INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

547

2020

Testing the sender: When signaling is not enough

Nicolás Figueroa y Carla Guadalupi.

www.economia.puc.cl • ISSN (edición impresa) 0716-7334 • ISSN (edición electrónica) 0717-7593

Testing the sender: When signaling is not enough

Nicolás Figueroa[†] Carla Guadalupi[†]

August 11, 2020

Abstract

We study signaling in the presence of endogenous information acquisition by the receiver. A firm, after observing the worker's costly action, may acquire further information by performing a test on the applicant, and then decide to hire him. We consider different models of information acquisition, including the rational inattention, a generalization of the "truth or noise" and a general grading model. We study test effectiveness as function of beliefs generated through signaling by the worker, and provide clear-cut predictions on the complementarity/substitutability between costly information transmission (signaling) and acquisition, and its implications for the equilibrium. We first show that test effectiveness is non-monotone in beliefs. It exhibits increasing regions, where beliefs and test effectiveness act as complements, with higher beliefs inducing more effective tests, and decreasing regions in which higher beliefs crowd out the firm's information acquisition. We then show that, when beliefs and test effectiveness are complements, the equilibrium involves (at least partial) separation between workers' types. Since the high type is more willing to face a more exacting test than a low type, he will exert costly effort to improve the firm's opinion. When beliefs and test effectiveness are substitutes, any signaling attempt by the high type will be mimicked by the low one who benefits more from relaxed standards and indiscriminate hiring, and the only plausible equilibrium involves both types pooling.

[†]Instituto de Economía, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile, E-mail: nicolasf@uc.cl.

[†]Facultad de Economía y Negocios, Universidad Andres Bello, Av. Fernández Concha 700, Santiago, Chile, E-mail: carla.guadalupi@unab.cl.

1 Introduction

Information is produced and transmitted by many different agents in an economy. Agents with private information may signal it through the choice of a costly action. At the same time, uninformed agents may costly acquire further information. Moreover, agents influence each other, with the amount of information provided by an agent influencing the amount acquired by another, and viceversa. One could argue that, in many cases, the main purpose of signaling private information is to influence the level of information acquisition by others.

Consider the labor market. High-skilled workers may choose to signal their type through education. Firms might complement this information with their own, through interviews, tests, internships or other expensive screening procedures. Something similar happens if the worker is being screened through a probationary period. He can go the extra mile, in order to signal his good fit with the firm, anticipating that his performance will affect the thoroughness of the review that precedes a permanent hiring. In both cases, the firm's opinion about the worker, influenced by his signaling effort, determines the way it acquires information. This, in turn, is anticipated by the worker when choosing the costly action.

In this paper, we study signaling in the presence of endogenous information acquisition by the receiver. Specifically, the receiver optimally chooses the accuracy of information to acquire conditional on the action chosen by the sender. The main objective is to understand the complementarity/substitutability between costly information transmission (signaling) and acquisition, and its implications for the equilibrium.

We consider a simple setting where a firm is deciding whether to hire a worker. The worker's productivity depends on his fit with the firm (specific skills), which is initially unknown to both. Since the market is characterized by wage rigidities, only a worker with a good fit is worth hiring. Nevertheless, the worker is privatly informed about his type (general skills), which makes him more or less likely to have a good fit with the firm.

The worker might perform an action, whose cost is type-independent, to signal his type. After observing this action, the firm updates beliefs and chooses an infomation strategy (test). Given the test results, it makes a hiring decision. A more accurate test is also more expensive, therefore the firm must trade-off accuracy and cost. In particular, we define test effectiveness as the difference between the probabilities of hiring a worker with a good and bad fit.

A test serves two main purposes: it identifies workers with a good fit (searching for diamonds), while allowing the firm to make less mistakes (avoiding lemons). Higher beliefs generated through signaling by the worker then have two countervailing effects on the test chosen by the firm. Since there are more good-fit workers on the market, the searching for diamonds motive becomes more important, and the firm improves test effectiveness as beliefs get better. At the same time, avoiding lemons becomes less important, since the fraction of bad fits is lower, so that the firm decreases test effectiveness. The overall effect is then ambiguous.

We first show that test effectiveness is non-monotone in beliefs. It exhibits increasing regions, where beliefs and test effectiveness act as complements, with higher beliefs inducing more effective tests, and decreasing regions in which higher beliefs crowd out the firm's information acquisition. This pattern in turn affects the worker's incentives to signal.

In our model, the high type has no cost advantage, therefore the worker's payoff does not show supermodularity in type and action as in Spence (1973). Instead, the rationale for signaling comes from the fact that beliefs change the effectiveness of the test performed by the firm. Specifically, the worker's payoffs are supermodular in type and test effectiveness, therefore a more effective test is more valuable to the high type. This implies that the worker's payoffs are supermodular in type and beliefs as long as beliefs and test effectiveness are complements.

We then show that, when beliefs and test effectiveness are complements, the equilibrium involves (at least partial) separation between workers' types. Since the high type is more willing to face a more exacting (but more rewarding) test than a low type, being less likely to flunk it, he will exert costly effort to improve the firm's opinion. When beliefs and test effectiveness are substitutes, any signaling attempt by the high type will be mimicked by the low one who benefits more from relaxed standards and indiscriminate hiring. Therefore the only plausible equilibrium involves both types pooling at no effort.

Finally, we consider three models of information acquisition, each one capturing different tools available to the firm at the moment of designing a hiring policy. For each setting, we study how beliefs affect test effectiveness and the implications for the signaling equilibrium.

When the firm may tailor the information it acquires to perfectly complement the one provided by workers through signaling, it seems reasonable to put no restriction on the firm's information strategy. In this case, we follow the Rational Inattention approach proposed by Matějka and McKay (2015), where the cost of a test is proportional to the expected entropy reduction between prior and posterior. We show that test effectiveness is inverse U-shaped on beliefs, and high-type workers signal as long as test effectiveness is increasing in beliefs. In particular, when the labor force is weak, and even high-skilled workers have few chances of having a good match with the firm, information is revealed through both signaling and information acquisition. The high type signals his private information up to the point in which test effectiveness is maximized. This could involve total or partial separation due to the existence of a hard limit on information revelation by workers, which is created by the firm endogenously choosing the test. When, on the other hand, the labor force is strong, and firm's prior is high, test effectiveness is decreasing in beliefs and any signaling effort made by the high type would be mimicked by the low type.

In other circumstances, however, the set of tools available to the firm is limited, and the test cannot be perfectly tailored to complement the information already available from signaling. A test might belong to a fixed family, with a cost increasing in the precision of the information it provides (grading models, Daley and Green, 2014). We first consider the "grade or noise" model, a generalization of truth-or-noise framework used, for example, in Lewis and Sappington (1994). With some probability, the test provides an informative grade; with the complementary probability, no news are revealed. The firm chooses the probability of receiving a meaningful grade. Again, we show that test effectiveness is inverse U-shaped in beliefs (though discontinuous), with the same implications for the equilibrium as in the case of rational inattention.

Finally, we consider a general grading model, in which each available test is defined as in Daley and Green (2014), but the firm endogenously chooses its informativeness. We first establish that test precision - as defined in Ganuza and Penalva (2010) for example - is inverse U-shaped, but test effectiveness is not. The difference stems from the fact that test precision is related to the amount of information provided by the test, while effectiveness also considers the use that the firm makes of it. Ceteris paribus, effectiveness increases with precision. The firm not only chooses precision, but also a “passing grade”, by setting a minimum grade a worker must obtain in order to be hired. Higher precision can force this "passing grade" down, making a test less effective. We show that, for low initial beliefs, test effectiveness and beliefs are complements, while the opposite happens when beliefs are high. There is an intermediate region of beliefs, however, where an arbitrary number of zones of complementarity/substitutability may exist. As before, signaling by the workers will occur up to the point of maximal test effectiveness.

In all three models, the situation in which there is no match-specific component to the worker-firm relationship exhibits a quite realistic pattern. In equilibrium, there is a hard limit on the information the worker is able to transmit via signaling. High effort is chosen not only by high types, but also by a fraction of low types, making its observation inconclusive to the firm. Upon observing good credentials, the firm then performs an additional test to determine the suitability of a candidate.

This paper is organized as follows. In subsection 1.1. we review the related literature. In Section 2 we present the model, the firm's information acquisition and the worker's signaling problem. In Section 3 we define the equilibrium concept and refinement and show our main results. We discuss the Rational Inattention and the Grading models in Section 4 and 5. Section 6 concludes. All the proofs can be found in Appendix A.

1.1 Related Literature

This paper contributes to the literature that studies the effect of complementary sources of information in signaling models. Most papers have considered an exogenous source of information. Alos-Ferrer and Prat (2012) consider a model in which the worker signals his productivity via education, but the firm also receives reports about the worker's performance on the job after hiring him. Daley and Green (2014) allow firms to observe a stochastic grade correlated with the worker's type and his educational choice, before making a hiring decision. These papers show that when external information is sufficiently informative, pooling and semi-pooling equilibria may survive refinement criteria (D1 and Intuitive Criterium) rather than the Riley equilibrium. This is so because high types prefer to reduce their investment in education and let their productivity be revealed by their performance on tests or on the job. Kurlat and Scheuer (2019) study competitive markets in which firms are heterogeneous in their expertise when evaluating additional information about the worker's type. They characterize wage dispersion and show the existence of equilibria in which high types do not signal and are still hired by sufficiently informed firms. We contribute to this branch of the literature by allowing the firm to strategically choose the amount and accuracy of the information acquired.

In the context of costly, and endogenously chosen information by the uninformed party, Bester and Ritzberger (2001), Martin (2017) and Boyaci and Akçay (2017) study a market in which a privately informed monopolist sets the price while uninformed consumers may acquire information about product quality. In the first paper, consumers may pay a fixed cost to access an external source of information, which fully revelas product quality. In the second and third, consumers are rational inattentive and may acquire information at a cost proportional to the entropy reduction between prior and posterior. These papers show that the only equilibrium is semipooling with the high-quality firm setting a high price, while the low-quality firm mixing between a high and low price. In Mayzlin and Shin (2011) the firm may signal quality through the choice of the informational content of advertising, and consumers can engage in costly search about product quality. They show that the high-type firm chooses uninformative advertising to invite consumers to search. Garfagnini (2017) investigates the worker's incentives to signal through overtime at work. At the same time the firm might exercise managerial oversight, depending on the effort made by the worker. When the worker's initial reputation is sufficiently high, then separation is the only equilibrium and oversight never occurs. When the worker's reputation is low, a good worker knows that the firm will acquire some information in order to avoid coslty mistakes, which makes exerting less effort optimal for the worker. Perez-Richet and Prady (2011) consider a sender choosing the complexity of the message he sends, which has an effect on the costs of the receiver's test. In equilibrium messages can be more or less complex than its "natural" level, due to signaling motives. Finally, Bester et al. (2019) consider competing firms that might acquire information about the workers' type at a fixed cost, as a complementary source of information. When auditing costs are sufficiently low, the unique equilibrium involves pooling (or semi-pooling) and no education.

We contribute to this literature by allowing the firm to choose from a very general set of information structures. Our setting encompasses both binary settings (the firm either acquires perfect information or not), and the rational inattention approach. Moreover, we fully characterize the set of Perfect Bayesian Equilibria, establishing the precise conditions under which pooling, separation or semi-pooling can arise.

2 The Model

There is a worker (sender) being considered for a job at a firm (receiver). The worker's productivity depends on his fit with the firm, which can be good or bad, $q \in \{G, B\}$, and it is initially unknown. Nevertheless, the worker has general skills which affect his productivity at any firm. In particular, the worker is privately informed about his type, $\theta \in \{H, L\}$, the probability of being a good fit for the firm, i.e. $\theta = \mathbb{P}(q = G)$, with $0 \leq L < H \leq 1$. Both types can potentially be a good fit for a firm, but a high type worker is more likely to be so.

The worker chooses a costly action $x \in \mathbb{R}_+$ (e.g. effort) at a cost cx, the same for both types.[1] The firm observes the worker's action x, updates beliefs about his fit q and might acquire information (e.g. perform a test) to complement what it has learned through signaling. After the test results are revealed, the firm decides whether to hire or not the worker, $a \in \{Y, N\}$. The firm pays a fix wage w if hiring, with $B < w < G$, so that only a worker with a good fit is worth hiring.

The timing of the game is the following. The worker learns θ and chooses x (the worker's signaling problem). Given x, the firm updates beliefs about q, and chooses a test (the firm's information acquisition

[1] Here, as opposed to Spence (1973), effort cost does not provide a rationale for signaling.

problem). After observing the test result, the firm updates beliefs and makes a hiring decision.

Beliefs. We denote by $z := \mathbb{P}(\theta = H)$ the firm's beliefs about the worker's type and $\mu := \mathbb{P}(q = G) = zH + (1 - z)L$ the beliefs about the worker's fit. Prior beliefs are given by $z_0 \in [0, 1]$ and $\mu_0 \in [L, H]$, respectively. After observing x, the firm forms interim beliefs $z(x) := \mathbb{P}(\theta = H|x)$ and $\mu(x) = \mathbb{P}(q = G|x) = zH + (1 - z)\,L$, where the latter is the relevant variable for the firm's information acquisition problem.

