Journal of Industrial Organization Education

Volume 5, *Issue* 1 | 2010 | *Article* 7

The Intuitive and Divinity Criterion: Interpretation and Step-by-Step Examples

Felix Munoz-Garcia, *Washington State University*
Ana Espinola-Arredondo, *Washington State University*

Recommended Citation:
Munoz-Garcia, Felix and Espinola-Arredondo, Ana (2010) "The Intuitive and Divinity Criterion: Interpretation and Step-by-Step Examples," *Journal of Industrial Organization Education*: Vol. 5: Iss. 1, Article 7.
DOI: 10.2202/1935-5041.1024

The Intuitive and Divinity Criterion: Interpretation and Step-by-Step Examples

Felix Munoz-Garcia and Ana Espinola-Arredondo

Abstract

The paper presents an intuitive explanation of the Cho and Kreps' (1987) Intuitive Criterion, and the Banks and Sobel's (1987) Divinity Criterion (also referred as D1-Criterion). We provide multiple examples in which students can understand, step-by-step, the two main phases involved in these refinement criteria. Furthermore, we present economic settings in which the Cho and Kreps' (1987) Intuitive Criterion does not restrict the set of equilibria, while the Banks and Sobel's (1987) Divinity Criterion help refine the set of admissible separating equilibria in signaling games.

KEYWORDS: signaling games, refinement criteria, intuitive criterion, divinity criterion

Author Notes: We thank the editors for their helpful suggestions, and Kim Dupin for her great assistance.

Introduction

Many economic contexts can be understood as sequential-move games involving elements of incomplete information among firms, consumers, etc., since in few cases every agent knows all the relevant information about other agents in the economy. This situation has been extensively analyzed in economics using signaling games, whereby one agent, privately informed about some relevant characteristic, chooses an action that might reveal information to other agents. Signaling games are then an excellent tool to explain a wide array of economic situations from the role of education in the labor market (Spence, 1973) to the practice of limit pricing by firms (Battacharya, 1979, and Kose and Williams, 1985), and from dividend policy (Milgrom and Roberts, 1982) to the type of warranties firms offer to their customers (Gal-Or, 1989). However, one of the main drawbacks of this class of games is that the set of strategy profiles that can be supported as Perfect Bayesian equilibria is usually relatively large, limiting the predictive power of the model. In addition, a second disadvantage is that some of these equilibria predict insensible behavior from the players. Refinement criteria as the Cho and Kreps' (1987) "Intuitive Criterion" and the Banks and Sobel's (1987) "Universal Divinity" Criterion (also referred as the D_1-Criterion) help overcome these potential disadvantages. In fact, multiple results in the industrial organization literature rely on the application of some of these refinement criteria.

Few game theory or industrial organization textbooks, however, offer an intuitive and applied approach to refinement criteria in signaling games. One of the objectives of this paper is to provide a gentle introduction to the Cho and Kreps' (1987) Intuitive Criterion and the Banks and Sobel's (1987) Divinity Criterion. We use multiple step-by-step examples to help understand the two main stages involved in both of these refinement criteria. In particular, the analysis focuses on the Spence's labor signaling model, assuming two types of workers, and discusses how the application of the Cho and Kreps' (1987) Intuitive Criterion is sufficient to eliminate all but one equilibria. We then show that under more than two types of workers, in contrast, the former criterion does not eliminate any equilibria, and we must rely on a more powerful refinement criterion, such as the Divinity Criterion, in order to restrict the set of equilibria in the signaling game.

The paper is structured as follows. First, we describe the Cho and Kreps' (1987) Intuitive Criterion, providing two examples: the signaling game that a monetary authority plays with a labor union, and the labor market signaling game. Afterwards, section four presents the Banks and Sobel's (1987) Divinity Criterion, with an example of its application. Section five answers the question "When do we need to apply the Divinity Criterion?" by providing an example of the Spence's labor market signaling game with three types of workers (in which the Intuitive

Criterion does not restrict the set of equilibria).

Signaling games

Consider a sequential-move game with the following time structure:

1. Nature reveals to player i some piece of private information (e.g., cost structure, the state of market demand, etc.). We denote this information as player i's type θ_i where $\theta_i \in \Theta$. In the previous examples, the set of types Θ might be $\Theta = \{$High costs, Low costs$\}$ for production costs or $\Theta = \{$High demand, Low demand$\}$ for market demand.[1]

2. Then, player i, who privately observes θ_i, chooses an action which is observed by all players moving afterwards. Player i's action may reveal information about his type to player j. For this reason, this action is normally referred to as message m. The player sending such message (player i) is referred to as the "sender," while the player receiving such message is the "receiver."

3. Player j observes message m, but does not know player i's type. He knows the prior probability distribution with which nature selects a given type θ_i from Θ, $\mu(\theta_i) \in [0, 1]$. Player j, observing player i's message, 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 a particular message m.

4. Given his beliefs about player i's type $\mu(\theta_i|m)$, player j selects an optimal action, a, as a best response to player i's message, m.

In a Perfect Bayesian equilibrium of this signaling game, given equilibrium message m^* chosen by the sender, equilibrium action a^* chosen by the receiver, and the sender's type being θ_i, player i's equilibrium payoff is $u_i(m^*, a^*, \theta_i)$, where for convenience $u_i^*(\theta) \equiv u_i(m^*, a^*, \theta_i)$. And similarly, player j's utility when player i's type is θ_i is $u_j(m^*, a^*, \theta_i)$. Finally, let $a \in A^*(\Theta, m)$ denote the action that the receiver optimally selects, after observing message m from the sender, and given that the set of possible types which can potentially send message m is[2] Θ.

[1] For simplicity, we assume a discrete set of types.

[2] Note that this set does not need to coincide with Θ, but it might be restricted to a subset of types, depending on the receiver's updated beliefs about the sender's type after observing message m.

