The Intuitive and Divinity Criterion:
Explanation and Step-by-step examples

Ana Espinola-Arredondo and Felix Munoz-Garcia
School of Economic Sciences - Washington State University

Motivation

- Many economic contexts can be understood as sequential games involving elements of incomplete information.

- Signaling games are an excellent tool to explain a wide array of economic situations:
    - Labor market [Spence, 1973]
    - Limit pricing [Battacharya, 1979 and Kose and Williams, 1985]
    - Dividend policy [Milgrom and Roberts, 1982]
    - Warranties [Gal-Or, 1989]

Motivation

- Problems with Signaling games: the set of PBE is usually large.
    - In addition, some equilibria are insensible ("crazy").
- Hence, how can we restrict the set of equilibria to those prescribing sensible behavior?
- Solutions to refine the set of PBE:
    - Intuitive criterion [Cho and Kreps, 1987], and
    - "Universal Divinity" criterion [Banks and Sobel, 1987] (also referred as the D_1-criterion).

Outline of the presentation

- Time structure of signaling games.
- Intuitive Criterion: first and second step.
  - Examples.
- Divinity Criterion: first and second step.
  - Examples.
- Similarities and differences between the Intuitive and the D_1-Criterion.

Signaling games

- One player is privately informed.
  - For example, he knows information about market demand, his production costs, etc.
- He uses his actions (e.g., his production decisions, investment in capacity, etc.) to communicate/conceal this information to other uninformed player.

Time Structure

In particular, let us precisely describe the time structure of the game:

1. Nature reveals to player i some piece of private information, $\theta_i \in \Theta$.
2. Then, player i, who privately observes θ_i, chooses an action (or message m) which is observed by other player j.
3. Player j observes message m, but does not know player i's type. He knows the prior probability distribution that nature selects a given type θ_i from Θ, $\mu(\theta_i) \in [0, 1]$.
   - For example, the prior probability for $\Theta = \{\theta_L, \theta_H\}$ can be $\mu(\theta_L) = p$ and $\mu(\theta_H) = 1 - p$.

Time Structure

Continues:

4. After observing player i's message, player j updates his beliefs about player i's type. Let $\mu(\theta_i|m)$ denote player j's beliefs about player i's type being exactly $\theta = \theta_i$ after observing message m.
5. Given these beliefs, player j selects an optimal action, a, as a best response to player i's message, m, given his own beliefs about player i's type $\mu(\theta_i|m)$.

Outline of the Intuitive Criterion

Consider a particular PBE with its corresponding equilibrium payoffs $u_i^*(\theta)$.

Application of the Intuitive Criterion in two steps:

1. **First Step:** Which type of senders could benefit by deviating from their equilibrium message?
2. **Second Step:** If deviations can only come from the senders identified in the First Step, is the lowest payoff from deviating higher than their equilibrium payoff?
    1. If the answer is *yes*, then the equilibrium *violates* the Intuitive Criterion.
    2. If the answer is *no*, then the equilibrium *survives* the Intuitive Criterion.

Formal definition: First Step

Let us focus on those types of senders who can obtain a higher utility level by deviating than by keeping their equilibrium message unaltered. That is,

$$\Theta^{**}(m) = \left\{ \theta \in \Theta \mid \underbrace{u_i^*(\theta)}_{\text{Equil. Payoff}} \leq \underbrace{\max_{a \in A^*(\Theta, m)} u_i(m, a, \theta)}_{\text{Highest util. from deviating to } m} \right\} \tag{1}$$

Intuitively: we restrict our attention to those types of agents for which sending the off-the-equilibrium message m **could** give them a higher utility level than that in equilibrium, $u_i^*(\theta)$. If m does *not* satisfy this inequality, we say that m is "equilibrium dominated."

Formal definition: Second Step

Then, take the subset of types for which the off-the-equilibrium message m is not equilibrium dominated, $\Theta^{**}(m)$, and check if the equilibrium strategy profile (m^*, a^*), with associated equilibrium payoff for the sender $u_i^*(\theta)$, satisfies:

$$\underbrace{\min_{a \in A^*(\Theta^{**}(m), m)} u_i(m, a, \theta)}_{\text{Lowest payoff from deviating to } m} > \underbrace{u_i^*(\theta)}_{\text{Equil. payoff}} \qquad (2)$$

If there is a type for which this condition holds, then the equilibrium strategy profile (m^*, a^*) violates the Intuitive Criterion.