2.1 The firm's information acquisition problem

After signaling, the firm holds beliefs $\mu \in [L, H]$ about the worker's fit q, and may acquire further information. In particular, it chooses an information strategy (test) consisting of a signal space S and a joint distribution $\{F(q, s)\}_{q\in\{G,B\},s\in S}$ at a cost $\hat{c}(F, \mu)$. Given a signal realization s, the firm updates beliefs to $b_F(s) \in [0, 1]$ through Bayes rule. Given b_F, the firm hires the worker if and only if $b_F(G - w) + (1 - b_F)(B - w) \geq 0$, or equivalently, $b_F \geq \hat{b} := \frac{w-B}{G-B}$. Therefore, the firm obtains $V(b) = 0$ if $b \leq \hat{b}$ and $V(b) = b\,(G - w) + (1 - b)\,(B - w)$ otherwise.

The firm solves

$$\max_{F\in\mathcal{F}} \mu \int_S V\left(b_F(s)\right) F(ds|G) + (1 - \mu) \int_S V\left(b_F(s)\right) F(ds|B) - \hat{c}(F, \mu)$$

where $\mathcal{F}$ is the set of available distributions. Different models have been proposed in the literature, considering alternative constraints on the set $\mathcal{F}$ and the cost function $\hat{c}(F, \mu)$. For example:

- The rational inattention model, where $\mathcal{F}$ is unrestricted and $\hat{c}(F, \mu)$ are entropic costs (Sims, 2006; Matějka and McKay, 2015).
- Models where the feasible set $\mathcal{F}$ can be summarized by a single-dimensional parameter σ, $\mathcal{F} = \{F(q, s; \sigma)\}_{\sigma\in\Sigma}$ and the cost function is given by $\hat{c}(F) = C(\sigma)$. These include the "truth-or-noise" model used in Lewis and Sappington (1994), or families of experiments ranked according to their precision, as in Ganuza-Penalva (2010).

In some circumstances, the firm may tailor the information it acquires to perfectly complement the one provided by the workers. For any belief μ about the worker's fit, the firm can optimally design a test, specifically chosen to complete the information transmitted through signaling. If the worker's effort fully signals his general skills, the firm may perform a test to evaluate a specific skill in a delimited area. On the other hand, if signaling is less informative, the firm will have to implement a more complex test to evaluate the suitability of a candidate across many dimensions. In this case, it seems reasonable to model the information acquisition cost as the expected entropy reduction between prior and posterior, where the prior is the belief generated through signaling and the entropy measures how much uncertainty or "missing information" there is in the state distribution.[2]

In other circumstances, however, the set of tools available to the firm is limited, and the information provided by a test does not necessarily complement, in an optimally designed way, the information already available from signaling. The firm might have, for example, access to an expert that either reveals a valuable signal or provides no new insight. Or it could have access to a particular family of standardized tests, which it can make more or less precise, but not adaptable to its a priori knowledge. In these cases, information acquisition takes a different form, with feasible tests belonging to a fixed family, and its cost depending only on the precision of the information they provide, and not on the level of the firm's a priori knowledge. Such is the case for the "grade or noise" and grading model presented later on.

Denoting by S_Y the set of signals such that the firm hires the worker $\left(b\,(s) \geq \hat{b}\right)$, the conditional probability of hiring a worker with fit q is given by $P_{Y|q}\,(\mu) = \int_{S_Y} dF(s \mid q)$. The solution to the firm's information acquisition problem leads to optimal hiring probabilities, for the good and bad-fit workers.

Definition 1. Test effectiveness can be defined as $\Delta P\,(\mu) := P_{Y|G}\,(\mu) - P_{Y|B}\,(\mu)$.

We assume that test effectiveness is upper semicontinuous in order to guarantee existence of equilibrium.

Assumption 2. *Test effectiveness $\Delta P(\mu)$ is upper semicontinuous.*

Ideally, the firm would choose $P_{Y|G} = 1$ and $P_{Y|B} = 0$, therefore hiring only workers with a good fit ($\Delta P = 1$). Without a test, however, the firm would be limited to $P_{Y|G} = P_{Y|B}$ ($\Delta(P) = 0$). Note that $\Delta P(\mu)$ depends on the technological features of the market: not only the set $\mathcal{F}$ of available distributions

[2] It is well known (Cover, 1999) that up to rounding, this difference captures the number of questions needed to learn the true state, when questions are designed to take advantage of a priori knowledge.

available to the firm and its cost, but also the potential gains or losses form hiring a worker ($G - w$ and $w - B$). It is, however, independent of the particular characteristics of the labor force, namely L, H and μ_0.

The relationship between beliefs generated through signaling μ and the (endogenously chosen) test effectiveness ΔP is important for our analysis: At the signaling stage, the incentives to signal through higher effort (inducing higher beliefs μ) are determined by the impact of those beliefs on the effectiveness of the test.

Definition 3. The beliefs generated through signaling by the worker and the firm's test effectiveness act as complements (substitutes) if and only if $\Delta P(\mu)$ is increasing (decreasing) in μ.

A test serves two main purposes: it identifies workers with a good fit (*searching for diamonds*), while allowing the firm to make less mistakes (*avoiding lemons*). Higher beliefs generated through signaling by the worker then have two countervailing effects on the test chosen by the firm. Since there are more good-fit workers on the market, the *searching for diamonds* motive becomes more important, and the firm improves test effectiveness as beliefs get better. At the same time, *avoiding lemons* becomes less important, since the fraction of bad fits is lower, so that the firm decreases test effectiveness. The overall effect is then ambiguous.

When the *searching for diamonds* effect dominates, beliefs generated through signaling and test effectiveness act as complements. When the *avoiding lemons* effect dominates, on the other hand, signaling and tests are substitutes.

Finally, it is important to highlight the difference between information precision and effectiveness. Precision is based on the property that more informative signals lead to a more disperse distribution of conditional expectations (Ganuza and Penalva, 2010), therefore being related to the amount of information provided by the signal structure. Effectiveness, however, measures not only the amount of information, but also the use of it made by the firm. A firm might choose a more precise test, then hiring under a wider set of signals. In this case precision would be high but effectiveness low, since the firm will be hiring indiscriminately. When considering the canonical rational inattention and "grade or noise" (where effectiveness is monotonically related to precision) and the grading model (where it is not) we will analyze in details the relationship between these two concepts.

2.2 The worker's signaling problem

Given the firm's beliefs μ about the worker's fit, a type-θ worker faces the following expected probability of being hired:

$$\begin{aligned} P_{Y|\theta}(\mu) &:= \theta P_{Y|G}(\mu) + (1-\theta) P_{Y|B}(\mu) \\ &= \theta \Delta P(\mu) + P_{Y|B}(\mu) \end{aligned}$$

The worker's payoff, given his type, costly action and induced beliefs can then be written as:

$$\begin{aligned} \Pi(\theta, x, \mu) &= w P_{Y|\theta}(\mu) - cx \\ &= w[\theta \Delta P(\mu) + P_{Y|B}(\mu)] - cx \end{aligned}$$

The worker's incentives to signal are critically determined by $\theta \Delta P(\mu)$, the interaction between his type and the effectiveness of the test chosen by the firm. A more effective test is more valuable to the high-type worker, since he is more likely to be a good fit. Therefore, when beliefs generated through signaling and test effectiveness are complements, he has more incentives to signal his type in order to increase the firm's beliefs. Such incentives disappear, however, when beliefs and test effectiveness are substitutes. In this case higher beliefs induce a less effective test, which benefits more low types, less likely to be a good fit.

Worker's payoffs exhibit no complementarity between type and action, since costs are type-independent. However, worker's payoffs are supermodular in type and test effectiveness, and therefore in beliefs as long as effectiveness is increasing. In this case, a single-crossing property is satisfied, allowing (at least partial) separation in equilibrium. The high type is more willing to pay for improving firm's beliefs, not because it is cheaper, but because it is more beneficial for him to face a more effective test.

3 Equilibrium Analysis

3.1 Set of PBE

We now characterize the set of Perfect Bayesian Equilibria. We consider separating, pooling, and semi-pooling equilibria, which are defined in the standard way:

Definition 4. A pooling equilibrium is a pair of strategies and beliefs (x_0, μ_0) such that both types choose the same education level x_0, so that the posterior belief μ equals the prior μ_0. The equilibrium x_0 must satisfy

1. $\Pi(L, x_0, \mu = \mu_0) \geq \Pi(L, x = 0, \mu = L) \Leftrightarrow wP_{Y|L}(\mu_0) - cx_0 \geq wP_{Y|L}(L)$
2. $\Pi(H, x_0, \mu = \mu_0) \geq \Pi(H, x = 0, \mu = L) \Leftrightarrow wP_{Y|H}(\mu_0) - cx_0 \geq wP_{Y|H}(L)$

Both types should be better off by pooling at education level x_0 given the prior belief μ_0 rather than following their most profitable deviation to $x = 0$ which induces beliefs $\mu = L$.

Definition 5. A separating equilibrium is a pair of strategies and beliefs $[(x_L, x_H), (L, H)]$ such that the firm perfectly infers the worker's type by observing his education level. It is direct that the low type cannot do better than choosing $x_L = 0$. Moreover, it should not find profitable to mimic the high type by deviating to x_H, even if the copycat behavior induces beliefs $\mu = H$. Analogously, the high type must not be willing to deviate to his most profitable deviation, which occurs at 0 while inducing beliefs $\mu = L$. Therefore x_H must satisfy:

1. $\Pi(L, x_L = 0, \mu = L) \geq \Pi(L, x_H, \mu = H) \Leftrightarrow wP_{Y|L}(L) \geq wP_{Y|L}(H) - cx_H$
2. $\Pi(H, x_H, \mu = H) \geq \Pi(H, x_L, \mu = L) \Leftrightarrow wP_{Y|H}(L) \leq wP_{Y|H}(H) - cx_H$.

Definition 6. A semi-pooling equilibrium is a pair of strategies $(x_H, [(0, x_H), (p, 1-p)])$ and beliefs $(L, \mu'(p))$ such that the high type chooses a positive level of education $x_H > 0$, while the low type randomizes between 0 and x_H (with probability p of choosing 0). Beliefs are given by:

$$\mu = \begin{cases} L & if \quad x = 0 \\ \mu'(p) = zH + (1-z)L & if \quad x = x_H \end{cases}$$

where $z := \mathbb{P}(\theta = H|x_H) = \frac{z_0}{z_0+(1-z_0)(1-p)}$ (and $z_0 := \mathbb{P}(\theta = H)$). Note that any $\mu'(p) \in [\mu_0, H]$ can be generated. For this to be an equilibrium the low type should be indifferent when mixing between $x = 0$ and $x = x_H$, while the high type should not find profitable to deviate to $x = 0$, while inducing beliefs $\mu = L$. Therefore:

1. $wP_{Y|L}(L) = wP_{Y|L}(\mu'(p)) - cx_H$
2. $wP_{Y|H}(\mu'(p)) - cx_H > wP_{Y|H}(L)$.

The following proposition shows the existence of the PBE.

Proposition 7.

1. There is a pooling equilibrium (x_0, μ_0)*, for every* $x_0 \le \frac{w}{c}\min\{P_{Y|L}(\mu_0) - P_{Y|L}(L), P_{Y|H}(\mu_0) - P_{Y|H}(L)\}$

2. A separating equilibrium exists if and only if $P_{Y|G}(H) - P_{Y|G}(L) \geq P_{Y|B}(H) - P_{Y|B}(L)$ *or equivalently* $\Delta P(H) \geq \Delta P(L)$*. The separating equilibrium is given by* $x_L = 0$ *and* x_H *such that* $\frac{w}{c}\left[P_{Y|L}(H) - P_{Y|L}(L)\right] \le x_H \le \frac{w}{c}\left[P_{Y|H}(H) - P_{Y|H}(L)\right]$*, with associated beliefs* $\mu(x_L) = L$ *and* $\mu(x_H) = H$*.*

3. A semi-pooling equilibrium exists if and only if $P_{Y|G}(\mu'(0)) - P_{Y|G}(L) \geq P_{Y|B}(\mu'(0)) - P_{Y|B}(L)$ *or equivalently* $\Delta P(\mu'(0)) \geq \Delta P(L)$*. The semipooling equilibrium is given by* $x_H = \frac{w}{c}\left[P_{Y|L}(\mu'(p)) - P_{Y|L}(L)\right]$ *for the high type and the low-type mixing between* 0 *and* x_H*, with beliefs calculated according to definition 6.*

3.2 D1 Equilibrium

Signaling games are characterized by multiplicity of equilibria, supported by pessimistic beliefs off-the-equilibrium path. Following the literature, we consider off-equilibrium path beliefs such that any deviation x is attributed to a low type, $\mu(x) = L$ and then apply D1 as a selection criterium (Banks and Sobel, 1987; Cho and Kreps, 1987) to eliminate implausible equilibria and provide satisfying predictions.

In this setting, the D1 can be stated as follows. Consider an equilibrium with associated payoffs $\{\Pi^*_H, \Pi^*_L\}$, and a potential deviation x for either type. Define the following sets of beliefs

$$B_L(x, \Pi_L^*) \quad := \{\mu \mid \Pi(L, x, \mu) > \Pi_L^*\}$$
$$B_H(x, \Pi_H^*) \quad := \{\mu \mid \Pi(H, x, \mu) > \Pi_H^*\}$$

as the set of beliefs under which either type will be better off by deviating to x rather than following the equilibrium strategy. If the set of beliefs such that a given type would like to deviate contains those of another type, a deviation must have come from the former. If $B_L(x, \Pi_L^*) \subset B_H(x, \Pi_H^*)$ for a deviation x, then D1 requires that $\mu(x) = H$, attributing any deviation to the high type. Therefore the considered equilibrium fails the D1 criterium. If instead, $B_L(x, \Pi_L^*) \supseteq B_H(x, \Pi_H^*)$, we have that $\mu(x) = L$ and off-equilibrium path beliefs are well-specified in the first place, so that the equilibrium satisfies the D1.

