CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 724

Flexibility, Rigidity, and Competitive Experimentation

Luca Picariello and Alexander Rodivilov

June 2024

University of Naples Federico II **University of Salerno** **Bocconi University, Milan**

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS AND STATISTICS – UNIVERSITY OF NAPLES FEDERICO II
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

WORKING PAPER NO. 724

Flexibility, Rigidity, and Competitive Experimentation

Luca Picariello* **and Alexander Rodivilov**†

Abstract

We study a framework in which agents can generate signals to increase their expected productivity. Such signals can be generated in heterogeneous environments: a flexible system in which the agent can freely allocate effort across different tasks, and a rigid system in which the agent must devote effort to all tasks. We provide sufficient and necessary conditions for optimal experimentation in each system. Experimentation is less likely if the agent has high bargaining power. Competition within the Flexible system makes specialization more likely. When agents from different systems compete, there is a unique equilibrium where both agents experiment if the Rigid System is restrictive enough.

JEL Classification: D61, D83, I21, I23, I28, J63, J65.

Keywords: Career Concerns, Experimentation, Learning.

Acknowledgments: We thank for helpful comments Guido Friebel, Ayush Gupta, Emmanuel Raynaud, Alexandre Volle and participants in seminars at the 2024 Spring Midwest Economic Theory Conference, the 21st session of the Institutional and Organizational Economics Academy, and the 35th Stony Brook International Conference on Game Theory.

* University of Naples Federico II and CSEF. Email: luca.picariello@unina.it

† School of Business, Stevens Institute of Technology. Email: arodivil@stevens.edu

1 Introduction

It is common for economic agents to discover their talent (productivity) before entering the labor market. For example, students choose majors and take classes to increase their attractiveness to potential employers.[1] During a probation period, job applicants undergo additional training to increase their chances of being hired. Moreover, either by design or due to historical reasons, economic agents often operate in different institutional environments. For instance, in some countries, students are restricted in their choices of subjects as there are mandatory classes everybody must take. In other countries, students' choices are more flexible. During the probation, some companies require job applicants to devote effort to tasks not perfectly correlated with their future duties. In other companies, job applicants devote their time and effort to a narrow task set directly related to their future jobs.

How are the agents' incentives to discover their productivity shaped by the institutional environment? How do these incentives depend on agents' bargaining powers and competition? What if agents from one institutional environment (for example, more flexible) compete with agents from another (i.e., more restrictive)? How should an employer treat applicants from different systems in the global international market?

We develop a model where a principal (she) hires an agent (he). The agent's productivity (type) could be either high or low. The agent's type is initially unknown. The prior probability that the agent is a high type is common knowledge. The high type is a productive agent who can generate a high profit if hired. The low type is not productive. Thus, the principal would hire only the high type if the type was known. Consequently, the principal hires the agent only if the belief that he is the high type is sufficiently high.

Before the hiring decision, the agent generates an informative signal observed by the principal. The agent is endowed with a fixed amount of time (effort) and can allocate it towards two symmetric tasks. If the agent allocates some of his time toward a task, he either succeeds or fails.[2] Signaling technology is taking the form of looking for good news. That is, the probability of success is increasing in both the agent's type and the devoted effort. Success in any of the two tasks increases the principal's belief that the agent is a high-type.

Our model captures the idea that markets operate with incomplete information regarding the applicants' productivity. Additional evidence, such as informative signals, updates beliefs regarding the applicant's type. To sharpen the intuition, we assume the principal hires the agent only if he succeeds in at least one task.[3]

[1]Similarly, high schoolers choose classes to signal their potential to universities.

[2]Rodina and Farragut (2016) show that the optimal grading policy has a threshold form in a similar framework under symmetric information. Bonatti and Hörner (2017) model success as a breakthrough in a professional career that reveals the worker's type.

[3]We also assume that the initial belief is sufficiently low so that if the agent chooses not to exert any

We distinguish two scenarios depending on how restricted the agent is in allocating his effort. In the ***Flexible*** system, the agent picks any allocation of his effort endowment across the two tasks. That is, the agent can choose to exert some effort toward both tasks or concentrate on one task only. The agent *experiments* if he splits his effort equally across the two tasks. The agent *specializes* if he devotes his entire effort endowment toward one task only. In the ***Rigid*** system, the agent must devote some (non-zero) effort level to both tasks. Therefore, specialization is not feasible.

After observing the signal generated by the agent's chosen effort allocation, the principal updates her belief that the agent is of high type. If the belief is sufficiently high (the agent succeeds at least in one task), the agent is hired. His wage is then increasing in his expected productivity. To capture various competitive environments, we assume the agent is paid a portion (i.e., bargaining power) of his expected productivity.

We begin the analysis by describing the efficient allocation of effort across the tasks. Efficiency requires experimentation (equally splitting the effort endowment across the two tasks) if either the two types are sufficiently close in terms of productivity, the ex-ante share of high types is relatively low, or loss in production imposed by a low-type is quite significant compared to the productive benefit generated by a high-quality agent. To see the intuition, note the following trade-off. On the one hand, experimentation is more "volatile" as it generates more dispersed posterior beliefs: high beliefs after two successes, intermediate after one success on either task and low beliefs after two failures. On the other hand, specialization is less "risky" as the posterior it generates are closer to each other: either one success or one failure. There is a lower risk if the two types are sufficiently close in productivity. Hence, agents are willing to experiment in this scenario as this would improve their expected productivity, though increasing risk.

In the Flexible system, where the agent can pick any allocation of his effort across the two tasks, he faces the following trade-off. On the one hand, the expected wage is higher if he succeeds in both tasks. Hence, there is a possible benefit from experimentation. On the other hand, experimentation increases the chance of failing (on one or even both of the tasks), which entails a lower expected wage. Thus, there is a possible benefit from specialization. We find that experimentation remains optimal under conditions similar to those in case of the efficient effort allocation. The reason is that, if the agent has all the bargaining power vis-á-vis the principal, his preferences are perfectly aligned with the social ones. A corollary from this observation is that experimentation is less likely in an environment that gives the agent sufficiently large bargaining power.

In the Rigid system, the agent must devote some effort to both tasks. As a result,

effort on both tasks, he is not hired.

specialization is not possible. Therefore, if specialization is optimal under some conditions in the Flexible system, the agent in the Rigid system is worse off due to the imposed constraints.

Having characterized the optimal choice of one agent, we next consider the case of competition between two agents. When two agents compete, an agent is hired not only if he succeeds in one task, but also if the posterior beliefs that he is a high type are higher than the beliefs about his rival.

