Signaling Games and Forward Induction

Songzi Du

UCSD
Econ 200C

March 31, 2023
Signaling Game

- Two players, sender (player 1) and receiver (player 2)
- Nature picks \( t \in T \) type of sender. \( p(t) \) is probability that type is \( t \).
- Sender observes \( t \), selects signal \( s \in S \).
Two players, sender (player 1) and receiver (player 2)

Nature picks $t \in T$ type of sender. $p(t)$ is probability that type is $t$.

Sender observes $t$, selects signal $s \in S$.

Receiver observes $s$ (but not $t$), selects action $a$.

$U_i(a, t, s)$ payoff function.
Beer-quiche game

"Real men don't eat quiche"

```
Beer  Quiche  Beer  Quiche
  1       2       1       2
  0  2   -1  0    1  3   1  0
  1  3   0  2
  1  0
```

```
Strong

9/10

Weak

1/10

Chance

1

1

Beer  Quiche

Fight  Retreat

Fight  Retreat

0  2   -1  0    1  3   1  0

\[ \text{Cho, Kreps, and AJE} \]
Labor market signaling (Spence)

- $S$ is worker, $t \in \mathbb{R}_+$ is ability, $s \in \mathbb{R}_+$ is education
- $R$ is market, $a \in \mathbb{R}_+$ is market wage
- $S$ preferences: $U_1(a, t, s) = a - c(s, t)$
Labor market signaling (Spence)

- $S$ is worker, $t \in \mathbb{R}_+$ is ability, $s \in \mathbb{R}_+$ is education
- $R$ is market, $a \in \mathbb{R}_+$ is market wage

$S$ preferences: $U_1(a, t, s) = a - c(s, t)$

$R$ preferences: $U_2(a, t, s) = -(a - t)^2$ \quad \Rightarrow \quad a_t = \mathbb{E}[t | s]

- $R$ should be thought of as firms who are competing (in a Bertrand fashion) to hire the worker.
- Each firm gets a payoff $t - a$ if it hires the worker at wage $a$. 
Other signaling models

1. Verifiable information: $S = \mathcal{P}(T)$, $U_1(a, t, s) = -\infty$ if $t \notin s$.

2. Cheap-talk: $U_i$ independent of $s$. 
Perfect Bayesian Equilibrium (for finite signaling game)

- Sender’s strategy: \( \sigma : T \rightarrow \Delta(S) \)  
- Receiver’s strategy: \( \alpha : S \rightarrow \Delta(A) \)  
- Receiver’s belief (assessment): \( \mu(\cdot | s) \in \Delta(T) \), \( \forall s \in S \).
Perfect Bayesian Equilibrium (for finite signaling game)

- Sender’s strategy: \( \sigma : T \to \Delta(S) \).
- Receiver’s strategy: \( \alpha : S \to \Delta(A) \).
- Receiver’s belief (assessment): \( \mu(\cdot | s) \in \Delta(T), \forall s \in S \).

\((\sigma^*, \alpha^*, \mu^*)\) is a **perfect Bayesian equilibrium** (PBE) if:

1. \( \sigma^*(t) \) solves
   \[
   \max_{s \in S} U_1(\alpha^*(s), t, s)
   \]
   for all \( t \),

2. \( \alpha^*(s) \) solves
   \[
   \max_{a \in A} \sum_{t \in T} \mu^*(t | s) U_2(a, t, s)
   \]
   for all \( s \),

3. \( \mu^* \) derives from prior and \( \sigma^* \) using Bayes’s Rule whenever possible,
   \[
   \frac{\mu^*(t | s)}{\sum_{t' \in T} p(t') \sigma^*(t')(s)} = \frac{p(t) \sigma^*(t)(s) \mathbb{P}(t, s)}{\sum_{t' \in T} p(t') \sigma^*(t')(s)}.
   \]
Perfect Bayesian Equilibrium (based on appraisals)

- Receiver’s belief (appraisal) about sender’s strategy: 
  \( \tilde{\mu}(\cdot | s) \in \Delta(Z(s)) \), where \( Z(s) \) is the set of pure strategies that sends \( s \) for some type \( (z(t) = s \text{ for some } t) \).