Suppose that the firm observes the effort level x, which is not part of any equilibrium strategy. Then of all the firm's possible beliefs after observing x, for which subset of these beliefs would the high and the low type prefer to deviate to x rather than following the equilibrium strategy? If there are beliefs such that the high type would prefer the deviation while the low type would not, then the firm should attribute the deviation to the high type, in contrast to what the existence of the equilibrium required, that any deviation should be thought of as coming from the low type. We denote the unique equilibrium that satisfies the D1 criterium by D1 equilibrium.

3.2.1 Regular case

We first consider the case, that applies to the rational inattention (section 4) and the "grade or noise" model (section 5.1), in which the information acquisition problem shows a well-defined pattern: test effectiveness is inverse U-shaped on beliefs.

Assumption 8. *There exists a belief μ^* such that the beliefs generated through signaling by the worker μ and test effectiveness $\Delta P(\mu)$ are complements if and only if $\mu \leq \mu^*$.*

Under this assumption, for $\mu \leq \mu^*$ the *searching for diamonds* effect dominates, and higher beliefs make the firm improve the test effectiveness. The opposite holds for $\mu \geq \mu^*$, where beliefs are sufficiently high and the test chosen by the firm mainly serves the purpose of *avoiding lemons.*

When beliefs and test effectiveness are complements, the equilibrium involves (at least partial) separation between workers' types. Since the high type is more willing to face a more exacting (but more rewarding) test than a low type, being less likely to flunk it, he will exert costly effort to improve the firm's opinion. When beliefs and test effectiveness are substitutes, any signaling attempt by the high type will be mimicked by the low one who benefits more from relaxed standards and indiscriminate hiring. Therefore the only plausible equilibrium involves both types pooling at no effort.

If the worker's expected fit with the firm is low ($\mu_0 < \mu^*$), information is revealed through both information acquisition and signaling. In particular, separation can be total or partial depending on whether H is lower or higher than μ^*. If $H \leq \mu^*$- even talented agents are not a sure thing when performing a specific task at a firm - worker's effort fully signals his type. If $H \geq \mu^*$- a high type is likely to perform well at a firm - a semi-pooling equilibrium, where the high type chooses an action that is imitated with positive probability by the low type, inducing the belief μ^*, is the only D1 equilibrium.

The belief μ^* becomes a hard limit on information revelation by workers. Even if a worker knows he is very likely to be a good fit for a firm (with a probability above μ^*), he would never tell his whole story in equilibrium. This hard limit is created by firms endogenously choosing the test they perform. It is not only that a signal becomes more or less informative when priors change (as in models of exogenous tests), but also tests become less effective as beliefs grow, because of a relaxation of the firm's incentives to screen.

If, on the other hand, the worker's expected fit is high, ($\mu_0 > \mu^*$), any attempt to induce higher beliefs is profitably mimicked by the low type because of its effect on the test effectiveness. Then the only equilibrium that survives the D1 is pooling at $x = 0$ and information is generated exclusively by the firm through tests.

This intuition is formally formulated in the next proposition:

Theorem 9.

1. If $\mu_0 > \mu^*$ the unique D1 equilibrium is pooling at $x_0 = 0$, and beliefs generated through signaling by the worker act as substitutes to the effectiveness of the test chosen by the firm.

2. If $\mu_0 < \mu^*$, the worker signals his information up to μ^*:

a) If $H < \mu^*$, the unique D1 equilibrium is the LCSE.

b) If $\mu^* < H$, the unique D1 equilibrium is a semi-pooling equilibrium such that x^* induces the belief μ^*. In this case, beliefs generated through signaling and test effectiveness act as complements.

3. In the non-generic case where $\mu^* = \mu_0$, pooling strategies such that $x_0 \leq \frac{w}{c}\min\{P_{Y|L}(\mu_0) - P_{Y|L}(L), P_{Y|H}(\mu_0) - P_{Y|H}(L)\}$ constitute an equilibrium that survives the D1 refinement.

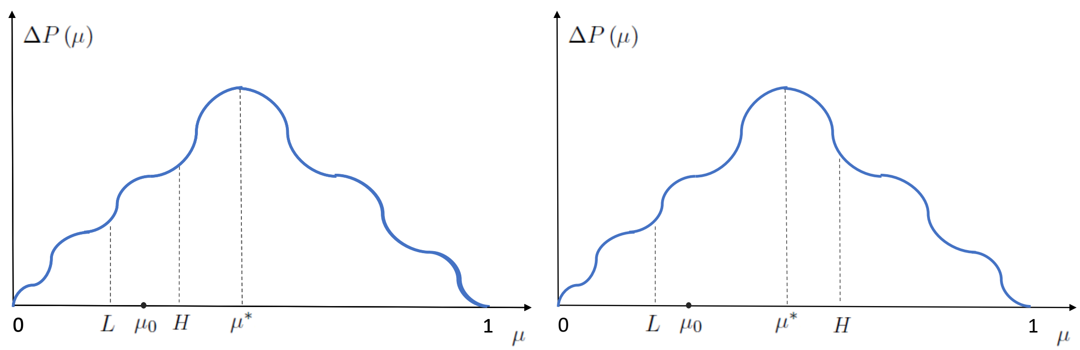

Figure 1: Separating and semi-separating equilibria (case 2a and 2b from Theorem 9).

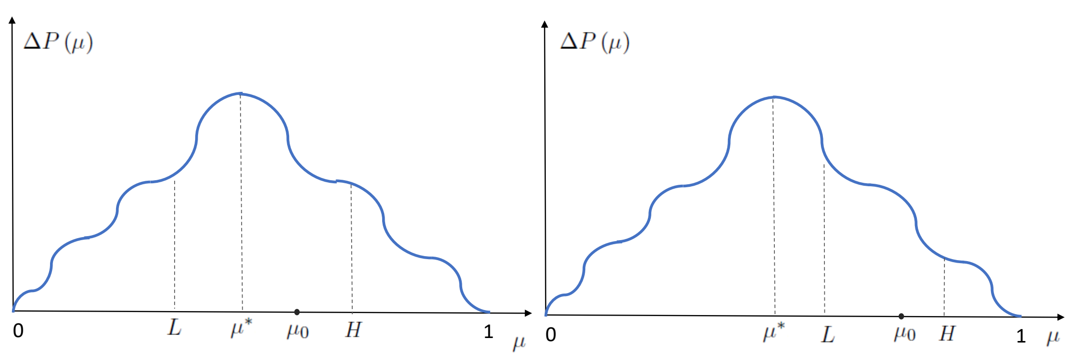

Figure 2: Pooling equilibria

When the labor force is weak (low μ_0), or equivalently the task is difficult for the "average Joe", we can expect signaling to complementinformation acquisition. High types will go the extra mile, revealing valuable information about their abilities through effort. In turn, firms will perform a test conditional on the workers' performance on the probationary period. Workers who exerted low effort face a cursory, not very precise evaluation, entailing a low probability of being hired. High-effort workers, on the other hand, face a more effective test, but involving a higher probability of being hired.

When the labor force is strong relative to the task at hand (high μ_0), we should expect signaling not playing any role in information revelation. Since higher beliefs only relax the firm's incentives to screen, leading to less effective tests, any signaling effort by the high type will be easily mimicked by the low type. Therefore, for "run of the mill" activities, firms do not rely on the observation of effort (or any other costly signal), but on their internal evaluation procedures.

3.2.2 Non-regular case

Test effectiveness, as a function of μ, does not necessarily show the regular pattern described in assumption 8, as we illustrate in the grading model in Section 5.2. However, the basic intuition holds: the high type will signal his private information up to the point where test effectiveness is maximized, since he benefits more from higher beliefs. As before, this can involve a separating or semi-pooling equilibrium.

Let assume that $\Delta P(\mu)$ attains a unique maximum in $[L, H]$ and define:

$$\mu_{max} = \arg \max_{\mu \in [L,H]} \Delta P(\mu)$$

Theorem 10.

If $\mu_0 \leq \mu_{max}$, the only D1 equilibrium involves the high type signaling up to μ_{max}. In particular:

1. If $\mu_{max} = H$, the only D1 equilibrium is the LCSE.

2. If $\mu_0 < \mu_{max} < H$, the only D1 equilibrium is a semi-pooling equilibrium such that x^* induces the belief μ_{max}.

3. If $\mu_{max} = \mu_0$, then any pooling equilibrium satisfies the D1 criterium.

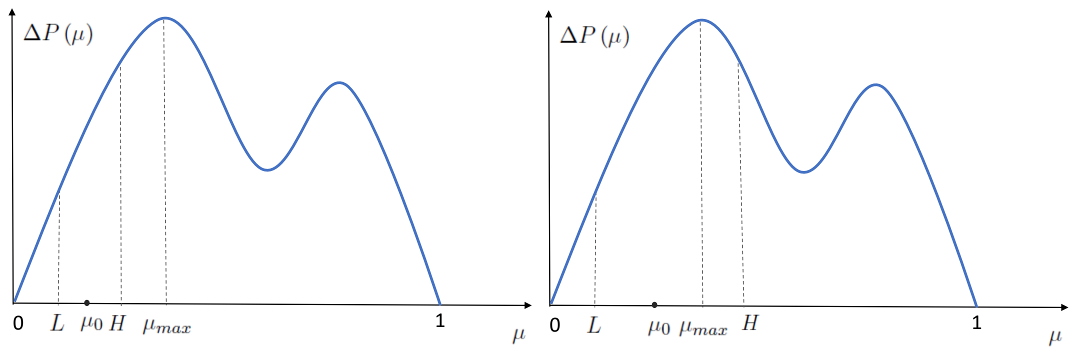

Figure 3: Separating and semi-separating equilibria (Theorem 10)

Any other equilibrium, where the action of the high type leads to a belief μ such that $\Delta P(\mu) < \Delta P(\mu_{max})$, involves off-equilibrium path beliefs that are non-credible. A profitable deviation for the high type that generates the belief μ_{max} is less profitable for the low-quality type exactly because test effectivness is maximized at μ_{max}, so that off-equilibrium path beliefs $\mu = L$ are not plausible.

A slightly different situation arises when $\mu_0 > \mu_{max}$. Here, to induce beliefs μ_{max}, the high-type worker would need to worsen the initial beliefs, but this behavior is never part of an equilibrium. If $\Delta P(\mu) < \Delta P(\mu_0)$ for all $\mu > \mu_0$, the only equilibrium is pooling at $x_0 = 0$. Any costly action chosen by the high type could be costlessly mimicked by the low type, therefore carrying no informational content.

Theorem 11. *Suppose that $\mu_{max} < \mu_0$, then*

1. *If $\mu_0 = \arg\max_{\mu \geq \mu_0} \Delta P(\mu)$, the only D1 equilibrium involves pooling at $x_0 = 0$.*

2. *If $\mu_0 \neq \arg\max_{\mu \geq \mu_0} \Delta P(\mu)$, there is no D1 equilibrium.*

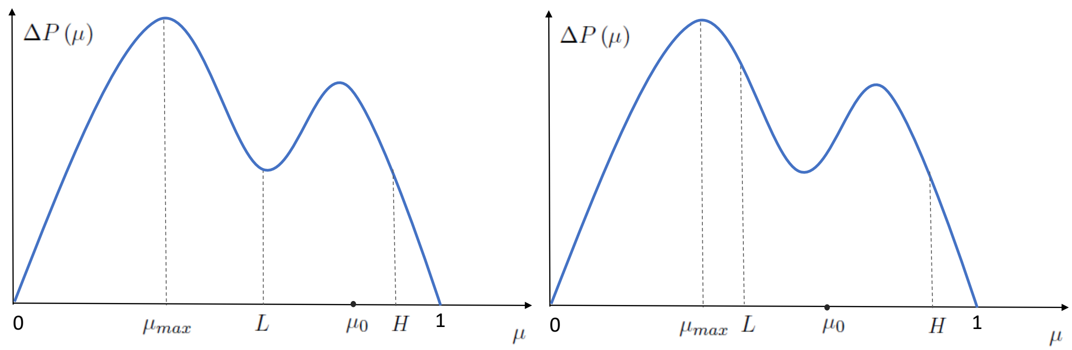

Figure 4: Theorem 11

4 Rational Inattention

We first consider the information acquisition problem arising from the rational inattention approach (Sims, 2006; Matějka and McKay, 2015), in which costs are entropic and there are no restrictions on the set $\mathcal{F}$ of available distributions to the firm.

The relevant state of nature is the worker's fit $q \in \{G, B\}$, about which the firm holds beliefs μ. The firm's action belongs to $\{Y, N\}$ (hiring or not) and its payoffs are state-dependent: $u(Y|q) = q - w$, $u(N|q) = 0$. In this setting a test is given by an information structure: a signal space S and a joint distribution $\{F(q,s)\}_{q\in\{G,B\},s\in S}$, such that the firm's posteriors are consistent, i.e. the marginal distribution over states equals the prior μ. Given a signal realization s, the firm updates beliefs to $b_F(s) \in [0,1]$ through Bayes rule, and makes a hiring decision. The cost of a test is proportional to the expected entropy reduction between prior and posterior:

$$\hat{c}(F,\mu) = \lambda \{H(\mu) - E[H(b_F(s))]\}$$

where entropy $H(\cdot)$ measures how much uncertainty or "missing information" there is in the state distribution:

$$\begin{aligned} H(\mu) &= -\mu log(\mu) - (1-\mu) log(1-\mu) \\ E[H(b_F(s))] &= \sum_{s\in S} Pr(s)[-b_F(s) log(b_F(s)) - (1-b_F(s)) log(1-b_F(s))] \end{aligned}$$

The firm determines the information structure $\{F(q,s)\}_{q\in\{G,B\},s\in S}$ and therefore the distribution of posterior beliefs $b_F(s)$, trading off more effective information, that leads to an action a better tailored to the state q, against higher costs. Since a signal realization is relevant only inasmuch as it translates into an action recommendation[3], the firm's problem can be reformulated as one in which the firm chooses the probability of hiring a worker conditional on his fit. In particular, let $P_{a|q}$ denote the probability of action a conditional on state q, and $P_a = \mu P_{a|G} + (1-\mu) P_{a|B}$ the unconditional probability. Using the techniques developed in Matějka and McKay (2015), we obtain conditional probabilities of being hired for a good and a bad-fit worker:

$$\begin{aligned} P_{Y|G}(\mu) &= \frac{P_Y(\mu)\exp\left\{\frac{G-w}{\lambda}\right\}}{P_Y(\mu)\exp\left\{\frac{G-w}{\lambda}\right\}+P_N(\mu)} \\ P_{Y|B}(\mu) &= \frac{P_Y(\mu)\exp\left\{\frac{B-w}{\lambda}\right\}}{P_Y(\mu)\exp\left\{\frac{B-w}{\lambda}\right\}+P_N(\mu)} \end{aligned} \tag{1}$$

where the unconditional probabilities $P_Y(\mu)$, $P_N(\mu)$ can be explicitly computed (see Appendix A).