To sharpen the effect of competition, we first discuss the case of competition *within* the Flexible System. That is when the two agents from the Flexible system compete with each other. Multiple equilibria emerge. First, if the two types are sufficiently close in productivity, there is an equilibrium where both the competing agents experiment. Intuitively, experimentation increases the risk of not being hired, but it increases the expected wage conditional on being hired. If the two types are close in productivity, the posterior beliefs from experimentation are close to each other. As a result, the benefit of the higher expected wage outweighs the cost of not being hired. Moreover, if the two types are sufficiently close in productivity, the best response to even a specializing rival is to experiment. Second, we characterize necessary and sufficient conditions for both agents to specialize in equilibrium. An agent who experiments might fail on both tasks while facing a rival who succeeds at least once. As a result, the presence of a rival makes experimentation riskier. Consequently, competition makes specialization more likely. Third, if the two types are sufficiently different in productivity, there is a symmetric equilibrium in which both agents allocate some effort to both tasks.

We then move to a case of competition *between* the two systems. That is when one agent from the Flexible system competes with an agent from the Rigid system. The main conclusion is two-fold. First, if the Rigid System is sufficiently restrictive, there is a unique equilibrium where both agents experiment. Second, if the Rigid System is not very restrictive, then there is a symmetric equilibrium with both agents allocating some effort to both tasks.

In Section 7.1, we discuss how the excessive experimentation our model predicts might be linked to the increase of college double majors in the USA. We also review the connection of our model to the differences in the distribution of talents across countries and the trade patterns discussed in Grossman and Maggi (2000). In Section 7.2, we discuss our results in light of the job classification (stars, guardians, and foot soldiers) by Baron and Kreps (1999). We describe inefficiencies in talent discoveries due to domestic and international competition. We then discuss how additional signals put in place by the companies, such as those generated during a probation period, alleviate those inefficiencies.

2 Related Literature

The idea that task allocation affects future job prospects is not novel. A strand of the literature started by Gibbons and Waldman (1999a,b) emphasized that the worker's ability to discover and display talent to the market depends on the tasks they are assigned within a firm. The intuition is that different task allocations to a worker within a firm might lead to different revelations of his talent (success/failure) and, as a result, different market perceptions of a worker's productivity. Picariello (2020) allows agents to accumulate portable human capital while learning from the tasks they perform. More recently, Bar-Isaac and Lévy (2022) study how workers' effort and companies' task allocation interact through the extent of labor market competition. Pagano and Picariello (2023) show that if firms compete for talent, the more risk-averse workers might choose less quality-revealing jobs.[4]

Our paper is related to the literature on career concerns pioneered by Holmström (1999). The literature has explored how career concerns might lead to over-provision of effort. Milbourn et al. (2001) consider a manager's choice of signal's precision on the project quality. The main finding is that the manager's investment in signal precision is excessive in the presence of career concerns.[5] We complement the literature by studying the experimentation/specialization trade-off. In particular, we highlight novel channels for excessive experimentation - an additional rigidity constraint and competition.

In a related paper, Brunner et al. (2021) study the experimentation-specialization trade-off for an agent seeking to discover his talent. The main difference is that there are two job sectors, and the agent's talent is sector-specific (whereas productivity is transferable across the sectors in our model). Experimentation (trying to discover talent in different sectors) is always efficient. The intuition is that, after trying his talent in one sector, the agent has "nothing to lose" by experimenting in a different sector. The reason is that the agent retains an option of working for the initial sector if the signal from the newly tried one is lower. In our paper, experimentation is not always efficient, and we provide necessary and sufficient conditions for specialization. One of the findings of Brunner et al. (2021) is that agents with higher bargaining power experiment more. We complement this finding by pointing out that if the talent is not sector-specific, the result is the opposite: an environment that gives the agent sufficiently large bargaining power makes experimentation less likely (see Corollary 1). In addition, we study the effect of within and across the systems competition on the agents'

[4]See also Kurino and Kurokawa (2023) and Canidio and Legros (2023).

[5]Hirshleifer and Thakor (1992) show that workers with career concerns may invest in excessively safe projects to avoid early project failure. Hermalin (1993) studies how a manager's career concern can affect his choice of project riskiness. Rodivilov et al. (2022) explore career concerns when the agent's private information is multidimensional.

incentives to discover productivity. Therefore, the agents' choices in our model are strategic.

We also contribute to the literature on certification (see Ali et al. (2022) for a recent literature review). The literature has explored both profit-maximizing and efficient certification. Lizzeri (1999) showed that a profit-maximizing certification intermediary might optimally use a threshold rule. DeMarzo et al. (2019) analyze optimal certification with ex-ante symmetric information when the agent can conceal a negative outcome. In our model, certification (signals generated by success/failures) is chosen strategically by an agent given the exogenous constraints. We also contribute to the literature by studying the effect of completion between agents with heterogeneous certification policies.

Finally, our paper is related to the literature on strategic experimentation and, in particular, to the model of looking for good news (see Khalil et al. (2020) for a literature review). The key novelty is that we consider competition across two heterogeneous systems and study how this leads to excessive experimentation.

3 Model

The agent's type. A principal (firm) would like to hire an agent (employee). It is commonly known that the agent's productivity is $\theta = \{\theta_L, \theta_H\}$ with $0 \leqslant \theta_L < \theta_H \leqslant 1$. The principal and the agent are symmetrically but incompletely informed about the latter's type. The agent's productivity is θ_H with probability $\mu_0 \in (0, 1)$ and θ_L with probability $1 - \mu_0$.

The signal structure. Before the hiring decision is made, the agent can generate an informative signal that is correlated with his productivity. In particular, we assume that the agent can perform two tasks: "A" and "B". By exerting effort, the agent can either succeed ($y_i = S$) or fail ($y_i = F$) in each task $i = \{A, B\}$. We denote by $e_i \in [0, 1]$ the effort level devoted to task i. Let the total amount of effort be normalized to one: $e_A + e_B = 1$. The available signaling technology is taking the form of *looking for good news*. Therefore, we define the probability of success in task i given agent's type θ and effort level e_i for $i = \{A, B\}$ as follows:

$$Pr(y_i = S|\theta) = \theta e_i. \tag{1}$$

Our model captures the idea that markets operate with incomplete information regarding the applicants' types, and some evidence (i.e., signals) is used to screen the applicant's type. Since the applicants differ in their productivity, hiring decisions are optimally made after observing signals correlated with the agent's type.

The posterior beliefs. An effort allocation (e_A, e_B) such that $e_A, e_B \geqslant 0$ with cost

$$C(e_A, e_B) = \frac{e_A^2}{2} + \frac{e_B^2}{2}, \tag{2}$$

generates a vector of signals

$$m = (m_A, m_B), \tag{3}$$

where

$$m_i \in \{S, F, \emptyset\} \text{ for } i = \{A, B\}. \tag{4}$$

That is, the agent can either succeed (S), fail (F), or not to perform the task at all, thus delivering no signal ($\emptyset$).

