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 $s \in S$.

Signaling Game

- Two players, sender (player 1) and receiver (player 2)

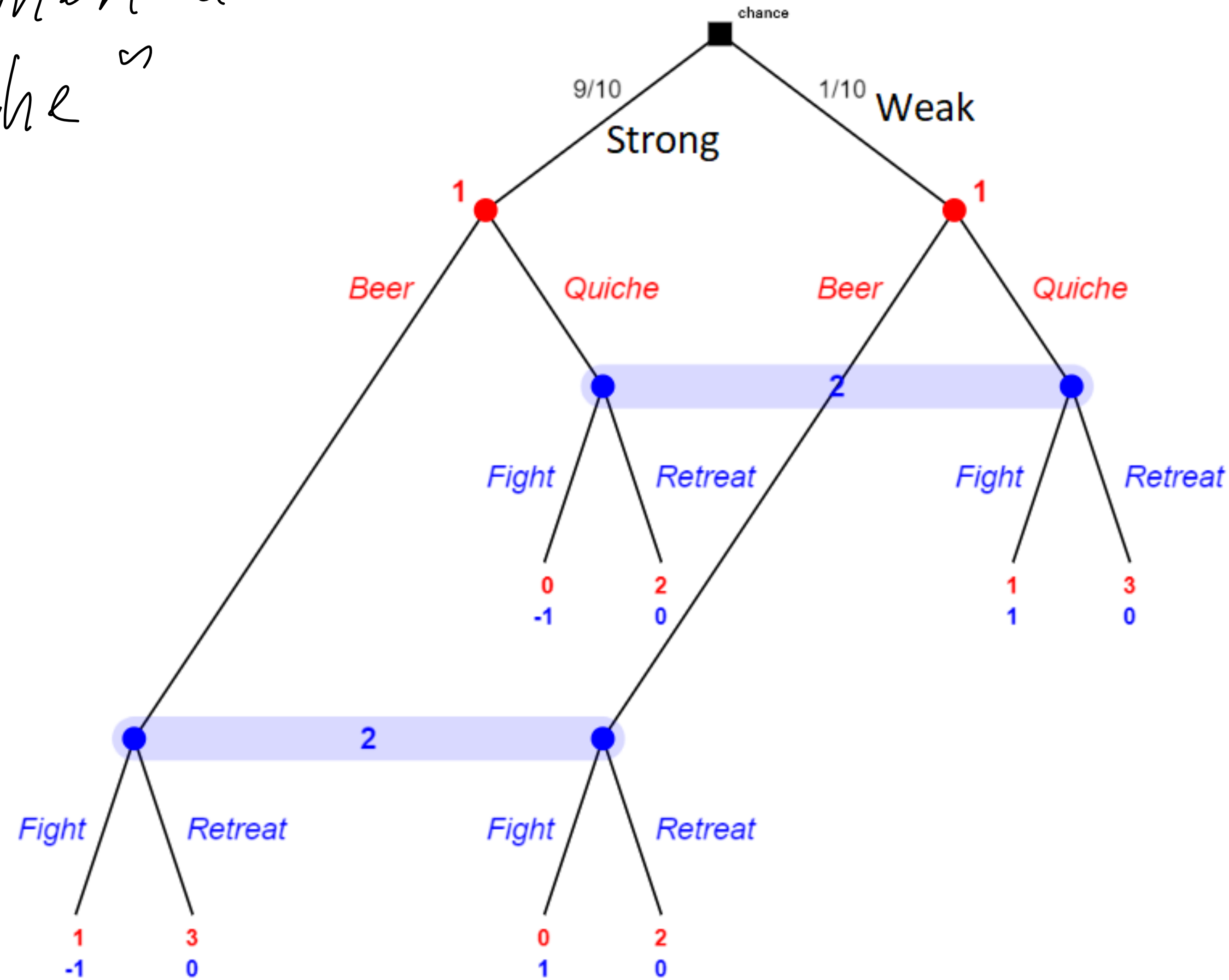
- Nature picks $t \in T$ type of sender. $p(t)$ is probability that type is t.
- Sender observes t, selects signal $s \in S$.
- Receiver observes s (but not t), selects action a.

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

Beer-quiche game

Cho Kreps QJE

"Real men don't eat quiche"



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: $U_2(a, t, s) = -(a - t)^2$ $\Rightarrow a = \mathbb{E}[t \mid s]$
  - 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: $\sigma : T \to \Delta(S)$. behaviral strategy
- Receiver's strategy: $\alpha : S \to \Delta(A)$.
- Receiver's belief (assessment): $\mu(\cdot \mid s) \in \Delta(T), \forall s \in S$.

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): $\mu(\cdot \mid s) \in \Delta(T)$, $\forall s \in S$.

$(\sigma^*, \alpha^*, \mu^*)$ 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
$$\max_{a \in A} \sum_{t \in T} \mu^*(t \mid s) U_2(a, t, s)$$
for all s, $\sum_s \sigma^*(t)(s) = 1$ anything goes off equilibrium
3. μ^* derives from prior and σ^* using Bayes's Rule whenever possible,

$\sigma^*(t) \in \Delta(S)$

$\sigma^*(t)(s) \in [0,1]$

$$\mu^*(t \mid s) = \frac{\overbrace{p(t)\,\sigma^*(t)(s)}^{\mathbb{P}(t,s)}}{\underbrace{\sum_{t' \in T} p(t')\,\sigma^*(t')(s)}_{\mathbb{P}(s)}} .$$
> 0

Perfect Bayesian Equilibrium (based on appraisals) $S(h)$

$h = s$ $\quad$ $\shortparallel$

$Z(s)$

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

Perfect Bayesian Equilibrium (based on appraisals)

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

$\Updownarrow$ $z(t)(s) = 1$

$(\sigma^*, \alpha^*, \tilde{\mu}^*)$ 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
$$\max_{a \in A} \sum_{z \in Z(s)} \tilde{\mu}^*(z \mid s) \frac{\sum_{t \in T} p(t)\, z(t)(s)\, U_2(a, t, s)}{\sum_{t' \in T} p(t')\, z(t')(s)}$$
for all s,

$\mathbb{P}(t \mid s, z)$

Perfect Bayesian Equilibrium (based on appraisals)

- Receiver's belief (appraisal) *about sender's strategy*: $\tilde{\mu}(\cdot \mid s) \in \Delta(Z(s))$, 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
$$\max_{a \in A} \sum_{z \in Z(s)} \tilde{\mu}^*(z \mid s) \frac{\sum_{t \in T} p(t)\, z(t)(s)\, U_2(a, t, s)}{\sum_{t' \in T} p(t')\, z(t')(s)}$$
for all s,
3. $\tilde{\mu}^*$ derives from prior and σ^* using Bayes's Rule whenever possible,
$$\tilde{\mu}^*(z \mid s) = \frac{\left(\prod_{t \in T} \overset{\mathbb{P}(z)}{\sigma^*}(t)(z(t))\right) \left(\sum_{t \in T} p(t)\, z(t)(s)\right) \quad \mathbb{P}(s \mid z)}{\sum_{t \in T} p(t)\, \sigma^*(t)(s) \quad \mathbb{P}(s)}.$$

Example in Beer-Quiche game

- Suppose $\underbrace{\sigma^*(S)(B) = 0.9 \text{ and } \sigma^*(W)(B) = 0.3.}_{\text{behaviral strategy}}$ What is $\tilde{\mu}(\cdot \mid B)$?