The Intuitive Criterion

First Step. Let us start analyzing the Intuitive Criterion. The first step focuses on those types of senders who can obtain a higher utility level by deviating (i.e., when they send off-the-equilibrium messages) than by keeping their equilibrium message m^* unaltered. Specifically, let us denote this set of agents as the subset of types for which a given off-the-equilibrium message is not equilibrium dominated (i.e., for which the equilibrium payoff does not dominate the highest payoff they could obtain by sending such an off-the-equilibrium message). Formally, for any off-the-equilibrium message m, we construct a subset of types $\Theta^{**}(m) \subset \Theta$ for which m cannot be equilibrium dominated. That is,

$$\Theta^{**}(m) = \left\{ \theta \in \Theta \mid u_i^*(\theta) \leq \max_{a \in A^*(\Theta,m)} u_i(m,a,\theta) \right\} \tag{1}$$

Intuitively, expression (1) states that, from all types in Θ, we restrict our attention to those types of agents for which sending the off-the-equilibrium message *could* give them a utility level higher than that in equilibrium, $u_i^*(\theta)$. Note the emphasis on "could" since $\max_{a \in A^*(\Theta,m)} u_i(m,a,\theta)$ represents the highest payoff that a θ-type can achieve by sending the off-the-equilibrium message[3] m. In short, we can interpret $\Theta^{**}(m)$ as the subset of senders who *could achieve* a higher utility level by sending the off-the-equilibrium message m rather than their equilibrium message m^*.

Second Step. The second step of the Intuitive Criterion[4] considers the subset of types for which the off-the-equilibrium message m is not equilibrium dominated, $\Theta^{**}(m)$, and checks if the equilibrium strategy profile (m^*,a^*), with associated equilibrium payoff for the sender $u_i^*(\theta)$, satisfies

$$\min_{a \in A^*(\Theta^{**}(m),m)} u_i(m,a,\theta) > u_i^*(\theta) \quad \text{for some } \theta \in \Theta^{**}(m) \tag{2}$$

Let us interpret the former inequality: once beliefs are restricted to $\Theta^{**}(m)$, the originally proposed equilibrium with payoff $u_i^*(\theta)$ cannot survive the Intuitive Criterion if there is a type of agent, θ, and a message he can send, m, that improves

[3]Note that the maximization problem is with respect to the follower's response, a, among the set of best responses to message, m, and given the set of all possible senders, Θ, for all possible off-the-equilibrium beliefs.

[4]As section 4 discusses, the Intuitive and D_1-Criterion *share* their second step (after restricting the set of types who could have sent a given off-the-equilibrium message). In contrast, these refinement criteria *differ* in the first step which determines the subset of types who could benefit from sending a given off-the-equilibrium message m.

his equilibrium payoff, $u_i^*(\theta)$, even if message m is responded with the action providing him the lowest possible payoff, i.e., $\min_{a \in A^*(\Theta^{**}(m),m)} u_i(m, a, \theta)$. That is, there is at least one type of sender who prefers to deviate to a message m which provides him with a higher utility level than his equilibrium message m^*, *regardless* of the response of the receiver.

Formally, an equilibrium strategy profile (m^*,a^*) *violates* the Intuitive Criterion if there is a type of agent θ and an action he can take m such that condition (2) is satisfied. Otherwise, we say that the equilibrium strategy profile *survives* the Intuitive Criterion. As suggested by Vega-Redondo (2003), the deviation by this type of agent can be conceived as if he explains the following to the receiver:

> 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.

Let us next analyze how to apply the Intuitive Criterion in a game with two types of agents and only two responses for the receiver. Afterwards, we extend this analysis to more general games.

Example 1 - Discrete messages

Let us consider the following sequential-move game with incomplete information, where a monetary authority decides whether to announce that the expectation of inflation for the upcoming year is High or Low, and a labor union which reacts to this announcement, demanding high or low wage raises. For simplicity, we assume that the monetary authority is Strong with probability 0.6 or Weak with probability 0.4, where this prior probability distribution is common knowledge among all players.[5] For convenience, we denote by μ the labor union's beliefs that the monetary authority is strong after observing a high inflation announcement, and let γ denote these beliefs after observing that the monetary authority announced a low inflation forecast (see figure 1). Only two strategy profiles can be supported as a Perfect

[5]This signaling game is analogous to the standard "Beer-Quiche game." We prefer to analyze this application to monetary announcements because of its stronger economic content. The same analysis can nonetheless be carried out in games such as the "Beer-Quiche game" applying the same steps used here. We include an application to the Beer-Quiche game in the accompanying homework assignment.

DOI: 10.2202/1935-5041.1024

Bayesian equilibria (PBE) in this signaling game: a pooling PBE with both types of monetary authorities announcing a high level of inflation (High, High); and a separating PBE in which the strong monetary authority announces low inflation, while the weak monetary authority announces high inflation (Low, High). The following figure represents the pooling equilibrium in which both types of monetary authorities send a message of high inflation.Note that this pooling equilibrium seems to predict a relatively insensible behavior from the Strong monetary authority. Indeed, announcing High expectation of inflation for the upcoming year provides a lower payoff than Low inflation, for a given response of the labor union to that announcement. Let us next check if this "insensible" pooling PBE survives the Intuitive Criterion.

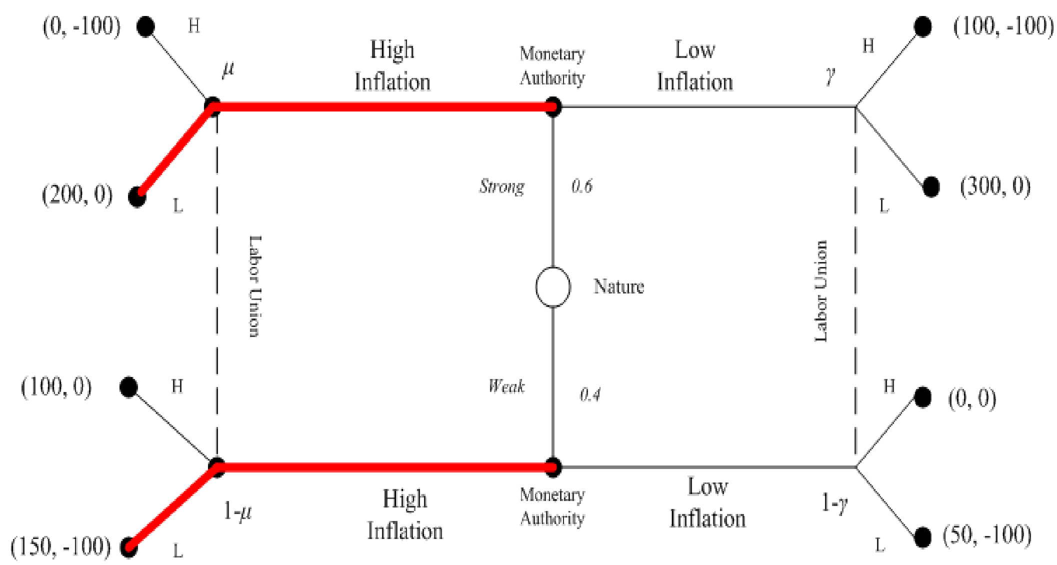

Figure 1. Monetary authority announcements game.