After observing a signal m, the principal believes that the agent has productivity θ_H with probability μ_m, generated using the Bayes' rule:[6]

$$\mu_m = \frac{g(m|\theta_H)\mu_0}{g(m|\theta_H)\mu_0 + g(m|\theta_L)(1-\mu_0)}. \tag{5}$$

Accordingly, let us define the ex-ante distribution over posterior beliefs $\tau \in (0,1)$ such that

$$\tau(\mu_m) = g(m|\theta_H)\mu_0 + g(m|\theta_L)(1-\mu_0). \tag{6}$$

The posterior beliefs defined in (5) can then be rewritten for all $\tau(\mu_m) > 0$ as

$$\mu_m = \frac{g(m|\theta_H)\mu_0}{\tau(\mu_m)}. \tag{7}$$

Note that if the agent succeeds in both tasks, the posterior belief does not depend on the effort allocation:

$$\mu_{(S,S)} = \frac{\theta_H^2\mu_0}{\theta_H^2\mu_0 + \theta_L^2(1-\mu_0)}. \tag{8}$$

However, if the agent succeeds in one task only, the exact effort allocation (which is correctly anticipated by the principal in equilibrium) is necessary to determine the posterior beliefs.

[6]Note that the conditional probabilities $g(m|\theta)$ depend on the chosen effort levels e_A and e_B. For interior effort choices, the probabilities are:

$$g\big((S,S)|\theta\big) = \theta e_A \theta e_B, \quad g\big((F,F)|\theta\big) = [1-\theta e_A][1-\theta e_B],$$
$$g\big((S,F)|\theta\big) = \theta e_A[1-\theta e_B], \quad g\big((F,S)|\theta\big) = [1-\theta e_A]\theta e_B.$$

For the corner choices $e_A, e_B \in \{0,1\}$, the probabilities are given by:

$$g((S,\emptyset)|\theta) = g((\emptyset,S)|\theta) = \theta \text{ and } g((F,\emptyset)|\theta) = g((\emptyset,F)|\theta) = 1-\theta.$$

The signal $(\emptyset,\emptyset)$ is irrelevant since the agent chooses positive effort for at least one task in equilibrium.

For example, if the agent succeeds in task A only, the posterior belief $\mu_{(S,F)}$ is decreasing in e_A, as can be seen in Figure 1. Similarly, if the agent succeeds in task B only, the posterior belief $\mu_{(F,S)}$ is increasing in e_A. Intuitively, the agent is more likely to be highly productive if he succeeds in a certain task while devoting less effort to that task. Finally, since the two tasks are symmetric, the posterior beliefs $\mu_{(S,F)}$ and $\mu_{(S,F)}$ are symmetric over $e_A = \frac{1}{2}$.

The production function. We assume that the agent's output, denoted by $f(\theta)$, is perfectly correlated with his productivity. To sharpen the intuition, we assume the following production function:

$$f(\theta) = \begin{cases} \overline{y} > 0, & \text{if } \theta = \theta_H \\ -k < 0, & \text{if } \theta = \theta_L \end{cases}$$

That is, if the principal knew the agent's type perfectly, she would employ only highly productive ($\theta = \theta_H$) agents.[7]

The Flexible and the Rigid systems. We compare two frameworks (i.e., systems) differing in the restrictions on the effort choice. In a *Flexible* System, the agent is free to choose any combination of non-negative e_A and e_B such that $e_A + e_B = 1$. In such a framework, agents have the option of not performing one task and devoting their entire effort to the other one. In the *Rigid* system the agent must explore (allocate a strictly positive effort towards) each of the two tasks; hence, a signal "$\emptyset$" is not feasible. Formally, the agent's effort choice in the Rigid system has to satisfy an additional assumption: $e_A, e_B \geqslant \overline{e} > 0$.

Assumptions on the primitives. We assume that the initial belief μ_0 is sufficiently low so that if the agent chooses not to experiment with any task, he is not hired. Formally, the ex-ante expected productivity is strictly negative:

$$\mu_0 \overline{y} - (1 - \mu_0)k = \mu_0(\overline{y} + k) - k < 0 \iff \mu_0 < \frac{k}{(\overline{y} + k)}. \tag{\textbf{A0}}$$

Therefore, the agent is hired only if the posterior belief μ_m is sufficiently large:

$$\mu_m \geq \hat{\mu} := \frac{k}{k + \overline{y}} \in (0, 1). \tag{9}$$

To focus on the most interesting results of the model, let us consider a set of parameters such that $\hat{\mu}$ is sufficiently large. In particular, we assume that the agent is hired only if he

[7]Assuming a more general production function would not change our qualitative results as long as the highly productive agent is sufficiently more productive than the low productivity type.

succeeds in at least one task. If the agent was hired regardless of the generated signal, there would be no incentive for him to generate a (costly) signal to begin with. The reason is that with the Bayesian updating, the expected posterior productivity must be equal to the prior (i.e., the "Bayes plausibility").[8]

Let us assume that the lowest belief generated when the agent succeeds in at least one task is greater than $\hat{\mu}$. Recall that

$$\mu_{(S,F)} = \frac{g((S,F)|\theta_H)\mu_0}{g((S,F)|\theta_H)\mu_0 + g((S,F)|\theta_L)(1-\mu_0)}, \tag{10}$$

where $g((S,F)|\theta) = \theta e_A(1-\theta e_B)$. The lowest value of $\mu_{(S,F)}$, which is achieved if $e_A = 1$ and $e_B = 0$, is formally $\mu_{(S,\emptyset)}$. Therefore, the condition becomes $\frac{\theta_H\mu_0}{\theta_H\mu_0+\theta_L(1-\mu_0)} > \frac{k}{k+\overline{y}}$, which can be rewritten as

$$\theta_H \geqslant \underline{\theta_H} \equiv \frac{(1-\mu_0)k}{\mu_0\overline{y}}\theta_L. \tag{\textbf{A1}}$$

Furthermore, notice that (**A0**) implies that the largest belief generated when the agent fails in both tasks is smaller than $\hat{\mu}$. To see this, recall that

$$\mu_{(F,F)} = \frac{g((F,F)|\theta_H)\mu_0}{g((F,F)|\theta_H)\mu_0 + g((F,F)|\theta_L)(1-\mu_0)}, \tag{11}$$

where $g((F,F)|\theta) = (1-\theta e_A)(1-\theta e_B)$. The largest value of $\mu_{(F,F)}$ is achieved if $e_A = \frac{1}{2} = e_B$ and it is smaller than $\frac{k}{k+\overline{y}}$ if

$$\mu_0 < \frac{\frac{k}{k+\overline{y}}\left(1-\frac{\theta_L}{2}\right)^2}{\frac{k}{k+\overline{y}}\left(1-\frac{\theta_L}{2}\right)^2 + \frac{\overline{y}}{k+\overline{y}}\left(1-\frac{\theta_H}{2}\right)^2}. \tag{12}$$

which is automatically satisfied since $\mu_0 < \frac{k}{(\overline{y}+k)}$.[9] Intuitively, since the agent is not hired if he chooses not to perform any task, the posterior belief after he fails on both tasks becomes even smaller.

