(4)</th>
<th>×</th>
</tr>
</thead>
</table>

(5) unraveling: Grossman–O. Hart 80, Grossman 81, Milgrom 81

voluntary disclosure: Dye 85, Shin 03, 06
### Literature

**A’s payoff depends**

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>yes</th>
<th>no</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>on A’s type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>on A’s move</strong></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>restricted</td>
</tr>
</tbody>
</table>

**GAME**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(3)</th>
<th>×</th>
<th>(5)</th>
</tr>
</thead>
</table>

**MECHANISM**

<table>
<thead>
<tr>
<th></th>
<th>(2)</th>
<th>(4)</th>
<th>×</th>
<th>(6)</th>
</tr>
</thead>
</table>

(5) unraveling: Grossman–O. Hart 80, Grossman 81, Milgrom 81
voluntary disclosure: Dye 85, Shin 03, 06
(6) Green–Laffont 86
## Literature

<table>
<thead>
<tr>
<th>A’s payoff depends</th>
<th>yes</th>
<th>yes</th>
<th>no</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>on A’s type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on A’s move</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>restricted</td>
</tr>
</tbody>
</table>

| GAME   | (1) | (3) | ×  | (5) |
| MECHANISM | (2) | (4) | ×  | (6) |

(5) unraveling: Grossman–O. Hart 80, Grossman 81, Milgrom 81
voluntary disclosure: Dye 85, Shin 03, 06
(6) Green–Laffont 86
(5-6) persuasion: Glazer–Rubinstein 04, 06
### Literature

#### A’s payoff depends

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>yes</th>
<th>no</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>on A’s type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on A’s move</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>restricted</td>
</tr>
</tbody>
</table>

#### GAME

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(3)</th>
<th>X</th>
<th>(5)</th>
</tr>
</thead>
</table>

#### MECHANISM

<table>
<thead>
<tr>
<th></th>
<th>(2)</th>
<th>(4)</th>
<th>X</th>
<th>(6)</th>
</tr>
</thead>
</table>

(5) unraveling: Grossman–O. Hart 80, Grossman 81, Milgrom 81
voluntary disclosure: Dye 85, Shin 03, 06

(6) Green–Laffont 86

(5-6) persuasion: Glazer–Rubinstein 04, 06

(5-6) EVIDENCE GAMES
In EVIDENCE GAMES
In EVIDENCE GAMES

the EQUILIBRIUM outcome obtained without commitment
In EVIDENCE GAMES

the EQUILIBRIUM outcome obtained without commitment

and the OPTIMAL MECHANISM outcome obtained with commitment
In EVIDENCE GAMES

the EQUILIBRIUM outcome obtained without commitment

and the OPTIMAL MECHANISM outcome obtained with commitment

COINCIDE
Main Result: Equivalence

In EVIDENCE GAMES

the EQUILIBRIUM outcome obtained without commitment

and the OPTIMAL MECHANISM outcome obtained with commitment

COINCIDE
Example 1

Professor wants salary as high as possible
Example 1

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value
Example 1

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value
- Professor’s evidence (verifiable):
Example 1

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value
- Professor’s evidence (verifiable):
  \[ t_0 \] 50%: no evidence
  \[ t_+ \] 25%: positive evidence
  \[ t_- \] 25%: negative evidence
Example 1

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value
- Professor’s evidence (verifiable):
  \[ t_0 \] 50%: no evidence \rightarrow value = 60
  \[ t_+ \] 25%: positive evidence \rightarrow value = 90
  \[ t_- \] 25%: negative evidence \rightarrow value = 30
Example 1

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value
- Professor’s evidence (verifiable):
  \[
  [t_0] \quad 50\%: \text{ no evidence} \quad \rightarrow \quad \text{value} = 60
  \]
  \[
  [t_+] \quad 25\%: \text{ positive evidence} \quad \rightarrow \quad \text{value} = 90
  \]
  \[
  [t_-] \quad 25\%: \text{ negative evidence} \quad \rightarrow \quad \text{value} = 30
  \]
Example 1: Equilibrium

- $t_+ : 25\% \quad 90$
- $t_0 : 50\% \quad 60$
- $t_- : 25\% \quad 30$
Example 1: Equilibrium

GAME: (G1) Professor provides evidence
      (G2) *then* Dean sets salary
Example 1: Equilibrium

<table>
<thead>
<tr>
<th></th>
<th>( t_+ )</th>
<th>25%</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_0 )</td>
<td>50%</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>( t_- )</td>
<td>25%</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

**GAME:**

- (G1) Professor provides evidence
- (G2) *then* Dean sets salary

**EQUILIBRIUM**
Example 1: Equilibrium

GAME: (G1) Professor provides evidence
      (G2) then Dean sets salary

EQUILIBRIUM

Professor:
- $t_+$ provides positive evidence
- $t_0, t_-$ provide no evidence
Example 1: Equilibrium

GAME: (G1) Professor provides evidence
(G2) then Dean sets salary

EQUILIBRIUM

Professor:
- $t_+$ provides positive evidence
- $t_0, t_-$ provide no evidence

Dean:
- positive evidence gets salary $= 90$
- negative evidence gets salary $= 30$
Example 1: Equilibrium

GAME: (G1) Professor provides evidence
(G2) then Dean sets salary

EQUILIBRIUM

Professor:
- \( t_+ \) provides positive evidence
- \( t_0, t_- \) provide no evidence

Dean:
- positive evidence gets salary = 90
- negative evidence gets salary = 30
- no evidence gets salary = 50

\[
= \left( 50\% \cdot 60 + 25\% \cdot 30 \right) / \left( 50\% + 25\% \right)
\]
Example 1: Equilibrium

**GAME**: (G1) Professor provides evidence
(G2) *then* Dean sets salary

unique sequential **EQUILIBRIUM**

- **Professor**:
  - $t_+$ provides positive evidence
  - $t_0, t_-$ provide no evidence

- **Dean**:
  - positive evidence gets salary = 90
  - negative evidence gets salary = 30
  - no evidence gets salary = 50

$$= (50\% \cdot 60 + 25\% \cdot 30)/(50\% + 25\%)$$

- $t_+ : 25\% \cdot 90$
- $t_0 : 50\% \cdot 60$
- $t_- : 25\% \cdot 30$
## Example 1: Equilibrium

<table>
<thead>
<tr>
<th>Time</th>
<th>Value 1</th>
<th>Value 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_+$</td>
<td>25%</td>
<td>90</td>
</tr>
<tr>
<td>$t_0$</td>
<td>50%</td>
<td>60</td>
</tr>
<tr>
<td>$t_-$</td>
<td>25%</td>
<td>30</td>
</tr>
</tbody>
</table>
Example 1: Equilibrium

\begin{itemize}
  \item $t_+ : 25\% \quad 90$
  \item $t_0 : 50\% \quad 60$
  \item $t_- : 25\% \quad 30$
\end{itemize}

value: 30 \quad 60 \quad 90
Example 1: Equilibrium

$t_+: 25\% \quad 90$
$t_0: 50\% \quad 60$
$t_-: 25\% \quad 30$

value: 30 60 90
Example 1: Equilibrium

$t_+ : 25\% \ 90$
$t_0 : 50\% \ 60$
$t_- : 25\% \ 30$

value: 30 60 90

partial truth:
Example 1: Equilibrium

prof says:

partial truth:

value:

$t_+ : 25\% 90$
$t_0 : 50\% 60$
$t_- : 25\% 30$
Example 1: Equilibrium

$t_+ : 25\% \quad 90$
$t_0 : 50\% \quad 60$
$t_- : 25\% \quad 30$

dean pays: 30 50 90
prof says: $t_-$
value: 30 60 90
partial truth:
Example 1: Mechanism

\[
\begin{align*}
  t_+ & : 25\% \quad 90 \\
  t_0 & : 50\% \quad 60 \\
  t_- & : 25\% \quad 30
\end{align*}
\]
Example 1: Mechanism

**MECHANISM:** (M1) Dean commits to salary *policy*
(M2) then Professor provides *evidence*
Example 1: Mechanism

**MECHANISM**: (M1) Dean commits to salary *policy*
(M2) then Professor provides evidence

**OPTIMAL MECHANISM**
Example 1: Mechanism

MECHANISM: (M1) Dean commits to salary policy
(M2) then Professor provides evidence

OPTIMAL MECHANISM

Dean:
- positive evidence gets salary = 90
- no evidence gets salary = 50
- negative evidence gets salary ≤ 50

\[
\begin{array}{lcc}
 t_+ & : & 25\% & 90 \\
 t_0 & : & 50\% & 60 \\
 t_- & : & 25\% & 30 \\
\end{array}
\]
Example 1: Mechanism

MECHANISM: (M1) Dean commits to salary \textit{policy}
(M2) then Professor provides \textit{evidence}

OPTIMAL MECHANISM

- Dean:
  - positive evidence gets salary \(= 90\)
  - no evidence gets salary \(= 50\)
  - negative evidence gets salary \(\leq 50\)

\[ t_+ : \quad 25\% \quad 90 \]
\[ t_0 : \quad 50\% \quad 60 \]
\[ t_- : \quad 25\% \quad 30 \]
Example 1: Explanation

\begin{itemize}
  \item \( t_+ \): 25% \quad 90
  \item \( t_0 \): 50% \quad 60
  \item \( t_- \): 25% \quad 30
\end{itemize}
Example 1: Explanation

in EQUILIBRIUM:

<table>
<thead>
<tr>
<th></th>
<th>t⁺</th>
<th>25%</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t₀</td>
<td>50%</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>t⁻</td>
<td>25%</td>
<td>30</td>
</tr>
</tbody>
</table>
Example 1: Explanation

in EQUILIBRIUM:

- \( t_- \) says \( t_0 \)
Example 1: Explanation

in EQUILIBRIUM:

- $t_-$ says $t_0$
- the value of $t_0$ is higher than the value of $t_-$

<table>
<thead>
<tr>
<th></th>
<th>$t_+$</th>
<th>25%</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0$</td>
<td>50%</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>$t_-$</td>
<td>25%</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
Example 1: Explanation

in **EQUILIBRIUM**:

- $t_-$ says $t_0$
- the value of $t_0$ is higher than the value of $t_-$

in **MECHANISM**:

$\begin{align*}
t_+ & : \quad 25\% \quad 90 \\
t_0 & : \quad 50\% \quad 60 \\
t_- & : \quad 25\% \quad 30
\end{align*}$
Example 1: Explanation

In EQUILIBRIUM:

- $t_-$ says $t_0$
- the value of $t_0$ is higher than the value of $t_-$

In MECHANISM:

- the only way to separate $t_-$ from $t_0$
  is to pay $t_-$ strictly more than to $t_0$
Example 1: Explanation

in EQUILIBRIUM:

- $t_-$ says $t_0$
- the value of $t_0$ is higher than the value of $t_-$

in MECHANISM:

- the only way to separate $t_-$ from $t_0$
  is to pay $t_-$ strictly more than to $t_0$
- this is not optimal
Example 1: Explanation

in **EQUILIBRIUM**:  
- $t_-$ says $t_0$
- the value of $t_0$ is higher than the value of $t_-$

in **MECHANISM**:  
- the *only way to separate* $t_-$ from $t_0$ is to pay $t_-$ strictly more than to $t_0$
- this is *not optimal*

OPTIMAL MECHANISM *does not separate more than* EQUILIBRIUM
**Example 1: Explanation**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_+$</td>
<td>25%</td>
</tr>
<tr>
<td>$t_0$</td>
<td>50%</td>
</tr>
<tr>
<td>$t_-$</td>
<td>25%</td>
</tr>
</tbody>
</table>

**OPTIMAL MECHANISM does not separate more than EQUILIBRIUM**
Example 1: Explanation

\[ t_+ : 25\% \quad 90 \]
\[ t_0 : 50\% \quad 60 \]
\[ t_- : 25\% \quad 30 \]

Partial truth:

**Optimal mechanism does not separate more than equilibrium.**
Example 2

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value
- Professor’s evidence (verifiable):
Example 2

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value
- Professor’s evidence (verifiable):
  \[^{t_0}\] 50\%: no evidence \rightarrow \text{value} = 60
  \[^{t_-}\] 25\%: negative evidence \rightarrow \text{value} = 30
Example 2

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value

Professor’s evidence (verifiable):

- $[t_0]$ 50%: no evidence $\rightarrow$ value = 60
- $[t_-]$ 25%: negative evidence $\rightarrow$ value = 30
- $[t_+]$ 20%: positive evidence $\rightarrow$ value = 110
- $[t_{\pm}]$ 5%: both evidences $\rightarrow$ value = 40
Example 2

- Professor wants salary as high as possible
- Dean wants salary to be as close as possible to the Professor’s value
- Professor’s evidence (verifiable):
  - $[t_0]$: 50%: no evidence → value = 60
  - $[t_-]$: 25%: negative evidence → value = 30
  - $[t_+]$: 20%: positive evidence → value = 110
  - $[t_{\pm}]$: 5%: both evidences → value = 40
Example 2: Equilibrium

- $t_+ : 20\% 110$
- $t_0 : 50\% 60$
- $t_{\pm} : 5\% 40$
- $t_- : 25\% 30$
Example 2: Equilibrium

EQUILIBRIUM

- $t_+ : 20\% \quad 110$
- $t_0 : 50\% \quad 60$
- $t_\pm : 5\% \quad 40$
- $t_- : 25\% \quad 30$
Example 2: Equilibrium

EQUILIBRIUM

Professor:

- $t_+$, $t_\pm$ provide positive evidence
- $t_0$, $t_-$ provide no evidence
Example 2: Equilibrium

Professor:
- $t_+ , t_{\pm}$ provide positive evidence
- $t_0 , t_-$ provide no evidence

Dean:
- positive evidence gets salary $= 90$
  $= (20\% \cdot 110 + 5\% \cdot 40)/25\%$
- no evidence gets salary $= 50$
  $= (50\% \cdot 60 + 25\% \cdot 30)/75\%$
- negative evidence gets salary $= 30$
- both evidences gets salary $= 40$
### Example 2: Equilibrium

<table>
<thead>
<tr>
<th>Time</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_+$</td>
<td>20%</td>
</tr>
<tr>
<td>$t_0$</td>
<td>50%</td>
</tr>
<tr>
<td>$t_{\pm}$</td>
<td>5%</td>
</tr>
<tr>
<td>$t_-$</td>
<td>25%</td>
</tr>
</tbody>
</table>
Example 2: Equilibrium

- $t_+: 20\% \quad 110$
- $t_0: 50\% \quad 60$
- $t_{\pm}: 5\% \quad 40$
- $t_-: 25\% \quad 30$

Value:

- 30
- 40
- 60
- 110

SERGIU HART © 2015 – p. 19
Example 2: Equilibrium

\[ t_+ : 20\% \quad 110 \]
\[ t_0 : 50\% \quad 60 \]
\[ t_{\pm} : 5\% \quad 40 \]
\[ t_- : 25\% \quad 30 \]

value: 30 40 60 110
Example 2: Equilibrium

$t_+ : 20\% \quad 110$
$t_0 : 50\% \quad 60$
$t_{\pm} : 5\% \quad 40$
$t_- : 25\% \quad 30$

value: 30 40 60 110

partial truth:
Example 2: Equilibrium

\[
\begin{align*}
t_+ : & \quad 20\% & 110 \\
t_0 : & \quad 50\% & 60 \\
t_{\pm} : & \quad 5\% & 40 \\
t_- : & \quad 25\% & 30 \\
\end{align*}
\]

prof says:

value:

\begin{align*}
t_- & \quad 30 \\
t_{\pm} & \quad 40 \\
t_0 & \quad 60 \\
t_+ & \quad 110 \\
\end{align*}

partial truth:
Example 2: Mechanism

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_+$</td>
<td>20%</td>
</tr>
<tr>
<td>$t_0$</td>
<td>50%</td>
</tr>
<tr>
<td>$t_{\pm}$</td>
<td>5%</td>
</tr>
<tr>
<td>$t_-$</td>
<td>25%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>30</td>
</tr>
</tbody>
</table>
Example 2: Mechanism

OPTIMAL MECHANISM

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_+$</td>
<td>20%</td>
</tr>
<tr>
<td>$t_0$</td>
<td>50%</td>
</tr>
<tr>
<td>$t_{\pm}$</td>
<td>5%</td>
</tr>
<tr>
<td>$t_-$</td>
<td>25%</td>
</tr>
</tbody>
</table>

110
60
40
30
Example 2: Mechanism

OPTIMAL MECHANISM

Dean:
- positive evidence gets salary = 90
- no evidence gets salary = 50
- negative evidence gets salary ≤ 50
- both evidences gets salary ≤ 90

<table>
<thead>
<tr>
<th></th>
<th>t⁺</th>
<th>20%</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>t₀</td>
<td>50%</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>t⁻</td>
<td>5%</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>t⁻⁻</td>
<td>25%</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>
Example 2: Mechanism

OPTIMAL MECHANISM

Dean:

- positive evidence gets salary = 90
- no evidence gets salary = 50
- negative evidence gets salary ≤ 50
- both evidences gets salary ≤ 90

OPTIMAL MECHANISM = EQUILIBRIUM
Example 2: Equilibrium

\[
\begin{align*}
  t_+ & : 20\% \quad 110 \\
  t_0 & : 50\% \quad 60 \\
  t_{\pm} & : 5\% \quad 40 \\
  t_- & : 25\% \quad 30
\end{align*}
\]
Example 2: Equilibrium

Another equilibrium

| t_+  | 20%  | 110 |
| t_0  | 50%  | 60  |
| t_-  | 5%   | 40  |
| t_-  | 25%  | 30  |
Example 2: Equilibrium

Another equilibrium

Professor:
- always provides no evidence

| $t_+$ | 20% | 110 |
| $t_0$ | 50% | 60  |
| $t_\pm$ | 5% | 40  |
| $t_-$ | 25% | 30  |
Another equilibrium

- Professor:
  - always provides no evidence

- Dean:
  - ignores all evidence and sets salary = 60

\[ \text{salary} = 50\% \cdot 60 + 25\% \cdot 30 + 20\% \cdot 110 + 5\% \cdot 40 \]
Example 2: Equilibrium

Another equilibrium

- **Professor:**
  - always provides **no evidence**

- **Dean:**
  - ignores all **evidence** and sets **salary = 60**
    
    \[
    = 50\% \cdot 60 + 25\% \cdot 30 + 20\% \cdot 110 + 5\% \cdot 40
    \]

supported by the belief of the Dean when receiving the out-of-equilibrium **positive evidence** that it mostly comes from \( t_{\pm} \) rather than from \( t_{+} \)
Example 2: Equilibrium

Another equilibrium ("babbling")

- Professor:
  - always provides no evidence

- Dean:
  - ignores all evidence and sets salary = $60$
    $$= 50\% \cdot 60 + 25\% \cdot 30 + 20\% \cdot 110 + 5\% \cdot 40$$

supported by the belief of the Dean when receiving the out-of-equilibrium positive evidence that it mostly comes from $t_{\pm}$ rather than from $t_{+}$
Another equilibrium ("babbling") SATISFIES ALL STANDARD REFINEMENTS

- Professor:
  - always provides no evidence

- Dean:
  - ignores all evidence and sets salary = 60
  \[
  = 50\% \cdot 60 + 25\% \cdot 30 + 20\% \cdot 110 + 5\% \cdot 40
  \]

supported by the belief of the Dean when receiving the out-of-equilibrium positive evidence that it mostly comes from \( t_\pm \) rather than from \( t_+ \)
Example 2: Equilibrium

Another equilibrium ("babbling") satisfies all standard refinements.
Example 2: Equilibrium

Another equilibrium ("babbling") SATISFIES ALL STANDARD REFINEMENTS

$t_+: 20\% \quad 110$
$t_0: 50\% \quad 60$
$t_-: 25\% \quad 30$

$L:$
Example 2: Equilibrium

Another equilibrium ("babbling") SATISFIES ALL STANDARD REFINEMENTS

prof says:

$L$:
Example 2: Equilibrium

Another equilibrium ("babbling") Satisfies all standard refinements

dean pays:  
\[ \begin{array}{ccc}
30 & 40 & 60 \\
\end{array} \]

prof says:

\[ L: \]

\[ \begin{array}{c}
t_- \\
30 \\
\end{array} \quad \begin{array}{c}
t_\pm \\
40 \\
\end{array} \quad \begin{array}{c}
t_0 \\
60 \\
\end{array} \quad \begin{array}{c}
t_+ \\
110 \\
\end{array} \]

- \[ t_+ : 20\% \quad 110 \]
- \[ t_0 : 50\% \quad 60 \]
- \[ t_\pm : 5\% \quad 40 \]
- \[ t_- : 25\% \quad 30 \]
Main Result

In EVIDENCE GAMES
In EVIDENCE GAMES

the EQUILIBRIUM outcome obtained without commitment
In EVIDENCE GAMES

the EQUILIBRIUM outcome obtained without commitment

and the OPTIMAL MECHANISM outcome obtained with commitment
In EVIDENCE GAMES

the EQUILIBRIUM outcome obtained without commitment

and the OPTIMAL MECHANISM outcome obtained with commitment

COINCIDE
Main Result: Equivalence

In EVIDENCE GAMES

the EQUILIBRIUM outcome obtained without commitment

and the OPTIMAL MECHANISM outcome obtained with commitment

COINCIDE
AGENT ($A$)

PRINCIPAL ($P$) (= "market")
Model

- AGENT \((A)\)
- PRINCIPAL \((P)\) (= "market")
- (finite) set of TYPES: \(T\)
- the type \(t \in T\) is chosen according to a probability distribution \(p \in \Delta(T)\)
Model

- **AGENT** \((A)\)
- **PRINCIPAL** \((P)\) (= "market")
- (finite) set of **TYPES**: \(T\)
- the type \(t \in T\) is chosen according to a probability distribution \(p \in \Delta(T)\)
- the type \(t \in T\) is revealed to Agent and not to Principal
Model

AGENT (\(A\))

PRINCIPAL (\(P\)) (= "market")

(finite) set of TYPES: \(T\)

the type \(t \in T\) is chosen according to a probability distribution \(p \in \Delta(T)\)

the type \(t \in T\) is revealed to Agent and not to Principal

Agent’s MESSAGE: \(s \in T\)
Model

AGENT $(A)$

PRINCIPAL $(P) (= "market")$

(finite) set of TYPES: $T$

the type $t \in T$ is chosen according to a probability distribution $p \in \Delta(T)$

the type $t \in T$ is revealed to Agent and not to Principal

Agent’s MESSAGE: $s \in T$

Principal’s DECISION: REWARD $x \in \mathbb{R}$
$U^A$ and $U^P$ do not depend on the message $s$. 
Payoffs / Utilities

- $U^A$ and $U^P$ do **not** depend on the message $s$
- $U^A$ does **not** depend on the type $t$

$$U^A(s, x; t) = x$$
Payoffs / Utilities

- $U^A$ and $U^P$ do not depend on the message $s$
- $U^A$ does not depend on the type $t$

$$U^A(s, x; t) = x$$

- $U^P$: "Canonical" example

$$h_t(x) := U^P(s, x; t) = -(x - \nu(t))^2$$

$\nu(t)$ = the "value" of type $t$
Payoffs / Utilities

- \( U^A \) and \( U^P \) do not depend on the message \( s \)
- \( U^A \) does not depend on the type \( t \)

\[
U^A(s, x; t) = x
\]

- \( U^P \): "Canonical" example

\[
h_t(x) := U^P(s, x; t) = -(x - v(t))^2
\]

\( v(t) \) = the "value" of type \( t \)

- General assumption:
  \((SP)\) \( U^P \) is single-peaked w.r.t. \( U^A \)
Single Peakedness (SP)
For every distribution of types (belief) $q \in \Delta(T)$, the principal’s expected utility

$$h_q(x) = \sum_{t \in T} q_t h_t(x)$$

is a *single-peaked* function of the reward $x$. 
Single Peakedness (SP)

For every distribution of types (belief) $q \in \Delta(T)$ the principal’s expected utility

$$h_q(x) = \sum_{t \in T} q_t h_t(x)$$

is a single-peaked function of the reward $x$

$\iff$ There exists $v(q)$ such that

$$h'_q(x) > 0 \quad \text{for } x < v(q)$$
$$h'_q(x) = 0 \quad \text{for } x = v(q)$$
$$h'_q(x) < 0 \quad \text{for } x > v(q)$$
Single Peakedness (SP)

- Canonical example:
  \[ h_t(x) = -(x - \nu(t))^2 \]
Single Peakedness (SP)

Canonical example:

\[ h_t(x) = -(x - v(t))^2 \]

\[ v(q) = E_q[v(t)] = \sum_t q_t v(t) \]
Single Peakedness (SP)

- Canonical example:
  \[ h_t(x) = -(x - v(t))^2 \]
  \[ v(q) = E_q[v(t)] = \sum_t q_t v(t) \]

- More general:
  \[ h_t(x) \text{ is a differentiable strictly concave function of } x, \text{ for each } t \]
Single Peakedness (SP)

- Canonical example:
  
  \( h_t(x) = -(x - v(t))^2 \)
  
  \( v(q) = E_q[v(t)] = \sum_t q_t v(t) \)

- More general:
  
  \( h_t(x) \) is a differentiable strictly concave function of \( x \), for each \( t \)

- (SP) is more general than concavity
Agent reveals:

"the truth, nothing but the truth"
Agent reveals:

"the truth, nothing but the truth"

NOT necessarily "the whole truth"
Information and Truth

Agent reveals:

- "the truth, nothing but the truth"

  all the evidence that the agent reveals must be true (it is verifiable)

- NOT necessarily "the whole truth"
Agent reveals:

- "the truth, nothing but the truth"
  all the evidence that the agent reveals must be true (it is verifiable)

- NOT necessarily "the whole truth"
  the agent does not have to reveal all the evidence that he has
Agent reveals:

- **"the truth, nothing but the truth"**
  
  all the evidence that the agent reveals must be true (it is verifiable)

- **NOT necessarily "the whole truth"**
  
  the agent does not have to reveal all the evidence that he has

⇒ Agent can *pretend* to be a type that has *less information (less evidence)*
Information and Truth

\[ t \rightarrow s \] (for types \( t, s \in T \)):

- \( s \) has less information than \( t \)
- \( s \) is a possible partial truth for \( t \)
\[ t \rightarrow s \] (for types \( t, s \in T \)):

- \( s \) has less information than \( t \)
- \( s \) is a possible partial truth for \( t \)
- \( \cdot \rightarrow \cdot \) is a WEAK ORDER on \( T \)
$t \rightarrow s$ (for types $t, s \in T$):

- $s$ has less information than $t$
- $s$ is a possible partial truth for $t$

* is a weak order on $T$:

- (L1) Reflexive: $t \rightarrow t$
  revealing the whole truth is always possible
Information and Truth

$t \rightarrow s$ (for types $t, s \in T$):
- $s$ has less information than $t$
- $s$ is a possible partial truth for $t$

$\rightarrow$ is a WEAK ORDER on $T$:

(L1) REFLEXIVE: $t \rightarrow t$
revealing the whole truth is always possible

(L2) TRANSITIVE:
$t \rightarrow s$ and $s \rightarrow r$ imply $t \rightarrow r$
if $s$ is a partial truth for $t$ and $r$ is a partial truth for $s$ then $r$ is a partial truth for $t$
**Information and Truth**

- $t \rightarrow s$ (for types $t, s \in T$):
  - $s$ has less information than $t$
  - $s$ is a possible partial truth for $t$

- $\rightarrow$ is a WEAK ORDER on $T$:

  (L1) REFLEXIVE: $t \rightarrow t$
  revealing the whole truth is always possible

  (L2) TRANSITIVE:
  $t \rightarrow s$ and $s \rightarrow r$ imply $t \rightarrow r$
  if $s$ is a partial truth for $t$ and $r$ is a partial truth for $s$ then $r$ is a partial truth for $t$

- need not be a complete order
Information and Truth

\[ t \rightarrow s \] (for types \( t, s \in T \)):

- \( s \) has less information than \( t \)
- \( s \) is a possible partial truth for \( t \)

\( \rightarrow \cdot \) is a weak order on \( T \):

- (L1) Reflexive: \( t \rightarrow t \)
- (L2) Transitive:
  \[ t \rightarrow s \] and \( s \rightarrow r \) imply \( t \rightarrow r \)
Information and Truth

$t \rightarrow s$ (for types $t, s \in T$):
- $s$ has less information than $t$
- $s$ is a possible partial truth for $t$

$\rightarrow$ is a weak order on $T$:
- (L1) Reflexive: $t \rightarrow t$
- (L2) Transitive:
  - $t \rightarrow s$ and $s \rightarrow r$ imply $t \rightarrow r$

$L(t) = \{ s \in T : t \rightarrow s \}$
Information and Truth

- $t \rightarrow s$ (for types $t, s \in T$):
  - $s$ has less information than $t$
  - $s$ is a possible partial truth for $t$

- $\rightarrow$ is a WEAK ORDER on $T$:
  - (L1) REFLEXIVE: $t \rightarrow t$
  - (L2) TRANSITIVE:
    - $t \rightarrow s$ and $s \rightarrow r$ imply $t \rightarrow r$

- $L(t) = \{ s \in T : t \rightarrow s \}$
  - the set of possible messages of type $t$
  - the set of types that $t$ can pretend to be
Information and Truth: Examples
Information and Truth: Examples

Evidences
Evidences

\[ T \subseteq 2^E \]

\[ t \rightarrow s \text{ iff } t \supseteq s \]
Information and Truth: Examples

- **Evidences**
  - $T \subseteq 2^E$
  - $t \rightarrow s$ iff $t \supseteq s$

- **Partitions**
**Information and Truth: Examples**

- **Evidences**
  - \( T \subseteq 2^E \)
  - \( t \rightarrow s \) iff \( t \supseteq s \)

- **Partitions**
  - \( T = \text{kens in a sequence of partitions} \)
  - \( t \rightarrow s \) iff \( t \subseteq s \)
Information and Truth: Examples