First Step

The first step of the Intuitive Criterion eliminates those off-the-equilibrium messages that are equilibrium dominated. In this case, a message of Low inflation is an off-the-equilibrium message.[6] In order to check if Low is equilibrium dominated,

[6]Note that in the separating PBE (Low,High) all messages are sent in equilibrium by some type of monetary authority. Hence, there are no off-the-equilibrium messages. When no off-the-equilibrium messages can be identified in a given PBE, such PBE survives the Intuitive and the D_1-Criterion. This is a useful result when checking which PBE survives these refinement criteria in signaling games.

we need to find what types of monetary authorities prefer to keep sending High (in equilibrium) rather than deviating by sending Low (off-the-equilibrium message). In particular

$$\begin{aligned} 200 &= u^*_{Mon}(High|Strong) < \max_{a_{Labor}} u_{Mon}(Low|Strong) = 300 \\ 150 &= u^*_{Mon}(High|Weak) > \max_{a_{Labor}} u_{Mon}(Low|Weak) = 50 \end{aligned}$$

The first inequality is indeed satisfied since 200<300 when the monetary authority is Strong. Specifically, the strong monetary authority obtains a payoff of 200 in equilibrium (by sending a high inflation announcement, which is responded with Low). But it could obtain a higher payoff by deviating towards a Low inflation announcement, 300, which arises when the labor union responds with Low wage demands. In contrast, when the monetary authority is Weak, its equilibrium payoff in the pooling PBE, 150, is higher than the maximum that it could obtain by deviating, 50 (which also occurs when the labor union responds to a Low inflation announcement choosing Low wage demands). Hence, the Strong monetary authority could indeed deviate to Low announcements of inflation but the Weak type could not. As a consequence, the subset of types for which the off-the-equilibrium message (Low inflation) is not equilibrium dominated is $\Theta^{**}(Low) = \{Strong\}$, since such a message can only come from the Strong monetary authority. Hence, the labor union's beliefs when observing such a message are $\gamma = 1$ (at the upper right-hand corner of the figure).

Second Step

The second step uses the above restriction on beliefs ($\gamma = 1$) to study if there is a type of monetary authority and a message it could send such that condition (2) is satisfied (i.e., obtaining a higher utility than in equilibrium, regardless of the labor union's response). First, when the labor union observes the off-the-equilibrium message of Low inflation, it responds with Low wage demands, since it concentrates all its beliefs in the node at the upper right-hand side corner of the game tree, i.e., $\gamma = 1$. By sequential rationality, and given this labor union response, the Strong monetary authority prefers to make an announcement of Low inflation levels. Indeed, this announcement is responded by the labor union with Low, providing a payoff of 300 to the monetary authority, which is higher than its equilibrium payoff of 200. Note that the second step of the Intuitive Criterion involves

$$300 = \min_{a_{Labor}} u_{Mon}(High|Strong) > u^*_{Mon}(Low|Strong) = 200$$

conditional on the belief that the Low inflation announcement can only come from the Strong monetary authority, i.e., $a_{Labor} \in A^*(Strong, Low)$. Then, the Strong monetary authority prefers to deviate from the pooling PBE of (High, High). Therefore, the pooling PBE of (High, High) *violates* the Intuitive Criterion given that there exist a type of sender (Strong monetary authority) and a message (Low) which gives that sender a higher utility level than in equilibrium, regardless of the response of the follower (labor union). □

Example 2 - Continuous messages
Let us now analyze the traditional Spence's (1973) signaling game with two types of workers, one with a high productivity level, and the other with a low productivity, $\Theta = \{\theta_H, \theta_L\}$, and a continuum of wage offers $w \in [0,1]$. The worker acts as the sender in this game because he acquires a particular education level that is observed by the firm which is potentially interested in hiring him. Education is, nonetheless, not enhancing the worker's productivity, and hence it serves only as a signal about the worker's productivity level. In particular, the firm's profit function is $\pi(w, \theta) = \theta - w$, and the worker's utility function is $u_i(e, w, \theta_K) = w - c(e, \theta_K)$, where $c(e, \theta_K)$ represents the worker's cost of acquiring education level e. Consider that acquiring no education is costless, $c(0, \theta_K) = 0$ for both types of workers. Additionally, assume that the marginal cost of acquiring an additional year of education, $c_e(e, \theta_K)$, is decreasing in the worker's productivity, i.e., $c_e(e, \theta_H) \leq c_e(e, \theta_L)$, and therefore worker's indifference curves satisfy the single-crossing property. Let us analyze one of the separating Perfect Bayesian equilibria of this game, such as that represented in the figure below, where the θ_L-type of worker sends a message of $e_L^* = 0$ years of education, while the θ_H-type of worker acquires $e_H^* = e_2$ years of education. In this case, education "fully reveals" the worker's type, since the firm can perfectly infer the worker's productivity level by observing the education he acquires. As a consequence, the firm offers a low wage offer to workers who acquire no education, $w(e_L^*) = \theta_L$, and a high wage to workers with e_2 years of education,[7] $w(e_H^*) = \theta_H$. In figure 2, IC_L and IC_H denote the indifference curves for the low and high-productivity workers, respectively, in this equilibrium. Since higher wages increase worker's utility and education is costly to acquire, indifference curves to the northwest (higher wages and less education) are associated to higher utility levels.

[7] Note that this separating PBE can be supported if off-the-equilibrium education levels $e \neq e_L^*, e_H^*$ are responded with wage offers such as $w(e) = \theta_L$.