In Figure 1, we illustrate an example of possible posterior beliefs. Figure 2 illustrates the probabilities $\tau(\mu_m)$ of generating posterior beliefs.

[8]See Kamenica and Gentzkow (2011) for details.
[9]The statement immediately follows from the fact that $\theta_H > \theta_L$.

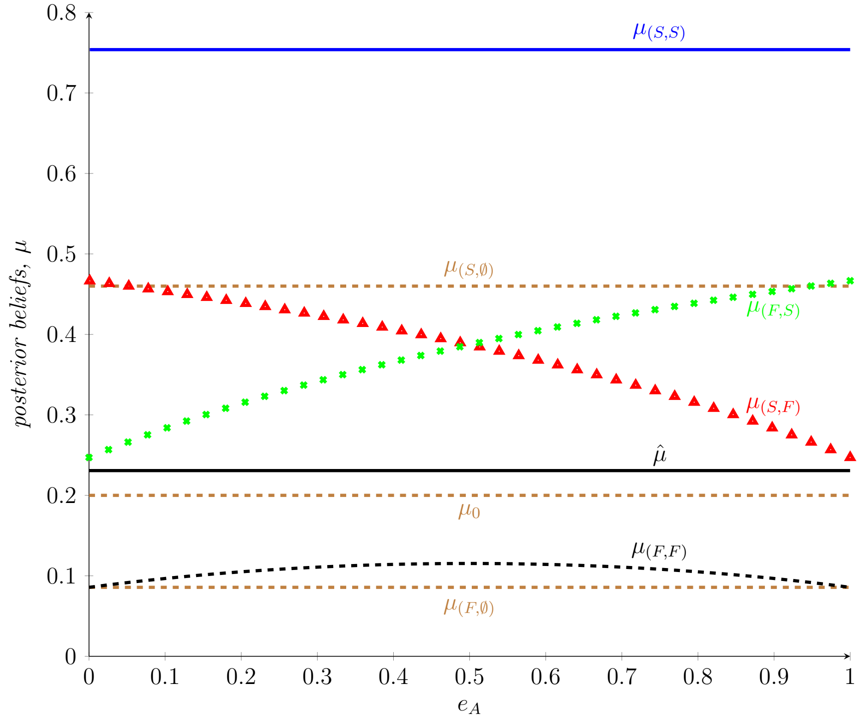

Figure 1. Posterior beliefs if $\mu_0 = 0.2$, $\theta_H = 0.7$, $\theta_L = 0.2$, $k = 3$, and $\overline{y} = 10$.
Both assumptions are satisfied; $(A0)$: $\mu_0 = 0.2 < \frac{k}{k+\overline{y}} = 0.23$ and $(A1)$: $\theta_H = 0.7 > \frac{(1-\mu_0)k}{\mu_0 \overline{y}}\theta_L = 0.24$.
Note that $\mu_{(F,\emptyset)} < \mu_{(F,F)} < \hat{\mu} = 0.23$ for any effort level.

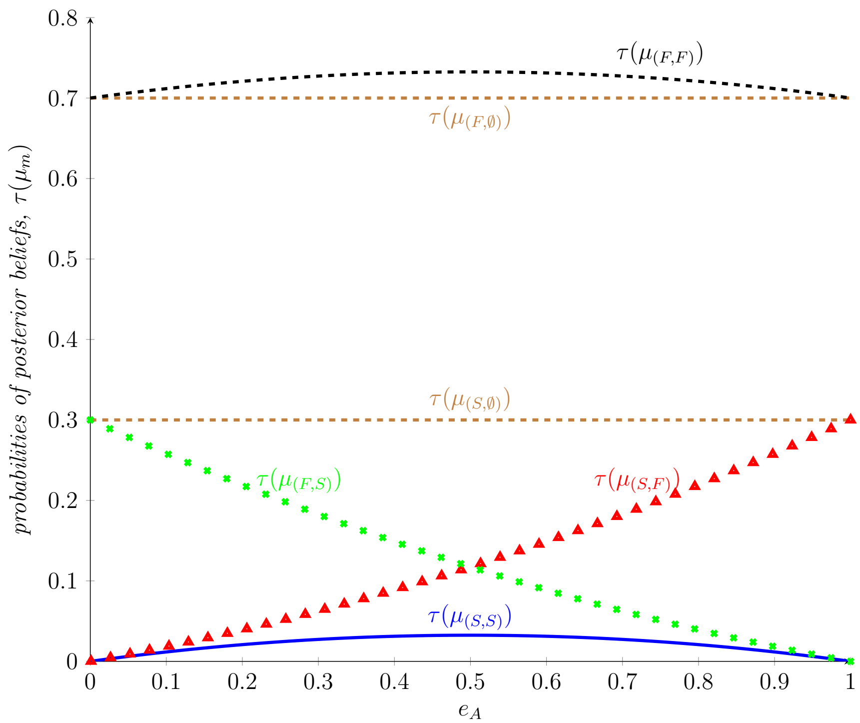

Figure 2. The probabilities of generating posterior beliefs if $\mu_0 = 0.2$, $\theta_H = 0.7$, and $\theta_L = 0.2$.