Possible speech from the sender with incentives to deviate:

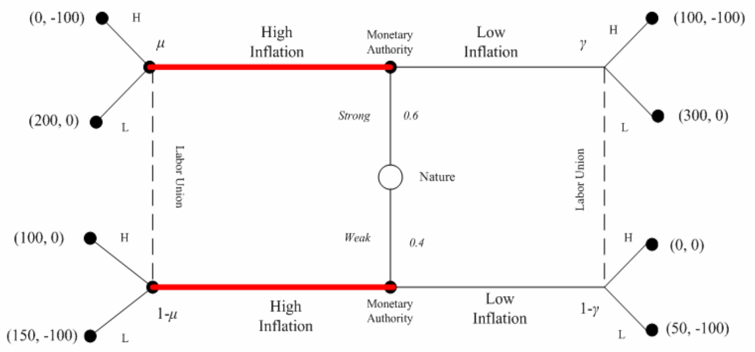
"It is clear that my type is in $\Theta^{**}(m)$. If my type was outside $\Theta^{**}(m)$ I would have no chance of improving my payoff over what I can obtain at the equilibrium (condition (1)). We can therefore agree that my type is in $\Theta^{**}(m)$. Hence, update your believes as you wish, but restricting my type to be in $\Theta^{**}(m)$. Given these beliefs, any of your best responses to my message improves my payoff over what I would obtain with my equilibrium strategy (condition (2)). For this reason, I am sending you such off-the-equilibrium message."

Example 1 - Discrete Messages

- Let us consider the following sequential game with incomplete information:
    - A monetary authority (such as the Federal Reserve Bank) privately observes its real degree of commitment with maintaining low inflation levels.
    - After knowing its type (either Strong or Weak), the monetary authority decides whether to announce that the expectation for inflation is High or Low.
    - A labor union, observing the message sent by the monetary authority, decides whether to ask for high or low salary raises (denoted as H or L, respectively)

Example 1 - Discrete Messages

- The only two strategy profiles that can be supported as a PBE of this signaling game are:
  - A polling PBE with both types choosing (High, High); and
  - A separating PBE with (Low, High).
- Let us check if (High, High) survives the Intuitive Criterion.



First Step

- **First Step:** Which types of monetary authority have incentives to deviate towards Low inflation?
  - *Low inflation* is an off-the-equilibrium message.
- Let us first apply condition (1) to the Strong type,

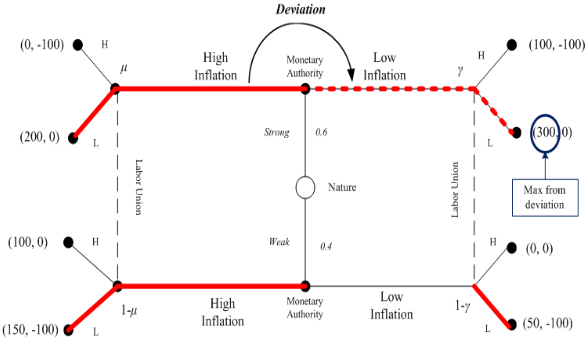
$$\underbrace{u^*_{Mon}\left(High|Strong\right)}_{\text{Equil. Payoff}} < \underbrace{\max_{a_{Labor}} u_{Mon}\left(Low|Strong\right)}_{\text{Highest payoff from deviating to Low}}$$

$$200 < 300$$

- Hence, the Strong type of monetary authority has incentives to deviate towards Low inflation.

First Step

- Graphically, we can represent the incentives of the Strong monetary authority to deviate towards Low inflation as follows:



First Step

- Let us now check if the Weak type also has incentives to deviate towards Low:

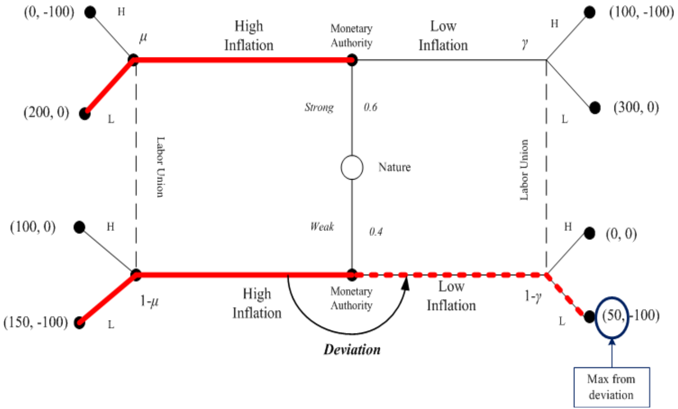
$$\underbrace{u^*_{Mon}\left(High|Weak\right)}_{\text{Equil. Payoff}} < \underbrace{\max_{a_{Labor}} u_{Mon}\left(Low|Weak\right)}_{\text{Highest payoff from deviating to Low}}$$

$$150 > 50$$

- Thus, the Weak type of monetary authority does *not* have incentives to deviate towards Low inflation.

