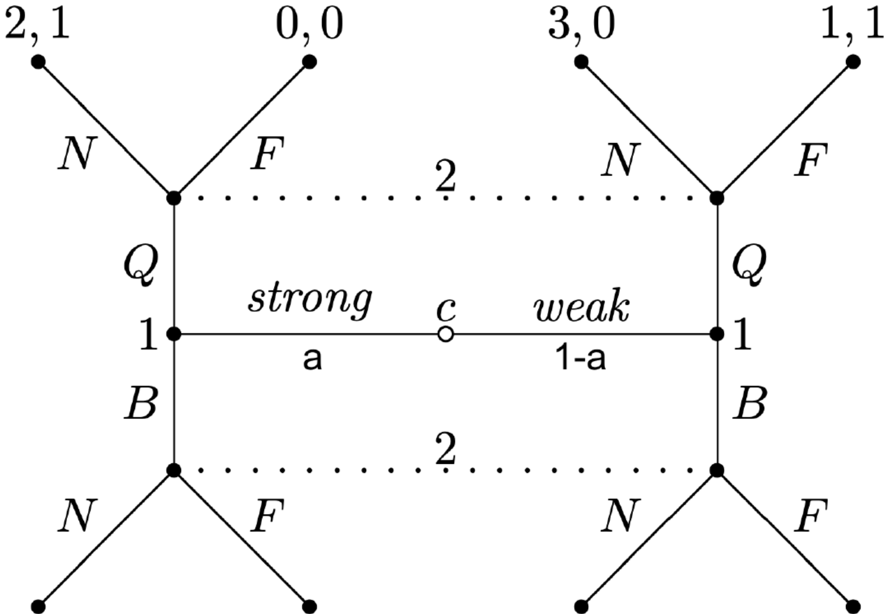
Section 1 – Refinements of PBE

Juan Dodyk

Gov 2006 – Harvard

January 28, 2020

Beer and Quiche



More schematically

1. Chance chooses $\theta \in \{S, W\}$ with $\Pr(\theta = S) = a$.
2. Player 1 knows θ and chooses $a_1 \in \{B, Q\}$.

   – If $\theta = S$, he gets 1 iff $a_1 = B$.
   – If $\theta = W$, he gets 1 iff $a_1 = Q$.

3. Player 2 observes a_1 but doesn't know θ, and chooses $a_2 \in \{F, N\}$.

   – If $\theta = S$, he gets 1 iff $a_2 = N$.
   – If $\theta = W$, he gets 1 iff $a_2 = F$.

4. Player 1 gets 2 iff $a_2 = N$.

Motivation

Beer and quiche may seem to be just an example to illustrate the intuitive criterion.

But it conveys the basic intuition of signaling, which is essentially the same in richer models.

In fact, we can interpret beer-quiche as a political accountability model:

1. An incumbent politician can be diligent (with probability a) or lazy. She knows her type.
2. The politician chooses whether to be accountable or shirk.

   - If she is diligent, she gets 1 iff she is accountable.
   - If she is lazy, she gets 1 iff she shirks.

3. The voters choose whether to reelect the politician. They see the politician's action but don't know her type.

   - If the politician is diligent, the voter gets 1 iff they reelect her.
   - If the incumbent is lazy, the voters get 1 iff they don't reelect her.

4. The incumbent gets 2 iff the voters reelect her.

Result: if $a > \frac{1}{2}$, the politician is accountable, even if she is lazy. Democracy works!

Another political application of the beer-quiche model is Gretchen Helmke's 2002 APSR, "The Logic of Strategic Defection: Court-Executive Relations in Argentina under Dictatorship and Democracy". (Actually the model I'll talk about is from her book, but it's the same idea.)

Question: why do justices who lack judicial independence rule against the president who appointed them?

Player 1 is a judge of the Supreme Court. She can be unbiased [strong] or loyal [weak].

She can rule against the incumbent [beer] or not [quiche].

Player 2 is the next incumbent. She doesn't know the judge's type. She can remove the judge [fight] or not. She wants to remove the judges who are loyal to the previous incumbent.

Result: loyal justices rule against the incumbent if they think that she is politically weak. Why? To signal to the opposition that they are unbiased.

Strategies for Player 1

Player 1 chooses $a_1 \in A_1 = \{B, Q\}$.

Histories: S, W. He knows the history. So $\mathcal{I}_1 = \{S, W\}$.

A strategy is a function $\sigma_1 : \mathcal{I}_1 \to \Delta(A_1)$.

So, we have to specify $\sigma_1(S)(B), \sigma_1(W)(B) \in [0, 1]$.