- **Evidences**
  - $T \subseteq 2^E$
  - $t \rightarrow s \iff t \supseteq s$

- **Partitions**
  - $T = \text{kens in a sequence of partitions}$
  - $t \rightarrow s \iff t \subseteq s$

- **Signals**
**Evidences**

- \( T \subseteq 2^E \)
- \( t \rightarrow s \) iff \( t \supseteq s \)

**Partitions**

- \( T = \) kens in a sequence of partitions
- \( t \rightarrow s \) iff \( t \subseteq s \)

**Signals**

- \( Z_1, Z_2, \ldots, Z_n \) random variables
- \( T \subseteq \mathcal{F}(Z_1, Z_2, \ldots, Z_n) \)
- \( t \rightarrow s \) iff \( t \subseteq s \)
(G1) Agent sends message $s \in L(t)$ to Principal
(G1) Agent sends message $s \in L(t)$ to Principal

(G2) Then Principal sets reward $x \in \mathbb{R}$
(G1) Agent sends message $s \in L(t)$ to Principal

(G2) Then Principal sets reward $x \in \mathbb{R}$
(G1) Agent sends message $s \in L(t)$ to Principal

(G2) Then Principal sets reward $x \in \mathbb{R}$

STRATEGIES

(Agent) $\sigma(s|t)$ = probability that type $t$ sends message $s$ in $L(t)$
(G1) Agent sends message \( s \in L(t) \) to Principal

(G2) Then Principal sets reward \( x \in \mathbb{R} \)

**STRATEGIES**

(Agent) \( \sigma(s | t) = \) probability that type \( t \) sends message \( s \) in \( L(t) \)

(Principal) \( \rho(s) \in \mathbb{R} = \) reward to message \( s \)
Equilibrium

$(\sigma, \rho)$ is a \textbf{NASH EQUILIBRIUM} if
(σ, ρ) is a **NASH EQUILIBRIUM** if

(A) \( \sigma(r|t) > 0 \Rightarrow \rho(r) = \max_{s \in L(t)} \rho(s) \)
Equilibrium

\((\sigma, \rho)\) is a **NASH EQUILIBRIUM** if

1. **(A)** \(\sigma(r|t) > 0 \Rightarrow \rho(r) = \max_{s \in L(t)} \rho(s)\)

2. **(P)** \(\bar{\sigma}(s) > 0 \Rightarrow \rho(s) = \nu(q(s))\)

where

- \(\bar{\sigma}(s) = \) total probability of message \(s\)
- \(q(s) \in \Delta(T) = \) the posterior distribution on types conditional on message \(s\)
Equilibrium

\((\sigma, \rho)\) is a **NASH EQUILIBRIUM** if

(A) \(\sigma(r|t) > 0 \Rightarrow \rho(r) = \max_{s \in L(t)} \rho(s)\)

(P) \(\bar{\sigma}(s) > 0 \Rightarrow \rho(s) = \nu(q(s))\)

where

- \(\bar{\sigma}(s) = \text{total probability of message } s\)
- \(q(s) \in \Delta(T) = \text{the posterior distribution on types conditional on message } s\)

**Outcome:** \(\pi_t = \max_{s \in L(t)} \rho(s) = \rho(\sigma(\cdot|t))\)
Equilibrium

$(\sigma, \rho)$ is a **Nash equilibrium** if

(A) $\sigma(r|t) > 0 \Rightarrow \rho(r) = \max_{s \in L(t)} \rho(s)$

(P) $\bar{\sigma}(s) > 0 \Rightarrow \rho(s) = v(q(s))$

where

- $\bar{\sigma}(s)$ = total probability of message $s$
- $q(s) \in \Delta(T)$ = the posterior distribution on types conditional on message $s$

**Outcome:** $\pi_t = \max_{s \in L(t)} \rho(s) = \rho(\sigma(\cdot|t))$

$\pi = (\pi_t)_{t \in T} \in \mathbb{R}^T$
Truth-Leaning Equilibrium

Revealing the whole truth gets a slight (= infinitesimal) boost in payoff and probability.
Revealing *the whole truth* gets a slight (= infinitesimal) boost in payoff and probability

(T1) Revealing *the whole truth* is preferable when the reward is the same
Revealing the whole truth gets a slight (= infinitesimal) boost in payoff and probability.

(T1) Revealing the whole truth is preferable when the reward is the same (lexicographic preference)
Truth-Leaning Equilibrium

Revealing the whole truth gets a slight (= infinitesimal) boost in payoff and probability

(T1) Revealing the whole truth is preferable when the reward is the same (lexicographic preference)

(T2) The whole truth is revealed with infinitesimal positive probability
Revealing *the whole truth* gets a slight (= infinitesimal) boost in payoff and probability

(T1) Revealing *the whole truth* is preferable when the reward is the same (lexicographic preference)

(T2) *The whole truth* is revealed with infinitesimal positive probability (by mistake, or because the agent may be non-strategic, or ... [UK])
A Nash equilibrium is **TRUTH-LEANING** if it satisfies:
A Nash equilibrium is **Truth-Leaning** if it satisfies:

\[(T1) \quad \rho(t) = \max_{s \in L(t)} \rho(s) \Rightarrow \sigma(t|t) = 1\]
A Nash equilibrium is **TRUTH-LEANING** if it satisfies:

\[(T1) \quad \rho(t) = \max_{s \in L(t)} \rho(s) \Rightarrow \sigma(t|t) = 1\]

(if message $t$ is a best reply for type $t$ then it is used for sure)
A Nash equilibrium is **TRUTH-LEANING** if it satisfies:

\[ (T1) \quad \rho(t) = \max_{s \in L(t)} \rho(s) \quad \Rightarrow \quad \sigma(t|t) = 1 \]

(if message \( t \) is a best reply for type \( t \) then it is used for sure)

\[ (T2) \quad \bar{\sigma}(t) = 0 \quad \Rightarrow \quad \rho(t) = v(t) \]
A Nash equilibrium is **Truth-Leaning** if it satisfies:

\[(T1) \quad \rho(t) = \max_{s \in L(t)} \rho(s) \Rightarrow \sigma(t|t) = 1 \]

(if message \(t\) is a best reply for type \(t\) then it is used for sure)

\[(T2) \quad \bar{\sigma}(t) = 0 \Rightarrow \rho(t) = v(t) \]

(if message \(t\) is not used then the reward equals the value of type \(t\); i.e., the belief is that the [unexpected] message \(t\) comes from type \(t\))
Truth-Leaning Equilibrium
Truth-Leaning equilibria:
Truth-Leaning equilibria:

- coincide with the equilibria selected in the "voluntary disclosure" literature
Truth-Leaning equilibria:

- coincide with the equilibria selected in the "voluntary disclosure" literature
- satisfy all the refinement conditions in the literature
Truth-Leaning Equilibrium

Truth-Leaning equilibria:

- coincide with the equilibria selected in the "voluntary disclosure" literature
- satisfy all the refinement conditions in the literature
- eliminate "unreasonable" equilibria (such as "babbling" in Example 2)
Truth-Leaning Equilibrium

Truth-Leaning equilibria:

- coincide with the equilibria selected in the "voluntary disclosure" literature
- satisfy all the refinement conditions in the literature
- eliminate "unreasonable" equilibria (such as "babbling" in Example 2)
- ...

Sergiu Hart © 2015 – p. 36
MECHANISM:
MECHANISM: Reward scheme $\rho : T \rightarrow \mathbb{R}$

$(\rho(s) = \text{reward to message } s)$
MECHANISM: Reward scheme $\rho : T \rightarrow \mathbb{R}$

$(\rho(s) = \text{reward to message } s)$

Agent’s payoff when type is $t$:

$$\pi_t = \max_{s \in L(t)} \rho(s)$$
MECHANISM: Reward scheme $\rho : T \rightarrow \mathbb{R}$

$(\rho(s) = \text{reward to message } s)$

- **Agent's payoff** when type is $t$:

$$\pi_t = \max_{s \in L(t)} \rho(s)$$

- **Outcome**: $\pi = (\pi_t)_{t \in T} \in \mathbb{R}^T$
**Mechanism**

**MECHANISM:** Reward scheme $\rho : T \rightarrow \mathbb{R}$

$(\rho(s) = \text{reward to message } s)$

- **Agent's payoff** when type is $t$:
  
  $$\pi_t = \max_{s \in L(t)} \rho(s)$$

- **Outcome:** $\pi = (\pi_t)_{t \in T} \in \mathbb{R}^T$

- **Principal's payoff**:
  
  $$H(\pi) = \sum_{t \in T} p_t h_t(\pi_t)$$
Incentive Compatibility
Outcome $\pi = (\pi_t)_{t \in T} \in \mathbb{R}^T$ is generated by a mechanism $\rho$
Outcome \( \pi = (\pi_t)_{t \in T} \in \mathbb{R}^T \) is generated by a mechanism \( \rho \) if and only if
Incentive Compatibility

Outcome $\pi = (\pi_t)_{t \in T} \in \mathbb{R}^T$ is generated by a mechanism $\rho$ if and only if

$$\pi_t \geq \pi_s$$

for all $s, t \in T$ with $s \in L(t)$
Incentive Compatibility

Outcome $\pi = (\pi_t)_{t \in T} \in \mathbb{R}^T$ is generated by a mechanism $\rho$
if and only if

$$\pi_t \geq \pi_s$$
for all $s, t \in T$ with $s \in L(t)$

Immediate because $L$ satisfies reflexivity (L1) and transitivity (L2)
Incentive Compatibility

Outcome $\pi = (\pi_t)_{t \in T} \in \mathbb{R}^T$ is generated by a mechanism $\rho$ if and only if

$$\pi_t \geq \pi_s$$

for all $s, t \in T$ with $s \in L(t)$

- Immediate because $L$ satisfies reflexivity (L1) and transitivity (L2)
- "direct" mechanism: $\rho(t) = \pi_t$
Incentive Compatibility

Outcome \( \pi = (\pi_t)_{t \in T} \in \mathbb{R}^T \) is generated by a mechanism \( \rho \) if and only if

\[
\pi_t \geq \pi_s \quad \text{for all } s, t \in T \text{ with } s \in L(t)
\]

- Immediate because \( L \) satisfies reflexivity (L1) and transitivity (L2)
- "direct" mechanism: \( \rho(t) = \pi_t \)

Green–Laffont 86
OPTIMAL MECHANISM:
OPTIMAL MECHANISM:

Maximize $H(\pi) = \sum_{t \in T} p_t h_t(\pi_t)$
OPTIMAL MECHANISM:

Maximize $H(\pi) = \sum_{t \in T} p_t h_t(\pi_t)$

subject to (IC):

$$\pi_t \geq \pi_s$$

for all $s, t \in T$ with $s \in L(t)$
OPTIMAL MECHANISM:

Maximize $H(\pi) = \sum_{t \in T} p_t h_t(\pi_t)$

subject to (IC):

$$\pi_t \geq \pi_s$$

for all $s, t \in T$ with $s \in L(t)$

**OPTIMAL MECHANISM** = Maximum under Incentive Constraints
Assume that the payoff functions $h_t$ for all $t \in T$ are differentiable and satisfy the single-peaked condition.
Assume that the payoff functions $h_t$ for all $t \in T$ are differentiable and satisfy the single-peaked condition.