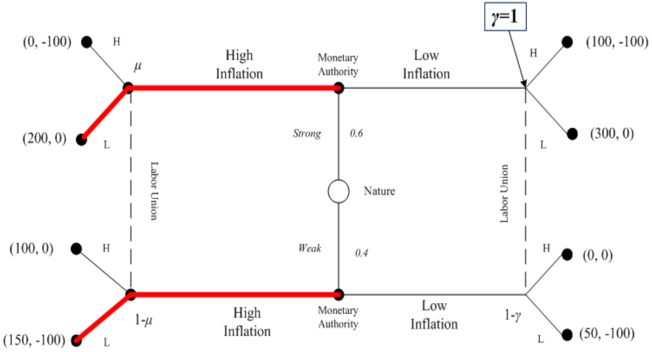
First Step

- Graphically, we can represent the lack of incentives of the Weak monetary authority to deviate towards Low inflation as follows:



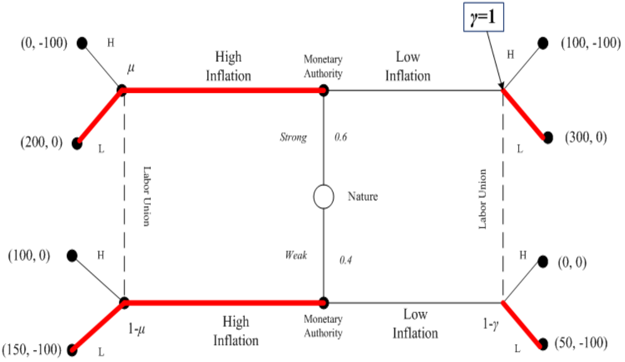
First Step

- Hence, the only type of Monetary authority with incentives to deviate is the Strong type, $\Theta^{**}(Low) = \{Strong\}$.
- Thus, the labor union beliefs after observing *Low inflation* are restricted to $\gamma = 1$.



First Step

- This implies that the labor union chooses *Low wage demands* after observing *Low inflation*. (0 is larger than −100, in the upper right-hand node).

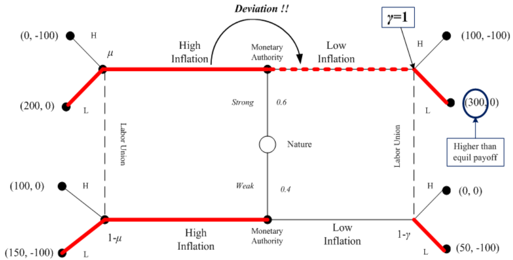


Second Step

- Study if there is a type of monetary authority and a message it could send such that condition (2) is satisfied:

$$\min_{a \in A^*(\Theta^{**}(m), m)} u_i(m, a, \theta) > u_i^*(\theta).$$

which is indeed satisfied since $300 > 200$ for the Strong monetary authority.



As a result...

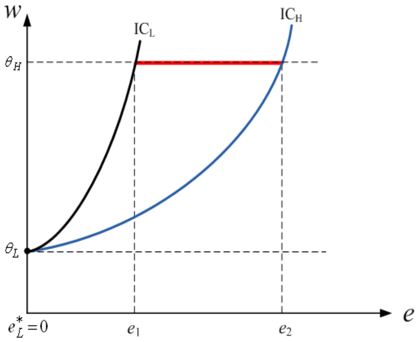
- The pooling PBE of (High, High) *violates* the Intuitive Criterion:
    - there exists a type of sender (Strong monetary authority) and
    - a message (Low)
    - which gives this sender a higher utility level than in equilibrium, regardless of the response of the follower (labor union).