Test effectiveness shows the regular pattern described in assumption 8 (see Figure 5):

Lemma 12. *There exists a cutoff belief μ^* such that $\frac{\partial}{\partial\mu}[\Delta P(\mu)] \geq 0 \iff \mu \leq \mu^*$. In other words, for $\mu \leq \mu^*$ ($\mu \geq \mu^*$) beliefs generated through signaling and test effectiveness act as complements (substitutes).*

For low levels of μ, the *searching for diamonds* effect dominates, and firms improve test effectiveness as μ grows. The opposite holds for high levels of μ, where beliefs are already high and the main purpose of a test is *avoiding lemons*. In this canonical model the test effectiveness and precision, given by the expected

[3]It is inefficient to have more than one signal leading to the same action, since it increases costs without any benefit.

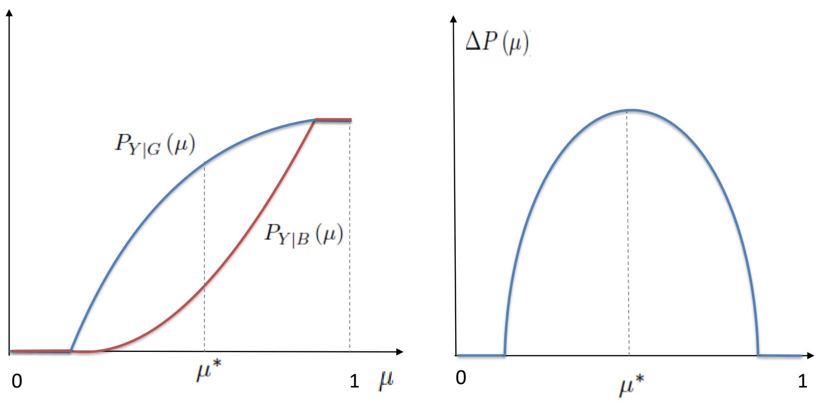

Figure 5:

reduction in entropy, coincide. The firm only buys the information it will later use (in fact the optimal test only uses two signals), and precision increases only when it is used to increase effectiveness. We use theorem 9 to characterize the worker's optimal signaling strategy:

Corollary 13. *If the firm's information acquisition problem is formulated according to the rational inattention paradigm, then the worker's optimal signaling can be described as follows:*

- *If $\mu^* < L$, then the only D1 equilibrium is pooling at $x = 0$.*
- *If $\mu^* > H$, then the only D1 equilibrium is the LCSE.*
- *If $\mu^* \in [L, H]$, then the only D1 equilibrium is semi-pooling such that x^* induces μ^*.*

When the labor force is sufficiently strong - even the low type is expected to be a good fit for the firm, so that $\mu_0 > \mu^*$ - information is generated exclusively by the firm through tests. Higher signaling beliefs would lead to a less effective test by the firm, which in turn will result more beneficial to low types. Any signaling effort by the high type will be mimicked by the low-type worker, therefore the only equilibrium is pooling at no education.

On the other hand, when the labor force is weak ($\mu_0 < \mu^*$), information is revealed through both information acquisition and signaling. In particular, the high type will signal his private information up to the point in which test effectiveness is maximal. In this case separation can be total or partial depending on whether H is lower or higher than μ^*.

5 Grading Models

We now present two models where the feasible set $\mathcal{F}$ can be summarized by a single-dimensional parameter σ, $\mathcal{F} = \{F_q(s;\sigma)\}_{\sigma\in\Sigma}$ and the cost function is given by $C(\sigma)$, which only depends on tests and not on the prior μ generated through signaling. These models include, for example, the truth-or-noise model used in Lewis and Sappington (1994), or families of experiments ranked according to their precision, as in Ganuza-Penalva (2010).

Following Daley and Green (2014), a test is a pair $\{f_G(s;\sigma), f_B(s;\sigma)\}$, where $f_q(s;\sigma)$ governs the distribution of grades $s \in S$ obtained by a worker whose fit with the firm is $q \in \{G, B\}$. The parameter σ represents the precision of the test.

We assume that higher grades improve beliefs (they are "good news" about the worker's fit), and an increase in precision σ makes the informational content of grades higher.

Assumption 14. *A test satisfies the following assumptions:*

i) Monotone Likelihood Ratio Property (MLRP): $\frac{f_G(s;\sigma)}{f_B(s;\sigma)}$ increasing in s.

ii) Increasing Informativeness: $[F_G(s;\sigma) - F_B(s;\sigma)]$ is decreasing in σ.

iii) $C(\sigma)$ increasing and convex.

The first assumption, MLRP, is standard and implies that higher grades are "good news" in the language of Milgrom (1981). The second states that, as precision increases, the distribution of grades becomes more different (in the FOSD sense) between workers with a good and bad fit. The third one just implies that precision is costly.

Given μ, generated through signaling, the firm updates beliefs after observing a grade s to

$$\mu'(\mu, s, \sigma) = \frac{\mu f_G(s;\sigma)}{\mu f_G(s;\sigma) + (1-\mu) f_B(s;\sigma)} \tag{2}$$

The firm hires the worker if and only if $\mu'(\mu, \overline{s}, \sigma) G + (1 - \mu'(\mu, \overline{s}, \sigma)) B - w \geq 0$. Because of MLRP, this is equivalent to hiring if and only if $s \geq \overline{s}$, where the passing grade $\overline{s}$ satisfies

$$\mu'(\mu, \overline{s}, \sigma) G + \left(1 - \mu'(\mu, \overline{s}, \sigma)\right) B = w \tag{3}$$

Therefore, the firm solves:

$$\max_{\sigma} \int_{s \geq \overline{s}(\mu,\sigma)} \left[\mu f_G(s;\sigma)(G-w) + (1-\mu) f_B(s;\sigma)(B-w)\right] ds - C(\sigma)$$

which leads to an optimal precision $\sigma(\mu)$ and passing grade $\overline{s}(\mu, \sigma(\mu))$. The conditional probability of hiring a worker with fit q can be expressed as $P_{Y|q}(\mu) = 1 - F_q(\overline{s}(\mu, \sigma(\mu)); \sigma(\mu))$, and test effectiveness is then given by :

$$\Delta P(\mu) = F_B(\overline{s}(\mu, \sigma(\mu)); \sigma(\mu)) - F_G(\overline{s}(\mu, \sigma(\mu)); \sigma(\mu))$$

5.1 Grade or Noise

We first present a model, in which the precision σ affects the probability of obtaining a meaningful grade, but not its distribution. We denote this information structure "grade or noise" :[4] With probability σ, the grade s is drawn from $\{f_G(s), f_B(s)\}$ and with probability $1-\sigma$ no information is obtained.[5]

[4]It could be thought of as a combination of the "truth or noise" information structure, as first defined by Lewis and Sappington (1994) - or linear experiment, as defined by Ganuza and Penalva (2010) - with the exogenous grades and tests as proposed by Daley and Green (2014).

[5]The test $\{f_G(s;\sigma), f_B(s;\sigma)\}$ can also be defined as in section 5:

$$\begin{cases} f_G(s;\sigma) = & \sigma \widetilde{f}_G(s) + (1-\sigma)\delta_{\hat{s}} \\ f_B(s;\sigma) = & \sigma \widetilde{f}_B(s) + (1-\sigma)\delta_{\hat{s}} \end{cases}$$

With probability σ, the grade s is drawn from $\left\{\widetilde{f}_G(s), \widetilde{f}_B(s)\right\}$ (grade), and with probability $1-\sigma$, the grade is $\hat{s}$ such that $\frac{\widetilde{f}_G(\hat{s})}{\widetilde{f}_B(\hat{s})} = 1$, so the posterior equals the prior (noise). It is easy to verify that $\{f_G(s;\sigma), f_B(s;\sigma)\}$ satisfies assumption 14

Without loss of generality we assume that $f_B(s)$ is uniform on $[0,1]$. Given a grade s, the firm hires the worker if $s \geq \overline{s}(\mu)$, and we can rewrite equation 3 as

$$\mu f_G(\overline{s})(G-w) + (1-\mu)(B-w) = 0$$

It is important to remark that the passing grade $\bar{s}$ does not depend on σ, since the informational content of a grade is unaffected by the test precision. If no grade is generated, the worker is hired if $\mu \geq \bar{\mu}$, where $\overline{\mu}$ is such that

$$\bar{\mu}(G-w) + (1-\bar{\mu})(B-w) = 0$$

Therefore we can write the firm's problem as:

$$\max_{\sigma} \sigma \int\limits_{s \geq \overline{s}(\mu)} [\mu f_G(s)(G-w) + (1-\mu)(B-w)]\, ds + (1-\sigma)\, \mathbb{I}_{\{\mu \geq \bar{\mu}\}} [\mu (G-w) + (1-\mu)(B-w)] - C(\sigma)$$

Lemma 15. *The firm's objective function is supermodular in (σ, μ) if $\mu \leq \bar{\mu}$, and supermodular in $(\sigma, -\mu)$ if $\mu \geq \bar{\mu}$. Therefore σ is increasing (decreasing) in μ if $\mu \leq \bar{\mu}(\mu \geq \bar{\mu})$.*

For low initial beliefs ($\mu \leq \bar{\mu}$), the *searching for diamonds* effect dominates, and firms improve test precision as μ grows. Since the status quo is not hiring, a more precise test serves the purpose of identifying good-fit workers. The opposite holds for $\mu \geq \bar{\mu}$, when the main purpose is *avoiding lemons*, and firms buys less information as beliefs get better. Here the status quo is to hire, and precision is used to avoid hiring bad-fit workers.

Test effectiveness is given by $\Delta P(\mu) = \sigma(\mu)[\overline{s}(\mu) - F_G(\overline{s}(\mu))]$, and we can write

above if we accept the convention that $\frac{\tilde{f}_G(\hat{s})}{\tilde{f}_B(\hat{s})} = 1$. This is in the spirit of our definition of "grade or noise" since the same grade is obtained with the same probability by the good and bad fit worker.

$$\frac{\partial}{\partial \mu}\left[\Delta P\left(\mu\right)\right] = \underbrace{\sigma'\left(\mu\right)\left[\overline{s}\left(\mu\right) - F_G\left(\overline{s}\left(\mu\right)\right)\right]}_{\textbf{Precision Effect}} + \underbrace{\frac{d\overline{s}\left(\mu\right)}{d\mu}\sigma\left(\mu\right)\left[1 - f_G\left(\overline{s}\left(\mu\right)\right)\right]}_{\textbf{Passing Grade Effect}}$$

The complementarity (or substitutability) between beliefs generated through signaling by the worker and the effectiveness of the test chosen by the firm depends on the effect of μ on both precision σ (precision effect) and passing grade $\bar{s}$ (passing grade effect).

Ceteris paribus, as precision increases the test becomes more effective. Therefore, the precision effect has the same sign as $\sigma'(\mu)$, which is inverse U-shaped (Lemma 15). The passing grade effect can either increase or decrease test effectiveness. More optimistic beliefs about the worker's fit always make the firm relax standards, by reducing the passing grade $\bar{s}$, which improves test effectiveness when the passing grade is high ($f_G(\bar{s}) > 1$). In this case, good-fit workers are more likely to obtain grades at the cutoff, and therefore benefit more from a lower passing grade. The opposite happens when the passing grade is low ($f_G(\bar{s}) < 1$).

In this model, higher beliefs make both the precision and passing grade effects work in the same direction.When beliefs are low ($\mu \leq \overline{\mu}$), both are positive. Precision increases with μ, making tests more effective. At the same time, since the firm is pessimistic about worker's fit, the passing grade $\bar{s}(\mu)$ is high ($f_G(\bar{s}) > 1$). The passing grade decreases with μ, which in turn is more beneficial to workers with a good fit. When beliefs are high ($\mu \geq \overline{\mu}$), the opposite happens: An increase in μ reduces both test precision and passing grade, which benefits workers with a bad fit more, since $\bar{s}$ is low ($f_G(\bar{s}) < 1$). In this case, beliefs generated through signaling and test effectiveness act as substitutes.

Lemma 16. *Beliefs generated through signaling μ and test effectiveness $\Delta P(\mu)$ act as complements if and only if $\mu \leq \bar{\mu}$.*

Test effectiveness shows the well-defined pattern described in Assumption 8 (Figure 6), that allows us to use Theorem 9 to characterize the worker's signaling strategy.[6]

When the worker's expected fit is low ($\mu_0 \leq \overline{\mu}$) the high type will signal his type up to $\overline{\mu}$, where test effectiveness is maximal. Higher signaling beliefs in this case induce a more effective test (by increasing its

[6] Indeed the belief that makes the firm indifferent about hiring the worker ($\overline{\mu}$) plays the rol of μ^* in Assumption 8.