\[
\begin{align*}
\text{Songzi Du (UCSD Econ 200C)} & \quad \text{Signaling Games} \\
\text{March 31, 2023} & \quad 7 / 19
\end{align*}
\]
Perfect Bayesian Equilibrium (based on appraisals)

- Receiver’s belief (appraisal) about sender’s strategy: 
  \( \tilde{\mu}(\cdot | s) \in \Delta(Z(s)) \), where \( Z(s) \) is the set of pure strategies that sends \( s \) for some type \( (z(t) = s \text{ for some } t) \).

\( (\sigma^*, \alpha^*, \tilde{\mu}^*) \) is a perfect Bayesian equilibrium (PBE) if:

1. \( \sigma^*(t) \) solves

\[
\max_{s \in S} U_1(\alpha^*(s), t, s)
\]

for all \( t \),

2. \( \alpha^*(s) \) solves

\[
\max_{a \in A} \sum_{z \in Z(s)} \tilde{\mu}^*(z | s) \frac{\sum_{t \in T} p(t) z(t)(s) U_2(a, t, s)}{\sum_{t' \in T} p(t') z(t')(s)}
\]

for all \( s \),

\[ P(t | s, z) \]
Perfect Bayesian Equilibrium (based on appraisals)

- Receiver’s belief (appraisal) about sender’s strategy: 
  \( \tilde{\mu}(\cdot | s) \in \Delta(Z(s)) \), where \( Z(s) \) is the set of pure strategies that sends \( s \) for some type \( (z(t) = s \) for some \( t \)).

\((\sigma^*, \alpha^*, \tilde{\mu}^*)\) is a perfect Bayesian equilibrium (PBE) if:

1. \( \sigma^*(t) \) solves

\[
\max_{s \in S} \ U_1(\alpha^*(s), t, s)
\]

for all \( t \),

2. \( \alpha^*(s) \) solves

\[
\max_{a \in A} \sum_{z \in Z(s)} \tilde{\mu}^*(z | s) \frac{\sum_{t \in T} p(t) z(t)(s) U_2(a, t, s)}{\sum_{t' \in T} p(t') z(t')(s)}
\]

for all \( s \),

3. \( \tilde{\mu}^* \) derives from prior and \( \sigma^* \) using Bayes’s Rule whenever possible,

\[
\tilde{\mu}^*(z | s) = \frac{\left(\prod_{t \in T} \sigma^*(t)(z(t))\right) \left(\sum_{t \in T} p(t) z(t)(s)\right)}{\sum_{t \in T} p(t) \sigma^*(t)(s) \mathbb{P}(s | z)}.
\]
Example in Beer-Quiche game

- Suppose $\sigma^*(S)(B) = 0.9$ and $\sigma^*(W)(B) = 0.3$. What is $\tilde{\mu}(\cdot | B)$?

Equivalent mixed strategy:

$\sigma^*(BB) = 0.9 \cdot 0.3 = 0.27$

$\sigma^*(BQ) = 0.9 \cdot 0.7 = 0.63$

$\sigma^*(QB) = 0.1 \cdot 0.3 = 0.03$

$\sigma^*(QQ) = 0.1 \cdot 0.7 = 0.07$

$$
\tilde{\mu}(BB | B) = \frac{0.27 \cdot P(B \mid BB)}{P(B)} = \frac{0.27 \cdot 1}{0.9 \cdot 0.9 + 0.1 \cdot 0.3} = \frac{0.27}{0.84}
$$
\[ \hat{m}(BQ | B) = \frac{0.63 \cdot P(B | BQ)}{0.94} = \frac{0.63 \cdot 0.9}{0.84} \]

\[ \hat{m}(QB | B) = \frac{0.03 \cdot 0.1}{0.94} \]

\[ \hat{m}(QA | B) = 0 \]

\[ \hat{m}(C | B) \in \Delta(2(B)) \]

\[ Z(B) = \{ BB, BQ, QB \} \]

\[ m(S | B) = \frac{0.63 \cdot 0.9}{0.94} \cdot \frac{P(S | BQ, B)}{1} + \frac{0.27}{0.84} \cdot \frac{P(S | BB, B)}{0.9} + \frac{0.03 \cdot 0.1}{0.94} \cdot \frac{P(S | QB, B)}{0.9} \]

\[ = \frac{P(S, B)}{P(B)} = \frac{0.9 \cdot 0.9}{0.84} \]
Terminology

- $s$ is on the equilibrium path if $\sigma^*(t)(s) > 0$ for some $t$.