Equivalent mixed strategy : $\sigma^*(BB) = 0.9 \cdot 0.3 = 0.27$

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

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

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

$$\tilde{\mu}(BB \mid B) = \frac{0.27 \cdot \mathbb{P}(B \mid BB)}{\underbrace{\mathbb{P}(B)}_{\mathbb{P}(BB)\mathbb{P}(B \mid BB) + \mathbb{P}(BQ)\mathbb{P}(B \mid BQ) + \cdots}} = \frac{0.27 \cdot 1}{0.9 \cdot 0.9 + 0.1 \cdot 0.3} = \frac{0.27}{0.84}$$

$$\tilde{\mu}(BQ \mid B) = \frac{0.63 \cdot \mathbb{P}(B \mid BQ)}{0.84} = \frac{0.63 \cdot 0.9}{0.84}$$

$$\tilde{\mu}(QB \mid B) = \frac{0.03 \cdot 0.1}{0.84}$$

$$\tilde{\mu}(QQ \mid B) = 0$$

$$\tilde{\mu}(\cdot \mid B) \in \Delta(Z(B))$$

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

$$\mu(S \mid B) = \frac{0.63 \cdot 0.9}{0.84} \cdot \underbrace{\mathbb{P}(S \mid BQ, B)}_{1}$$

$$+ \frac{0.27}{0.84} \cdot \underbrace{\mathbb{P}(S \mid BB, B)}_{0.9}$$

$$+ \frac{0.03 \cdot 0.1}{0.84} \cdot \underbrace{\mathbb{P}(S \mid QB, B)}_{0}$$

$$= \frac{\mathbb{P}(S, B)}{\mathbb{P}(B)} = \frac{0.9 \cdot 0.9}{0.84}$$

Terminology

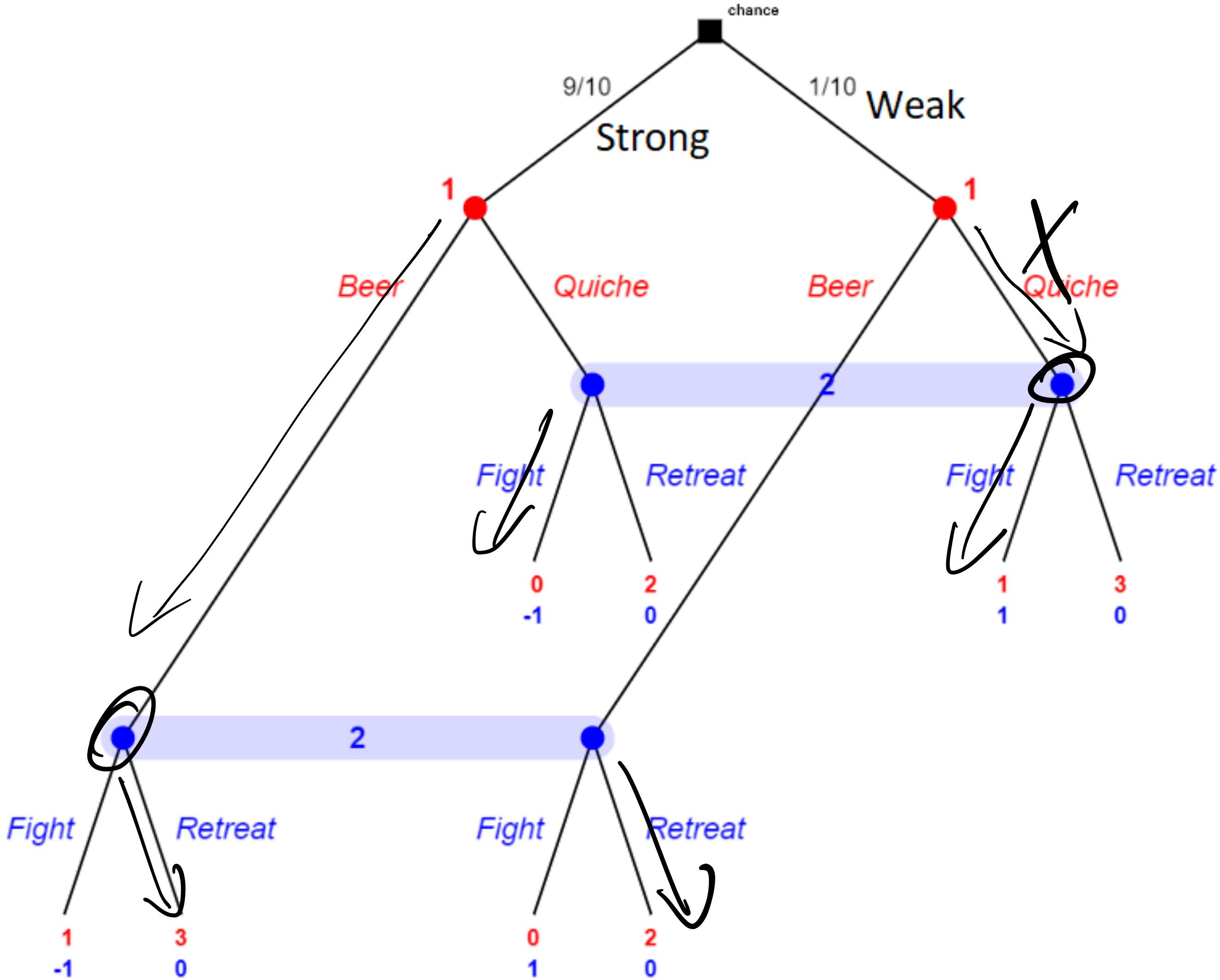
- s is on the equilibrium path if $\sigma^*(t)(s) > 0$ for some t.
- Equilibrium outcome $\pi \in \Delta(T \times S \times A)$, induced by some PBE.