Payoffs. Given the posterior beliefs μ_m, an agent's expected output is

$$\mathbb{E}\big[f(\theta) \bigm| \mu_m\big] := \mu_m \overline{y} - (1-\mu_m)k = \mu_m(\overline{y}+k) - k. \tag{13}$$

The principal hires an agent and offers a competitive wage depending on the observed signal. Therefore, the agent's wage is increasing in his expected productivity. We apply the generalized Nash bargaining solution and assume that the agent is paid a portion $\gamma \in (0,1]$ of the expected output. Consequently, the principal retains the portion $1-\gamma$ of the expected output. For example, if $\gamma = 1$, the market is perfectly competitive and the agent collects the entire expected surplus. Given the posterior belief μ_m, the agent is paid a wage defined as

$$w(\mu_m) := \gamma \underbrace{\mathbb{E}\big[f(\theta) \bigm| \mu_m\big]}_{\text{expected output}} = \gamma\big[\mu_m(\overline{y}+k) - k\big]. \tag{14}$$

For a given wage schedule $w(\mu_m)$, an agent who generates posterior beliefs μ_m with effort profile (e_A, e_B) has an expected utility function

$$U(w, e_A, e_B) := \underbrace{\sum_{\mu_m} \tau(\mu_m) w(\mu_m)}_{\text{expected wage}} - \underbrace{C(e_A, e_B)}_{\text{cost of effort}}. \tag{15}$$

Finally, given the posterior belief μ_m, the principal has the following expected profit if she hires an agent:

$$\pi(w, \mu_m) := (1-\gamma) \underbrace{\big[\mu_m(\overline{y}+k) - k\big]}_{\text{expected output}} - \underbrace{w(\mu_m)}_{\text{agent's wage}} \tag{16}$$

and earns zero profits when not hiring any agent.

The timing. The timing of the model is as follows:
$t = 1$: The agent allocates his effort across the two tasks given the common prior beliefs.
$t = 2$: Principal and agent observe the signals generated and form the posterior beliefs.
$t = 3$: The agent is hired if the posterior belief is high enough.
$t = 4$: Payoffs are collected (i.e., wages are paid).

3.1 Social Planner's Problem

In order to provide a benchmark against which to compare all our next results, let us first consider the efficient allocation of effort across tasks, i.e. the one that maximizes social welfare. To do so, we analyze the maximization program for a utilitarian social planner. Formally, the problem can be described as

$$\max_{\{e_A \geqslant 0, e_B \geqslant 0\}} \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) \Big[\mu_m f(\theta_H) + (1-\mu_m) f(\theta_L)\Big] - C(e_A, e_B)$$

$$\text{subject to } e_A + e_B = 1.$$

The solution to the problem is summarized in the following proposition.

Proposition 1. ***Efficient Effort Allocation.***
Efficiency requires to equally split effort (***experimentation***), $e_A^* = e_B^* = \frac{1}{2}$, if

$$\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2 < 1; \qquad \textbf{(EE)}$$

$e_A^* = 1 - e_B^* \in \{0, 1\}$ (***specialization***) is the efficient strategy if

$$\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2 \geqslant 1. \qquad \textbf{(ES)}$$

Proof: See Appendix A.

Any of the following three conditions are sufficient for experimentation, rather than specialization, to be efficient. First, the two types are close in terms of productivity (θ_H close to θ_L). Recall that assumption (**A0**) implies that an agent who does not perform any task is not hired (the expected productivity is negative). Comparing the left-had-side of (**EE**) with the left-had-side of (**A0**), it is immediate to see that the former is negative if θ_H is close enough to θ_L. In this case, (**EE**) holds true. Intuitively, if the two types are sufficiently close in terms of productivity, there is low volatility, i.e. low risk ex-ante. Hence, agents are willing to experiment in this scenario as this would improve their expected productivity, though increasing risk. To see, note that experimentation is more "volatile" in the sense that it generates more dispersed posterior beliefs. As shown in Figure 1, experimentation might generate high posterior $\mu_{(S,S)}$, low posterior $\mu_{(F,F)}$, and intermediate levels $\mu_{(S,F)}$ and $\mu_{(F,S)}$. Specialization, on the other hand, is less "risky" as the posterior it generates are closer to each other. If the two types are getting closer in terms of productivity, experimentation becomes less volatile and, as a result, leads to a higher social surplus.[10] For example, consider Figure 3 that illustrates the social surplus for $\mu_0 = 0.2$, $\theta_L = 0.2$, $k = 3$, and $\overline{y} = 10$. If the

[10]For instance, if θ_H is getting closer to its lower bound $\underline{\theta_H}$ introduced in (**A1**), then $\lim_{\theta_H \to \underline{\theta_H}} \mu_{(S,S)} = \hat{\mu}$. Similarly, as θ_H approaches $\underline{\theta_H}$, if the agent generates a good signal before being hired, the posterior belief about his type will be such that $\mu_{(S,\emptyset)}$ is closer to $\hat{\mu}$ the lower θ_H. If θ_H is getting closer to θ_L, the difference

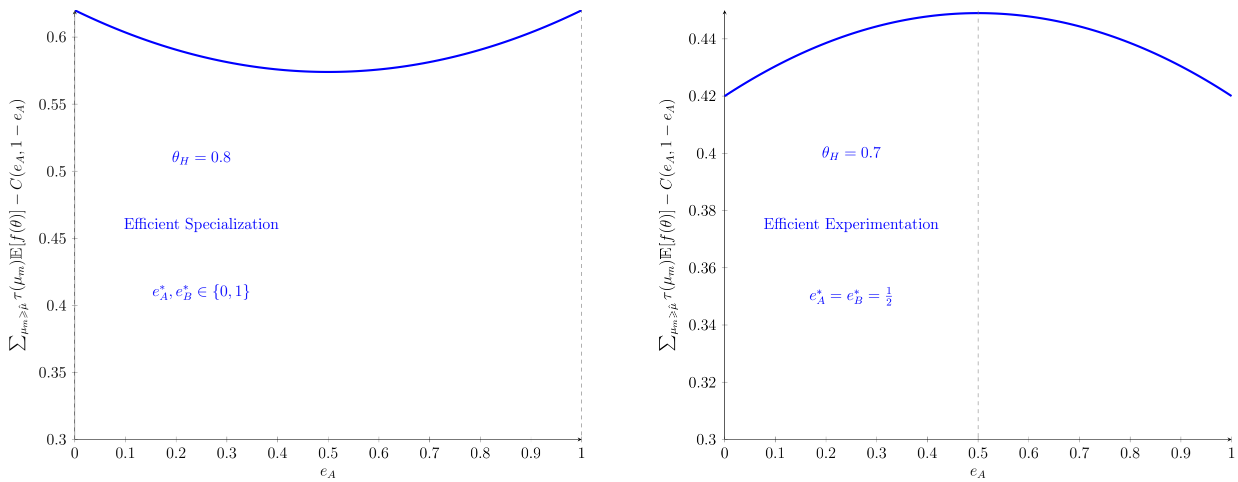

Figure 3. Efficient specialization and experimentation if $\mu_0 = 0.2$, $\theta_L = 0.2$, $k = 3$, and $\overline{y} = 10$.

two types are far apart in terms of productivity, then specialization is efficient ($\theta_H = 0.8$ on the left-hand side of Figure 3). If the two types become closer in terms of productivity, then experimentation becomes efficient ($\theta_H = 0.7$ on the right-hand side of Figure 3).