Then there is a unique TRUTH-LEANING EQUILIBRIUM outcome,
Main Result

Assume that the payoff functions $h_t$ for all $t \in T$ are differentiable and satisfy the single-peaked condition.

Then there is a unique

**TRUTH-LEANING EQUILIBRIUM** outcome,

and a unique **OPTIMAL MECHANISM** outcome,
Assume that the payoff functions \( h_t \) for all \( t \in T \) are differentiable and satisfy the single-peaked condition.

Then there is a unique **TRUTH-LEANING EQUILIBRIUM** outcome, a unique **OPTIMAL MECHANISM** outcome, and these two outcomes **COINCIDE**.
Main Result: Equivalence
Main Result: Equivalence

The equilibrium strategies need not be unique (happens only when the Agent is indifferent—and then the Principal is also indifferent)
Main Result: Equivalence
Main Result: Equivalence

All the conditions are indispensable:
Main Result: Equivalence

All the conditions are indispensable:

- Truth structure: reflexivity
Main Result: Equivalence

*All the conditions are indispensable:*

- Truth structure: reflexivity
- Truth structure: transitivity
Main Result: Equivalence

All the conditions are indispensable:

- Truth structure: reflexivity
- Truth structure: transitivity
- Truth Leaning: whole truth slightly better
Main Result: Equivalence

All the conditions are indispensable:

- Truth structure: reflexivity
- Truth structure: transitivity
- Truth Leaning: whole truth slightly better
- Truth Leaning: whole truth slightly possible
Main Result: Equivalence

All the conditions are indispensable:

- Truth structure: reflexivity
- Truth structure: transitivity
- Truth Leaning: whole truth slightly better
- Truth Leaning: whole truth slightly possible
- Agent’s utility: independent of type
Main Result: Equivalence

All the conditions are indispensable:

- Truth structure: reflexivity
- Truth structure: transitivity
- Truth Leaning: whole truth slightly better
- Truth Leaning: whole truth slightly possible
- Agent’s utility: independent of type
- Principal’s utility: single-peaked with respect to Agent’s utility
Main Result: Equivalence

All the conditions are indispensable:

- Truth structure: reflexivity
- Truth structure: transitivity
- Truth Leaning: whole truth slightly better
- Truth Leaning: whole truth slightly possible
- Agent’s utility: independent of type
- Principal’s utility: single-peaked with respect to Agent’s utility
- Principal’s utility: differentiable
Main Result: Equivalence
Main Result: Equivalence

**EQUILIBRIUM** (without commitment) yields the same as **COMMITMENT**
Main Result: Equivalence

- EQUILIBRIUM (without commitment) yields the same as COMMITMENT

- EQUILIBRIUM yields OPTIMAL SEPARATION (for the principal / "market")
Main Result: Equivalence

- EQUILIBRIUM (without commitment) yields the same as COMMITMENT

- EQUILIBRIUM yields OPTIMAL SEPARATION (for the principal / "market")

- EQUILIBRIUM yields PARETO EFFICIENCY (in the canonical setup)
Main Result: Equivalence

under Incentive Constraints

- **EQUILIBRIUM** (without commitment) yields the same as **COMMITMENT**

- **EQUILIBRIUM** yields **OPTIMAL SEPARATION** (for the principal / "market")

- **EQUILIBRIUM** yields **PARETO EFFICIENCY** (in the canonical setup)
Disclosure by public firms
Applications: Finance

*Disclosure by public firms*

- Disclosing false information is a criminal act
Applications: Finance

Disclosure by public firms

- Disclosing false information is a criminal act
- Withholding information is allowed in some cases
Applications: Finance

Disclosure by public firms

- Disclosing false information is a criminal act
- Withholding information is allowed in some cases, and is difficult (if not impossible) to detect
Applications: Finance

Disclosure by public firms

- Disclosing false information is a criminal act
- Withholding information is allowed in some cases, and is difficult (if not impossible) to detect
- Impacts asset prices (e.g.: quarterly reports)
U.S.: "You have the right to remain silent. Anything you say can and will be used against you in a court of law ..."
(Miranda Warning, following the 1966 Miranda v. Arizona Supreme Court decision)
Law: Right to Remain Silent

- **U.S.**: "You have the right to remain silent. Anything you say can and will be used against you in a court of law ..."
  (Miranda Warning, following the 1966 Miranda v. Arizona Supreme Court decision)

- **U.K.**: "You do not have to say anything. But it may harm your defence if you do not mention when questioned something which you later rely on in court. Anything you do say may be given in evidence ..."
  (The Criminal Justice and Public Order Act 1994)
Law: Right to Remain Silent

- **U.S.**: "You have the right to remain silent. Anything you say can and will be used against you in a court of law ..." 
  *(Miranda Warning, following the 1966 Miranda v. Arizona Supreme Court decision)*

- **U.K.**: "You do not have to say anything. But it may harm your defence if you do not mention when questioned something which you later rely on in court. Anything you do say may be given in evidence ..." 
  *(The Criminal Justice and Public Order Act 1994)*
Law: Right to Remain Silent

- **U.S.**: "You have the right to remain silent. Anything you say can and will be used against you in a court of law ..." *(Miranda Warning, following the 1966 Miranda v. Arizona Supreme Court decision)*

- **U.K.**: "You do not have to say anything. But it may harm your defence if you do not mention when questioned something which you later rely on in court. Anything you do say may be given in evidence ..." *(The Criminal Justice and Public Order Act 1994)*
ask Kobi ...
Equivalence Theorem: Intuition
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \ (\neq t) \) then:
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \) \((\neq t)\) then:

- \( s \) reveals his type (i.e., says \( s \))
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \) \((\neq t)\) then:

- \( s \) reveals his type (i.e., says \( s \))

If \( s \) had something better, so would \( t \)
In a **TRUTH-LEANING EQUILIBRIUM** if $t$ pretends to be $s$ ($\neq t$) then:

- $s$ reveals his type (i.e., says $s$)
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \) \((\neq t)\) then:

- \( s \) reveals his type (i.e., says \( s \))
- \( v(t) < v(s) \)
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM**
if \( t \) pretends to be \( s \) (\( \neq t \)) then:
- \( s \) reveals his type (i.e., says \( s \))
- \( v(t) < v(s) \)

No one pretends to be worth less than they are
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \) \((\neq t)\) then:

- \( s \) reveals his type (i.e., says \( s \))
- \( v(t) < v(s) \)

**NOTE**: These conclusions need **not** hold for equilibria that are **not** truth-leanin
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \) \((\neq t)\) then:

- \( s \) reveals his type (i.e., says \( s \))
- \( v(t) < v(s) \)
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if $t$ pretends to be $s \ (\neq t)$ then:

- $s$ reveals his type (i.e., says $s$)
- $v(t) < v(s)$

$\Rightarrow$ $t$ and $s$ cannot be separated in any **OPTIMAL MECHANISM**
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \) \((\neq t)\) then:

- \( s \) reveals his type (i.e., says \( s \))
- \( v(t) < v(s) \)

\[ \Rightarrow \] \( t \) and \( s \) cannot be separated in any **OPTIMAL MECHANISM**

- To separate: \( \rho(t) > \rho(s) \) (else \( t \) says \( s \))
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \) \((\neq t)\) then:

- \( s \) reveals his type (i.e., says \( s \))
- \( v(t) < v(s) \)

\[ \Rightarrow \] \( t \) and \( s \) cannot be separated in any **OPTIMAL MECHANISM**

- To separate: \( \rho(t) > \rho(s) \) (else \( t \) says \( s \))
- Not optimal: decreasing \( \rho(t) \) or increasing \( \rho(s) \) brings rewards closer to values
In a **TRUTH-LEANING EQUILIBRIUM** if $t$ pretends to be $s$ ($\neq t$) then:

- $s$ reveals his type (i.e., says $s$)
- $v(t) < v(s)$

$\Rightarrow t$ and $s$ cannot be separated in any **OPTIMAL MECHANISM**
Equivalence Theorem: Intuition

In a **TRUTH-LEANING EQUILIBRIUM** if \( t \) pretends to be \( s \) \((\neq t)\) then:

- \( s \) reveals his type (i.e., says \( s \))
- \( v(t) < v(s) \)

\( \Rightarrow \) \( t \) and \( s \) cannot be separated in any **OPTIMAL MECHANISM**

**CONCLUSION:**

**OPTIMAL MECHANISM** cannot separate **more than** **TRUTH-LEANING EQUILIBRIUM**
0. Preliminaries
Equivalence Theorem: Proof

0. Preliminaries

1. Every TL-EQUILIBRIUM outcome equals the unique OPTIMAL MECHANISM outcome
Equivalence Theorem: Proof

0. Preliminaries

1. Every **TL-EQUILIBRIUM** outcome equals the **unique** **OPTIMAL MECHANISM** outcome

2. A **TL-EQUILIBRIUM** exists
"In betweenness": $v(t_1) \leq v(t_2)$ implies

$$v(t_1) \leq v(\{t_1, t_2\}) \leq v(t_2)$$
"In betweenness": \( v(t_1) \leq v(t_2) \) implies

\[
v(t_1) \leq v(\{t_1, t_2\}) \leq v(t_2)
\]