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

Terminology

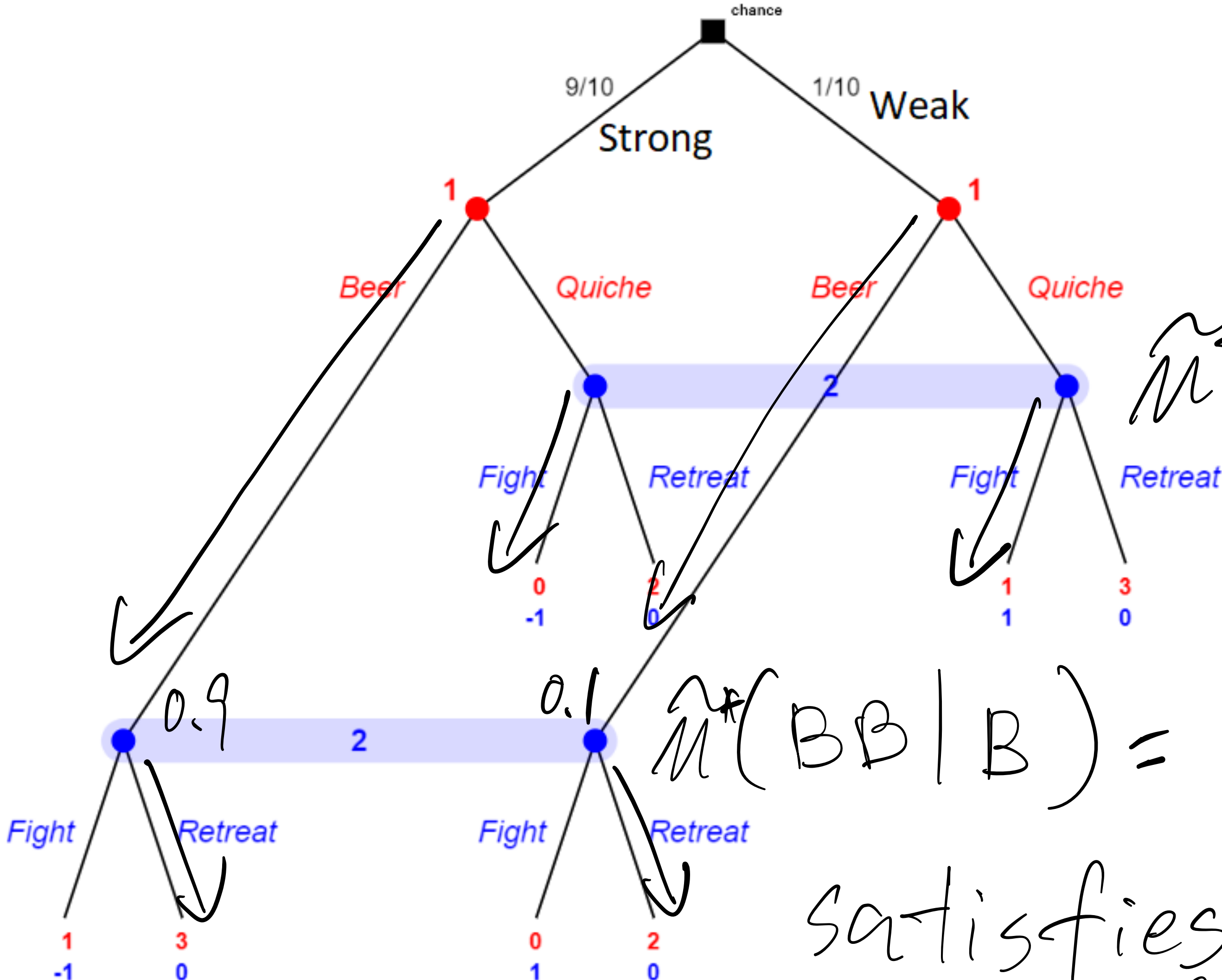
- s is on the equilibrium path if $\sigma^*(t)(s) > 0$ for some t.
- Equilibrium outcome $\pi \in \Delta(T \times S \times A)$, induced by some PBE.
- Pooling equilibrium: $\sigma^*(t)$ constant in t
  - $\mu^*(\cdot \mid s) = p(\cdot)$ whenever s is on the equilibrium path
- Separating equilibrium: the support of $\sigma^*(t)$ does not intersect the support of $\sigma^*(t')$ if $t \neq t'$
  - $\mu^*(t \mid s) = 1$ 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 $\sigma^* = BB$