Second, the ex-ante expected share of high productivity agents, i.e. μ_0, is relatively low. The left-hand side of (**EE**) is increasing in μ_0 and is strictly negative if $\mu_0 = 0$. Thus, if the prior belief is small, experimentation is efficient. We define a threshold value μ_0^E such that experimentation is efficient if:

$$\mu_0 < \mu_0^E := \frac{1 + k\theta_L^2}{\overline{y}\theta_H^2 + k\theta_L^2}. \tag{17}$$

Intuitively, if μ_0 is small, the agent's ex-ante expected wage is small. Then, it is efficient to experiment, i.e. to take more risk, thus implementing higher volatility of the expected output. In such a scenario, a negative signal would not entail too large a productive loss, whereas a positive signal would generate high expected productivity.

Third, the loss in production imposed by a low-type agent is quite significant compared to the productive benefit generated by a high-type agent. Rearranging (**EE**), experimentation is efficient if:

$$k > k^E := \frac{\overline{y}\mu_0\theta_H^2 - 1}{(1-\mu_0)\theta_L^2}. \tag{18}$$

Intuitively, experimentation, by reducing the variance between expected outputs in any possible scenario with at least one success, should efficiently be used as a "hedging device"

in the posterior beliefs generated with success and with failure under specialization is getting smaller:

$$\mu_{(S,\emptyset)} - \mu_{(F,\emptyset)} = \frac{\mu_0(1-\mu_0)}{\big(\mathbb{E}[\theta]\big)\big(\mathbb{E}[1-\theta]\big)}(\theta_H - \theta_L).$$

against excessive output volatility.

If either of these three conditions does not hold, it is efficient for the agent to focus on just one task (specialization).

4 The Flexible System

Let us now consider a framework in which the agent allocates his effort as he sees fit. In particular, he can choose any combination of non-negative e_A and e_B such that $e_A + e_B = 1$. Optimally chosen e_A and e_B generate a signal $m = (m_A, m_B)$ from the set

$$\mathcal{M}_F = \big\{(S,F),(F,S),(S,S),(F,F),(S,\emptyset),(F,\emptyset),(\emptyset,S),(\emptyset,F)\big\}. \tag{19}$$

In this scenario, the agent solves

$$[P^F] \qquad \max_{\{e_A \geqslant 0, e_B \geqslant 0\}} \textstyle\sum_{\mu_m \in \mathcal{M}_F | \mu_m \geqslant \hat{\mu}} \tau(\mu_m) w(\mu_m) - C(e_A, e_B)$$

$$\text{subject to } e_A + e_B = 1.$$

Specifically, the agent maximizes his expected wage conditional on the posterior beliefs about his type being sufficiently high, net of the cost of effort. While doing so, he faces the following trade-off. On the one hand, the expected wage is higher if he succeeds in both tasks, hence there is a possible benefit from experimentation. On the other hand, experimentation increases the chance of failing (on one or even both of them), which entails a lower expected wage, thus, there is a possible benefit from specialization. The following proposition characterizes the optimal allocation of effort:

Proposition 2. Optimal Effort in the *Flexible* System.
The agent equally splits effort (***experimentation***) across the two tasks, $e_A^* = e_B^* = \frac{1}{2}$, if

$$\gamma\big[\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2\big] < 1; \tag{\textbf{FE}}$$

the agent ***specializes*** in either of the two tasks, $e_A^* = 1 - e_B^* \in \{0,1\}$, if

$$\gamma\big[\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2\big] \geqslant 1. \tag{\textbf{FS}}$$

Proof: See Appendix B.

Comparing the results in Proposition 2 and Proposition 1, it is immediate to see that, on top of the parameter conditions described in Section 4.1, bargaining power also plays a role in the optimal choice of effort. Specifically, in the Flexible system effort is efficient if the agent has all the bargaining power vis-á-vis the principal (i.e., $\gamma = 1$). Intuitively, if the agent captures all the surplus generated, he aims at maximizing the expected surplus net of the cost of effort.

Proposition 2 shows how the market structure and the consequent sharing rule for surplus affect incentives to experiment. In particular, since $\gamma\big[\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2\big]$ is positive[11] and, therefore, increasing in γ, specialization is more likely if the agent captures a larger portion of the expected output (if γ becomes higher). We summarize the results in Corollary 1 below:

Corollary 1 (Experimentation and Bargaining Power). An environment that gives the agent sufficiently large bargaining power makes experimentation less likely.

Also note that if an agent chooses to specialize, it does not matter which task (A or B) he picks. Since signal realization across the two tasks is independent, there are two possible equilibria: one in which $e_A^* = 1$ and $e_B^* = 0$, and another where $e_A^* = 0$ and $e_B^* = 1$.

5 The Rigid System

We now analyze the Rigid system. The agent is constrained in his effort choice and can no longer choose the corner solutions of $e_A \in \{0, 1\}$. As a result, a signal "$\emptyset$" can not be generated. In particular, e_A and e_B generate a signal $m = (m_A, m_B)$ from a set

$$\mathcal{M}_R = \big\{(S,F),(F,S),(S,S),(F,F)\big\}. \tag{20}$$

The agent's optimization problem becomes the following

$$[P^R] \quad \max_{e_A,e_B} \textstyle\sum_{\mu_m \in \mathcal{M}_R | \mu_m \geqslant \hat{\mu}} \tau(\mu_m) w(\mu_m) - C(e_A, e_B) \text{ subject to}$$

$(C1^R)$ $$e_A, e_B \geqslant \overline{e} > 0,$$

$(C2)$ $$e_A + e_B = 1.$$

[11] To see that $\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2$ is positive, we evaluate it at the smallest value of θ_H to obtain:

$$\lim_{\theta_H \to \underline{\theta_H}} \big[\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2\big] = \overline{y}\mu_0\big(\tfrac{(1-\mu_0)k}{\mu_0\overline{y}}\big)^2\theta_L^2 - k(1-\mu_0)\theta_L^2 = k(1-\mu_0)\theta_L^2\big[\tfrac{(1-\mu_0)k}{\mu_0\overline{y}} - 1\big] > 0,$$

where the last inequality follows from assumption (**A0**), i.e., $(1-\mu_0)k > \mu_0\overline{y}$.

Without loss of generality, we assume that the threshold level $\overline{e}$ is such that

$$0 < \overline{e} < \frac{1}{2}. \tag{21}$$

The novel constraint ($C1^R$) captures the agent's restriction in the Rigid system. Essentially, an agent who wants to specialize in one task must devote some resources to the second task as well. If the agent devotes to one task just enough effort to satisfy ($C1^R$), he risks failing. In this case, sending a signal "F" diminishes the posterior beliefs. This was not the case in the Flexible system, where signal $\emptyset$ was available.