- Equilibrium outcome $\pi \in \Delta(T \times S \times A)$, induced by some PBE.

$$
\pi(t, s, \alpha) = p(t) \cdot \sigma^*(t)(s) \cdot \alpha^*(s)(\alpha)
$$

There are also “hybrid” equilibria in which some types pool and some types separate.
Terminology

- $s$ is on the equilibrium path if $\sigma^*(t)(s) > 0$ for some $t$.

- Equilibrium outcome $\pi \in \Delta(T \times S \times A)$, induced by some PBE.

- Pooling equilibrium: $\sigma^*(t)$ constant in $t$
  - $\mu^*(\cdot | s) = p(\cdot)$ whenever $s$ is on the equilibrium path

- Separating equilibrium: the support of $\sigma^*(t)$ does not intersect the support of $\sigma^*(t')$ if $t \neq t'$
  - $\mu^*(t | s) = 1$ if $\sigma^*(t)(s) > 0$

- There are also “hybrid” equilibria in which some types pool and some types separate.
Beer-quiche game

BQ is not a part of a PBE
PBE#1 in beer-quiche game

\[ \sigma^* = BB \]

\[ \mu^*(BQ|Q) = 1 \]
\[ \Rightarrow \mu^*(S|Q) = 0 \]

\[ \mu^*(BB|B) = 1 \]

satisfies no strictly dominated strategy refinement
PBE#2 in beer-quiche game  \( \sigma^* = QA \)

This PBE does not survive "no strictly dominated strategy" refinement.
### Beer-quiche game, normal form

<table>
<thead>
<tr>
<th></th>
<th>FF</th>
<th>FR</th>
<th>RF</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>-4/5</td>
<td>-4/5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>9/10</td>
<td>9/10</td>
<td>29/10</td>
<td>29/10</td>
</tr>
<tr>
<td>BQ</td>
<td>-4/5</td>
<td>-9/10</td>
<td>1/10</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6/5</td>
<td>14/5</td>
<td>3</td>
</tr>
<tr>
<td>QB</td>
<td>-4/5</td>
<td>1/10</td>
<td>-9/10</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>9/5</td>
<td>1/5</td>
<td>2</td>
</tr>
<tr>
<td>QQ</td>
<td>-4/5</td>
<td>0</td>
<td>-4/5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1/10</td>
<td>21/10</td>
<td>1/10</td>
<td>21/10</td>
</tr>
</tbody>
</table>

\[
\frac{3}{4} \text{QQ} + \frac{1}{4} \text{BQ}
\]

strictly dominates

\[
\text{A B}
\]
A “no strictly dominated strategy” refinement of PBE

- For every off-equilibrium signal $s$, receiver’s appraisal $\tilde{\mu}(\cdot | s)$ places probability 0 on strictly dominated strategies whenever possible (i.e., when there is a strategy in $Z(s)$ that is not strictly dominated).

\[
\text{Every strategy in } Z(s) \text{ is strictly dominated}
\]

\[
\Rightarrow \text{ anything goes for } \tilde{\mu}(\cdot | s)
\]
$(QA, FR, \tilde{\mu}^*)$ is a PBE

$\tilde{\mu}^*(QB \mid B) = 0$, QB is strictly dominated

$\tilde{\mu}^*(BB \mid B) + \tilde{\mu}^*(BA \mid B) = 1$

$\downarrow$

$\tilde{\mu}^*(s \mid B) = 0.9$

$\downarrow$

$\tilde{\mu}^*(s \mid B) = 1$

$\Rightarrow \tilde{\mu}^*(s \mid B) \geq 0.9 \Rightarrow$ Retreat after observing B

$\Rightarrow$ Contradicts S type choosing A.