Example 2 - Continuous Messages

- Consider the following sequential-move game between a worker and a firm; Spence (1973).
  - First, nature selects the type of a worker, either θ_H (high productivity) or θ_L (low productivity), such that $\theta_H > \theta_L$.
  - The worker observes his own productivity level, but the firm does not. Observing his type, the worker chooses an education level, $e \geq 0$.
  - Observing the education level of the worker, e, the firm offers a wage $w(e)$.
  - The worker's utility function is $u_{worker}(w, e|\theta) = w - \frac{e}{2\theta}$ if he accepts a wage offer, and zero if he rejects. (Note that θ only affects the worker's cost of acquiring education).

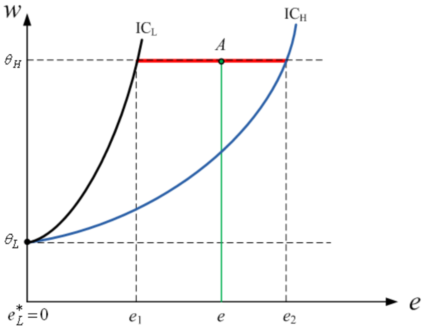
Example 2 - Continuous Messages

- The figure represents separating equilibria where $e_L^* = 0$ and $e_H^* \in [e_1, e_2]$ and $w(e_L^*) = \theta_L$ and $w(e_H^*) = \theta_H$.



First Step

- Let us check if the separating equilibrium $e_L^* = 0$ and $e_H^* = e_2$ survives the Intuitive Criterion.
- Hence, let us consider any off-the-equilibrium message $e \in (e_1, e_2)$. (Green color).

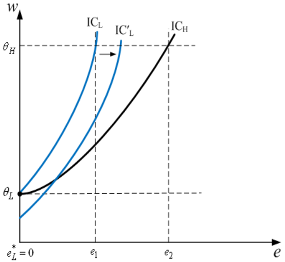


First Step

- The θ_L-type of worker doesn't have incentives to deviate towards e since:

$$\underbrace{u_L^*\left(\theta_L\right)}_{\text{Equil. Payoff}} > \underbrace{\max_{w \in W^*(\theta,m)} u_L\left(e, w, \theta_L\right)}_{\text{Highest payoff from deviating towards } e}$$

[Given that any Indifference Curve passing through any $e \in \left(e_1, e_2\right)$ and $w = \theta_H$ lies *below* the equilibrium IC_L].

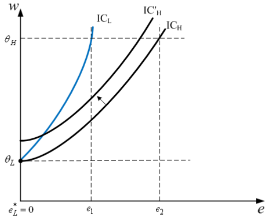


First Step

- But the θ_H-type of worker has incentives to deviate:

$$\underbrace{u_i^*(\theta_H)}_{\text{Equil. payoff}} < \underbrace{\max_{w \in W^*(\theta,e)} u_H(e, w, \theta_H)}_{\text{Highest payoff of deviating towards } e}$$

[since any Indifference Curve passing through any $e \in (e_1, e_2)$ and $w = \theta_H$ lies *above* the equilibrium IC_H].



- Therefore, education levels in $e \in (e_1, e_2)$ can only come from the θ_H-worker, and $\Theta^{**}(e) = \{\theta_H\}$.

Second Step

- Given that e only comes from θ_H, the firm offers a wage $w(e) = \theta_H$ after observing e.

$$\underbrace{\min_{w \in W^*(\Theta^{**}(e), e)} u_H(e, w, \theta_H)}_{\theta_H - c(e, \theta_H)} > \underbrace{u_H^*(\theta_H)}_{\theta_H - c(e_2, \theta_H)}$$

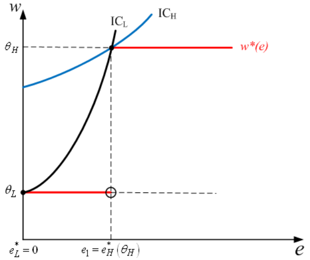
- since, $e_2 > e$, then $c(e_2, \theta_H) > c(e, \theta_H)$. Hence $\theta_H - c(e, \theta_H) > \theta_H - c(e_2, \theta_H)$.
- Intuitively, the lowest payoff that the θ_H-worker obtains by deviating towards e is higher than his equilibrium payoff.
- Therefore, the separating PBE

$$\{e_L^*(\theta_L), e_H^*(\theta_H)\} = \{0, e_2\}$$

violates the Intuitive Criterion.

Example 2 - Final remarks

- All separating equilibria in which the θ_H-worker sends $e \in (e_1, e_2)$ violate the Intuitive Criterion. (Practice).
- The *unique* separating equilibrium surviving the Intuitive Criterion is that in which the θ_H-worker sends $e = e_1$. This equilibria is usually referred as the efficient outcome (or *Riley* outcome).
- Prove the above results in the Homework assignment.