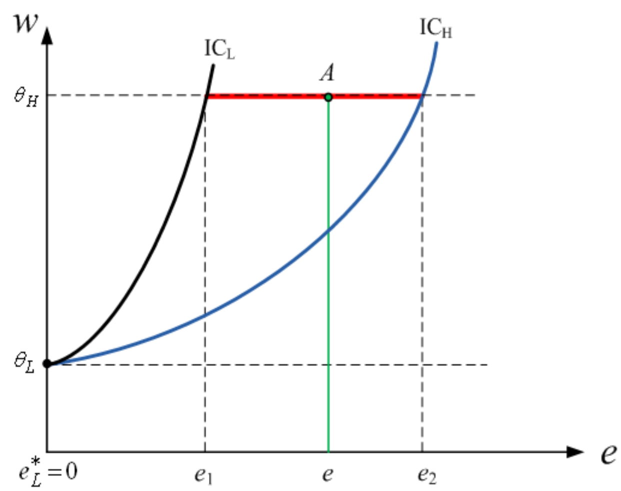

Figure 2. Labor market game with two types.

First step

Consider now that the firm observes an off-the-equilibrium message $e \in (e_1, e_2)$, as indicated in figure 2. In order to study what type of worker might have sent such a message, let us apply the previous analysis of equilibrium dominance. In particular, for the θ_L-type of worker, we have that

$$u_L^*(\theta_L) > \max_{w \in W^*(\theta,m)} u_L(e, w, \theta_L)$$

That is, his equilibrium payoff $u_L^*(\theta_L)$ is higher than the maximal utility he could obtain if the firm offered him the highest possible salary. In other words, his equilibrium payoff from sending $e_L^* = 0$, $u_L^*(\theta_L) = \theta_L - c(0, \theta_L) = \theta_L$, is higher than the *highest* payoff he could obtain by sending the off-the-equilibrium message e, $\theta_H - c(e, \theta_L)$ (when the firm believes that the worker is a θ_H-type and pays him a salary of $w(e) = \theta_H$). Therefore, the above inequality implies that for any off-the-equilibrium message $e \in (e_1, e_2)$,

$$c(e, \theta_L) > \theta_H - \theta_L$$

Intuitively, the cost from acquiring e years of education for the low-productivity worker, $c(e, \theta_L)$, exceeds the wage increase, $\theta_H - \theta_L$, he can experience if the firm believes that, because of acquiring education level e, he must be a high productivity worker, paying him $w(e) = \theta_H$. Graphically, the θ_L-worker's indifference curve when he acquires the equilibrium education level $e_L^* = 0$ is represented by IC_L,

DOI: 10.2202/1935-5041.1024

and the indifference curve from acquiring the (off-the-equilibrium) education level e and receiving a salary of $w(e) = \theta_H$ would cross point A. Clearly, the indifference curve associated to education level e_L^* implies a higher utility level than that associated with e, even when the salary the worker receives is $w(e) = \theta_H$. This process can then be repeated for any off-the-equilibrium message $e \in (e_1, e_2)$, concluding that the θ_L-type of worker does not send such a message, because it is *equilibrium dominated*.

Let us now apply the same analysis of equilibrium dominance to the θ_H-worker. This type of worker can send the off-the-equilibrium message e since:

$$\begin{aligned} u_i^*(\theta_H) &< \max_{w \in W^*(\theta, e)} u_H(e, w, \theta_H) \\ \theta_H - c(e_2, \theta_H) &< \theta_H - c(e, \theta_H) \end{aligned}$$

Intuitively, he receives the same salary as in equilibrium ($w(e) = \theta_H$) but incurs fewer costs because of acquiring a lower education level, i.e., $c(e_2, \theta_H) > c(e, \theta_H)$ since $e_2 > e$. Hence, the equilibrium payoff of this worker is lower than the maximal payoff he could obtain if the firm manager offers him a salary of $w(e) = \theta_H$ after observing education level e. Graphically, indifference curves through point A (if he receives the high salary) are associated to higher utility levels than that in equilibrium, as represented by IC_H. Therefore, off-the-equilibrium message e is *not* equilibrium dominated for the θ_H-worker, but it *is* for the θ_L-worker. We can now state which is the subset of types that the receiver (firm) considers after observing the off-the-equilibrium message e. In particular, the firm concentrates its beliefs on the θ_H-type of worker, since he is the only type whose utility can increase by deviating from his equilibrium message. Formally, we state that the subset of types for which message e is not equilibrium dominated is given by $\Theta^{**}(e) = \{\theta_H\}$.

Second step

The subset of types who could have sent message e is $\Theta^{**}(e) = \{\theta_H\}$. Then, the firm offers a wage of $w(e) = \theta_H$ given that it assigns full probability to the worker being a high-productivity worker. Note that the minimal utility level that the worker can achieve from sending the off-the-equilibrium message e is

$$\min_{w \in W^*(\Theta^{**}(e), e)} u_H(e, w, \theta_H) = \theta_H - c(e, \theta_H)$$

and the equilibrium payoff for the equilibrium education level e_2 is $u_H^*(\theta_H) = \theta_H - c(e_2, \theta_H)$. Given that $c(e_2, \theta_H) > c(e, \theta_H)$, we have that $\theta_H - c(e, \theta_H) > \theta_H - c(e_2, \theta_H)$. Hence,

$$\min_{w \in W^*(\Theta^{**}(e), e)} u_H(e, a, \theta_H) > u_H^*(\theta_H)$$

Therefore, the separating PBE where workers acquire education levels $\{e_L^*(\theta_L), e_H^*(\theta_H)\}=\{0, e_2\}$ *violates* the Intuition Criterion because there exists a type of worker, θ_H, and an off-the-equilibrium message $e \in (e_1, e_2)$, for which the above inequality is satisfied. Intuitively, the θ_H-worker can signal his type (productivity level) to the firm by acquiring less education than in the separating equilibrium where he acquires $e_H^* = e_2$.

It can be verified that all separating equilibria can be eliminated following the above procedure, except for the equilibrium in which the low-type acquires zero education and the high-type acquires education level e_1. The surviving separating equilibria is usually referred to as the *efficient* equilibrium outcome (or Riley outcome, after Riley, 1979), since it is the equilibrium in which workers spend the least amount of resources in signaling to the firm their different productivity levels. Specifically, the θ_L-type acquires an education level of $e_L^* = 0$ and the θ_H-type acquires the minimal education level that allows him to separate himself from the θ_L-type, $e_H^* = e_1$. We illustrate this equilibrium in the following figure.