So $(QA, FR, \tilde{\mu}^*)$ cannot satisfy "no strictly dominated strategy refinement."
PBE in labor market signaling (two types)

- Suppose $\frac{\partial c}{\partial s} > 0$ and $\frac{\partial^2 c}{\partial t \partial s} < 0$.  

$$t_1 < t_2$$

$$\alpha(s) = E[t | s]$$

$$u_1(t, s, a) = a - c(s, t)$$
\[ a(s) - c(s, t_1) = \text{Constant} \]

\[ \frac{\partial h}{\partial s}(s) - \frac{\partial c}{\partial s}(s, t_1) = 0 \]

Example: \[ c(s, t) = \frac{s^2}{t} \]
Separating PBE:

\[ o^*(t_1) = s_1, \quad o^*(t_2) = s_2 \]
\[ s_1 \neq s_2 \]

\[ \forall s, \quad \alpha^*(s) \in [t_1, t_2] \]
\[ \alpha^*(s_1) = t_1, \quad \alpha^*(s_2) = t_2 \]

\[ s_1 > 0, \quad \alpha^*(s_1) = t_1 \implies \text{profitable deviation to } \]
\[ s_1 = 0 \]

\[ \implies s_1 = 0 \]

\[ t_1 - c(0, t_1) \geq t_2 - c(s_2, t_1) \]
\[ t_2 - c(s_2, t_2) \geq t_1 - c(0, t_2) \]
\[ c(s_2, t_2) - c(0, t_2) \leq t_2 - t_1 \leq c(s_2, t_1) - c(0, t_1) \]
$\sigma^*(t_1) = 0 \, , \, \sigma^*(t_2) = S_2$

$\alpha^*(s) = \left\{ \begin{array}{ll}
t_1 & \leq S_2 \\
t_2 & S_2 \leq s 
\end{array} \right. $

**Pooling PBE**

$\sigma^*(t_1) = \sigma^*(t_2) = S_0$

$\alpha^*(S_0) = pt_2 + (1-p)t_1$

\[ pt_2 + (1-p)t_1 - c(s_0, t_1) \geq t_1 - c(0, t_1) \]

\[ \Rightarrow \, pt_2 + (1-p)t_1 - c(s_0, t_2) \geq t_1 - c(0, t_2) \]
\[ \alpha^*(s) = \begin{cases} 0 & s = S_0 \\ \theta_1 & s \neq S_0 \end{cases} \]

\[ \alpha^*(s) = \begin{cases} \theta_1 & s < S_0 \\ pt_2 + (1-p)t_1 & s \geq S_0 \end{cases} \]
Forward Induction (Govindan and Wilson, 2009)

- Fix a (PBE) equilibrium outcome $\pi \in \Delta(T \times S \times A)$.

- A pure strategy $z : T \rightarrow S$ is relevant for $\pi$ if it is a best response to $\alpha^*$ of a PBE $(\sigma^*, \alpha^*, \tilde{\mu}^*)$ that induces $\pi$.

- A signal $s \in S$ is relevant for $\pi$ if $z(t) = s$ for some relevant $z$ and type $t$. 
Forward Induction (Govindan and Wilson, 2009)

- Fix a (PBE) equilibrium outcome \( \pi \in \Delta(T \times S \times A) \).

- A pure strategy \( z : T \rightarrow S \) is relevant for \( \pi \) if it is a best response to \( \alpha^* \) of a PBE \((\sigma^*, \alpha^*, \tilde{\mu}^*)\) that induces \( \pi \).

- A signal \( s \in S \) is relevant for \( \pi \) if \( z(t) = s \) for some relevant \( z \) and type \( t \).

If \( s \) is not \( \pi \)-relevant, then every strategy in \( Z(s) \) is not \( \pi \)-relevant.

**Definition**

An equilibrium outcome \( \pi \) satisfies **forward induction** if it is induced by a PBE \((\sigma^*, \alpha^*, \tilde{\mu}^*)\) such that at all \( \pi \)-relevant signals \( s \), \( \tilde{\mu}^*(\cdot|s) \) places probability one on \( \pi \)-relevant strategies.