We can use the solution from the flexible version and compare the unconstrained optima with $\overline{e} > 0$. In particular, when deriving the solution for the Flexible system in Appendix B, we establish conditions that guarantee an interior as well as a corner solution. Repeating exactly the same steps and replacing the corner solution $1 - e_B^* = e_A^* \in \{0, 1\}$ with $e_A^* = \overline{e}$ constitutes the proof. We summarize the results in Proposition 3 below.

Proposition 3. Optimal effort choice in the *Rigid* System.
The agent equally splits effort (***experimentation***) across the two tasks, $e_A^* = e_B^* = \frac{1}{2}$, if

$$\gamma\left[\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2\right] < 1; \tag{\textbf{RE}}$$

and $e_A^* = 1 - e_B^* = \overline{e}$ or $e_A^* = 1 - e_B^* = 1 - \overline{e}$ if

$$\gamma\left[\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2\right] \geqslant 1. \tag{\textbf{RS}}$$

Intuition is reminiscent of that in the case of the Flexible system. If either the two types are similar (θ_H close to θ_L), or the agent is unlikely to be of high type (μ_0 low), or the loss in production imposed by a low-type agent is significant (k high enough), it is optimal for the agent to experiment. This was the case in the Flexible system as well.

If either the two types are significantly different (θ_H significantly higher than θ_L), or the agent is likely to be of high type (μ_0 high enough), or the loss in production imposed by a low-type agent is significant (k high enough), it is optimal for the agent to specialize in one subject only. However, in the Rigid system, devoting the entire time to one task only is not feasible. As a result, the agent devotes the minimum required effort to one task, $e_A^* = \overline{e} > 0$ or $e_B^* = \overline{e} > 0$, so that ($C1^R$) becomes binding. Intuitively, in the Rigid system, the agent specializes to the extent it is feasible given the constraint.

6 The Effect of Competition

We now consider the case in which two agents compete for a job in both a closed and an open economy, i.e., within and between systems. The firm hires only one agent, thus agents compete against each other for the position. In particular, if two agents form identical posterior beliefs, each is hired with equal probability. To distinguish the two agents from different systems, we use a superscript R (Rigid) and F (Flexible), respectively. That is, we denote by (e_A^F, e_B^F) and (e_A^R, e_B^R) the effort choices of an agent from the Flexible and the Rigid system, respectively. The posterior belief that the agent from system $s = \{F, R\}$ has productivity θ_H generated by signal m^s is denoted by $\mu^s_{m^s}$.

We now define the equilibrium effort choice with competition.

Definition. ***The Equilibrium effort profiles with competition.***
Effort profiles $(\hat{e}_A^F, \hat{e}_B^F)$ and $(\hat{e}_A^R, \hat{e}_B^R)$ constitute an equilibrium if

1. $(\hat{e}_A^F, \hat{e}_B^F)$ solves

$$\max_{e_A^F \geqslant 0,\, e_B^F \geqslant 0,\, e_A^F + e_B^F = 1} \quad \sum_{\mu^F_{m^F} \geqslant max\{\mu^R_{m^R}, \hat{\mu}\}} \tau(\mu^F_{m^F}) w(\mu^F_{m^F}) - C(e_A^F, e_B^F).$$

2. $(\hat{e}_A^R, \hat{e}_B^R)$ solves

$$\max_{e_A^R \geqslant \overline{e},\, e_B^R \geqslant \overline{e},\, e_A^R + e_B^R = 1} \quad \sum_{\mu^R_{m^R} \geqslant max\{\mu^F_{m^F}, \hat{\mu}\}} \tau(\mu^R_{m^R}) w(\mu^R_{m^R}) - C(e_A^R, e_B^R).$$

Two features distinguish the current scenario from the previously considered benchmarks without competition. First, the agent from the Flexible System is now facing competition from another agent. As can be seen under the summation sign in the optimization problem above, in addition to the posterior beliefs being sufficiently high ($\mu^F_{m^F} \geqslant \hat{\mu}$), now the agent is hired only if the posterior belief $\mu^F_{m^F}$ is no smaller than the belief about his rival. Therefore, the agent from the Flexible system is hired only if the posterior belief $\mu^F_{m^F}$ is higher than both the threshold level $\hat{\mu}$ and the rival's posterior belief $\mu^R_{m^R}$:

$$\mu^F_{m^F} \geqslant max\{\mu^R_{m^R}, \hat{\mu}\}. \tag{22}$$

Second, the two competing agents are coming from different systems and, therefore, are facing a different set of constraints. For instance, the agent from the flexible system can choose any $e_i^F \geqslant 0$ whereas his competitor must respect the additional $e_i^R \geqslant \overline{e} > 0$ constraint for $i = \{A, B\}$. Both these effects shape the equilibrium emerging in this environment. To decompose these two effects, we first consider a scenario where one agent from the Flexible System competes with another agent from the same Flexible System.

6.1 Competition *within* the Flexible System

To sharpen the intuition, we first ignore the fact that the agent from the Rigid system must respect the additional $e_i^R \geqslant \overline{e} > 0$ constraint for $i = \{A, B\}$. Therefore, in this subsection, we study the effect of competition *within* the Flexible system. Multiple equilibria emerge in this scenario: the "experimentation" and "specialization" equilibria, as well as an equilibrium where the agents mimic each other by devoting some effort to both tasks. We discuss each of them and provide the intuition below. The main takeaway from this section is that specialization is more likely with competition. The reason is that an attempt to succeed in two tasks becomes a riskier option in the presence of a rival.

Experimentation equilibrium. Recall that experimentation is socially efficient if the two types are sufficiently close in productivity (see Section 4.1). Intuitively, if θ_H is close to θ_L, the posterior beliefs from experimentation are close to each other. As a result, experimentation increases the expected wage although increases the risk. This intuition remains intact in case of competition within the Flexible system. If the two types are sufficiently close in terms of productivity, the best response to even a specializing rival is to experiment. Moreover, experimentation might be a dominant strategy. For instance, consider Figure 4A where we plot the best response effort allocation if $\theta_L = 0.2$ and $\theta_H = 0.6$. As can be seen, experimentation ($\hat{e}_A^F = \frac{1}{2}$ for all choices of a rival) is a dominant strategy. However, if the two types are getting farther apart in productivity, experimentation becomes a riskier best-response strategy. As can be seen in Figures 4A-4D, as θ_H increases, experimentation remains the best response only if the rival's effort is close to experimentation as well.