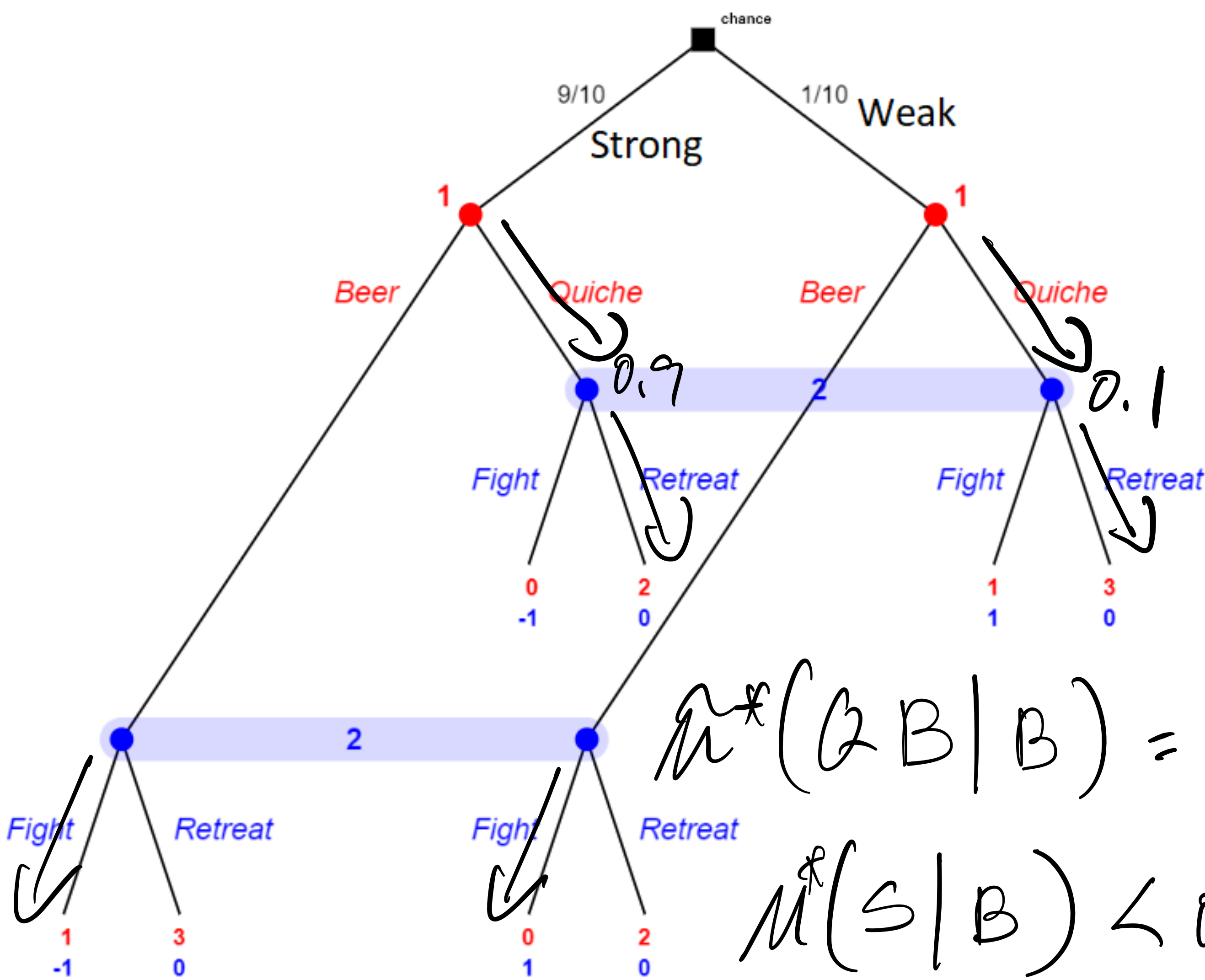
$\hat{\mu}^*(BQ \mid Q) = 1$

$\Rightarrow \mu^*(S \mid Q) = 0$

$\hat{\mu}^*(BB \mid B) = 1$

satisfies "no strictly dominated strategy" refinement

PBE#2 in beer-quiche game $\sigma^* = QQ$



$\tilde{\mu}^*(QQ \mid Q) = 1$

$\tilde{\mu}^*(QB \mid B) = 1 \Rightarrow \mu^*(S \mid B) = 0$

$\mu^*(S \mid B) < 0.5 \Rightarrow \tilde{\mu}^*(QB \mid B) > 0$

This PBE does not survive "no strictly dominated strategy" refinement.

Beer-quiche game, normal form

1 \ 2	F F	F R	R F	R R
B B	9/10, -4/5	9/10, -4/5	$\boxed{29/10}$, $\boxed{0}$	29/10, $\boxed{0}$
B Q	$\boxed{1}$, -4/5	6/5, -9/10	14/5, $\boxed{1/10}$	$\boxed{3}$, 0
Q B	0, -4/5	9/5, $\boxed{1/10}$	1/5, -9/10	2, 0
Q Q	1/10, -4/5	$\boxed{21/10}$, $\boxed{0}$	1/10, -4/5	21/10, $\boxed{0}$

$\frac{3}{4} QQ + \frac{1}{4} BQ$

strictly dominates

QB

A "no strictly dominated strategy" refinement of PBE

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

Every strategy in $Z(s)$ is strictly dominated

$\Rightarrow$ anything goes for $\tilde{\mu}(\cdot \mid s)$

* $(QQ, FR, \tilde{\mu}^*)$ is a PBE

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

$\underbrace{\tilde{\mu}^*(BB \mid B)}_{\Downarrow \; \mu^*(S \mid B) = 0.9} + \underbrace{\tilde{\mu}^*(BQ \mid B)}_{\Downarrow \; \mu^*(S \mid B) = 1} = 1$

$\Rightarrow \mu^*(S \mid B) \geq 0.9$ $\Rightarrow$ Retreat after observing B

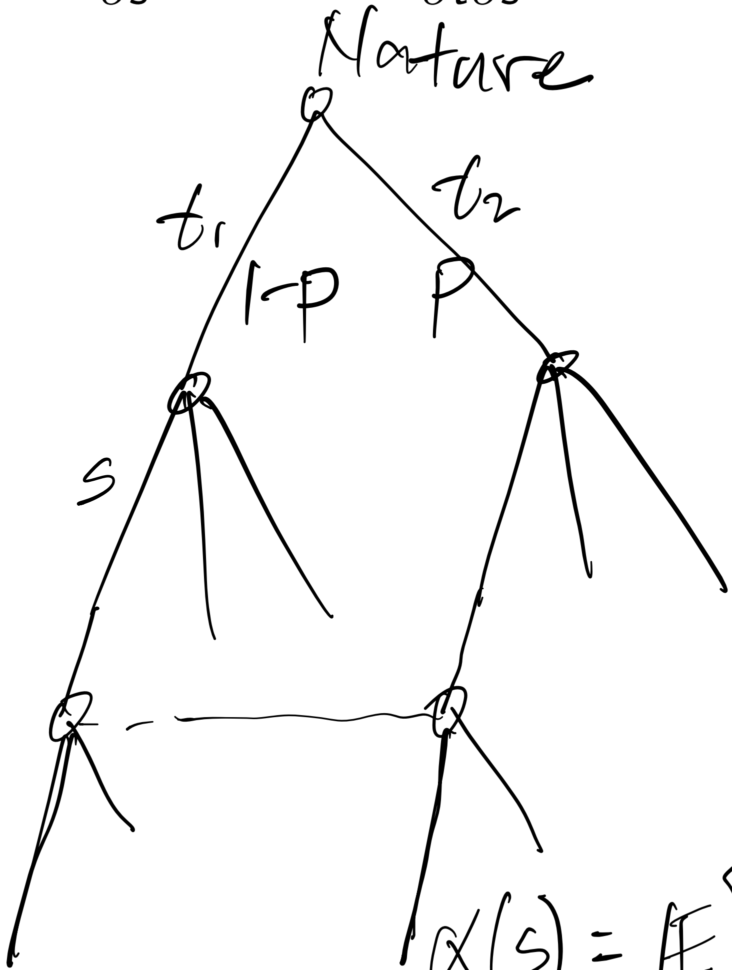
$\Rightarrow$ Contradicts S type choosing Q.

So $(QQ, FR, \tilde{\mu}^*)$ cannot satisfy "no strictly dominated strategy" refinement.

PBE in labor market signaling (two types)