Outline of the Divinity Criterion

Consider a particular PBE with its corresponding equilibrium payoffs.

Application of the D_1-Criterion in two steps:

1. **First Step:** Which type of senders are *more likely* to deviate from their equilibrium message?
    1. In particular, for which type of senders are most of the responder's actions beneficial?
2. **Second Step:** If deviations can only come from the senders identified in the First Step, is the lowest payoff from deviating higher than their equilibrium payoff?
    1. If the answer is *yes*, then the equilibrium *violates* the D_1-Criterion.
    2. If the answer is *no*, then the equilibrium *survives* the D_1-Criterion.

Formal definition: First Step

- Let us first introduce some notation:

$$D\left(\theta, \widehat{\Theta}, m\right) := \bigcup_{\mu:\mu\left(\widehat{\Theta}|m\right)=1} \{a \in MBR\left(\mu, m\right) \mid u_i^*\left(\theta\right) < u_i\left(m, a, \theta\right)\}$$

Intuition: $D\left(\theta, \widehat{\Theta}, m\right)$ is the set of mixed best responses (MBR) of the receiver such that the θ-type of sender is better-off by sending message m than the equilibrium message m^*. [Note that $\mu\left(\widehat{\Theta} \mid m\right) = 1$ represents that the receiver believes that message m only comes from types in the subset $\widehat{\Theta} \in \Theta$].

Similarly for MBR that make the sender indifferent

$$D^{\circ}\left(\theta, \widehat{\Theta}, m\right) := \bigcup_{\mu:\mu\left(\widehat{\Theta}|m\right)=1} \{a \in MBR\left(\mu, m\right) \mid u_i^*\left(\theta\right) = u_i\left(m, a, \theta\right)\}$$

Formal definition: First Step

- Let us now identify which type of senders are **more likely** to deviate from their equilibrium message:

$$\left[D\left(\theta,\widehat{\Theta},m\right)\cup D^{\circ}\left(\theta,\widehat{\Theta},m\right)\right]\subset D\left(\theta',\widehat{\Theta},m\right)$$

- That is, for a given message m, the set of receiver's actions which make the θ'-type of sender better off (relative to equilibrium), $D\left(\theta',\widehat{\Theta},m\right)$, is larger than those actions making the θ-type of sender strictly better off, $D\left(\theta,\widehat{\Theta},m\right)$, or indifferent, $D^{\circ}\left(\theta,\widehat{\Theta},m\right)$.
- The set of types that cannot be deleted after using the above procedure is denoted by $\Theta^{**}\left(m\right)$.

Formal definition: Second Step

- Given the subset of types in $\Theta^{**}(m)$, check if the equilibrium strategy profile (m^*, a^*), with associated equilibrium payoff $u_i^*(\theta)$, satisfies

$$\underbrace{\min_{a \in A^*(\Theta^{**}(m), m)} u_i(m, a, \theta)}_{\text{Lowest payoff from deviating to } m} > \underbrace{u_i^*(\theta)}_{\text{Equil. payoff}}$$

- **Intuition:** if deviations can only come from the senders identified in the First Step, is the lowest payoff from deviating higher than their equilibrium payoff?
- Should be familiar: indeed, it coincides with the 2^{nd} step of the Intuitive Criterion.

Similarities and Differences

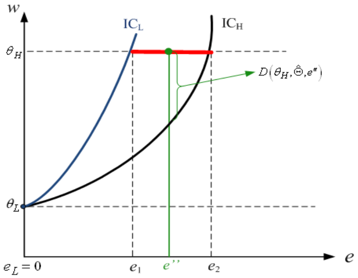
- Both refinement criteria coincide in their 2nd step.
- The 1st step of the D_1-Criterion and the Intuitive Criterion are different.
- In particular, they differ in how they determine the set of senders who can benefit by deviating from their equilibrium message:
  - **Intuitive Criterion:** For which senders *there is at least one action* of the responder that is beneficial?
  - D_1-**Criterion:** For which senders *most of the actions* of the responder are beneficial?
- Hence, the types of senders who benefit from deviating according to the D_1-Criterion are a subset of those who benefit from deviating according to the Intuitive Criterion.