More generally: if \( q \in \Delta(T) \) is a weighted average of \( q_1, q_2, \ldots, q_n \in \Delta(T) \) then

\[
\min_i v(q_i) \leq v(q) \leq \max_i v(q_i)
\]
"In betweenness": $v(t_1) \leq v(t_2)$ implies

$$v(t_1) \leq v(\{t_1, t_2\}) \leq v(t_2)$$

More generally: if $q \in \Delta(T)$ is a weighted average of $q_1, q_2, \ldots, q_n \in \Delta(T)$ then

$$\min_i v(q_i) \leq v(q) \leq \max_i v(q_i)$$

Proof: Follows from single-peakedness (SP) and differentiability
Proof: 0. Preliminaries
Let \((\sigma, \rho)\) be a \textbf{TL-EQUILIBRIUM} with outcome \(\pi\). Then:
Proof: 0. Preliminaries

Let \((\sigma, \rho)\) be a \textbf{TL-EQUILIBRIUM} with outcome \(\pi\). Then:

- **message** \(t\) is \textbf{used} in equilibrium:
  \[
  \bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff \nu(t) \geq \pi_t = \rho(t)
  \]

- **message** \(t\) is \textbf{not used} in equilibrium:
  \[
  \bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > \nu(t) = \rho(t)
  \]
Proof: 0. Preliminaries

Let \((\sigma, \rho)\) be a TL-EQUILIBRIUM with outcome \(\pi\). Then:

- message \(t\) is used in equilibrium:
  \[ \bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t) \]

- message \(t\) is not used in equilibrium:
  \[ \bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t) \]

Pf.

If \(\sigma(t|t) = 0\) then \(\pi(t) > \rho(t) = v(t)\) (by (T2)).

If \(\sigma(t|t) > 0\) then: \(\sigma(t|s) > 0\) for \(s \neq t\) implies \(\pi_t = \pi_s > v(s)\); but \(\pi_t = v(q(t))\) and so \(\pi_t \leq v(t)\) by in-betweeness. \(\square\)
Proof: 0. Preliminaries

Let \((\sigma, \rho)\) be a **TL-EQUILIBRIUM** with outcome \(\pi\). Then:

- **message** \(t\) is **used** in equilibrium:
  \[
  \bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t)
  \]

- **message** \(t\) is **not used** in equilibrium:
  \[
  \bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t)
  \]

**Corollary.**

Let \(s \neq t\). If \(\sigma(s|t) > 0\) then \(v(s) > v(t)\).
Proof: 0. Preliminaries

Let \((\sigma, \rho)\) be a \textbf{TL-EQUILIBRIUM} with outcome \(\pi\). Then:

- message \(t\) is \textbf{used} in equilibrium:
  \[
  \bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff \nu(t) \geq \pi_t = \rho(t)
  \]

- message \(t\) is \textbf{not used} in equilibrium:
  \[
  \bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > \nu(t) = \rho(t)
  \]

**Corollary.**

Let \(s \neq t\). If \(\sigma(s|t) > 0\) then \(\nu(s) > \nu(t)\).

**Pf.** \(\nu(s) \geq \rho(s)\)

\((s\) is used)
Proof: 0. Preliminaries

Let \((\sigma, \rho)\) be a **TL-EQUILIBRIUM** with outcome \(\pi\). Then:

- **message** \(t\) **is used** in equilibrium:
  \[ \bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t) \]

- **message** \(t\) **is not used** in equilibrium:
  \[ \bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t) \]

**Corollary.**

Let \(s \neq t\). If \(\sigma(s|t) > 0\) then \(v(s) > v(t)\).

**Pf.** \(v(s) \geq \rho(s) = \pi_t\)

\((s\ \text{is optimal for } t)\)
Proof: 0. Preliminaries

Let \((\sigma, \rho)\) be a **TL-EQUILIBRIUM** with outcome \(\pi\). Then:

- **message** \(t\) is **used** in equilibrium:
  \[\bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t)\]

- **message** \(t\) is **not used** in equilibrium:
  \[\bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t)\]

**Corollary.**

Let \(s \neq t\). If \(\sigma(s|t) > 0\) then \(v(s) > v(t)\).

**Pf.** \(v(s) \geq \rho(s) = \pi_t > v(t)\)

\((t\) is not used)
Proof: 0. Preliminaries

Let \((\sigma, \rho)\) be a TL-EQUILIBRIUM with outcome \(\pi\). Then:

- message \(t\) is used in equilibrium:
  \[ \bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t) \]

- message \(t\) is not used in equilibrium:
  \[ \bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t) \]

**Corollary.**

Let \(s \neq t\). If \(\sigma(s|t) > 0\) then \(v(s) > v(t)\).

**Pf.**
\[ v(s) \geq \rho(s) = \pi_t > v(t) \]
Proof: 0. Preliminaries

Let $(\sigma, \rho)$ be a **TL-EQUILIBRIUM** with outcome $\pi$. Then:

- **message $t$ is used** in equilibrium:
  \[ \bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff \nu(t) \geq \pi_t = \rho(t) \]

- **message $t$ is not used** in equilibrium:
  \[ \bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > \nu(t) = \rho(t) \]

**Corollary.**

Let $s \neq t$. If $\sigma(s|t) > 0$ then $\nu(s) > \nu(t)$.

**Pf.** $\nu(s) \geq \rho(s) = \pi_t > \nu(t)$  □

**Note.** May *not* hold for **NON-TL**-equilibria.
Proof: 1. equilibrium $\rightarrow$ mechanism
Proof: 1. equilibrium $\rightarrow$ mechanism

Let $(\sigma, \rho)$ be **TL-EQUILIBRIUM**, with outcome $\pi$. 
Proof: 1. equilibrium → mechanism

Let \((\sigma, \rho)\) be TL-EQUILIBRIUM, with outcome \(\pi\).

- **Special Case:**
  - There is a single message \(s\) that is used (i.e., \(\sigma(s|t) = 1\) for all \(t\)).
Proof: 1. equilibrium $\rightarrow$ mechanism

Let $(\sigma, \rho)$ be TL-EQUILIBRIUM, with outcome $\pi$.

\textbf{Special Case:} \\
There is a single message $s$ that is used (i.e., $\sigma(s|t) = 1$ for all $t$).

$\Rightarrow \quad \pi_t = \rho(s) = v(T) \quad \text{for all } t$
Proof: 1. equilibrium $\rightarrow$ mechanism

Let $(\sigma, \rho)$ be TL-EQUILIBRIUM, with outcome $\pi$.

- **Special Case:**
  
  There is a single message $s$ that is used (i.e., $\sigma(s|t) = 1$ for all $t$).

  \[\Rightarrow \pi_t = \rho(s) = v(T) \quad \text{for all } t\]

  \[\Rightarrow v(t) < v(T) \leq v(s) \quad \text{for all } t \neq s\]
Proof: 1. equilibrium → mechanism

Let \((\sigma, \rho)\) be TL-EQUILIBRIUM, with outcome \(\pi\).

- **Special Case:**
  
  *There is a single message \(s\) that is used (i.e., \(\sigma(s|t) = 1\) for all \(t\)).*

  \[\pi_t = \rho(s) = v(T)\] for all \(t\)
  
  \[v(t) < v(T) \leq v(s)\] for all \(t \neq s\)

  \[\Rightarrow \pi \text{ is the unique OPTIMAL MECHANISM}\]

  **Pf.** \(\pi\) is optimal even if we keep only the (IC) constraints \(\pi_t \geq \pi_s\) for all \(t \neq s\),
Proof: 1. equilibrium $\rightarrow$ mechanism

Let $(\sigma, \rho)$ be TL-EQUILIBRIUM, with outcome $\pi$.

- **Special Case:**
  
  *There is a single message $s$ that is used (i.e., $\sigma(s|t) = 1$ for all $t$).*

  $\Rightarrow \pi_t = \rho(s) = v(T)$ for all $t$

  $\Rightarrow v(t) < v(T) \leq v(s)$ for all $t \neq s$

  $\Rightarrow \pi$ is the unique OPTIMAL MECHANISM

**Pf.** $\pi$ is optimal even if we keep only the (IC) constraints $\pi_t \geq \pi_s$ for all $t \neq s$, because $v(t) < v(T) \leq v(s)$ for all $t \neq s$
Proof: 1. equilibrium $\rightarrow$ mechanism
Proof: 1. equilibrium $\rightarrow$ mechanism

- General Case.
Proof: 1. equilibrium $\rightarrow$ mechanism

- **General Case.** For each message $s$ that is used (i.e., $\bar{\sigma}(s) > 0$)
Proof: 1. equilibrium → mechanism

- **General Case.** For each message $s$ that is used (i.e., $\bar{\sigma}(s) > 0$) apply the Special Case with:
**Proof: 1. equilibrium \( \rightarrow \) mechanism**

- **General Case.** For each message \( s \) that is used (i.e., \( \bar{\sigma}(s) > 0 \)) apply the Special Case with:

  - set of types \( T_s := \{ t : \sigma(s|t) > 0 \} \)
  
  (the types that use message \( s \))
Proof: 1. equilibrium $\rightarrow$ mechanism

- **General Case.** For each message $s$ that is used (i.e., $\bar{\sigma}(s) > 0$) apply the Special Case with:
  - set of types $= T_s := \{ t : \sigma(s|t) > 0 \}$ (the types that use message $s$)
  - probability distribution $= q(s)$ (the posterior given message $s$)
Proof: 1. equilibrium → mechanism

- **General Case.** For each message \( s \) that is used (i.e., \( \bar{\sigma}(s) > 0 \)) apply the Special Case with:

  - set of types \( = T_s := \{ t : \sigma(s|t) > 0 \} \)
  - probability distribution \( = q(s) \)

  \( \Rightarrow \) \( \pi \) restricted to \( T_s \) is the unique OPTIMAL MECHANISM, for each \( s \)
Proof: 1. equilibrium $\rightarrow$ mechanism