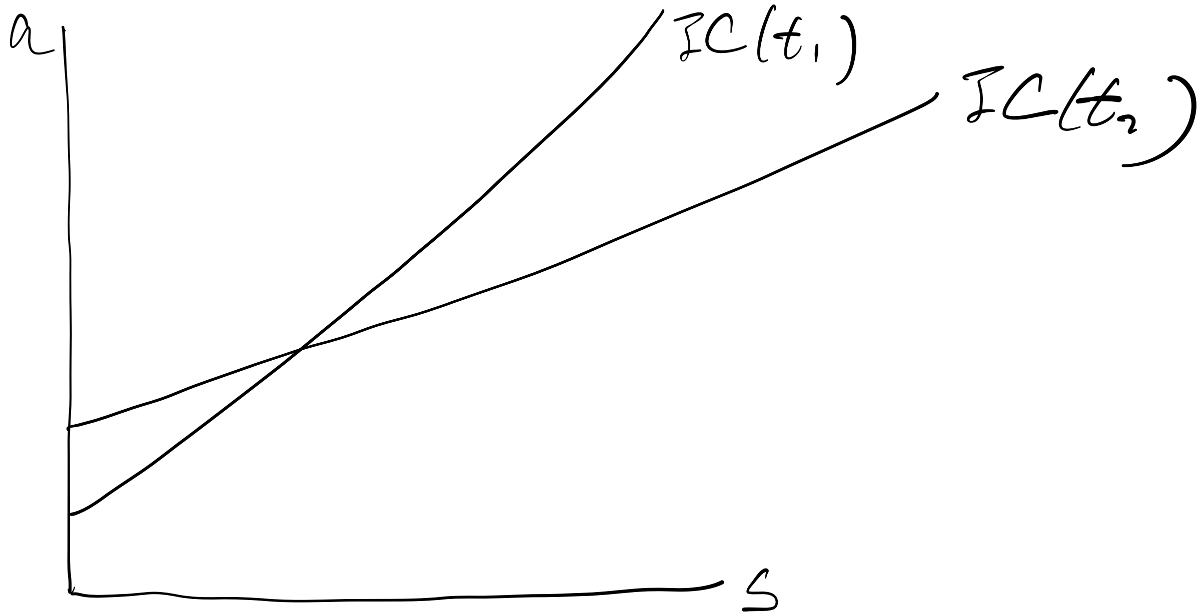
- Suppose $\frac{\partial c}{\partial s} > 0$ and $\frac{\partial^2 c}{\partial t \partial s} < 0$.

$t_1 < t_2$



$\alpha(s) = \mathbb{E}[t \mid s]$

$u_1(t, s, a) = a - c(s, t)$



$$a(s) - c(s, t_1) = \text{constant}$$

$$\frac{\partial a}{\partial s}(s) - \frac{\partial c}{\partial s}(s, t_1) = 0$$

Example: $c(s, t) = \dfrac{s^2}{t}$

* Separating PBE :

$\sigma^*(t_1) = s_1$, $\sigma^*(t_2) = s_2$

$s_1 \neq s_2$

$\forall s$, $\alpha^*(s) \in [t_1, t_2]$

$\alpha^*(s_1) = t_1$, $\alpha^*(s_2) = t_2$

$s_1 > 0$, $\alpha^*(s_1) = t_1 \Rightarrow$ profitable deviation to $s_1 = 0$

$\Rightarrow s_1 = 0$

$$t_1 - c(0, t_1) \geq t_2 - c(s_2, t_1)$$

$$t_2 - c(s_2, t_2) \geq t_1 - c(0, t_2)$$

$$c(s_2, t_2) - c(0, t_2) \leq t_2 - t_1 \leq c(s_2, t_1) - c(0, t_1)$$

$\sigma^*(t_1) = 0, \quad \sigma^*(t_2) = s_2$

$$\alpha^*(s) = \begin{cases} t_1 & s < s_2 \\ t_2 & s \geq s_2 \end{cases}$$

Pooling PBE

$\sigma^*(t_1) = \sigma^*(t_2) = s_0$

$\alpha^*(s_0) = p t_2 + (1-p) t_1$

$$p t_2 + (1-p) t_1 - c(s_0, t_1) \geq t_1 - c(0, t_1)$$

$\Rightarrow p t_2 + (1-p) t_1 - c(s_0, t_2) \geq t_1 - c(0, t_2)$

$$\alpha^*(s) = \begin{cases} pt_2 + (1-p)t_1 & s = s_0 \\ t_1 & s \neq s_0 \end{cases}$$

$$\alpha^*(s) = \begin{cases} t_1 & s < s_0 \\ pt_2 + (1-p)t_1 & s \geq s_0 \end{cases}$$

Forward Induction (Govindan and Wilson, 2009)

NWBR

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

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

- A signal $s \in S$ is *relevant* for π 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 $z : T \to S$ is *relevant* for π if it is a best response to α^* of a PBE $(\sigma^*, \alpha^*, \tilde{\mu}^*)$ that induces π.
- A signal $s \in S$ is *relevant* for π if $z(t) = s$ for some relevant z and type t.

If s is not π-relevant, then every strategy in $Z(s)$ is not π-relevant.

Definition

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

- Forward induction implies intuitive criterion, D1, D2, etc.

Intuition of forward induction

An equilibrium theory of off-equilibrium signals

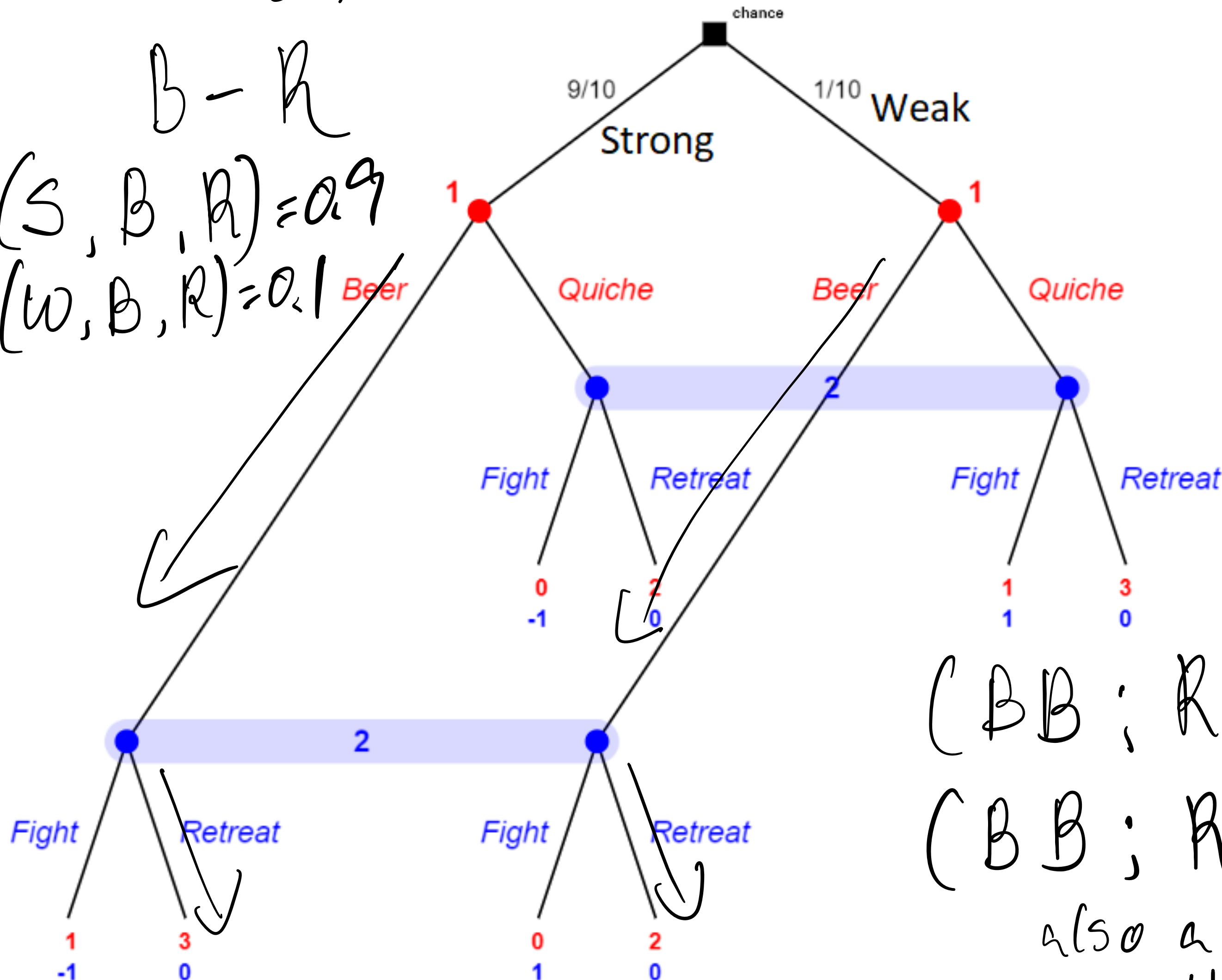
- At a PBE, receiver tries to rationalize an off-equilibrium s, believing that s is sender's best response to another PBE with the same outcome.
- Receiver thinks that:
  - Sender is confused about which PBE is in effect, which is understandable since the two PBE's result in the same outcome.
  - In other words, Sender is confused about receiver's off-equilibrium play.