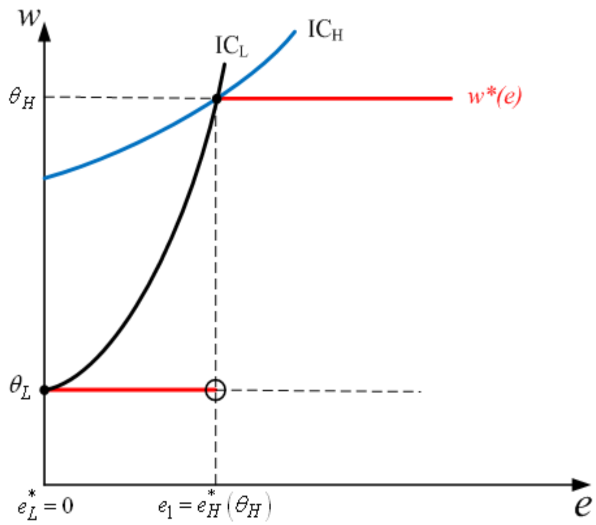

Figure 3. Efficient separating equlibrium.

DOI: 10.2202/1935-5041.1024

The Divinity Criterion

As described in the previous section, the Intuitive Criterion restricts the receiver's beliefs to those type of senders for which deviating towards a given off-the-equilibrium message could improve his equilibrium payoff. If more than one type of sender could benefit from such deviation, however, the Intuitive Criterion assumes that the receiver's beliefs assign the same weight to all potential deviators (as if they were all equally likely to deviate towards the off-the-equilibrium message). The D_1-Criterion, instead, considers that, among all potential deviators, the receiver restricts his beliefs to only those types of senders who *most likely* send the off-the-equilibrium message.

In particular, this restriction on beliefs is analyzed by focusing on the sender for whom most of the responder's actions provide a payoff above his equilibrium payoff. Formally,[8] for any off-the-equilibrium message m, let us define

$$D\left(\theta, \widehat{\Theta}, m\right) := \bigcup_{\mu:\mu\left(\widehat{\Theta}|m\right)=1} \{a \in MBR(\mu, m) \mid u_i^*(\theta) < u_i(m, a, \theta)\} \quad (3)$$

as the set of mixed best responses[9] (MBR) of the receiver for which the θ-type of sender is *strictly better-off* deviating towards message m than sending his equilibrium message m^*. Note that $\mu\left(\widehat{\Theta} \mid m\right) = 1$ in the previous definition represents that the receiver believes that message m only comes from types in the subset $\widehat{\Theta} \in \Theta$. Let us also define

$$D^\circ\left(\theta, \widehat{\Theta}, m\right) := \bigcup_{\mu:\mu\left(\widehat{\Theta}|m\right)=1} \{a \in MBR(\mu, m) \mid u_i^*(\theta) = u_i(m, a, \theta)\} \quad (4)$$

as the set of MBR of the receiver that make the θ-type *indifferent* between deviating towards message m and sending his equilibrium message m^*. Let us next describe the first step of the Divinity Criterion.

First Step. A θ-type can be *deleted* if there is another θ'-type such that, when the off-the-equilibrium message m is observed

$$\left[D\left(\theta, \widehat{\Theta}, m\right) \cup D^\circ\left(\theta, \widehat{\Theta}, m\right)\right] \subset D\left(\theta', \widehat{\Theta}, m\right) \quad (5)$$

[8]In this section, we follow Fudenberg and Tirole's (2002) notation (see pp. 452-453), but applying the Divinity Criterion to the Spence's labor market signaling game.

[9]The set of mixed best responses (MBR) of the receiver to a given message m from the sender includes both the actions that the receiver chooses using pure strategies, and those involving mixed strategies.

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, $D^{\circ}\left(\theta, \widehat{\Theta}, m\right)$. Intuitively, after receiving message m there are more best responses of the receiver that improve the θ'-type's equilibrium payoff than there are for the θ-type. As a consequence, the θ'-type is the sender who is most likely to deviate from his equilibrium message m^* to the off-the-equilibrium message m. We continue this comparison for all types of senders, deleting those for which there is another type of sender who is more likely to deviate towards m. Finally, the set of types that cannot be deleted after using this procedure is denoted by $\Theta^{**}(m)$.

Second Step. As discussed in the previous section, the second step of both the Intuitive and the D_1-Criterion, analyzes 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

$$\min_{a \in A^*(\Theta^{**}(m), m)} u_i(m, a, \theta) > u_i^*(\theta) \quad \text{for some } \theta \in \Theta^{**}(m) \tag{6}$$

Example 3 - Continuous messages.
Figure 4 represents the labor market signaling game described in example 2. Similarly to example 2, let us analyze if the separating PBE where $e_L^* = 0$ and $e_H^* = e_2$ survives the D_1-Criterion.

First step
First, after sending an off-the-equilibrium message e', the set of wage offers that improve the equilibrium payoff of the low-productivity worker, $D\left(\theta_L, \widehat{\Theta}, e'\right)$, is smaller than that for the high-productivity worker, $D\left(\theta_H, \widehat{\Theta}, e'\right)$, i.e., $D\left(\theta_L, \widehat{\Theta}, e'\right) \subset D\left(\theta_H, \widehat{\Theta}, e'\right)$. These two sets are represented in figure 4 below. Intuitively, after sending message e', there are more wage offers that improve the equilibrium payoff of the high-productivity worker than that of the low-productivity worker; see sets $D\left(\theta_H, \widehat{\Theta}, e'\right)$ and $D\left(\theta_L, \widehat{\Theta}, e'\right)$, respectively, in figure 4. Hence, the θ_H-type is more likely to send message e'. As a consequence, the firm, after receiving message e', restricts its beliefs to $\Theta^{**}(e') = \{\theta_H\}$.