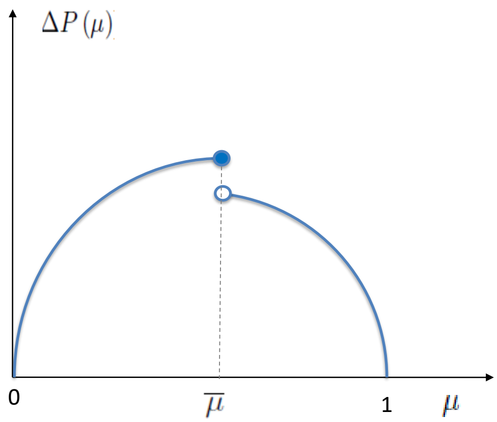

Figure 6:

precision and relaxing the passing grade) which benefits more the high type. On the other hand, when the worker's expected fit is high enough ($\mu_0 \geq \overline{\mu}$), there is no room for signaling and all information revelation occurs through tests. Higher signaling beliefs relax the test - by reducing its precision and relaxing the passing grade - which makes copycat behavior costless for the low type.

5.2 A grading model

We now consider the general case in which, by choosing a more precise test, the firm changes the distribution of grades. A test is then given by $\{f_G(s;\sigma), f_B(s;\sigma)\}$, with $f_B(s;\sigma)$ uniformly distributed on $[0,1]$, which implies that the likelihood ratio $\frac{f_G(s;\sigma)}{f_B(s;\sigma)} = f_G(s;\sigma)$.

For tractability, we assume that there exists a pivot grade $\hat{s}$ such that $f_G(\hat{s};\sigma) = 1$. Therefore, regardless of σ, a grade s is good news about the worker's fit if and only if $s > \hat{s}$, which, together with assumption 14, implies that

$$\begin{cases} \frac{\partial f_G(s;\sigma)}{\partial \sigma} < 0; \quad f_G(s;\sigma) < 1 \quad \forall s < \hat{s} \\ \frac{\partial f_G(s;\sigma)}{\partial \sigma} > 0; \quad f_G(s;\sigma) > 1 \quad \forall s > \hat{s} \end{cases}$$

Moreover, we assume that the likelihood ratio $(f_G(s;\sigma))$ becomes steeper in grades when precision increases, which implies that a higher grade is more informative when precision is higher. We also assume that there are decreasing returns to scale in information for higher grades (the likelihood ratio is concave).

Assumption 17. $\frac{\partial^2 f_G(s;\sigma)}{\partial s \partial \sigma} \geq 0$ *and* $\frac{\partial^2 f_G(s;\sigma)}{\partial^2 s} \leq 0$.

After choosing σ and observing s, the firm updates beliefs according to (2) which leads to an optimal passing grade $\overline{s}(\sigma(\mu), \mu)$ determined by (3). Since the precision σ changes the distribution of grades, $\overline{s}$ depends directly on beliefs generated through signaling μ but also indirectly through $\sigma(\mu)$.

The firm's information acquisition problem is given by:

$$\max_{\sigma} \int_{s \geq \overline{s}(\sigma(\mu),\mu)} \left[\mu f_G(s;\sigma)(G-w) + (1-\mu)(B-w)\right] ds - C(\sigma)$$

which leads to precision $\sigma(\mu)$ and passing grade $\overline{s}(\sigma(\mu), \mu)$. Test effectiveness can then be written as

$$\Delta P(\mu) = \left[\overline{s}(\sigma(\mu), \mu) - F_G(\overline{s}(\sigma(\mu), \mu); \sigma(\mu))\right]$$

and we have

$$\frac{\partial \Delta P(\mu)}{\partial \mu} = \underbrace{-\sigma'(\mu) \frac{\partial F_G(\overline{s}(\sigma(\mu), \mu); \sigma(\mu))}{\partial \sigma}}_{\textbf{Precision Effect}} + \underbrace{\left[1 - f_G(\overline{s}(\sigma(\mu), \mu); \sigma(\mu))\right] \frac{d\overline{s}(\sigma(\mu), \mu)}{d\mu}}_{\textbf{Passing Grade Effect}} \qquad (4)$$

As before, the complementarity (or substitutability) between beliefs and test effectiveness depends on both the precision and passing grade effect.

The precision effect has the same sign as $\sigma'(\mu)$, which is inverse U-shaped. When the firm is pessimistic about the worker's fit (initial low beliefs μ) the *searching for diamonds* effect dominates, and precision increases. When beliefs are sufficiently high, the *avoiding lemons* effect does and the firm reduces test precision.

The passing grade effect depends on $\bar{s}$: If high ($\bar{s} > \hat{s}$), good-fit workers are more likely to obtain grades around $\overline{s}$. Therefore they benefit marginally more from a lower passing grade, which contributes to improve test effectiveness. The opposite happens if the passing grade is low ($\bar{s} < \hat{s}$). Thus the passing grade effect has the same sign as $\frac{d\bar{s}}{d\mu}$ as long as $\bar{s} < \hat{s}$, and the opposite sign otherwise.

The passing grade depends on μ directly and indirectly through the choice of precision σ. The direct effect is simple: as the firm improves beliefs about the worker's fit, it relaxes standards and $\bar{s}$ decreases. The sign of the indirect effect depends on the level of $\bar{s}$. Low grades are more likely to come from workers with a bad fit, and higher precision intensifies this effect. Therefore, if $\bar{s} < \hat{s}$ a more precise test deteriorates beliefs at the cutoff, and makes the firm increase the passing grade. The opposite happens if $\bar{s} > \hat{s}$, where an increase in σ leads to a lower passing grade which benefits more good-fit workers, more likely to obtain high grades.

When initial beliefs are low, both the precision and passing grade effect are positive and test effectiveness increases. Intuitively, higher beliefs lead the firm to choose a more precise test, since the *searching for diamonds* effect dominates. At the same time, since the firm is pessimistic, it chooses a high passing grade ($\bar{s} > \hat{s}$), which goes down (both through the direct and indirect channel) as beliefs improve. A lower passing grade benefits disproportionally more workers with a good fit, so that beliefs and test effectiveness act as complements.

When initial beliefs are high, both the precision and passing grade effect are negative and test effectiveness decreases. As beliefs improve, test precision is reduced since the *avoiding lemons* effect dominates. Moreover, the passing grade is low ($\bar{s} > \hat{s}$), and higher beliefs induce laxer standards (through both the direct and indirect channel), which disproportionally benefit workers with a bad fit, also reducing test effectiveness. Beliefs and effectiveness are substitutes.

There is an intermediate region of beliefs where things are more complicated. In this region, precision is still increasing in μ, but the passing grade is low, $\bar{s} < \hat{s}$. The latter implies that $\bar{s}$ decreases through the direct channel, but indirectly *increases* due to a higher σ. If the indirect effect dominates, then both the precision and passing grade effect are positive and test effectiveness increases (beliefs generated through signaling and test effectiveness are complements). If the direct effect dominates ($\frac{d\bar{s}}{d\mu} < 0$), then the precision and passing grade effects go in opposite directions and the test could become either more or less effective. We are able to establish that, depending on the particular shape of the cost function C, any of the two effects can dominate at any level of μ. Figure 7 and Theorem 20 show formally our argument.

Theorem 18. *There exists a belief $\mu_1 \in (0, 1)$, and a belief $\mu_2 \in (\mu_1, 1)$ such that:*

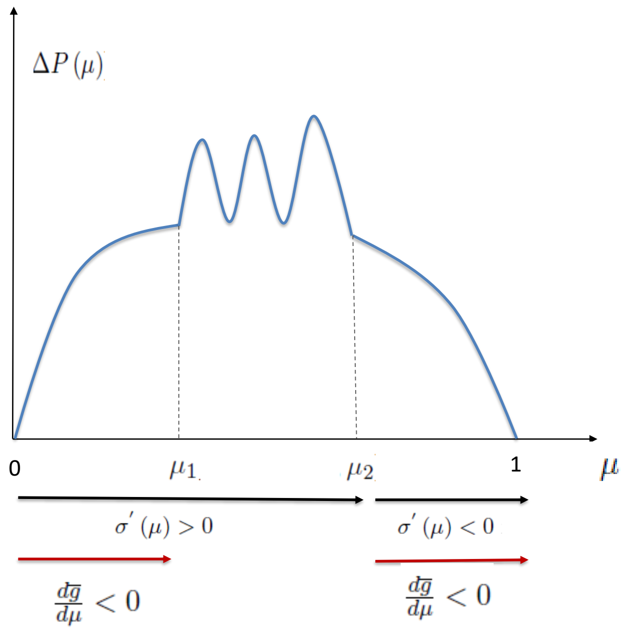

Figure 7:

- $\frac{\partial \Delta P(\mu)}{\partial \mu} > 0$ for any $\mu \in [0, \mu_1]$
- $\frac{\partial \Delta P(\mu)}{\partial \mu} < 0$ for any $\mu \in [\mu_2, 1]$
- $\frac{\partial \Delta P(\mu)}{\partial \mu} \gtreqless 0$ in $\mu \in [\mu_1, \mu_2]$. Moreover, it can change signs any countable number of times depending on the shape of the cost function $C(\cdot)$.

As before, this allows us to characterize the worker's signaling strategy:

Corollary 19. *If the firm's information acquisition problem corresponds to the grading model, then the worker's optimal signaling can be described as follows:*

- *If $[L, H] \subseteq [0, \mu_1]$, then the only D1 equilibrium is the LCSE*
- *If $[L, H] \subseteq [\mu_2, 1]$, then the only D1 equilibrium is pooling at $x_0 = 0$.*
- *Otherwise, we apply Theorem 10 to obtain:*

- *If* $\mu_0 < \mu_{max} := \arg\max_{\mu \in [L,H]} \Delta P(\mu)$ *, then the only D1 equilibrium involves the high type signaling up* μ_{max}
- *If* $\mu_0 \geq \mu_{max}$ *and* $\mu_0 = \arg\max_{\mu \geq \mu_0} \Delta P(\mu)$*, then the only D1 equilibrium involves pooling at* $x_0 = 0$

Again, when the task the worker is expected to perform requires little specialization (μ_0 is high), signaling plays no role in information revelation. On the other hand, more specialized tasks (μ_0 is low) lead to separation, with the high type exerting costly effort to induce a more effective test by the firm.

The situation in which there is no match-specific component to the worker-firm relationship is a special case: High-type workers have a good fit with the firm, while low types do not ($L = 0$ and $H = 1$). In this case, the equilibrium always involves partial separatation, since there is a hard limit on the information the worker is able to transmit via signaling. High effort is chosen not only by high types, but also by a fraction of low types, making its observation indicative but inconclusive evidence to the firm. Upon observing good credentials, the firm then performs an additional test to determine the suitability of a candidate. This is in line with the typical behavior of labor markets, where educational credentials or performing well during an internship are a necessary prerequisite for a job, but firms still design ad hoc tests on candidates before committing to a long-term relationship.

6 Conclusion

We analyze a strategic environment in which the information transmitted by an agent through signaling is complemented by the information acquired by another. In particular, we study the worker's signaling incentives in the presence of endogenous information acquisition by the firm. The worker's effort influences the firm's beliefs, which in turn determine the amount and accuracy of the test performed before making a hiring decision. As a relevant accuracy measure we use test effectiveness, defined as the difference between the probabilities of hiring a worker with a good and a bad fit with the firm. We show that test effectiveness is not monotone in beliefs and then provide clear-cut predictions on the complementarity/substitutability between costly information transmission (signaling) and acquisition, and its implications for the equilibrium.

When test effectiveness is increasing in beliefs (the firm is searching for diamonds), the worker's payoffs are supermodular in type and beliefs. Since the high type is more willing to face a more rigorous test than a low type, he will exert costly effort to improve the firm's opinion, so that the equilibrium involves total or partial separation. When, on the other hand, test effectiveness is decreasing in beliefs (the firm is avoiding lemons), any signaling attempt by the high type will be mimicked by the low one, and the only plausible equilibrium involves pooling. Here, any effort spent to improve firm's beliefs leads to indiscriminate hiring, which in turn is more attractive to the low-type worker.

We finally consider different models of information acquisition, including the rational inattention, a generalization of the "truth or noise" and a general grading model. For each model, we study test effectiveness as a function of beliefs and the equilibrium in the signaling stage. We find that when tasks require specialization, with the average worker having low chances of succeeding, signaling transmits relevant information in equilibrium. The firm conditions tests on the effort/education level of a candidate, and workers with better previous record face more precise tests, and a higher probability of being hired. When the task at hand requires more general skills, on the other hand, all information is revealed by the firm through a test, which is the same for all applicants.

The situation in which there is no match-specific component to the worker-firm relationship is a special case. High-type workers have a good fit with the firm, while low types do not. In equilibrium, there is a hard limit on the information the worker is able to transmit via signaling. High effort is chosen not only by high types, but also by a fraction of low types, making its observation inconclusive to the firm. Upon observing good credentials, the firm then performs an additional test to determine the suitability of a candidate.

Appendix A

Proof. of Proposition 7

1) The existence of a pooling equilibrium follows directly from Definition 4 4. In particular, x_0 should be lower than the most restrictive of the two no-deviation conditions in Definition 4.