Forward induction in beer-quiche game

π : S-B, W-B, B-R

$\pi(S, B, R) = 0.9$

$\pi(W, B, R) = 0.1$



BB : relevant

BQ : relevant

QB : irrelevant

QQ : irrelevant

(BB ; RF) is a PBE

(BB ; R, 0.5F+0.5R) is also a PBE

$\Downarrow$

BQ is a best response to (R, 0.5F+0.5R)

$(BB, RF, \tilde{\mu}(BB \mid B) = 1, \tilde{\mu}(BQ \mid Q) = 1)$

is a PBE.

$\Downarrow$

satisfies forward induction.

$\Rightarrow \Pi$ satisfies forward induction

$(S\text{-}Q, W\text{-}Q, Q\text{-}R)$ does not satisfy forward induction.

Forward induction in job market signaling (two types)

Pooling equilibrium outcome:

$$\pi : (\underset{\underset{t_1}{\uparrow}}{s_0}, \underset{\underset{t_2}{\uparrow}}{s_0}) = \sigma^*, \quad \alpha^*(s_0) = pt_2 + (1-p)t_1$$

Consider $s' > s_0$, slightly

Claim: s' is a π-relevant signal, (s_0, s') is π-relevant.

<u>Proof</u>: Pick a PBE $((s_0, s_0), \alpha^*)$

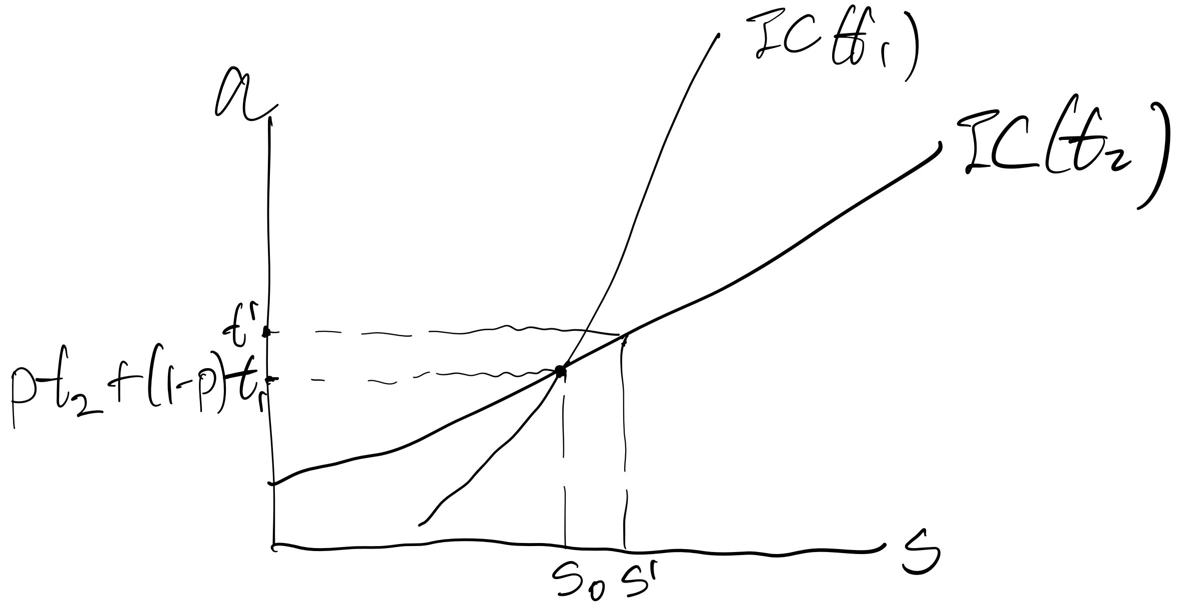
$$\alpha'(s) = \begin{cases} \alpha^*(s) & \text{if } s \neq s' \\ t' & \text{if } s = s' \end{cases}$$

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

$$pt_2 + (1-p)t_1 - c(s_0, t_1) > t' - c(s', t_1)$$

$$\left(c(s', t_1) - c(s_0, t_1) > c(s', t_2) - c(s_0, t_2) \right.$$

$$\left. \int_{s_0}^{s'} \frac{\partial c}{\partial s}(s, t_1)\, ds > \int_{s_0}^{s'} \frac{\partial c}{\partial s}(s, t_2)\, ds \right)$$



$((s_0, s_0), \alpha')$ is a PBE.

(s_0, s') is a best response to α'.

So (s_0, s') is π-relevant. $\square$

* Claim: (s', s'') is not Π-relevant for any s''.