- Forward induction implies intuitive criterion, D1, D2, etc.
Intuition of forward induction

An equilibrium theory of off-equilibrium signals

- At a PBE, receiver tries to rationalize an off-equilibrium \( s \), believing that \( s \) is sender’s best response to another PBE with the same outcome.

- Receiver thinks that:
  - Sender is confused about which PBE is in effect, which is understandable since the two PBE’s result in the same outcome.
  - In other words, Sender is confused about receiver’s off-equilibrium play.
Forward induction in beer-quiche game

\[ \pi : S - B, W - B, \]

\[ B - R \]

\[ \pi(S, B, R) = 0.9 \]

\[ \pi(W, B, R) = 0.1 \]

\((\text{BB}; \text{RF})\) is a PBE

\((\text{BB}; R, 0.5F + 0.5R)\) is also a PBE

\(BQ\) is a best response to \((R, 0.5F + 0.5R)\)
\((BB, BF, \overline{\pi}(BB|B) = 1, \overline{\pi}(BB|Q) = 1)\)

\(\rightarrow\) is a PBE.

\(\downarrow\)

\(satisfies\ \text{forward\ induction.}\)

\(\Rightarrow \Pi\ \text{satisfies forward induction}\)

\( (S-Q, W-Q, Q-K) \) does not satisfy forward induction.
Forward induction in job market signaling (two types)

Pooling equilibrium outcome:

\[ \Pi : (S_0, S_0) = o^*, \ \alpha^*(S_0) = pt_2 + (1-p)t_1 \]

Consider \( s' > s_0 \), slightly

Claim: \( s' \) is a \( \Pi \)-relevant signal, \( (S_0, s') \) is \( \Pi \)-relevant.

Proof: Pick a PBE \( (S_0, s_0), \alpha^* \)
\( \alpha'(s) = \begin{cases} \alpha^i(s) & \text{if } s \neq s' \\
 \ t' & \text{if } s = s' \end{cases} \)

\[ pt_2 + (1-p)t_1 - c(s_0, t_2) = t' - c(s', t_2) \]

\[ pt_2 + (1-p)t_1 - c(s_0, t_1) > t' - c(s', t_1) \]

\[ \left( c(s', t_1) - c(s_0, t_1) \right) > \left( c(s', t_2) - c(s_0, t_2) \right) \]

\[ \int_{s_0}^{s'} \frac{\partial c}{\partial s}(s, t_1) ds > \int_{s_0}^{s'} \frac{\partial c}{\partial s}(s, t_2) ds \]

\(( s_0, s'_0, \alpha' \) is a PBE.

\(( s_0, s'_1 \) is a best response to \( \alpha' \).

\( s_0 ( s_0, s'_1 \) is \( \alpha \)-relevant. \( \square \)
Claim: \((s', s'')\) is not \(\pi\)-relevant for any \(s''\).

Proof: Suppose \((s', s'')\) is \(\pi\)-relevant, i.e., \(\exists \alpha' \text{ s.t. } ((s_0, s_0), \alpha')\) is a PBE, and \((s', s'')\) is optimal given \(\alpha'\).

Then \(t_0\) is indifferent between \(s_0\) and \(s'\)

\[ \Rightarrow t_2 \text{ strictly prefers } s' \text{ to } s_0 \]

\[ \Rightarrow \text{contradiction of } ((s_0, s_0), \alpha') \text{ being PBE.} \]
Suppose \( \Pi \) satisfies FL, i.e.,

\[ \text{PBE} ( (s_0, s_0), \alpha^*, \mathcal{M}^* ) , \]

and \( \mathcal{M}^* (s', s'') | s' \) = 0 \( \forall s'' \)

\[ \Rightarrow \mathcal{M}^* (t_2 | s') = 1 \]

\[ \Rightarrow \alpha^* (s') = t_2 \]

\[ pt_2 + (1 - p)t_1 - c(s_0, t_2) \]

\[ < t_2 - c(s', t_2) \]

if \( s' \) is close to \( s_0 \),

i.e., a contradiction to \( ((s_0, s_0), \alpha^*) \) being PBE.