- **General Case.** For each message $s$ that is used (i.e., $\bar{\sigma}(s) > 0$) apply the Special Case with:
  - set of types $= T_s := \{t : \sigma(s|t) > 0\}$
    (the types that use message $s$)
  - probability distribution $= q(s)$
    (the posterior given message $s$)
  $\Rightarrow$ $\pi$ restricted to $T_s$ is the unique OPTIMAL MECHANISM, for each $s$
  $\Rightarrow$ $\pi$ is the unique OPTIMAL MECHANISM
Proof: 2. existence
Proof: 2. existence

For every $\varepsilon > 0$, let $\Gamma^\varepsilon$ be the perturbation of the game $\Gamma$: 
Proof: 2. existence

For every $\varepsilon > 0$, let $\Gamma^\varepsilon$ be the perturbation of the GAME $\Gamma$:

$$U^A = x + \varepsilon 1_{s=t}$$

(revealing the whole truth increases agent’s payoff by $\varepsilon$)
Proof: 2. existence

For every $\varepsilon > 0$, let $\Gamma^{\varepsilon}$ be the perturbation of the game $\Gamma$:

- $U^A = x + \varepsilon 1_{s=t}$
  (revealing the whole truth increases agent’s payoff by $\varepsilon$)

- $\sigma(t|t) \geq \varepsilon$
  (probability of revealing the whole truth is at least $\varepsilon$)
Proof: 2. existence
Proof: 2. existence

Proposition. $\Gamma^\varepsilon$ has a Nash equilibrium.
Proposition. $\Gamma^\varepsilon$ has a Nash equilibrium.

Proof. Standard use of Kakutani’s Fixed Point Theorem.
Proof: 2. existence

- **Proposition.** $\Gamma^\varepsilon$ has a Nash equilibrium.

**Proof.** Standard use of Kakutani’s Fixed Point Theorem.

- **Proposition.** A limit point $(\sigma, \rho)$ of Nash equilibria $(\sigma^\varepsilon, \rho^\varepsilon)$ of $\Gamma^\varepsilon$ yields a **TL-EQUILIBRIUM** $(\sigma', \rho)$ of $\Gamma$. 
Proof: 2. existence

Proposition. $\Gamma^\varepsilon$ has a Nash equilibrium.

Proof. Standard use of Kakutani’s Fixed Point Theorem.

Proposition. A limit point $(\sigma, \rho)$ of Nash equilibria $(\sigma^\varepsilon, \rho^\varepsilon)$ of $\Gamma^\varepsilon$ yields a TL-EQUILIBRIUM $(\sigma', \rho)$ of $\Gamma$.

Proof.
Proposition. $\Gamma^\varepsilon$ has a Nash equilibrium.


Proposition. A limit point $(\sigma, \rho)$ of Nash equilibria $(\sigma^\varepsilon, \rho^\varepsilon)$ of $\Gamma^\varepsilon$ yields a TL-EQUILIBRIUM $(\sigma', \rho)$ of $\Gamma$.

Proof.

$\sigma(t|t) < 1 \implies$
Proof: 2. existence

- **Proposition.** $\Gamma^\varepsilon$ has a Nash equilibrium.

**Proof.** Standard use of Kakutani’s Fixed Point Theorem.

- **Proposition.** A limit point $(\sigma, \rho)$ of Nash equilibria $(\sigma^\varepsilon, \rho^\varepsilon)$ of $\Gamma^\varepsilon$ yields a **TL-EQUILIBRIUM** $(\sigma', \rho)$ of $\Gamma$.

**Proof.**

- $\sigma(t|t) < 1 \Rightarrow \sigma(t|r) = 0$ for all $r$
Proof: 2. existence

Proposition. $\Gamma^\varepsilon$ has a Nash equilibrium.

Proof. Standard use of Kakutani’s Fixed Point Theorem.

Proposition. A limit point $(\sigma, \rho)$ of Nash equilibria $(\sigma^\varepsilon, \rho^\varepsilon)$ of $\Gamma^\varepsilon$ yields a TL-EQUILIBRIUM $(\sigma', \rho)$ of $\Gamma$.

Proof.

- $\sigma(t|t) < 1 \implies \sigma(t|r) = 0$ for all $r$
  \[ \implies q(t) = 1_t \]
Proof: 2. existence

**Proposition.** \( \Gamma^\varepsilon \) has a Nash equilibrium.

**Proof.** Standard use of Kakutani’s Fixed Point Theorem.

**Proposition.** A limit point \((\sigma, \rho)\) of Nash equilibria \((\sigma^\varepsilon, \rho^\varepsilon)\) of \(\Gamma^\varepsilon\) yields a TL-EQUILIBRIUM \((\sigma', \rho)\) of \(\Gamma\).

**Proof.**

- \( \sigma(t|t) < 1 \) \( \Rightarrow \) \( \sigma(t|r) = 0 \) for all \( r \)
- \( \Rightarrow \) \( q(t) = 1_t \)
- \( \Rightarrow \) \( \rho(t) = v(t) \)
Proof: 2. existence

**Proposition.** $\Gamma^\varepsilon$ has a Nash equilibrium.

**Proof.** Standard use of Kakutani’s Fixed Point Theorem.

**Proposition.** A limit point $(\sigma, \rho)$ of Nash equilibria $(\sigma^\varepsilon, \rho^\varepsilon)$ of $\Gamma^\varepsilon$ yields a TL-EQUILIBRIUM $(\sigma', \rho)$ of $\Gamma$.

**Proof.**

- $\sigma(t|t) < 1 \Rightarrow \sigma(t|r) = 0$ for all $r$
  $\Rightarrow q(t) = 1_t$
  $\Rightarrow \rho(t) = v(t)$

- If $t \in BR^A(t)$ then put $\sigma'(t|t) = 1$
Proof: 2’. mechanism $\rightarrow$ equilibrium
Proof: 2’. mechanism $\rightarrow$ equilibrium

Let $\pi$ be an OPTIMAL MECHANISM outcome.
Proof: 2’. mechanism $\rightarrow$ equilibrium

Let $\pi$ be an **OPTIMAL MECHANISM** outcome.
We will construct a **TL-EQUILIBRIUM** $(\sigma, \rho)$ with outcome $\pi$. 
Proof: 2’. mechanism $\rightarrow$ equilibrium
Proof: 2’. mechanism $\rightarrow$ equilibrium

Recall that in a **TL-EQUILIBRIUM** we have

$$\bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t)$$

$$\bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t)$$
Recall that in a **TL-EQUILIBRIUM** we have

\[
\bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t)
\]

\[
\bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t)
\]

Principal’s strategy: put

\[
\rho(t) = \min\{\pi_t, v(t)\} \quad \text{for each } t
\]
Proof: 2’. mechanism → equilibrium

Recall that in a **TL-EQUILIBRIUM** we have

\[
\bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t)
\]

\[
\bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t)
\]

Agent’s strategy:
Proof: 2’. mechanism → equilibrium

Recall that in a **TL-EQUILIBRIUM** we have

\[
\bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t)
\]

\[
\bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t)
\]

Agent’s strategy:

Let \( S := \{ t : v(t) \geq \pi_t \} \) – these are the messages that will be used in equilibrium
Recall that in a **TL-EQUILIBRIUM** we have

\[
\bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t)
\]

\[
\bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t)
\]

**Agent’s strategy:**

- Let \( S := \{ t : v(t) \geq \pi_t \} \) — these are the messages that will be used in equilibrium

- If \( t \in S \) then put \( \sigma(t|t) = 1 \)
Proof: 2’. mechanism → equilibrium

Recall that in a **TL-EQUILIBRIUM** we have
\[ \bar{\sigma}(t) > 0 \iff \sigma(t|t) = 1 \iff v(t) \geq \pi_t = \rho(t) \]
\[ \bar{\sigma}(t) = 0 \iff \sigma(t|t) = 0 \iff \pi_t > v(t) = \rho(t) \]

Agent’s strategy:

- Let \( S := \{ t : v(t) \geq \pi_t \} \) – these are the messages that will be used in equilibrium

- If \( t \in S \) then put \( \sigma(t|t) = 1 \)

- If \( t \not\in S \) then put \( \sigma(t|t) = 0 \)

We need to determine which messages (in \( S \)) type \( t \not\in S \) will choose
Agent’s strategy:

- $S := \{ t : \nu(t) \geq \pi_t \}$ (messages used)
- We need to determine which messages in $S$ types $t \notin S$ will use
Proof: 2’. mechanism $\rightarrow$ equilibrium

Agent’s strategy:

- $S := \{ t : v(t) \geq \pi_t \}$ (messages used)
- We need to determine which messages in $S$ types $t \not\in S$ will use
- For each $s \in S$ put
  
  \[ R_s := \{ t \not\in S : s \in L(t), \pi_t = \pi_s \} \cup \{ s \} \]
  
  (set of types that may use message $s$)
Proof: 2’. mechanism $\rightarrow$ equilibrium

Agent’s strategy:

- $S := \{t : v(t) \geq \pi_t \}$ (messages used)
- We need to determine which messages in $S$ types $t \notin S$ will use
- For each $s \in S$ put
  
  $R_s := \{t \notin S : s \in L(t), \pi_t = \pi_s\} \cup \{s\}$
  (set of types that may use message $s$)
- A simple case: $S$ is a singleton
Proof: 2’. mechanism $\rightarrow$ equilibrium

Agent’s strategy:

- $S := \{ t : v(t) \geq \pi_t \}$ (messages used)
- We need to determine which messages in $S$ types $t \notin S$ will use
- For each $s \in S$ put $R_s := \{ t \notin S : s \in L(t), \pi_t = \pi_s \} \cup \{ s \}$ (set of types that may use message $s$)
- A simple case: $S$ is a singleton
- The general case: Partition $T$ into disjoint sets $Q_s \subseteq R_s$ such that $v(Q_s) = \pi_s$ for every $s \in S$
Hall’s Marriage Theorem
Hall’s Marriage Theorem