2) The two no-deviation conditions in Definition 5 5 imply that x_H is such that

$$\frac{w}{c}\left[P_{Y|L}\left(H\right)-P_{Y|L}\left(L\right)\right]\leq x_H\leq\frac{w}{c}\left[P_{Y|H}\left(H\right)-P_{Y|H}\left(L\right)\right]$$

which is equivalent to

$$P_{Y|G}(H) - P_{Y|G}(L) \geq P_{Y|B}(H) - P_{Y|B}(L) \quad (\Delta P(H) \geq \Delta P(L))$$

3) In any semi-pooling equilibrium, $x_H = \frac{w}{c}\left[P_{Y|L}(\mu'(p)) - P_{Y|L}(L)\right]$, by the indifference condition for the low type (condition 1 in Definition 6 6). The no-deviation condition for the high type then becomes

$$w\left[P_{Y|H}\left(\mu'(p)\right) - P_{Y|L}\left(\mu'(p)\right)\right] > w\left[P_{Y|H}(L) - P_{Y|L}(L)\right]$$

$$P_{Y|G}\left(\mu'(p)\right) - P_{Y|G}(L) \geq P_{Y|B}\left(\mu'(p)\right) - P_{Y|B}(L)$$

Therefore a sufficient condition for the existence of a semi-pooling equilibrium is

$$P_{Y|G}\left(\mu'(0)\right) - P_{Y|G}(L) \geq P_{Y|B}\left(\mu'(0)\right) - P_{Y|B}(L) \quad \left(\Delta P\left(\mu'(0)\right) \geq \Delta P(L)\right)$$

□

Proof. of Theorem 9

1) Consider first $\mu_0 > \mu^*$, and $x_0 \leq \frac{w}{c}\min\{P_{Y|L}(\mu_0) - P_{Y|L}(L), P_{Y|H}(\mu_0) - P_{Y|H}(L)\}$, so that x_0 is a pooling equilibrium by Proposition 7. We can write

$$\begin{aligned} B_L(x, \Pi_L^*) &= \left\{\mu | P_{Y|L}(\mu) - P_{Y|L}(\mu_0) > \tfrac{c}{w}(x - x_0)\right\} \\ B_H(x, \Pi_H^*) &= \left\{\mu | P_{Y|H}(\mu) - P_{Y|H}(\mu_0) > \tfrac{c}{w}(x - x_0)\right\} \end{aligned}$$

For any pooling equilibrium $x_0 > 0$, consider a small deviation $x < x_0$. Then there exists a belief $\hat{\mu} \in (\mu^*, \mu_0)$ such that $P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$. Then $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Since $P_{Y|H}(\hat{\mu}) - P_{Y|H}(\mu_0) > P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$ by assumption 8 (ΔP is decreasing in μ for $\mu > \mu^*$), $\hat{\mu} \in intB_H(x, \Pi_H^*)$ and $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$, then any pooling at $x_0 > 0$ fails the D1 criterion.

Consider now $x_0 = 0$ and a deviation $x > 0$. Then there exists a belief $\hat{\mu} > \mu_0$ satisfying $P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. By assumption 8 $P_{Y|H}(\hat{\mu}) - P_{Y|H}(\mu_0) < P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0)$, and $B_L(x, \Pi_L^*) \supseteq B_H(x, \Pi_H^*)$. Then pooling at 0 is the only pooling equilibrium that satisfies the D1 criterion.

Assume now that condition 2 in Proposition 7 is satisfied, so that there is always a separating equilibrium. We first show that any separating equilibrium but the least-costly separating equilibrium (LCSE) can be ruled out by the D1 criterion. The LCSE is given by $\underline{x} = \frac{w}{c}\left[P_{Y|L}(H) - P_{Y|L}(L)\right]$. For any equilibrium $x_H > \underline{x}$, consider a deviation $x < x_H$. Since $B_L(x, \Pi_L^*) = \emptyset$ by no-deviation condition 1 (in definition 5), and $B_H(x, \Pi_H^*)$ contains at least $\mu = H$, then $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$, and any $x_H > \underline{x}$ can be ruled out by the D1. Consider now the LCSE and a small deviation $x < \underline{x}$. Then there exists a belief $\hat{\mu} < H$ satisfying $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(L)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Then $\hat{\mu} \in int B_H(x, \Pi_H^*)$ by assumption 8:

$$wP_{Y|H}(\hat{\mu}) - cx > wP_{Y|H}(H) - w[P_{Y|L}(H) - P_{Y|L}(L)]$$

$$P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) > P_{Y|H}(H) - P_{Y|L}(H)$$

and $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$, then any $x_H \in [\underline{x}, \overline{x}]$ fails the D1.

Assume now that condition condition 3 in Proposition 7 is satisfied. Note that any $\mu'(p) \in [\mu_0, H]$ can be generated. Consider a small deviation to $x < x_H$. Since $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$, with $\hat{\mu}$ satisfying $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(\mu'(p)) - cx_H$, it is easy to see that

$$\begin{aligned} wP_{Y|H}(\hat{\mu}) - cx &= wP_{Y|H}(\hat{\mu}) - wP_{Y|L}(\hat{\mu}) + wP_{Y|L}(\mu'(p)) - cx_H \\ P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) &> P_{Y|H}(\mu'(p)) - P_{Y|L}(\mu'(p)) \end{aligned}$$

by assumption 8.

Therefore $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$ and any equilibrium $\{(x_H, [(0, x_H), (p, 1-p)]), (L, \mu'(p))\}$ such that $\mu'(p) > \mu_0 > \mu^*$ fails the D1.

2) Consider now the case in which $\mu_0 < \mu^*$. For any pooling equilibrium $x_0 \leq \frac{w}{c}\min\{P_{Y|L}(\mu_0) - P_{Y|L}(L), P_{Y|H}(\mu_0) - P_{Y|H}(L)\}$, consider a small deviation $x > x_0$. Then there exists a belief $\hat{\mu}$ that satisfy

$P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Since $\mu_0 < \hat{\mu} < \mu^*$, it is direct that $P_{Y|H}(\hat{\mu}) - P_{Y|H}(\mu_0) > P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$ (by assumption 8). Therefore $\hat{\mu} \in intB_H(x, \Pi_H^*)$ and $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$, and any pooling equilibrium fails the D1.

Consider now the case in which $L < H < \mu^*$. It is straightforward to show that any separating equilibrium but the LCSE can be easily ruled out by the D1 criterion (as in 1)). Consider the LCSE $\underline{x} = \frac{w}{c}\left[P_{Y|L}(H) - P_{Y|L}(L)\right]$ and a potential deviation $x < \underline{x}$. Then there exists a belief $\hat{\mu}$ that satisfy $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(L)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. It is easy to see that $\hat{\mu} \notin intB_H(x, \Pi_H^*)$:

$$
\begin{aligned}
& wP_{Y|H}(\hat{\mu}) - cx < wP_{Y|H}(H) - w[P_{Y|L}(H) - P_{Y|L}(L)] \\
\iff \quad & P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) < P_{Y|H}(H) - P_{Y|L}(H)
\end{aligned}
$$

by assumption 8. Therefore $B_H(x, \Pi_H^*) \subseteq B_L(x, \Pi_L^*)$, and the only separating equilibrium that satisfies the D1 is the LCSE.

Consider now any semipooling equilibrium as defined in 6, and a deviation $x > x_H$. Then there exists a belief $\hat{\mu}$ satisfying $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(\mu'(p)) - cx_H$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$.It is easy to see that

$$
\begin{aligned}
wP_{Y|H}(\hat{\mu}) - cx &= wP_{Y|H}(\hat{\mu}) - wP_{Y|L}(\hat{\mu}) + wP_{Y|L}(\mu'(p)) - cx_H \\
P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) &> P_{Y|H}(\mu'(p)) - P_{Y|L}(\mu'(p))
\end{aligned}
$$

by assumption 8. Therefore $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$ and any semipooling equilibrium such that $\mu'(p) < \mu^*$ fails the D1.

Finally consider the case in which $L < \mu^* < H$. Assume that condition 2 in Proposition 7 is satisfied. Consider the LCSE and a potential deviation $x < \underline{x}$. Then there exists a belief $\hat{\mu}$ that satisfy $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(L)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Here we have that $\hat{\mu} \in intB_H(x, \Pi_H^*)$:

$$
\begin{aligned}
& wP_{Y|H}(\hat{\mu}) - cx > wP_{Y|H}(H) - w[P_{Y|L}(H) - P_{Y|L}(L)] \\
\iff \quad & P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) > P_{Y|H}(H) - P_{Y|L}(H)
\end{aligned}
$$

by assumption 8. Then $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$, so that no separating equilibrium satisfies the D1.

Consider now any semipooling equilibrium as defined in 6, and a deviation $x > x_H$. Then there exists a belief $\hat{\mu}$ satisfying $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(\mu'(p)) - cx_H$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. It is easy to see that

$$\begin{aligned} wP_{Y|H}(\hat{\mu}) - cx &= wP_{Y|H}(\hat{\mu}) - wP_{Y|L}(\hat{\mu}) + wP_{Y|L}(\mu'(p)) - cx_H \\ P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) &> P_{Y|H}(\mu'(p)) - P_{Y|L}(\mu'(p)) \end{aligned}$$

by assumption 8. Therefore $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$ and any semipooling equilibrium that induces $\mu'(p) < \mu^*$ fails the D1. The proof is analogous if the equilibrium involves $\mu'(p) > \mu^*$. Consider now a semi-pooling equilibrium such that x_H induces belief μ^*, and a potential deviation $x > x_H$. Then there exists a belief $\hat{\mu}$ satisfying $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(\mu'(p)) - cx_H$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Then

$$\begin{aligned} wP_{Y|H}(\hat{\mu}) - cx &= wP_{Y|H}(\hat{\mu}) - wP_{Y|L}(\hat{\mu}) + wP_{Y|L}(\mu^*) - cx_H \\ P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) &< P_{Y|H}(\mu^*) - P_{Y|L}(\mu^*) \end{aligned}$$

because μ^* is such that the $\Delta P(\mu)$ is maximal. The same argument applies if we consider a deviation $x < x_H$. Therefore $B_L(x, \Pi_L^*) \supseteq B_H(x, \Pi_H^*)$ and the only equilibrium that satisfies the D1 is a semipooling equilibrium such that x^* induces the belief μ^*.

3) We now consider the non-generic case where $\mu^* = \mu_0$. By the same argument in 1) we can rule out any separating and semi-pooling equilibria. For any pooling equilibrium x_0, consider a small deviation $x > x_0$. Then there exists a belief $\hat{\mu} > \mu^* = \mu_0$ such that $P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$. Then $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Since $P_{Y|H}(\hat{\mu}) - P_{Y|H}(\mu_0) < P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$ (ΔP is maximal at $\mu_0 = \mu^*$), $\hat{\mu} \notin int B_H(x, \Pi_H^*)$ and $B_L(x, \Pi_L^*) \supseteq B_H(x, \Pi_H^*)$. Then any pooling satisfies the D1 criterion.

□

Proof. of Theorem 10

1) If $\mu_{max} = H$, the only D1 equilibrium is the LCSE. In fact, Condition 2 in Proposition 7 is satisfied, and the set of separating equilibria is given by $x_H \in [\underline{x}, \overline{x}]$ where $\underline{x} = \frac{w}{c}\left[P_{Y|L}(H) - P_{Y|L}(L)\right]$ (the LCSE). As shown in Proof of Theorem 9, any separating equilibrium which is not the LCSE can be ruled out by

the D1. Consider now the LCSE and a small deviation $x < \underline{x}$ (the only relevant one). There exists a belief $\hat{\mu}$ that satisfy $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(L)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. It is easy to see that $\hat{\mu} \notin intB_H(x, \Pi_H^*)$:

$$
\begin{aligned}
& wP_{Y|H}(\hat{\mu}) - cx < wP_{Y|H}(H) - w[P_{Y|L}(H) - P_{Y|L}(L)] \\
\iff \quad & P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) < P_{Y|H}(H) - P_{Y|L}(H)
\end{aligned}
$$

since $\Delta P(\mu)$ is maximal at $\mu_{max} = H$. Then $B_H(x, \Pi_H^*) \subseteq B_L(x, \Pi_L^*)$, and a deviation should be attributed to type L, therefore the LCSE satisfies D1.

For any pooling equilibrium $x_0 \leq \frac{w}{c} \min\{P_{Y|L}(\mu_0) - P_{Y|L}(L), P_{Y|H}(\mu_0) - P_{Y|H}(L)\}$, consider a small deviation $x > x_0$ such that $P_{Y|L}(H) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Since $\mu_0 < H$,

$$
\hat{\mu} \in intB_H(x, \Pi_H^*) \iff P_{Y|H}(\hat{\mu}) - P_{Y|H}(\mu_0) > P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)
$$

therefore $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$, and the pooling equilibrium fails the D1.

Consider now semi-pooling equilibria. Then condition 3 in Proposition 7 is satisfied. Note that any $\mu'(p) \in [\mu_0, H]$ can be generated. Consider a small deviation $x > x_H$. Since $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$, with $H \geq \hat{\mu} > \mu'(p)$ satisfying $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(\mu'(p)) - cx_H$, it is easy to see that

$$
\begin{aligned}
wP_{Y|H}(\hat{\mu}) - cx &= wP_{Y|H}(\hat{\mu}) - wP_{Y|L}(\hat{\mu}) + wP_{Y|L}(\mu'(p)) - cx_H \\
P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) &> P_{Y|H}(\mu'(p)) - P_{Y|L}(\mu'(p))
\end{aligned}
$$

Therefore $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$ and any semi-pooling equilibrium fails the D1 criterium.