On the other hand, after observing the off-the-equilibrium message e'', the firm knows that sending such a message would never be payoff improving for the low-productivity worker, i.e., $D\left(\theta_L, \widehat{\Theta}, e''\right) = \emptyset$. However, sending e'' might be profitable for the high-productivity worker. Indeed, as the figure indicates, the high-productivity worker can receive some wage offers that would raise his utility level beyond his equilibrium payoff. Hence, when observing the off-the-equilibrium message e'', $D\left(\theta_L, \widehat{\Theta}, e''\right) \subset D\left(\theta_H, \widehat{\Theta}, e''\right)$, as figure 4 indicates. Therefore, message e'' is most likely to come from a θ_H-worker, $\Theta^{**}(e'') = \{\theta_H\}$. Repeating this process for any off-the-equilibrium message, we can prove that, after observing any education level e off-the-equilibrium path, firm's beliefs are restricted to $\Theta^{**}(e) = \{\theta_H\}$.

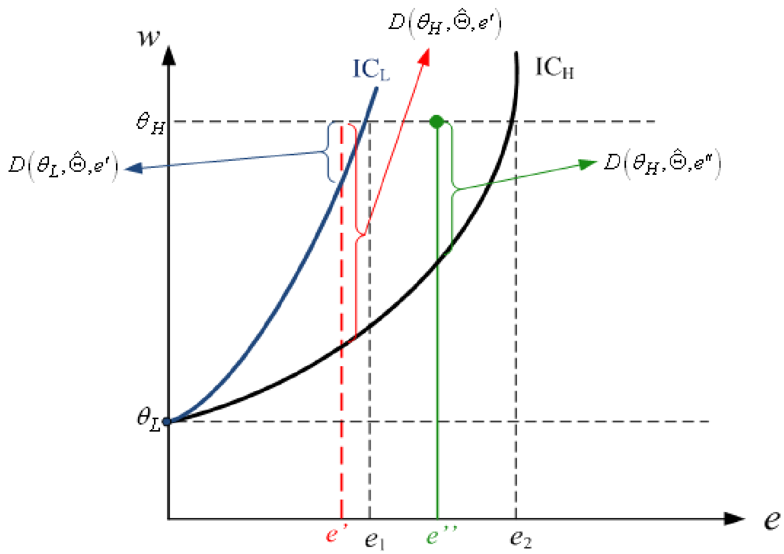

Figure 4. Applying the D_1-criterion to the labor market game.

Second step

After restricting the subset of types who could have sent any off-the-equilibrium message e to $\Theta^{**}(e) = \{\theta_H\}$, the firm offers a wage of $w(e) = \theta_H$ given that it assigns full probability to the worker being a high-productivity type. Now, let us apply the same methodology as in the second step of the Intuitive Criterion. First note that, the minimal utility level that the worker can achieve from sending the off-the-equilibrium message e is

$$\min_{w \in W^*(\Theta^{**}(e), e)} u_H(e, w, \theta_H) = \theta_H - c(e, \theta_H)$$

and the equilibrium payoff for the equilibrium education level e_2 is $u_H^*(\theta_H) = \theta_H - c(e_2, \theta_H)$. Given that $c(e_2, \theta_H) > c(e, \theta_H)$, we have that $\theta_H - c(e, \theta_H) > \theta_H - c(e_2, \theta_H)$. Hence,

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

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 because there exists a type of sender (θ_H-worker) and an off-the-equilibrium message, e, for which the above inequality is satisfied. Similarly to the Intuitive Criterion, one can show that all separating equilibria in this game can be eliminated using the D_1-Criterion, except for the efficient (Riley) outcome, where the low-productivity worker acquires an education level of $e_L^* = 0$ and the high-productivity worker only acquires education $e_H^* = e_1$.

When do we need to apply the D_1-Criterion?

In the previous section, we described the Intuitive and D_1-Criterion, and examined that, when there are only $n = 2$ types of senders, the equilibria that survive these two equilibrium refinement *coincide*. However, as we show in this section, this might not be the case when there are $n > 2$ types of senders (for instance, more than two types of workers in the labor market signaling game). First, we describe how the application of the Intuitive Criterion to signaling games with more than two senders might *not* help us restrict the set of equilibria, and then we show that the D_1-Criterion reduces the set of equilibria in this class of games.

Example 4 - Continuous messages with n= 3 types of workers. Intuitive Criterion.
Let us analyze if the separating PBE $\{e_L^*(\theta_L), e_M^*(\theta_M), e_H^*(\theta_H)\} = \{0, e_M, e_H\}$ survives the Intuitive Criterion. (This is one of the multiple separating equilibria in the Spence's signaling game with three types of workers).

DOI: 10.2202/1935-5041.1024

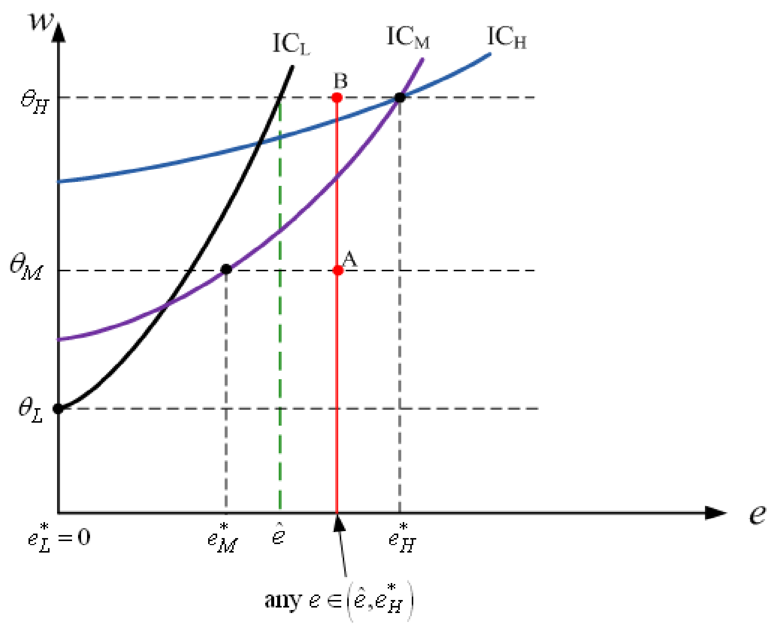

Figure 5. Intuitive Criterion with 3 types of workers.