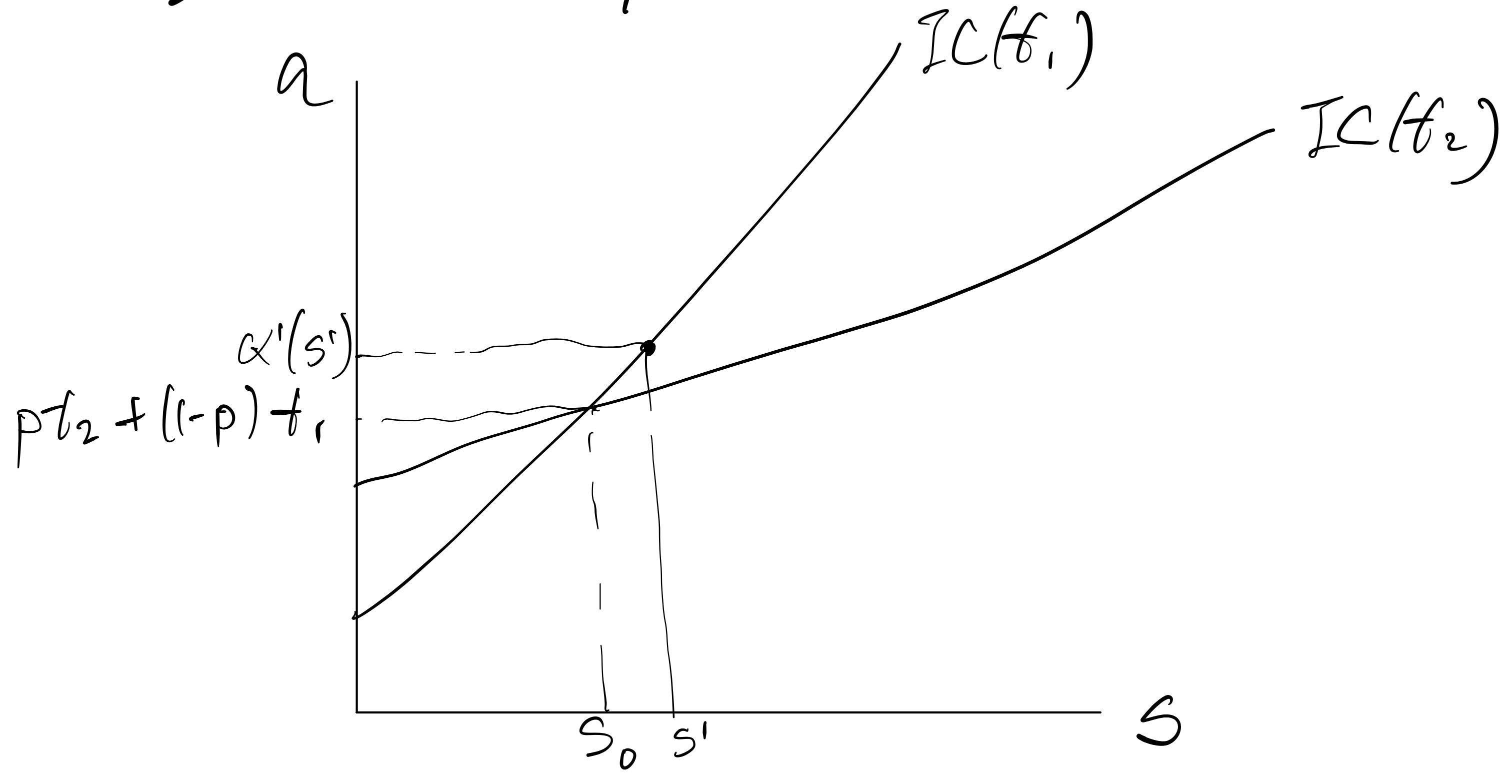
<u>Proof</u>: Suppose (s', s'') is Π-relevant,

i.e., $\exists\, \alpha'$ s.t. $((s_0, s_0), \alpha')$ is a PBE,

and (s', s'') is optimal given α'.

Then t_1 is indifferent between s_0 and s'

$\Rightarrow t_2$ strictly prefers s' to s_0



$\Rightarrow$ contradiction of $((s_0, s_0), \alpha')$ being PBE. $\square$

Suppose π satisfies FI, i.e.,

a PBE $((s_0, s_0), \alpha^*, \tilde{\mu}^*)$,

and $\tilde{\mu}^*((s', s'') \mid s') = 0 \ \forall s''$

$$\Rightarrow \mu^*(t_2 \mid s') = 1$$

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

$$p t_2 + (1-p) t_1 - c(s_0, t_2)$$

$$< t_2 - c(s', t_2)$$

if s' is close to s_0,

i.e. a contradiction to

$((s_0, s_0), \alpha^*)$ being PBE.

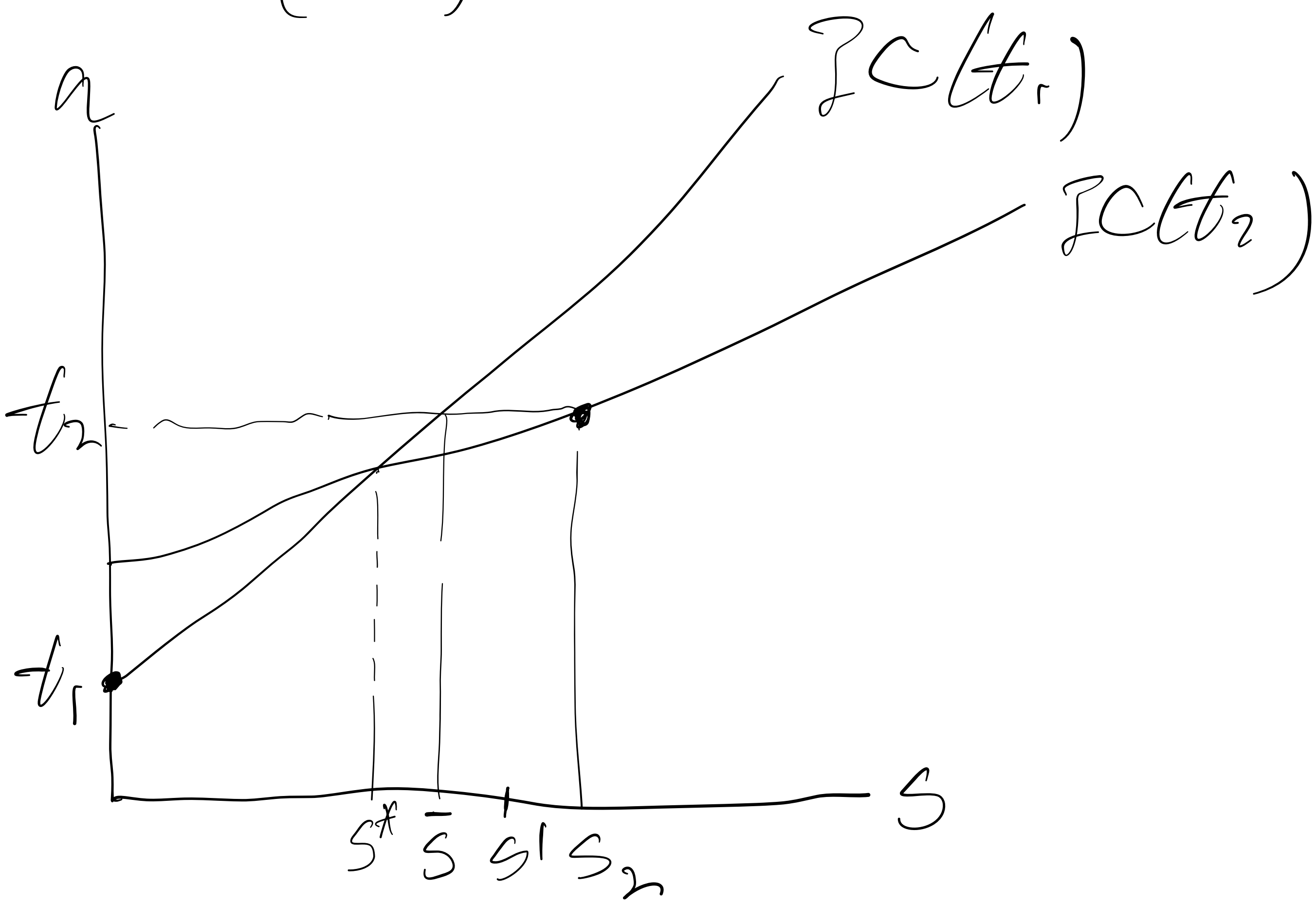
So π cannot satisfy FI.