2) Consider now $\mu_{max} < H$ and a semi-pooling equilibrium as defined in 6. Take a small deviation $x > x_H$. Then there exists a belief $\hat{\mu} \in (\mu_0, \mu_{max}]$ satisfying $wP_{Y|L}(\hat{\mu}) - cx = wP_{Y|L}(\mu'(p)) - cx_H$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. It is easy to see that

$$
\begin{aligned}
wP_{Y|H}(\hat{\mu}) - cx &= wP_{Y|H}(\hat{\mu}) - wP_{Y|L}(\hat{\mu}) + wP_{Y|L}(\mu'(p)) - cx_H \\
P_{Y|H}(\hat{\mu}) - P_{Y|L}(\hat{\mu}) &> P_{Y|H}(\mu'(p)) - P_{Y|L}(\mu'(p))
\end{aligned}
$$

since ΔP is maximal at μ_{max}. Therefore $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$ and any semipooling equilibrium that induces $\hat{\mu} > \mu'(p)$ fails the D1. The argument is analogous if we consider a deviation $x < x_H$ that induces $\hat{\mu} < \mu'(p)$. Therefore the only semi-pooling equilibrium that satisfies the D1 is such that x^* induces the belief μ_{max}.

Noting that ΔP is maximal at μ_{max} , with $\mu_{max} \in (\mu_0, H)$, by the same argument in 1) and in Proof of Theorem 9 we can rule out any pooling and separating equilibrium.

3) Consider now $\mu_{max} = \mu_0$. For any pooling equilibrium x_0, consider a small deviation $x > x_0$. Then there exists a belief $\hat{\mu} > \mu_0$ satisfying $P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Since $P_{Y|H}(\hat{\mu}) - P_{Y|H}(\mu_0) < P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0)$ (ΔP is maximal at μ_0), we have that $B_L(x, \Pi_L^*) \supseteq B_H(x, \Pi_H^*)$. Then any pooling satisfies the D1 criterion.

By the same argument in 1) and 2) and in Proof of Theorem 9 we can rule out any separating and semi-pooling equilibria.

□

Proof. of Theorem 11

1) Consider the case in which $\mu_{max} < \mu_0$ and $\mu_0 = \arg\max_{\mu \geq \mu_0} \Delta P(\mu)$. For any pooling equilibrium $x_0 > 0$, consider a small deviation $x < x_0$. Then there exists a belief $\hat{\mu} \in [L, \mu_0)$ such that $P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$. Then $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. Since $P_{Y|H}(\hat{\mu}) - P_{Y|H}(\mu_0) > P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$, $\hat{\mu} \in intB_H(x, \Pi_H^*)$ and $B_L(x, \Pi_L^*) \subsetneq B_H(x, \Pi_H^*)$, then any pooling at $x_0 > 0$ fails the D1 criterion (there exists at least a deviation inducing $\mu_{max} \in [L, \mu_0)$, such that $\Delta P(\mu_{max})$ is maximal).

Consider now $x_0 = 0$ and a deviation $x > 0$. Then there exists a belief $\hat{\mu} > \mu_0$ satisfying $P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0) = \frac{c}{w}(x - x_0)$, so that $B_L(x, \Pi_L^*) = (\hat{\mu}, H]$. It is easy to see that $P_{Y|H}(\hat{\mu}) - P_{Y|H}(\mu_0) < P_{Y|L}(\hat{\mu}) - P_{Y|L}(\mu_0)$ if and only if $\mu_0 = \arg\max_{\mu \geq \mu_0} \Delta P(\mu)$. Therefore $B_L(x, \Pi_L^*) \supseteq B_H(x, \Pi_H^*)$ and pooling at 0 is the only equilibrium that satisfies the D1 criterion.

In order to prove that there is no separating nor semi-pooling that satisfies the D1 follows logic and argument in Proofs of Theorem 9 and 10.

Consider now the case in which $\mu_0 \neq \arg\max_{\mu \geq \mu_0} \Delta P(\mu)$, any pooling, separating or semi-pooling equilibrium can be ruled out using the same argument in Proofs of Theorem 9 and 10. □

Rational Inattention - Analysis

Following Matějka and McKay (2015), the firm's problem can be reformulated as one in which the firm optimally chooses $P_{a|q}$. Abusing notation, write $\mu := \mu_G$ and $1-\mu := \mu_B$. Then the firm solves:

$$\max_{P_{a|q}} \sum_{a\in\{Y,N\}} \sum_{q\in\{G,B\}} u\left(a \mid q\right) P_{a|q}\mu_q$$

$$-\lambda \left[\sum_{a\in\{Y,N\}} P_a log\left(P_a\right) - \sum_{a\in\{Y,N\}} \sum_{q\in\{G,B\}} P_{a|q} log\left(P_{a|q}\right)\mu_q \right]$$

$$s.t. \begin{cases} P_{a|q} \geq 0, \forall q \\ \sum\limits_{a\in\{Y,N\}} P_{a|q} = 1 \end{cases}$$

The solution leads to the conditional probabilities in 1, such that $P_{Y|q}$ is increasing in μ for any $q \in \{G, B\}$. Moreover $P_{Y|G}$ is concave, while $P_{Y|B}$ is convex.

The auxiliary problem to solve for the unconditional choice probabilities P_a is the following:

$$\max_{P_a} \sum_{q\in\{G,B\}} \lambda log \left\{ \sum_{a\in\{Y,N\}} P_a \exp\left\{\frac{q-w}{\lambda}\right\} \right\} \mu_q$$

$$s.t. \begin{cases} P_a \geq 0, \forall a \\ \sum_{a\in\{Y,N\}} P_a = 1 \end{cases}$$

that leads to

$$P_Y = \begin{cases} 0 \qquad for\ \mu \leq \frac{1-\exp\left\{\frac{B-w}{\lambda}\right\}}{\exp\left\{\frac{G-w}{\lambda}\right\}-\exp\left\{\frac{B-w}{\lambda}\right\}} \\ -\frac{\mu\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]+(1-\mu)\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]}{\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]} \qquad for \quad \mu \in \left[\frac{1-\exp\left\{\frac{B-w}{\lambda}\right\}}{\exp\left\{\frac{G-w}{\lambda}\right\}-\exp\left\{\frac{B-w}{\lambda}\right\}}, \frac{\exp\left\{\frac{G-w}{\lambda}\right\}\left[1-\exp\left\{\frac{B-w}{\lambda}\right\}\right]}{\exp\left\{\frac{G-w}{\lambda}\right\}-\exp\left\{\frac{B-w}{\lambda}\right\}}\right] \\ 1 \qquad for\ \mu \geq \frac{\exp\left\{\frac{G-w}{\lambda}\right\}\left[1-\exp\left\{\frac{B-w}{\lambda}\right\}\right]}{\exp\left\{\frac{G-w}{\lambda}\right\}-\exp\left\{\frac{B-w}{\lambda}\right\}} \end{cases} \tag{5}$$

Replacing 5 in1, we obtain the following conditional probabilities:

$$\begin{cases} P_{Y|G} = & \frac{\left\{\mu\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]+(1-\mu)\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]\right\}\exp\left\{\frac{G-w}{\lambda}\right\}}{\left\{\mu\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]+(1-\mu)\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]\right\}\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]-\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]} \\ P_{Y|B} = & \frac{\left\{\mu\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]+(1-\mu)\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]\right\}\exp\left\{\frac{B-w}{\lambda}\right\}}{\left\{\mu\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]+(1-\mu)\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]\right\}\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]-\left[\exp\left\{\frac{G-w}{\lambda}\right\}-1\right]\left[\exp\left\{\frac{B-w}{\lambda}\right\}-1\right]} \end{cases}$$

Proof. of Lemma 12

Consider the conditional hiring probabilities in 1. Then $\frac{\partial}{\partial\mu}\left[\Delta P\left(\mu\right)\right] = \frac{\partial}{\partial\mu}\left[P_{Y|G}\left(\mu\right) - P_{Y|B}\left(\mu\right)\right] \geq 0$ can be written as:

$$\begin{aligned} \frac{\exp\left\{\frac{w-G}{\lambda}\right\}}{\left(1+\frac{1-P_Y(\mu)}{P_Y(\mu)}\exp\left\{\frac{w-G}{\lambda}\right\}\right)^2} &\geq \frac{\exp\left\{\frac{w-B}{\lambda}\right\}}{\left(1+\frac{1-P_Y(\mu)}{P_Y(\mu)}\exp\left\{\frac{w-B}{\lambda}\right\}\right)^2} \\ \frac{1-P_Y\left(\mu\right)}{P_Y\left(\mu\right)} &\geq \frac{\exp\left\{\frac{G-B}{2\lambda}\right\}-1}{\exp\left\{\frac{w-B}{\lambda}\right\}\left[1-\exp\left\{\frac{B-G}{2\lambda}\right\}\right]} \end{aligned}$$

Since $\frac{1-P_Y(\mu)}{P_Y(\mu)}$ is decreasing in μ, $\frac{1-P_Y(\mu)}{P_Y(\mu)} \geq \frac{\exp\left\{\frac{G-B}{2\lambda}\right\}-1}{\exp\left\{\frac{w-B}{\lambda}\right\}\left[1-\exp\left\{\frac{B-G}{2\lambda}\right\}\right]}$ if and only if $\mu \leq \mu^*$. ☐

Proof. of Lemma 15

Let denote by l the firm's objective function. Then $\frac{\partial l(\mu,\sigma)}{\partial\sigma}$

$$= \frac{\partial}{\partial\sigma}\left\{\sigma \int\limits_{s\geq\overline{s}(\mu)} \left[\mu f_G(s)(G-w)+(1-\mu)(B-w)\right] ds + (1-\sigma)_{\{\mu\geq\bar{\mu}\}}\left[\mu(G-w)+(1-\mu)(B-w)\right]\right\} - C^{'}(\sigma)$$

$$= \int\limits_{s\geq\overline{s}(\mu)} \left[\mu f_G(s)(G-w)+(1-\mu)(B-w)\right] ds -_{\{\mu\geq\bar{\mu}\}}\left[\mu(G-w)+(1-\mu)(B-w)\right] - C^{'}(\sigma)$$

and therefore

$$\frac{\partial^2 l(\mu,\sigma)}{\partial\sigma\partial\mu} = \int\limits_{s\geq\overline{s}(\mu)} \left[f_G(s)(G-w)-(B-w)\right] ds -_{\{\mu\geq\bar{\mu}\}}(G-B)$$

It is easy to see that:

$$\frac{\partial^2 l(\mu,\sigma)}{\partial\sigma\partial\mu} = \begin{cases} \int_{s\geq\overline{s}(\mu)} \left[f_G(s)(G-w)-(B-w)\right] ds \geq 0 \quad if \quad \mu\leq\bar{\mu} \\ \int_{s\geq\overline{s}(\mu)} \left[f_G(s)(G-w)\right] ds - (G-w) \leq 0 \quad if \quad \mu\geq\bar{\mu} \end{cases}$$

Therefore the firm's objective function is supermodular in (σ,μ) if $\mu\leq\bar{\mu}$, and supermodular in $(\sigma,-\mu)$ if $\mu\geq\bar{\mu}$. □

Proof. of Lemma 16

Consider

$$\frac{\partial}{\partial\mu}\left[\Delta P(\mu)\right] = \underbrace{\sigma'(\mu)\left[\overline{s}(\mu)-F_G(\overline{s}(\mu))\right]}_{\textbf{Precision Effect}} + \underbrace{\frac{d\overline{s}(\mu)}{d\mu}\sigma(\mu)\left[1-f_G(\overline{s}(\mu))\right]}_{\textbf{Passing Grade Effect}}$$

When $\mu\leq\bar{\mu}$, precision is increasing in beliefs (see Lemma 15). Moreover MLRP implies FOSD, so that $\overline{s}(\mu) > F_G(\overline{s}(\mu))$. Thus the precision effect is positive. The passing grade is decreasing in beliefs:

$$\frac{\partial\overline{s}}{\partial\mu} \quad = \quad \frac{(B-w)-(G-w)f_G(\overline{s})}{\mu(G-w)\frac{\partial f_G(\overline{s})}{\partial s}} < 0$$

Moreover $[1 - f_G(\overline{s}(\mu))] \leq 0$ for any $\mu \leq \bar{\mu}$ (since $\overline{\mu}$ is such that $\bar{\mu}(G-w) + (1-\bar{\mu})(B-w) = 0$, $\bar{\mu} f_G(\overline{s}(\mu))(G-w) + (1-\bar{\mu})(B-w) = 0$, and $f_G(\overline{s}(\mu))$ is increasing). Then the passing grade effect is positive. Therefore $\frac{\partial}{\partial \mu}[\Delta P(\mu)] \geq 0$.

Similarly when $\mu \geq \bar{\mu}$, the precision effect is negative. The passing grade effect is also negative since the passing grade is decreasing in beliefs and $[1 - f_G(\overline{s}(\mu))] \geq 0$. Therefore for any $\mu \geq \bar{\mu}$, $\frac{\partial}{\partial \mu}[\Delta P(\mu)] \leq 0$.

□

Proof. of Theorem 18

The test effectiveness responds to beliefs in the following way:

$$\frac{\partial \Delta P(\mu)}{\partial \mu} = -\sigma'(\mu) \frac{\partial F_G(\overline{s}(\sigma(\mu), \mu); \sigma(\mu))}{\partial \sigma} + [1 - f_G(\overline{s}(\sigma(\mu), \mu); \sigma(\mu))] \frac{d\overline{s}(\sigma(\mu), \mu)}{d\mu} \tag{6}$$

Note that $[1 - f_G(\overline{s}; \sigma)] > 0$ iff $\overline{s} < \hat{s}$. Moreover we know that $\frac{\partial F_G(\overline{s};\sigma)}{\partial \sigma} < 0$ by assumption 14. It is then critical to sign $\frac{d\overline{s}(\sigma(\mu),\mu)}{d\mu}$ and $\sigma'(\mu)$.