Similarities and Differences

- Therefore, the set of equilibria surviving the D_1-Criterion are a subset of those surviving the Intuitive Criterion.
- Examples about this result (next):
  - Example 3 will show a game where both refinement criteria lead to the same set of surviving PBEs.
  - Example 4 will show a game where the refinement criteria do not lead to the same set (one is a subset of the other).

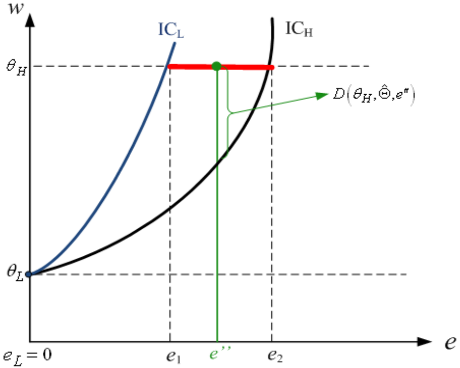
Example 3 - Continuous messages.

- Let us apply the D_1-Criterion in the Spence's (1973) labor market signaling game.
- As before, let us check if the separating equilibrium $e_L^* = 0$ and $e_H^* = e_2$ survives the D_1-Criterion.
- Let us consider the off-the equilibrium message $e'' \in (e_1, e_2)$ (in green color).



Example 3 - Continuous messages.

- **First Step:** $D\left(\theta_L, \widehat{\Theta}, e''\right) \subset D\left(\theta_H, \widehat{\Theta}, e''\right)$, where $D\left(\theta_L, \widehat{\Theta}, e''\right) = \varnothing$, and thus $\Theta^{**}(e'') = \{\theta_H\}$.
- Repeating this process for any off-the-equilibrium message, firm's beliefs are restricted to $\Theta^{**}(e'') = \{\theta_H\}$.



Example 3 - Continuous messages.

- **Second Step:** given $\Theta^{**}(e'') = \{\theta_H\}$, then $w(e'') = \theta_H$. Thus, the minimal utility level that the worker can achieve by sending the off-the-equilibrium message e is

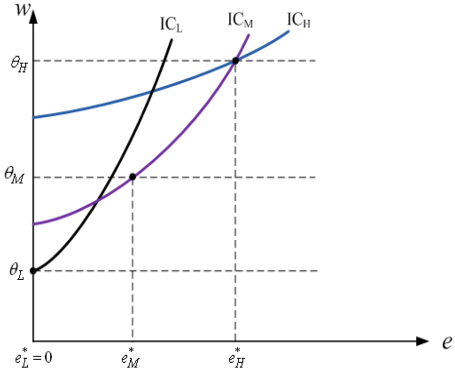
$$\underbrace{\min_{w \in W^*(\Theta^{**}(e''), e'')} u_H(e'', w, \theta_H)}_{\theta_H - c(e'', \theta_H)} > \underbrace{u_H^*(\theta_H)}_{\theta_H - c(e_2, \theta_H)}$$

  Given that, $e_2 > e''$ and $c_e(e'', \theta) > 0$, we have that $c(e_2, \theta_H) > c(e'', \theta_H)$, which ultimately implies $\theta_H - c(e'', \theta_H) > \theta_H - c(e_2, \theta_H)$.

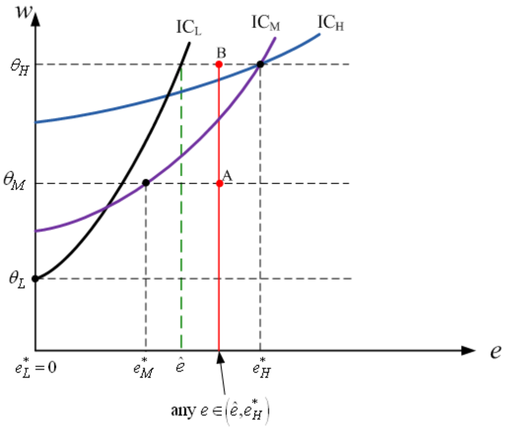
- Therefore, the separating PBE where workers acquire education levels $\{e_L^*(\theta_L), e_H^*(\theta_H)\} = \{0, e_2\}$ *violates* the D_1-Criterion.

- So far both refinement criteria deleted the same equilibria...
- Let us analyze an example where the Intuitive Criterion does not eliminate any equilibria, whereas the D_1-Criterion eliminates all but one.

- **Example 4** - Spence's (1973) education signaling game *but* with $n = 3$ types of workers.
  - The figure represents one of the multiple separating equilibria (e_L^*, e_M^*, e_H^*).