First step

First, we need to construct the subset of types $\Theta^{**}(e) \subset \Theta$ for which the off-the-equilibrium message, $e \in (\widehat{e}, e_H)$, is never equilibrium dominated (see message $e \in (\widehat{e}, e_H)$ in figure 5). That is,

$$\Theta^{**}(e) = \left\{ \theta \in \Theta \,|u_i^*(\theta) \underset{w \,\in\, W^*(\Theta,e)}{\leq \max} u_i(e, w, \theta) \right\}$$

Let us start checking this condition for the L-type. In particular, note that

$$u_L^*(\theta_L) > \max_{w \in W^*(\Theta,e)} u_L(e, w, \theta_L)$$

since $u^*(\theta_L) = \theta_L - c(0, \theta_L) = \theta_L$ and $\max_{w \in W^*(\Theta,e)} u_L(e, w, \theta_L) = \theta_H - c(e, \theta_L)$. That is, the above condition implies $c(e, \theta_L) > \theta_H - \theta_L$, indicating that the cost of acquiring e years of education for the L-type of worker exceeds his potential salary gain, $\theta_H - \theta_L$. Graphically, his equilibrium utility level, $u_L^*(\theta_L)$, is represented by the indifference curve IC_L, and $\max_{w \in W^*(\Theta,e)} u_L(e, w, \theta_L)$ would correspond to the downward shift of the indifference curve IC_L that passes through point B (when the worker is paid the high-productivity wage $w(e) = \theta_H$). So, θ_L does not send a message $e \in (\widehat{e}, e_H)$. In contrast, θ_M-workers could send such a message $e \in (\widehat{e}, e_H)$ because

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

since $\theta_M - c(e_M^*, \theta_M) < \theta_H - c(e, \theta_M)$, or alternatively, $c(e, \theta_M) - c(e_M^*, \theta_M) < \theta_H - \theta_M$, reflecting that the cost of acquiring $e - e_M^*$ additional years of education is offset by the increase in salary that the M-type of worker can obtain if the firm offers him a high-productivity wage $w(e) = \theta_H$. Graphically, $\max_{w \in W^*(\Theta, e)} u_M(e, w, \theta_M)$ is represented by the indifference curve of the M-type of worker, passing through point B, which is associated to a higher utility level than his equilibrium utility, as represented by IC_M. Similarly for the H-type of worker,

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

since $\theta_H - c(e_H^*, \theta_H) < \theta_H - c(e, \theta_H)$, given that $c(e_H^*, \theta_H) > c(e, \theta_H)$. Intuitively, by deviating towards e the H-type of worker does not modify his salary (if the firm maintains a wage offer of $w(e) = \theta_H$), but he does not incur so much education costs. In figure 5, $\max_{w \in W^*(\Theta, e)} u_H(e, w, \theta_H)$ is illustrated by the indifference curve of the H-type of worker passing through point B. This curve represents a utility level which is above the equilibrium payoff of the worker (see IC_H). Hence, education levels in the interval $e \in (\widehat{e}, e_H)$ are *not* equilibrium dominated for the θ_M and θ_H workers, since they both have incentives to deviate from their equilibrium messages. Therefore, 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:

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

Second step

Once we have determined $\Theta^{**}(e) = \{\theta_M, \theta_H\}$ for all $e \in (\widehat{e}, e_H)$, we need to find a type θ that can be tempted to send an education level in $e \in (\widehat{e}, e_H)$ anticipating that firms' best response to this education level will be a wage offer somewhere in between $w(e) = \theta_M$ and $w(e) = \theta_H$. First, for the θ_M-worker, if he thinks pessimistically, he can consider the case in which his deviation towards message $e \in (\widehat{e}, e_H)$ is interpreted by firms as coming from a θ_M-worker. Hence, the firm will offer $w(e) = \theta_M$ and θ_M-workers' indifference curve will pass through point A, being *below* the indifference curve corresponding to his equilibrium payoff. Therefore,

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

DOI: 10.2202/1935-5041.1024

and the M-type's equilibrium payoff exceeds the lowest payoff he can obtain from deviating towards e. Similarly for the θ_H-worker, if he thinks pessimistically, he can consider the same situation described above. That is, firms believe that any message $e \in (\widehat{e}, e_H)$ must come from a θ_M-worker, and as a consequence they offer $w(e) = \theta_M$. Therefore, θ_H-workers' indifference curve through point A is also *below* his indifference curve at the equilibrium payoff. So,

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

Therefore, there does not exist any type of worker in the set $\Theta^{**}(e) = \{\theta_M, \theta_H\}$ who would deviate towards the off-the-equilibrium message $e \in (\widehat{e}, e_H)$. Hence, the separating PBE specified in figure 5 does not violate the Intuitive Criterion. Thus, the application of the Intuitive Criterion does not necessarily eliminate separating PBE with $n > 2$ types of senders. $\square$

Example 4b - Continuous messages with $n=$ 3 types of workers. D_1-Criterion. We now show that the D_1-Criterion restricts the set of equilibria, even if the set of senders is strictly larger than $n = 2$.

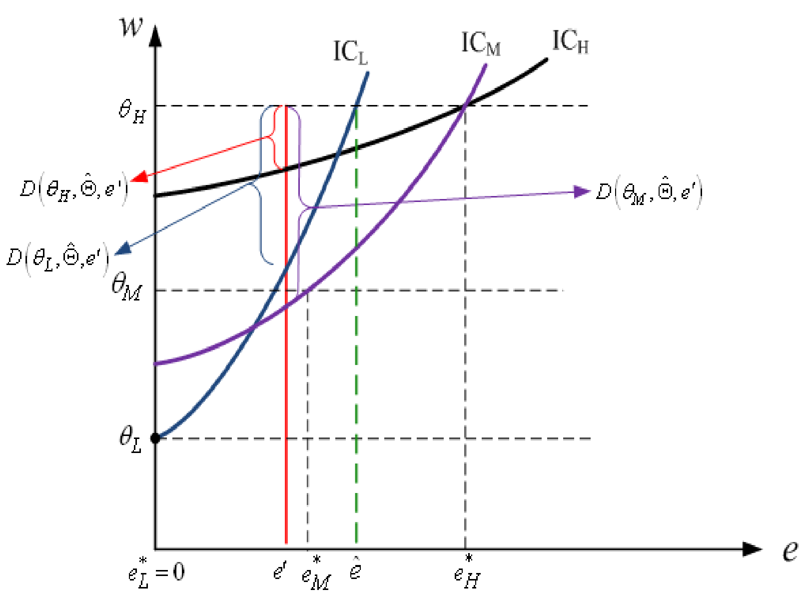

Figure 6. The D_1-Criterion with 3 types of workers.