1) We first analyze $\frac{d\overline{s}(\sigma(\mu),\mu)}{d\mu}$. The passing grade $\overline{s}$ is defined

$$\mu f_G(\overline{s}(\sigma(\mu), \mu); \sigma(\mu))(G-w) + (1-\mu)(B-w) = 0 \tag{7}$$

Therefore

$$\frac{d\overline{s}}{d\mu} = \frac{\partial \overline{s}}{\partial \mu} + \frac{\partial \overline{s}}{\partial \sigma}\sigma'(\mu)$$

It is direct that $\frac{\partial \overline{s}}{\partial \mu} < 0$: As the firm becomes more optimistic about the worker's fit, it relaxes standards. It is also easy to verify that $\frac{\partial \overline{s}}{\partial \sigma} > 0$ if and only if $\overline{s} < \hat{s}$. In this case grades at and above $\overline{s}$ are more likely to be obtained by workers with a bad fit, and, as the test becomes more precise, this effect is intensified. Therefore an increase in precision deteriorates beliefs at such cutoff, and it makes it necessary for the firm to increase the passing grade. The opposite happens when $\overline{s} > \hat{s}$.

Consider μ_1 such that $\overline{s}(\mu_1) = \hat{s}$ (that is $\mu_1 = \frac{w-B}{G-B}$). Then we have from equation 7 that $\mu > \mu_1$ requires $f_G(\overline{s}; \sigma) < 1$ and therefore $\overline{s} < \hat{s}$. This allows us to conclude that:

- $\mu < \mu_1 \Longleftrightarrow \overline{s} > \hat{s}$ and $\frac{\partial \overline{s}}{\partial \sigma} < 0$
- $\mu > \mu_1 \Longleftrightarrow \overline{s} < \hat{s}$ and $\frac{\partial \overline{s}}{\partial \sigma} > 0$

2) We now show that $\sigma'(\mu)$ is inverse U-shaped. More precisely there exists a belief $\mu_2 > \mu_1$ such that:

- $\sigma'(\mu) > 0$ if $\mu < \mu_2$
- $\sigma'(\mu) < 0$ if $\mu > \mu_2$

To do that, we show that the objective function satisfies a single-crossing property:

$$\begin{aligned}
\frac{\partial^2}{\partial \sigma \partial \mu}\left[\int_{s \geq \overline{s}(\sigma(\mu),\mu)} \left[\mu f_G(s;\sigma)(G-w) + (1-\mu)(B-w)\right] ds\right] &= \frac{\partial}{\partial \mu}\left\{\mu (G-w) \int_{s \geq \overline{s}(\sigma(\mu),\mu)} \frac{\partial f_G(s;\sigma)}{\partial \sigma} ds\right\} \\
&\simeq \int_{s \geq \overline{s}(\sigma(\mu),\mu)} \frac{\partial f_G(s;\sigma)}{\partial \sigma} ds - \mu \frac{\partial \overline{s}}{\partial \mu} \frac{\partial f_G(\overline{s};\sigma)}{\partial \sigma} \\
&= -\frac{\partial F_G(s;\sigma)}{\partial \sigma} - \mu \frac{\partial \overline{s}}{\partial \mu} \frac{\partial f_G(\overline{s};\sigma)}{\partial \sigma}
\end{aligned}$$

Note that if $\mu < \mu_1$, then $\overline{s} > \hat{s}$, and therefore $\frac{\partial f_G(s;\sigma)}{\partial \sigma} > 0$. Moreover $\frac{\partial F_G(s;\sigma)}{\partial \sigma} < 0$ by assumption 14 and $\frac{\partial \overline{s}}{\partial \mu} < 0$, as shown in 1. Therefore $\sigma'(\mu) > 0$. We now proceed to analyze $\mu > \mu_1$.

Define by μ_2 the belief such that $\frac{\partial^2}{\partial \sigma \partial \mu}\left[\int_{s \geq \overline{s}(\sigma(\mu),\mu)} \left[\mu f_G(s;\sigma)(G-w) + (1-\mu)(B-w)\right] ds\right] = 0$. Then $\sigma'(\mu_2) = 0$. We will prove that $\frac{\partial^3}{\partial \sigma \partial^2 \mu}\left[\int_{s \geq \overline{s}(\sigma(\mu_2),\mu_2)} \left[\mu_2 f_G(s;\sigma)(G-w) + (1-\mu_2)(B-w)\right] ds\right] < 0$, establishing a single-crossing property.

If $\frac{\partial^2}{\partial \sigma \partial \mu}\left[\int_{s \geq \overline{s}(\sigma(\mu),\mu)} \left[\mu f_G(s;\sigma)(G-w) + (1-\mu)(B-w)\right] ds\right] < 0$ at μ_2, it is negative for any $\mu > \mu_2$, which in turn implies the result.

We have that

$$\frac{\partial}{\partial \mu}\left[\int\limits_{s\geq\overline{s}(\sigma(\mu),\mu)} \frac{\partial f_G(s;\sigma)}{\partial\sigma}ds - \mu(G-w)\frac{\partial\overline{s}(\mu)}{\partial\mu}\frac{\partial f_G(\overline{s};\sigma)}{\partial\sigma}\right]|_{\mu=\mu_2} = \frac{\partial}{\partial\mu}\left[-\frac{\partial F_G(\overline{s};\sigma)}{\partial\sigma} - \mu\frac{\partial\overline{s}}{\partial\mu}\frac{\partial f_G(\overline{s};\sigma)}{\partial\sigma}\right]$$
$$= -\frac{\partial f_G(\overline{s};\sigma)}{\partial\sigma}\frac{\partial\overline{s}}{\partial\mu} - \frac{\partial\overline{s}}{\partial\mu}\frac{\partial f_G(\overline{s};\sigma)}{\partial\sigma} - \mu\frac{\partial^2\overline{s}}{\partial^2\mu}\frac{\partial f_G(\overline{s};\sigma)}{\partial\sigma} - \mu\left(\frac{\partial\overline{s}}{\partial\mu}\right)^2\frac{\partial^2 f_G(\overline{s};\sigma)}{\partial\sigma\partial s} \quad (8)$$

We now analyze each of these terms to prove the single-crossing property:

$$\frac{\partial^2\overline{s}}{\partial^2\mu} = -\frac{1}{\mu}\frac{\partial\overline{s}}{\partial\mu} + \frac{1}{\mu(G-w)}\frac{\partial}{\partial\mu}\left(\frac{(B-w)-(G-w)f_G(\overline{s};\sigma)}{\frac{\partial f_G(\overline{s};\sigma)}{\partial s}}\right)$$
$$= -\frac{1}{\mu}\frac{\partial\overline{s}}{\partial\mu} - \frac{1}{\mu}\frac{\partial\overline{s}}{\partial\mu} - \frac{1}{\mu(G-w)}\frac{\frac{\partial^2 f_G(\overline{s};\sigma)}{\partial^2 s}\frac{\partial\overline{s}}{\partial\mu}\left[(B-w)-(G-w)f_G(\overline{s};\sigma)\right]}{\left(\frac{\partial f_G(\overline{s};\sigma)}{\partial s}\right)^2} \quad (9)$$

Replacing 9 in 8 we obtain:

$$\frac{\partial}{\partial\mu}\left[\int_{s\geq\overline{s}(\sigma(\mu),\mu)}\frac{\partial f_G(s;\sigma)}{\partial\sigma}ds - \mu(G-w)\frac{\partial\overline{s}}{\partial\mu}\frac{\partial f_G(\overline{s};\sigma)}{\partial\sigma}\right]|_{\mu=\mu_2}$$
$$= \frac{1}{(G-w)}\frac{\frac{\partial^2 f_G(\overline{s};\sigma)}{\partial^2 s}\frac{\partial\overline{s}}{\partial\mu}\left[(B-w)-(G-w)f_G(\overline{s};\sigma)\right]}{\left(\frac{\partial f_G(\overline{s};\sigma)}{\partial s}\right)^2}\frac{\partial f_G(\overline{s};\sigma)}{\partial\sigma} - \mu\left(\frac{\partial\overline{s}}{\partial\mu}\right)^2\frac{\partial^2 f_G(\overline{s};\sigma)}{\partial\sigma\partial s}$$
$$< 0$$

where the last inequality comes from assumption 17.

We now conclude: If $\mu < \mu_1$, then $\frac{\partial\Delta P(\mu)}{\partial\mu} > 0$. By 2) and assumption 14, we have the first term in **??** is positive. Moreover we know that that $[1 - f_G(\overline{s};\sigma)] < 0$ for $\overline{s} > \hat{s}$. Finally by 1), we can conclude that the total derivative $\frac{d\overline{s}}{d\mu} < 0$.

Similarly, for any $\mu \in [\mu_2, 1]$, $\frac{\partial\Delta P(\mu)}{\partial\mu} < 0$. By 2) and assumption 14, we have the first term in **??** is negative. Moreover we know that that $[1 - f_G(\overline{s};\sigma)] > 0$ for $\overline{s} < \hat{s}$. By 1), we can conclude that the total derivative $\frac{d\bar{s}}{d\mu} < 0$. Therefore $\frac{\partial \Delta P(\mu)}{\partial \mu} < 0$.

Finally, for $\mu \in [\mu_1, \mu_2]$, we have that $\sigma'(\mu) > 0$, $\frac{\partial \bar{s}}{\partial \mu} < 0$ and $\frac{\partial \bar{s}}{\partial \sigma} > 0$. Going back to expression (**??**), we can write

$$\begin{aligned}
\frac{\partial \Delta P(\mu)}{\partial \mu} &= -\sigma'(\mu) \frac{\partial F_G(\bar{s};\sigma)}{\partial \sigma} + [1 - f_G(\bar{s};\sigma)] \frac{d\bar{s}}{d\mu} \\
&= \left[(1 - f_G(\bar{s};\sigma)) \frac{\partial \bar{s}}{\partial \sigma} - \frac{\partial F_G(\bar{s};\sigma)}{\partial \sigma} \right] \sigma'(\mu) \\
&\quad + [1 - f_G(\bar{s};\sigma)] \frac{\partial \bar{s}}{\partial \mu}
\end{aligned}$$

where the first term is positive and the second negative. Finally note that any arbitrary change in $C''(\sigma)$ will affect only the absolute value of σ' in $[\mu_1, \mu_2]$, which can be made arbitrarily large or small. Therefore the expression can turn positive or negative, and changes signs as often as wanted.

□

References

[1] Alós-Ferrer, C. and J. Prat (2012), Job Market Signaling and Employer Learning, Journal of Economic Theory, 147(5), 1787–1817.

[2] Banks, J. and J. Sobel (1987), Equilibrium Selection in Signaling games, Econometrica, 55, 647–662.

[3] Boyacı, T. and Y. Akçay (2017), Pricing when customers have limited attention, Management Science, 64 (7), 2973-3468.

[4] Bester, H., Lang, M. and J. Li (2019), Signaling versus Auditing, CESifo Working Paper Series No. 7183.

[5] Bester, H. and K. Ritzberger (2001), Strategic pricing, signalling, and costly information acquisition, International Journal of Industrial Organization, 19(9), 1347–1361.

[6] Cho, I.K. and D. M. Kreps (1987), Signaling Games and Stable Equilibria, Quarterly Journal of Economics, 102(2), 179–221.

[7] Cho, I.K. and J. Sobel (1990), Strategic Stability and Uniqueness in Signaling Games, Journal of Economic Theory, 50, 381–413.

[8] Daley, B. and B. Green (2014), Market signaling with grades, Journal of Economic Theory, 151, 114-145.

[9] Garfagnini, U. (2017), The Downsides of Managerial Oversight in Signaling Environment, Working Paper, University of Surrey.

[10] Ganuza, J.J and J.S. Penalva (2010), Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions, Econometrica, 78(3), 1007-1030.

[11] Kreps, D. M. and J. Sobel (1994), Signalling, Chapter 25 in Handbook of Game Theory with Economic Applications (2), North–Holland, Amsterdam, 849–867.

[12] Kurlat, P. and F. Scheuer (2019), Signaling to Experts, NBER Working Paper No. 23817.

[13] Lewis, T. R., and D. E. Sappington (1994), Supplying Information to Facilitate Price Discrimination, International Economic Review, 35 (2), 309–27.

[14] Martin, D. (2017), Strategic pricing with rational inattention to quality, Games and Economic Behavior, 104, 131-145.

[15] Matějka F. and A. McKay (2015), Rational inattention to discrete choices: A new foundation for the multinomial logit model, American Economic Review, 105(1), 272–298.

[16] Mayzlin, D. and J. Shin (2011), Uninformative Advertising as an Invitation to Search, Marketing Science, 30(4), 565-756.

[17] Milgrom, P.R. (1981), Good News and Bad News: Representation Theorems and Applications, The Bell Journal of Economics, 12(2), 380-391.

[18] Perez-Richet, E. and D. Prady (2011), Complicating to Persuade, SSRN Working Paper No. 1868066

[19] .Riley, J. (2001), Silver Signals: Twenty-Five Years of Screening and Signaling, Journal of Economic Literature 39(2), 432–478.

[20] Sims, C.A. (2003), Implications of rational inattention, Journal of Monetary Economics, 50(3), 665–690.

[21] Sims, C. A. (2006), Rational inattention: Beyond the linear-quadratic case, The American Economic Review, 96(2), 158–163.

[22] Spence, M.A.(1973), Job market signaling, Quarterly Journal of Economics, 90, 225–243.