So \( \Pi \) cannot satisfy FL.
Separating Equilibrium outcome:

\[ \Pi^* \left( 0, S_2 \right) = t^* \]

\[ x^*(0) = t_1 \]
\[ x^*(S_2) = t_2 \]

\( t_1 - c(0, t_1) = t_2 - c(S_1, t_1) \)
\( (s', s'') \) is not \( \text{II}-\text{relevant} \) for \( s'' \).

\( (s'', s') \) is \( \text{II}-\text{relevant} \) if \( 0 \leq s'' \leq s' \).

Suppose a PBE \((l_0, s_2), \alpha^*, \mathbb{M}^*\) satisfies FI, \( \mathbb{M}^* (\text{\text{II}-relevant } | s') = 1 \)

\( \Rightarrow \mathbb{M}^* (t_2 | s') = 1 \)

\( \Rightarrow \alpha^* (s') = t_2 \)

\( \Rightarrow t_2 - c(s_2, t_2) \prec t_2 - c(s', t_2) \)

\( \Rightarrow \) contradiction

\( \Rightarrow \) it does not satisfy FI.
Riley outcome:

Least costly separating equilibrium outcome satisfies FL.

* $s' < \bar{s}$, $(s', \bar{s})$ is $\Pi$-relevant.
* $s' > \bar{s}$, $s'$ is not $\Pi$-relevant signal.
$s' \prec s$, $(s'', s')$ is not $T$-relevant.

\[ \forall s'' \]

For $s' \prec s$, \( \hat{M^*}(s'' | s^r) = 0 \), \( \forall s'' \)

\[ \Rightarrow \hat{M^*}(t | s^r) = 1. \]

For $s' \succ s$, anything goes for

\[ \hat{M^*}(t | s^r). \]
A recipe for forward induction in signaling game

0. For simplicity, suppose the sender is playing a pure strategy.

1. Fix the equilibrium outcome $\pi$. (This tells you the sender’s equilibrium strategy $\sigma$ as well as the receiver’s equilibrium action $\alpha(s)$ for $s \in \{\sigma(t) : t \in T\}$.)
A recipe for forward induction in signaling game

0 For simplicity, suppose the sender is playing a pure strategy.

1 Fix the equilibrium outcome $\pi$. (This tells you the sender’s equilibrium strategy $\sigma$ as well as the receiver’s equilibrium action $\alpha(s)$ for $s \in \{\sigma(t) : t \in T\}$.)

2 For any off-equilibrium $s'$, it suffices to consider relevant strategy $z$ such that $z(t)$ is either $s'$ or $\sigma(t)$ for every $t$.
   - Make up $\alpha'(s')$ (and given $\alpha(\sigma(t))$) so that every $t$ weakly prefers $\sigma(t)$ to $s'$; and if $z(t) = s'$ then $t$ is indifferent between $\sigma(t)$ and $s'$. 

---

Songzi Du (UCSD Econ 200C)
A recipe for forward induction in signaling game

0. For simplicity, suppose the sender is playing a pure strategy.

1. Fix the equilibrium outcome $\pi$. (This tells you the sender’s equilibrium strategy $\sigma$ as well as the receiver’s equilibrium action $\alpha(s)$ for $s \in \{\sigma(t) : t \in T\}$.)

2. For any off-equilibrium $s'$, it suffices to consider relevant strategy $z$ such that $z(t)$ is either $s'$ or $\sigma(t)$ for every $t$.
   - Make up $\alpha'(s')$ (and given $\alpha(\sigma(t))$) so that every $t$ weakly prefers $\sigma(t)$ to $s'$; and if $z(t) = s'$ then $t$ is indifferent between $\sigma(t)$ and $s'$.

3. $\tilde{\mu}(\cdot | s')$ puts probability 1 on the relevant $z$ from (2), choose $\alpha(s')$ to best respond to $\tilde{\mu}(\cdot | s')$.
   - Make sure every $t$ prefers $\sigma(t)$ to $s'$. 