Separating Equilibrium outcome:

$$\pi : (0, s_2) = \sigma^*, \quad s_2 > 0$$

$$\alpha^*(0) = t_1$$

$$\alpha^*(s_2) = t_2$$



$$t_1 - c(0, t_1) = t_2 - c(\bar{s}, t_1)$$

* (s', s'') is not π-relevant, $\forall s''$.
* (s'', s') is π-relevant if $0 \leq s'' \leq s^*$.
* Suppose a PBE $((0, s_2), \alpha^*, \tilde{\mu}^*)$ satisfies FI, $\tilde{\mu}^*(\pi\text{-relevant strategies} \mid s') = 1$

$$\Rightarrow \mu^*(t_2 \mid s') = 1$$

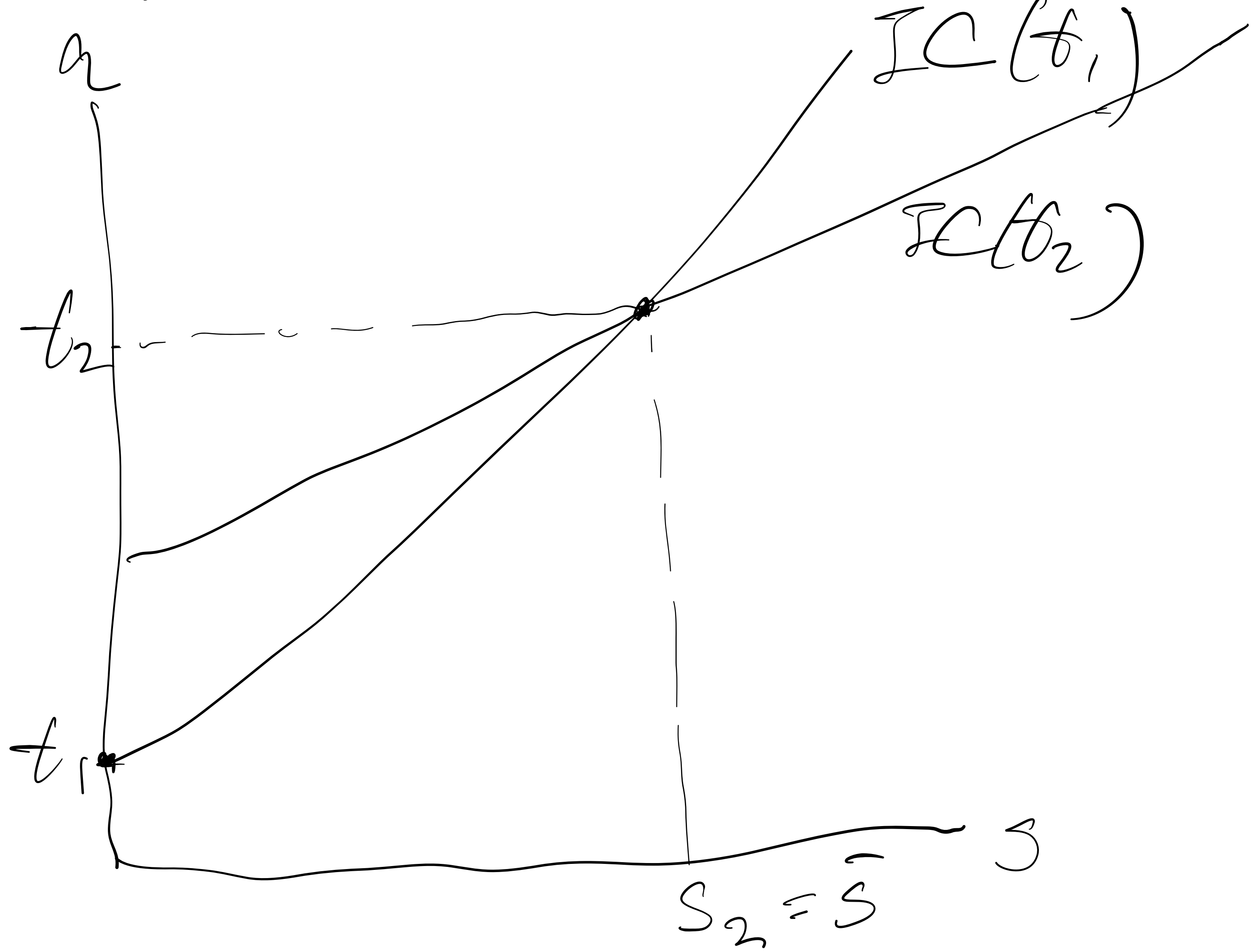
$$\Rightarrow \alpha^*(s') = t_2$$

$$\Rightarrow t_2 - c(s_2, t_2) < t_2 - c(s', t_2)$$

$\Rightarrow$ contradiction

$\Rightarrow$ π does not satisfy FI.

Riley outcome:



Least costly separating equilibrium outcome satisfies FI.

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

$\forall s' < \bar{s}$, (s'', s') is not π-relevant, $\forall s''$

For $s' < \bar{s}$, $\hat{\mu}^*((s'', s') \mid s') = 0$, $\forall s''$

$\Rightarrow \mu^*(t \mid s') = 1$.

For $s' > \bar{s}$, anything goes for $\hat{\mu}^*(\cdot \mid s')$.

A recipe for forward induction in signaling game

0. For simplicity, suppose the sender is playing a pure strategy.
1. Fix the equilibrium outcome π. (This tells you the sender's equilibrium strategy σ 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 π. (This tells you the sender's equilibrium strategy σ 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'.

A recipe for forward induction in signaling game

0. For simplicity, suppose the sender is playing a pure strategy.
1. Fix the equilibrium outcome π. (This tells you the sender's equilibrium strategy σ 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. $\tilde{\mu}(\cdot \mid s')$ puts probability 1 on the relevant z from (2), choose $\alpha(s')$ to best respond to $\tilde{\mu}(\cdot \mid s')$.
   - Make sure every t prefers $\sigma(t)$ to s'.