First step

We reduce the firms' beliefs by considering who is the type of worker who most probably sent message e'. We next define the set of wage offers for which a worker of type $i = \{L, M, H\}$ can improve his equilibrium utility level, $u_i^* (\theta_i)$, by acquiring education level e' rather than his equilibrium education of e_i^* (see figure 6).

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

Applying this concept to the L and M-types of workers, we have

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

Intuitively, the set of wage offers for which the M-type of worker improves his equilibrium utility is larger than those for which the L-type of worker improves his (see figure 6), making the former more likely to deviate towards e' than the latter. So, applying the D_1-Criterion, we can eliminate θ_L-type worker from having sent e'. Similarly,

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

and the M-type of worker is more likely to deviate towards education level e' than the H-type of worker. So, applying the D_1-Criterion, we can eliminate θ_H-type worker from having sent e'. Hence, firms beliefs when observing an education level of e' can be restricted to only the M-type of worker, $\Theta^{**}\left(e'\right) = \{\theta_M\}$.

Second step

Given $\Theta^{**}\left(e'\right) = \{\theta_M\}$, firms offer a wage $w\left(e'\right) = \theta_M$ when observing an education level of e'. Therefore, for the θ_M-worker we have that

$$\min_{a \in W^*(\Theta^{**}(e'), e')} u_M\left(e', w, \theta_M\right) = w\left(e'\right) - c\left(e', \theta_M\right) = \theta_M - c\left(e', \theta_M\right)$$

And his equilibrium payoff is

$$u_M^*\left(\theta_M\right) = w\left(e_M\right) - c\left(e_M, \theta_M\right) = \theta_M - c\left(e_M, \theta_M\right)$$

And given that $e' < e_M$ and $c_e\left(e, \theta\right) > 0$, we then have $c\left(e', \theta_M\right) < c\left(e_M, \theta_M\right)$; which implies

$$\theta_M - c\left(e', \theta_M\right) > \theta_M - c\left(e_M, \theta_M\right)$$

That is,

$$\min_{w \in W^*(\Theta^{**}(e'), e')} u_M\left(e', w, \theta_M\right) > u_M^*\left(\theta_M\right)$$

Hence, we have found a type of worker, θ_M, for whom deviating towards e' improves his equilibrium payoff, $u_M^*(\theta_M)$. Therefore, the separating PBE described in figure 6 above *violates* the D_1-Criterion. Repeating this process for all off-the-equilibrium messages, we can delete all separating PBEs, except for the efficient (Riley) equilibrium outcome described in the following figure.

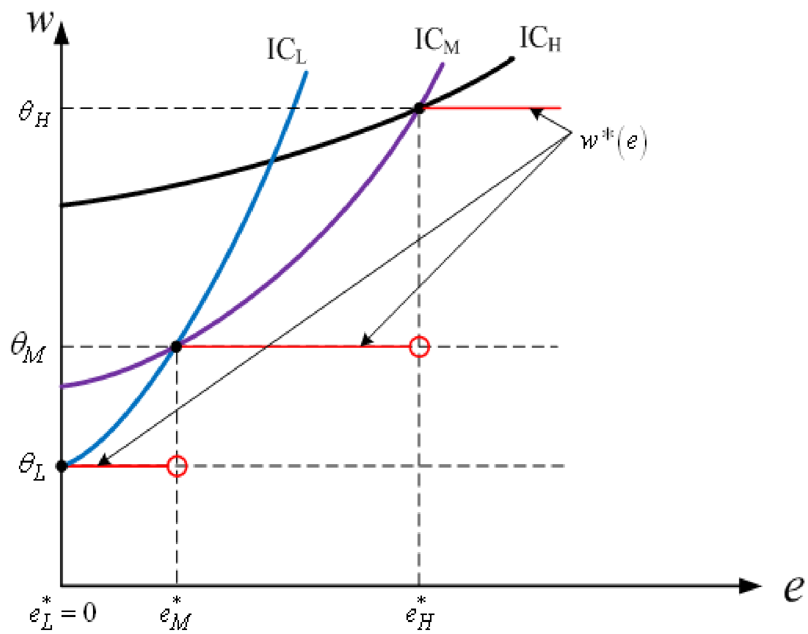

Figure 7. Efficient separating equilibrium with 3 types of workers.

References

1. BANKS, J. AND SOBEL, J. (1987) "Equilibrium selection in signaling games," Econometrica, 55, pp. 647-661.

2. BATTACHARYA, S. (1979) "Imperfect information, dividend policy and the 'bird in the hand' fallacy", Bell Journal of Economics, 9(1), pp. 259-70.

3. CHO, I. AND KREPS, D. (1987) "Signaling games and stable equilibria," Quarterly Journal of Economics, 102, pp. 179-221.

4. FUDENBERG, D. AND TIROLE, J. (2002) *Game Theory*, MIT Press.

5. GAL-OR, ESTHER (1989) "Warranties as a signal of quality", Canadian Journal of Economics, 22(1), pp. 30-51.

6. KOSE, JOHN AND JOSEPH WILLIAMS (1985) “Dividends, dilution and taxes: a signaling equilibrium” Journal of Finance, 40(4), pp. 1053-1069.

7. MILGROM, PAUL AND JOHN ROBERTS (1982), “Limit pricing and entry under incomplete information: an equilibrium analysis”. Econometrica, 50(2), pp. 443-459.

8. RILEY, JOHN G. (1979) “Informational equilibrium,” Econometrica, 47(2), pp. 331-359

9. SPENCE, M (1973) “Job market signaling,” Quarterly Journal of Economics, 87, pp. 355-374

10. VEGA-REDONDO, F. (2003) *Economics and the theory of games*, Cambridge University Press.