Then $\sigma_1(S)(Q) = 1 - \sigma_1(S)(B)$ and $\sigma_1(W)(Q) = 1 - \sigma_1(W)(B)$.

Meaning: $\sigma_1(\theta)(a_1) = \Pr(\text{Player 1 chooses } a_1 \mid \text{Player 1 is of type } \theta)$.

Strategies for Player 2

Player 2 chooses $a_2 \in A_2 = \{F, N\}$.

Histories: (θ, a_1), for $\theta \in \Theta = \{S, W\}$ and $a_1 \in A_1$.

Player 2 knows a_2 but doesn't know θ.

So, $\mathcal{I}_2 = \{\{(S,B),(W,B)\},\{(W,B),(W,Q)\}\}$.

A strategy is a function $\sigma_2 : \mathcal{I}_2 \to \Delta(A_2)$.

So, we have to specify

- $\sigma_2(\{(S,B),(W,B)\})(F) \in [0,1]$
- $\sigma_2(\{(S,Q),(W,Q)\})(F) \in [0,1]$

Too cumbersome! We can identify his information set by what he knows, namely $a_1 \in \{B, Q\}$.

So, let's write

- $\sigma_2(B)(F) := \sigma_2(\{(S, B), (W, B)\})(F)$
- $\sigma_2(Q)(F) := \sigma_2(\{(S, Q), (W, Q)\})(F)$

Meaning: $\sigma_2(a_1)(a_2) = \Pr(\text{Player 2 chooses } a_2 \mid \text{Player 1 chose } a_1)$.

Beliefs

Player 1 knows his history, so we don't have to specify beliefs.

Player 2 knows a_1 but has to form beliefs over θ.

Formally, for each information set $I \in \mathcal{I}_2$, we have to define $\mu_2(I) \in \Delta(I)$.

Two cases:

- $I = \{(S, B), (W, B)\}$ — P2 knows $a_1 = B$ and knows that $\theta \in \{S, W\}$
- $I = \{(S, Q), (W, Q)\}$ — P2 knows $a_1 = Q$ and knows that $\theta \in \{S, W\}$

$\mu_2(\{(S, B), (W, B)\})((S, B))$ means $\Pr_2(\theta = S \mid a_1 = B)$. Etc.

Let's make the notation more manageable. Let's write

- $\mu_2(\theta|a_1) := \mu_2(\{(S, a_1), (W, a_1)\})((\theta, a_1))$

Meaning: $\Pr_2(\text{Player 1 is of type } \theta \mid \text{Player 1 chose } a_1)$.

Bizarre pooling strategy

Player 1 always chooses Quiche regardless of type: $\sigma_1(\theta)(Q) = 1$ for both $\theta \in \{S, W\}$.

Player 2 doesn't fight if $a_1 = Q$, but fights if $a_1 = B$.

- $\sigma_2(Q)(F) = 0$
- $\sigma_2(B)(F) = 1$

Beliefs:

- If Player 1 chooses quiche, they are determined by Bayes rule.
- If Player 1 chooses beer, Player 2 thinks that Player 1 is strong with proba less than $\frac{1}{2}$.

Formally:

- $\mu_2(S|Q) = a$ (the prior)
- $\mu_2(S|B) < \frac{1}{2}$

Check that it is a SE

In class we checked that $\mu_2(S|B) = 0$ is a SE.

Let's check that $\mu_2(S|B) = q$ with $0 < q < \frac{1}{2}$ is also a SE.

We want σ^k, μ^k such that σ^k is totally mixed and $(\sigma^k, \mu^k) \to (\sigma, \mu)$.

μ^k is given by Bayes rule, so let's calculate it.

Check that it is a SE

In class we checked that $\mu_2(S|B) = 0$ is a SE.

Let's check that $\mu_2(S|B) = q$ with $0 < q < \frac{1}{2}$ is also a SE.

We want σ^k, μ^k such that σ^k is totally mixed and $(\sigma^k, \mu^k) \to (\sigma, \mu)$.