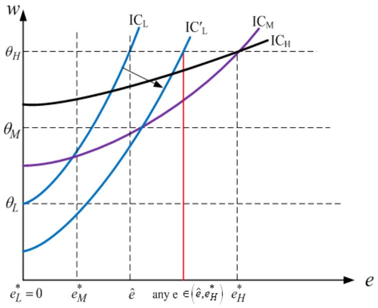
- Let us check if this separating equilibrium survives the Intuitive Criterion, by choosing an off-the-equilibrium message $e \in (\hat{e}, e_H^*)$.



Intuitive Criterion - First Step

- θ_L-type sending a message $e \in (\widehat{e}, e_H^*)$ is equilibrium dominated given that

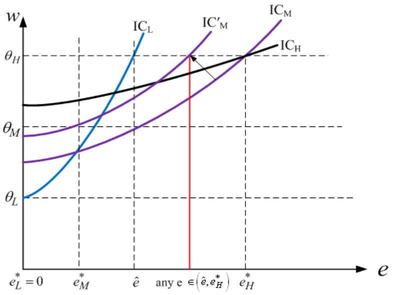
$$\underbrace{u_L^*(\theta_L)}_{\text{Equil. Payoff}} > \underbrace{\max_{w \in W^*(\Theta, e)} u_L(e, w, \theta_L)}_{\text{Highest payoff from deviating to } e}$$



Intuitive Criterion - First Step

- θ_M-workers could send a message $e \in (\hat{e}, e_H^*)$ because for the M-type of worker,

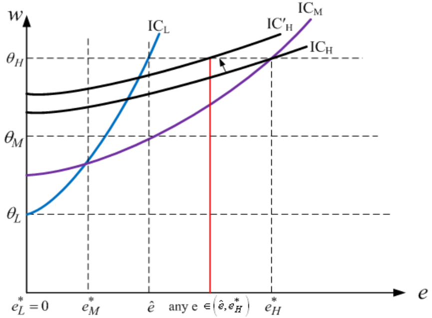
$$\underbrace{u_M^* (\theta_M)}_{\text{Equil. Payoff}} < \underbrace{\max_{w \in W^*(\Theta, e)} u_M (e, w, \theta_M)}_{\text{Highest payoff from deviating to } e}$$



Intuitive Criterion - First Step

- Similarly for the θ_H-type of worker,

$$\underbrace{u_H^*(\theta_H)}_{\text{Equil. Payoff}} < \underbrace{\max_{w \in W^*(\Theta, e)} u_H(e, w, \theta_H)}_{\text{Highest payoff from deviating to } e}$$



-

Intuitive Criterion - First Step

- Hence, when firms observe $e \in (\widehat{e}, e_H^*)$ they will concentrate their beliefs on those types of workers for which these education levels are *not* equilibrium dominated: θ_M and θ_H.
- That is,

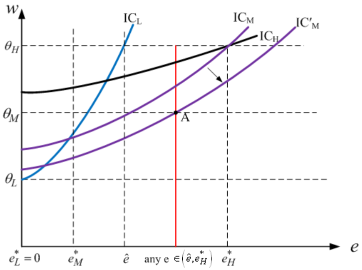
$$\Theta^{**}(e) = \{\theta_M, \theta_H\} \quad \text{for all } e \in (\widehat{e}, e_H^*)$$

Intuitive Criterion - Second Step

- For the θ_M-worker,

$$\min_{w \in W^*(\Theta^{**}(e),e)} u_M(e,w,\theta) < u_M^*(\theta)$$

- Hence, the θ_M-worker does not deviate towards $e \in (\widehat{e}, e_H^*)$. Graphically,