A set $B$ of $n$ boys, and a set $G$ of $n$ girls
Hall’s Marriage Theorem

- A set $B$ of $n$ boys, and a set $G$ of $n$ girls
- Each boy $b \in B$ knows a subset $G_b \subseteq G$ of girls
Hall’s Marriage Theorem

- A set \( B \) of \( n \) boys, and a set \( G \) of \( n \) girls
- Each boy \( b \in B \) knows a subset \( G_b \subseteq G \) of girls
- *Matching*: one-to-one pairing of boys with girls
Hall’s Marriage Theorem

- A set $B$ of $n$ boys, and a set $G$ of $n$ girls
- Each boy $b \in B$ knows a subset $G_b \subseteq G$ of girls
- *Matching*: one-to-one pairing of boys with girls
- *Necessary* condition for a matching to exist:
Hall’s Marriage Theorem

A set $B$ of $n$ boys, and a set $G$ of $n$ girls

Each boy $b \in B$ knows a subset $G_b \subseteq G$ of girls

Matching: one-to-one pairing of boys with girls

Necessary condition for a matching to exist:

Every set of $k$ boys knows at least $k$ girls

$$\left| \bigcup_{b \in C} G_b \right| \geq |C| \text{ for every } C \subseteq B$$
Hall’s Marriage Theorem

- A set $B$ of $n$ boys, and a set $G$ of $n$ girls
- Each boy $b \in B$ knows a subset $G_b \subseteq G$ of girls
- **Matching**: one-to-one pairing of boys with girls
- **Necessary** condition for a matching to exist:
  
  Every set of $k$ boys knows at least $k$ girls
  
  \[ |\bigcup_{b \in C} G_b| \geq |C| \text{ for every } C \subseteq B \]

- **Theorem** (Hall 1935). The condition is also **sufficient**.
The Hull of Hall’s Theorem
Finite sets $B$ and $G$
Finite sets $B$ and $G$

For each $b \in B$ a subset $G_b \subseteq G$
The Hull of Hall’s Theorem

- Finite sets $B$ and $G$
- For each $b \in B$ a subset $G_b \subseteq G$
- Measures $\beta$ on $B$ and $\gamma$ on $G$ such that...
The Hull of Hall’s Theorem

- Finite sets $B$ and $G$
- For each $b \in B$ a subset $G_b \subseteq G$
- Measures $\beta$ on $B$ and $\gamma$ on $G$ such that
  $$\gamma(\bigcup_{b \in C} G_b) \geq \beta(C)$$
  for every $C \subseteq B$
  with equality for $C = B$
The Hull of Hall’s Theorem

- Finite sets $B$ and $G$
- For each $b \in B$ a subset $G_b \subseteq G$
- Measures $\beta$ on $B$ and $\gamma$ on $G$ such that $\gamma(\bigcup_{b \in C} G_b) \geq \beta(C)$ for every $C \subseteq B$ with equality for $C = B$

**Theorem.** There exists a partition of $\bigcup_{b \in B} G_b$ into disjoint sets $(F_b)_{b \in B}$ such that for every $b \in B$
The Hull of Hall’s Theorem

- Finite sets $B$ and $G$
- For each $b \in B$ a subset $G_b \subseteq G$
- Measures $\beta$ on $B$ and $\gamma$ on $G$ such that $\gamma(\bigcup_{b \in C} G_b) \geq \beta(C)$ for every $C \subseteq B$ with equality for $C = B$

**Theorem.** There exists a partition of $\bigcup_{b \in B} G_b$ into disjoint sets $(F_b)_{b \in B}$ such that for every $b \in B$

- $F_b \subseteq G_b$
Finite sets $B$ and $G$

For each $b \in B$ a subset $G_b \subseteq G$

Measures $\beta$ on $B$ and $\gamma$ on $G$ such that

$\gamma(\bigcup_{b \in C} G_b) \geq \beta(C)$ for every $C \subseteq B$

with equality for $C = B$

**Theorem.** There exists a partition of $\bigcup_{b \in B} G_b$ into disjoint sets $(F_b)_{b \in B}$ such that for every $b \in B$

$F_b \subseteq G_b$

$\gamma(F_b) = \beta(\{b\})$
Finite sets $B$ and $G$

For each $b \in B$ a subset $G_b \subseteq G$

Measures $\beta$ on $B$ and $\gamma$ on $G$ such that

$$\gamma(\bigcup_{b \in C} G_b) \geq \beta(C)$$
for every $C \subseteq B$ with equality for $C = B$

**Theorem.** There exists a partition of $\bigcup_{b \in B} G_b$ into disjoint fractional sets $(F_b)_{b \in B}$ such that for every $b \in B$

- $F_b \subseteq G_b$
- $\gamma(F_b) = \beta(\{b\})$
Finite sets \( B \) and \( G \)

For each \( b \in B \) a subset \( G_b \subseteq G \)

Measures \( \beta \) on \( B \) and \( \gamma \) on \( G \) such that

\[
\gamma(\bigcup_{b \in C} G_b) \geq \beta(C) \quad \text{for every} \quad C \subseteq B \\
\text{with equality for} \quad C = B
\]

**Theorem.** There exists a partition of \( \bigcup_{b \in B} G_b \) into disjoint fractional sets \((F_b)_{b \in B}\) such that for every \( b \in B \)

\[
F_b \subseteq G_b \\
\gamma(F_b) = \beta(\{b\})
\]

**Proof.** Hart–Kohlberg 74
Proof: 2’. mechanism $\rightarrow$ equilibrium

Agent’s strategy:

- $S := \{ t : v(t) \geq \pi_t \}$ (messages used)
- We need to determine which messages in $S$ types $t \notin S$ will use
- For each $s \in S$ put
  $R_s := \{ t \notin S : s \in L(t), \pi_t = \pi_s \} \cup \{ s \}$
  (set of types that may use message $s$)
- Partition $T$ into disjoint sets $Q_s \subseteq R_s$ such that $v(Q_s) = \pi_s$ for every $s \in S$
Proof: 2’. mechanism $\rightarrow$ equilibrium

Agent’s strategy:

- $S := \{ t : v(t) \geq \pi_t \}$ (messages used)

- We need to determine which messages in $S$ types $t \notin S$ will use

- For each $s \in S$ put
  
  $R_s := \{ t \notin S : s \in L(t), \pi_t = \pi_s \} \cup \{ s \}$

  (set of types that may use message $s$)

- Partition $T$ into disjoint fractional sets $Q_s \subseteq R_s$ such that $v(Q_s) = \pi_s$ for every $s \in S$
Proof: 2’. mechanism → equilibrium

Agent’s strategy:

- \( S := \{ t : v(t) \geq \pi_t \} \) (messages used)
- We need to determine which messages in \( S \) types \( t \notin S \) will use
- For each \( s \in S \) put
  \[ R_s := \{ t \notin S : s \in L(t), \pi_t = \pi_s \} \cup \{ s \} \]
  (set of types that may use message \( s \))
- Partition \( T \) into disjoint fractional sets \( Q_s \subseteq R_s \) such that \( v(Q_s) = \pi_s \) for every \( s \in S \)
  \[ \iff \] the strategy \( \sigma \)

- **Messages**: arbitrary
Messages: arbitrary $\Rightarrow$ no "truth" structure

- **Messages**: arbitrary $\Rightarrow$ no "truth" structure
- **Result**: 
  \[
  \{\text{optimal mechanisms}\} \subseteq \{\text{equilibria}\}
  \]

- **Messages**: arbitrary $\Rightarrow$ no "truth" structure
- **Result**: 
  \[ \{ \text{optimal mechanisms} \} \subseteq \{ \text{equilibria} \} \]
- **Rewards**: mixtures of two outcomes

- **Messages**: arbitrary $\Rightarrow$ no "truth" structure

- **Result**: 
  \[
  \{\text{optimal mechanisms}\} \subseteq \{\text{equilibria}\}
  \]

- **Rewards**: mixtures of two outcomes 
  $\Rightarrow$ linear $h_t$

- **Messages**: arbitrary $\Rightarrow$ no "truth" structure
- **Result**:
  \[ \{ \text{optimal mechanisms} \} \subseteq \{ \text{equilibria} \} \]
- **Rewards**: mixtures of two outcomes
  $\Rightarrow$ linear $h_t$
  - Extended to *concave* $h_t$: Sher (2011)

- **Messages**: arbitrary $\Rightarrow$ no "truth" structure
- **Result**: 
  \[ \{ \text{optimal mechanisms} \} \subseteq \{ \text{equilibria} \} \]
- **Rewards**: mixtures of two outcomes $\Rightarrow$ linear $h_t$
- **Extended to concave $h_t$**: Sher (2011)
- **General condition**: 
  "**Generalized Single-Peakedness**"

- **Messages**: arbitrary $\Rightarrow$ no "truth" structure
- **Result**: 
  \[ \{ \text{optimal mechanisms} \} \subseteq \{ \text{equilibria} \} \]
- **Rewards**: mixtures of two outcomes $\Rightarrow$ linear $h_t$
  - Extended to *concave* $h_t$: Sher (2011)
- **General condition**: 
  "**Generalized Single-Peakedness**" $\iff$ "**Principal’s Uniform Best**" 
  (includes convex $h_t$, ...)

Sergiu Hart © 2015 – p. 62
EVIDENCE GAMES model very common setups
EVIDENCE GAMES model very common setups

In EVIDENCE GAMES there is EQUIVALENCE between EQUILIBRIUM (without commitment) and OPTIMAL MECHANISM (with commitment)
EVIDENCE GAMES model very common setups

In EVIDENCE GAMES there is EQUIVALENCE between EQUILIBRIUM (without commitment) and OPTIMAL MECHANISM (with commitment)

EQUILIBRIUM is CONSTRAINED EFFICIENT (in the canonical case)
EVIDENCE GAMES model very common setups

In EVIDENCE GAMES there is EQUIVALENCE between EQUILIBRIUM (without commitment) and OPTIMAL MECHANISM (with commitment)

EQUILIBRIUM is CONSTRAINED EFFICIENT (in the canonical case)

The conditions of EVIDENCE GAMES are indispensable for this EQUIVALENCE
And That Is The Whole Truth ...
"Do you swear to tell the truth, the whole truth, and nothing but the truth in the most entertaining way possible?"