μ^k is given by Bayes rule, so let's calculate it.

$$\begin{aligned}
\mu_2^k(S|a_1) &= \Pr_2(S|a_1) = \\
&= \frac{\Pr(a_1|S)\Pr(S)}{\Pr(a_1|S)\Pr(S) + \Pr(a_1|W)\Pr(W)} = \\
&= \frac{\sigma_1^k(S)(a_1) \times a}{\sigma_1^k(S)(a_1) \times a + \sigma_1^k(W)(a_1) \times (1-a)}
\end{aligned}$$

We want to find

- $\sigma_1^k(S)(B), \sigma_1^k(W)(B), \sigma_2^k(B)(F), \sigma_2^k(Q)(F) \in (0,1)$

and $\mu_2^k(B)(S), \mu_2^k(Q)(S)$ given by Bayes rule such that

- $\sigma_1^k(S)(B) \to 0$
- $\sigma_1^k(W)(B) \to 0$
- $\sigma_2^k(B)(F) \to 1$
- $\sigma_2^k(Q)(F) \to 0$
- $\mu_2^k(S|B) \to q$
- $\mu_2^k(S|Q) \to a$

$\sigma_2^k(B)(F), \sigma_2^k(Q)(F)$ can be whatever we want because they don't affect beliefs. Take $\sigma_2^k(B)(F) = 1 - \frac{1}{k}$ and $\sigma_2^k(Q)(F) = \frac{1}{k}$.

We must find $\sigma_1^k(S)(B) \to 0$ and $\sigma_1^k(W)(B) \to 0$ that make the beliefs converge to what we want.

$$\mu_2^k(S|B) \to q$$

$$\mu_2^k(S|Q) \to a$$

So, we need

$$\frac{\sigma_1^k(S)(B)}{\sigma_1^k(W)(B)} \to \frac{a(1-q)}{(1-a)q}$$

and

$$\frac{1-\sigma_1^k(S)(B)}{1-\sigma_1^k(W)(B)} \to 1.$$

So, we need

$$\frac{\sigma_1^k(S)(B)}{\sigma_1^k(W)(B)} \to \frac{a(1-q)}{(1-a)q}$$

and

$$\frac{1-\sigma_1^k(S)(B)}{1-\sigma_1^k(W)(B)} \to 1.$$

The second follows from $\sigma_1^k(S)(B) \to 0$ and $\sigma_1^k(W)(B) \to 0$. For the first, we can take

$$\sigma_1^k(S)(B) = \frac{a(1-q)}{k},$$
$$\sigma_1^k(W)(B) = \frac{(1-a)q}{k}.$$

Done!

(You can double-check on your own that everything works.)

The intuitive criterion

In a general signaling game, we have types $\theta \in \Theta$ and messages $a_1 \in A_1$.

Take an equilibrium, and take a message a_1 that no type chooses in equilibrium. In a PBE, Player 2 can believe whatever he wants about θ after observing a_1. There is no restriction. But it is "unreasonable" to expect a_1 for some types.

What we will do is to find the types θ that could not have any reason to deviate to a_1, and "prune" the pair (θ, a_1). We will assume that Player 2 makes this reasoning. And we will assume that Player 1 thinks that Player 2 will make this reasoning. So, when Player 1 considers the possibility of choosing a_1, she assumes that Player 2 will choose her best response assuming that θ is not a pruned type.

In the bizarre pooling equilibrium, who is unlikely to deviate? The weak type.

How do we know this? Because the best that can happen to him if he chooses beer is that Player 2 doesn't fight (he gets payoff 2). In equilibrium his payoff is 3. So, there is no reason for him to deviate and choose beer.

Thus, we can prune the pair (W, B). Therefore if Player 2 sees that Player 1 chooses beer, he infers that $\theta = S$. Hence Player 2 will not fight if $a_1 = B$. This contradicts the prediction of our equilibrium!

Let's formalize this intuition

Let $a_1 \in A_1$ be an off-equilibrium action.

Let $\theta \in \Theta$ be a type.

We will "prune" (θ, a_1) if the best that can happen to θ if she chooses a_1 is worse that what she actually gets in equilibrium:

$$u_1(\sigma; \theta) > \max_{a_2 \in \mathrm{BR}_2(\Theta, a_1)} u_1(a_1, a_2; \theta).$$

Let $S(a_1)$ be the set of "pruned" types.

We infer that if Player 1 chooses a_1, Player 2 will assume that $\theta \notin S(a_1)$ when she chooses her best response. Formally, $a_2 \in \mathrm{BR}_2(\Theta \setminus S(a_1), a_1)$.

Let $\theta \notin S(a_1)$. The worst that can happen to her if she chooses a_1 is $\min_{a_2 \in \mathrm{BR}_2(\Theta \smallsetminus S(a_1), a_1)} u_1(a_1, a_2; \theta)$. If this is better than what she gets in equilibrium, i.e., if

$$u_1(\sigma; \theta) < \min_{a_2 \in \mathrm{BR}_2(\Theta \smallsetminus S(a_1), a_1)} u_1(a_1, a_2; \theta),$$

then she will choose a_1. But we are assuming that a_1 is an off-equilibrium message. Hence if this happens, we rule out the equilibrium. We say that the equilibrium *fails the intuitive criterion.*

If this doesn't happen for any off-equilibrium $a_1 \in A_1$ and any type $\theta \in \Theta \smallsetminus S(a_1)$, then the equilibrium passes the criterion.