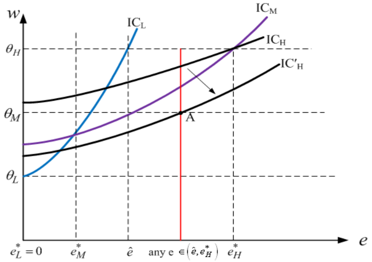


Intuitive Criterion - Second Step

- Similarly for the θ_H-worker,

$$\min_{w \in W^*(\Theta^{**}(e),e)} u_H(e,w,\theta) < u_H^*(\theta)$$

- Thus, the θ_H-worker does not deviate. Graphically,



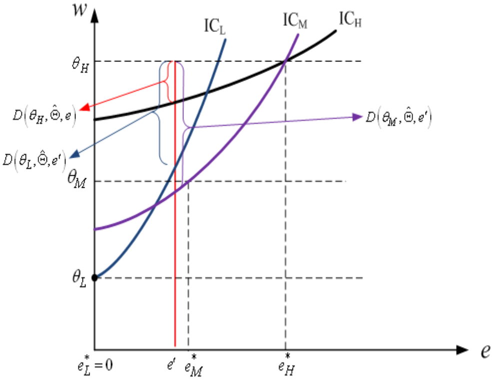
- Therefore, there does not exist any type of worker in the set $\Theta^{**}(e) = \{\theta_M, \theta_H\}$ for whom sending message $e \in (\widehat{e}, e_H^*)$ is beneficial.

D1-criterion. First Step

- Let us now check if the previous separating equilibrium (e_L^*, e_M^*, e_H^*) survives the D_1-Criterion.
- Let us consider the off-the-equilibrium message e' (in red color, in the following figure).
- First, we need to construct sets $D\left(\theta_K, \widehat{\Theta}, e'\right)$ for $K = \{L, M, H\}$, representing the set of wage offers for which a θ_K-worker is better-off when he deviates towards message e' than when he sends his equilibrium message.

D1-criterion. First Step

- Let us illustrate sets $D\left(\theta_K, \widehat{\Theta}, e'\right)$: wage offers for which a θ_K−worker is better-off by sending e' (in red color) than by sending his equilibrium message:



D1-criterion. First Step

- As we can check from the previous figure:

$$\left[D\left(\theta_H, \widehat{\Theta}, e'\right) \cup D^{\circ}\left(\theta_H, \widehat{\Theta}, e'\right)\right] \subset D\left(\theta_M, \widehat{\Theta}, e'\right)$$

Hence, the θ_M-worker has more incentives to deviate than the θ_H-worker. And similarly,

$$\left[D\left(\theta_L, \widehat{\Theta}, e'\right) \cup D^{\circ}\left(\theta_L, \widehat{\Theta}, e'\right)\right] \subset D\left(\theta_M, \widehat{\Theta}, e'\right)$$

the θ_M-worker has more incentives to deviate than the θ_L-worker.

- So, applying the D_1-criterion, the θ_M-worker is the most likely type of sending the message e'. Hence, $\Theta^{**}(e') = \{\theta_M\}$.

D1-criterion. Second Step

- Given $\Theta^{**}(e') = \{\theta_M\}$, firms offer $w(e') = \theta_M$ Therefore, for the θ_M-worker

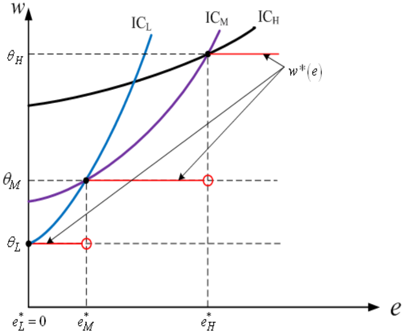
$$\underbrace{\min_{a \in W^*(\Theta^{**}(e'), e')} u_M(e', w, \theta_M)}_{\theta_M - c(e', \theta_M)} > \underbrace{u_M^*(\theta_M)}_{\theta_M - c(e_M, \theta_M)}$$

Since $e' < e_M$ and $c_e(e, \theta) > 0$, which implies $c(e', \theta_M) < c(e_M, \theta_M)$.

- Therefore, the separating PBE *violates* the D_1-criterion.

D1-criterion. Second Step

- Applying the D_1-Criterion to all separating equilibria of this game, we can delete all separating PBEs...
  - except for the efficient (Riley) equilibrium outcome (represented in the figure).



Conclusions

- The set of strategy profiles that can be supported as PBE in Signaling games is usually very large, and contains equilibria predicting "insensible" behaviors.
- The Intuitive and D_1-Criteria are a useful to eliminate multiple equilibria.

Conclusions

- In their application, they both share a *common* second step, but *differ* in their first step. In particular, they differ in how to restrict the set of senders who could benefit by deviating from their equilibrium message:
  - Intuitive Criterion: For which sender/s *there is at least one action* of the responder that is beneficial?
  - D_1-Criterion: For which sender/s *most of the actions* of the responder are beneficial?
- The set of equilibria surviving the Intuitive Criterion might coincide with those surviving the D_1-Criterion, but generally...
  - the latter is a subset of the former.