Back to the bizarre equilibrium

Beer is the only off-equilibrium action.

We suspect that the weak type should not choose it. We have to verify

$$u_1(\sigma; W) > \max_{a_2 \in \mathrm{BR}_2(\Theta, B)} u_1(B, a_2; W).$$

In equilibrium she gets $u_1(\sigma; W) = 3$. What is the best she can get if she chooses beer?

First, we need to calculate

$$\mathrm{BR}_2(\Theta, B) = \bigcup_{\mu \in \Delta(\Theta)} \operatorname*{argmax}_{a_2 \in A_2} \sum_{\theta \in \Theta} \mu(\theta) u_2(B, a_2; \theta).$$

First, we need to calculate

$$\mathrm{BR}_2(\Theta, B) = \bigcup_{\mu \in \Delta(\Theta)} \operatorname*{argmax}_{a_2 \in A_2} \sum_{\theta \in \Theta} \mu(\theta) u_2(B, a_2; \theta).$$

We can represent $\mu \in \Delta(\Theta)$ by a number $\mu(S) \in [0, 1]$. Of course, $\mu(W) = 1 - \mu(S)$.

We have

$$\operatorname*{argmax}_{a_2 \in A_2} \sum_{\theta \in \Theta} \mu(\theta) u_2(B, a_2; \theta) = \begin{cases} \{F\}, & \text{if } \mu(S) < \frac{1}{2}, \\ \{F, N\}, & \text{if } \mu(S) = \frac{1}{2}, \\ \{N\}, & \text{if } \mu(S) > \frac{1}{2}. \end{cases}$$

So, $\mathrm{BR}_2(\Theta, B) = \{F, N\}$. Recall: we have to verify

$$\begin{aligned} u_1(\sigma; W) &> \max_{a_2 \in \mathrm{BR}_2(\Theta, B)} u_1(B, a_2; W) = \\ &= \max\{u_1(B, F; W), u_1(B, N; W)\}. \end{aligned}$$

So, $\mathrm{BR}_2(\Theta, B) = \{F, N\}$. Recall: we have to verify

$$u_1(\sigma; W) > \max_{a_2 \in \mathrm{BR}_2(\Theta, B)} u_1(B, a_2; W) =$$
$$= \max\{u_1(B, F; W), u_1(B, N; W)\}.$$

Done!

So, Player 1 shouldn't choose beer.

What about the strong type? Let's see if

$$u_1(\sigma; S) > \max_{a_2 \in \mathrm{BR}_2(\Theta, B)} u_1(B, a_2; S) = \max\{u_1(B, F; S), u_1(B, N; S)\}.$$

So, $\mathrm{BR}_2(\Theta, B) = \{F, N\}$. Recall: we have to verify

$$u_1(\sigma; W) > \max_{a_2 \in \mathrm{BR}_2(\Theta,B)} u_1(B, a_2; W) =$$
$$= \max\{u_1(B, F; W), u_1(B, N; W)\}.$$

Done!

So, Player 1 shouldn't choose beer.

What about the strong type? Let's see if

$$u_1(\sigma; S) > \max_{a_2 \in \mathrm{BR}_2(\Theta,B)} u_1(B, a_2; S) = \max\{u_1(B, F; S), u_1(B, N; S)\}.$$

No!

Hence, $S(B) = \{W\}$.

So, assuming that Player 2 believes that the weak type can't choose beer, should the strong type deviate?

Formally, does this hold?

$$u_1(\sigma; S) < \min_{a_2 \in \mathrm{BR}_2(\Theta \setminus S(B), B)} u_1(B, a_2; S)$$

So, assuming that Player 2 believes that the weak type can't choose beer, should the strong type deviate?

Formally, does this hold?

$$u_1(\sigma; S) < \min_{a_2 \in \mathrm{BR}_2(\Theta \setminus S(B), B)} u_1(B, a_2; S)$$

We have $\mathrm{BR}_2(\Theta \setminus S(B), B) = \mathrm{BR}_2(\{S\}, B) = \{N\}$. So the condition is $u_1(\sigma; S) < u_1(B, N; S)$.

It holds!

Hence, the intuitive criterion removes the bizarre pooling equilibrium.