Note that, for the rival's effort in the range between specialization and experimentation, the best response is to experiment a bit more than the rival. Moreover, this tendency increases as the two types become farther apart in productivity. As can be seen in Figures 4C and 4D, the best response for the rival's effort strictly between $\hat{e}_A^F = 0$ (specialization) and $\hat{e}_A^F = \frac{1}{2}$ (experimentation) is slightly above the 45-degree line. Experimenting just a bit more than a rival leads to the following trade-off. On the one hand, it prevents the two competing agents from generating the same beliefs, which is beneficial. On the other hand, it makes the posterior beliefs more volatile. If the two types are getting farther apart in productivity, the former effect dominates.

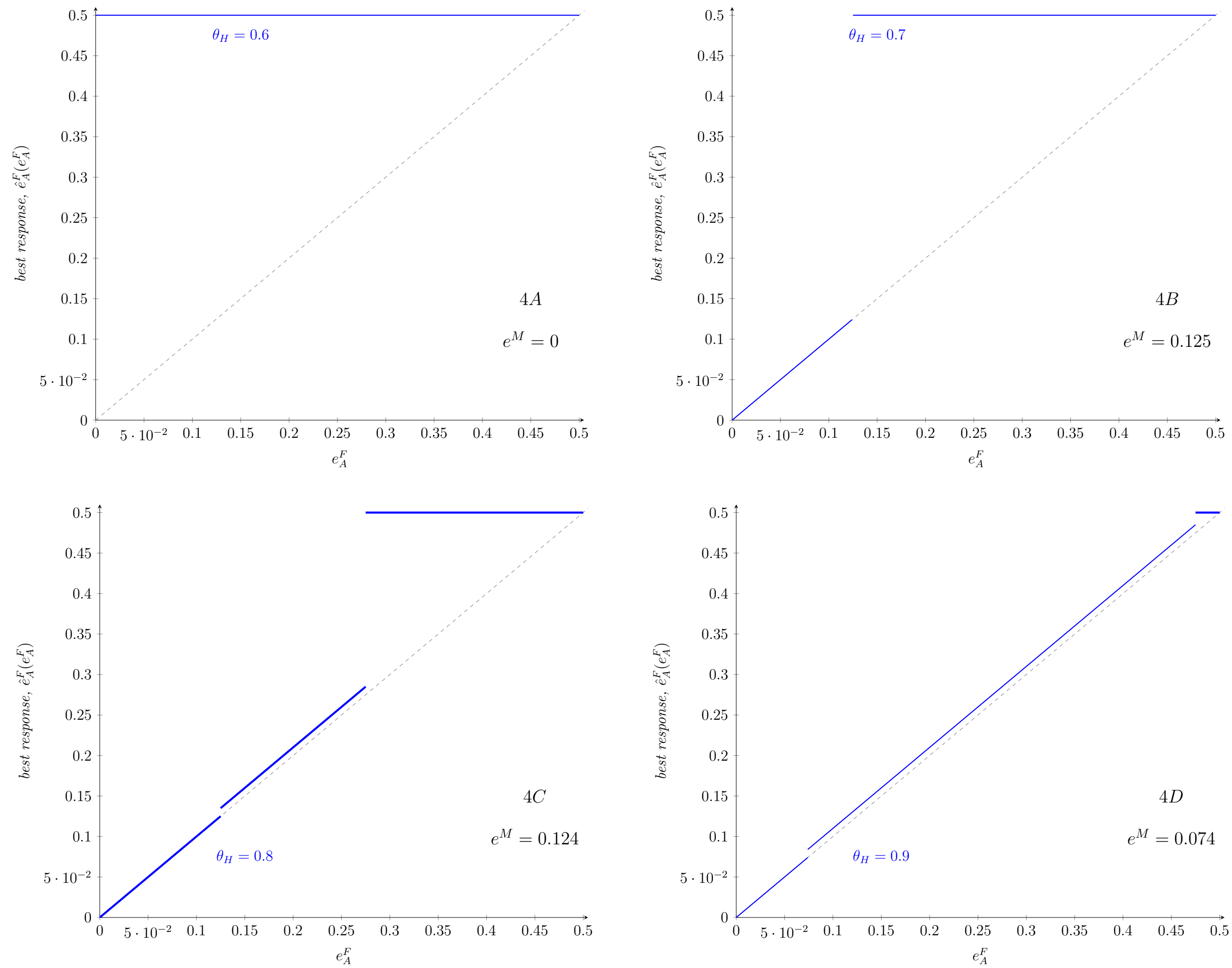

Figure 4. Best response correspondences (Flexible System) if $\mu_0 = 0.2$, $\theta_L = 0.2$, $k = 3$, and $\overline{y} = 10$.

Specialization. If the two types are significantly different in productivity, it is the best response strategy to mimic the rival's behavior. Matching the rival's choice of effort allocation leads to the following trade-off. On the one hand, it increases the chances of the competing agents generating identical posterior beliefs, which is beneficial, since then each is hired with some probability. On the other hand, it lowers the expected wage conditional on being hired. If the two types are getting farther apart in productivity, the latter effect dominates. As can be seen in Figures 4C and 4D, the best response for the rival's effort close to $\hat{e}_A^F = 0$ (specialization) coincides with the 45-degree line.

We next provide a necessary and sufficient condition for the specialization equilibrium.

Lemma 1. Necessary and sufficient condition for Specialization.
There exists an equilibrium where agents competing *within* the Flexible ***specialize*** if

$$\underbrace{\gamma\big[\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2\big]}_{\text{Specialization with single agent (FS)}} \geqslant 1 - \underbrace{2\gamma\big[\overline{y}\mu_0\theta_H - k(1-\mu_0)\theta_L\big]\mathbb{E}[\theta]}_{>0 \text{ by (A1)}}. \tag{WFS}$$

Proof: See Appendix C.

Competition within the Flexible system introduces a novel effect. In particular, it makes specialization more likely. To see, recall from condition (**FS**) in Proposition 2 that the agent in the Flexible system specializes in the absence of competition if $\gamma\big[\overline{y}\mu_0\theta_H^2 - k(1-\mu_0)\theta_L^2\big] \geqslant 1$. Given that the second term on the right-hand side of (**WFS**) is positive,[12]

$$2\gamma\big[\overline{y}\mu_0\theta_H - k(1-\mu_0)\theta_L\big]\mathbb{E}[\theta] > 0, \tag{23}$$

an agent facing a competitor from the same system is more likely to specialize.

Note that although the success/failure of one agent in each task is not explicitly affected by the presence of a rival, the wage and probability of being hired are indeed affected by competition. This is exactly what the term $2\gamma\big[\overline{y}\mu_0\theta_H - k(1-\mu_0)\theta_L\big]\mathbb{E}[\theta]$ captures. By altering the probability of being hired, competition increases incentives to specialize. Intuitively, if an experimenting agent fails at least once, he might face a rival who succeeds at least once while specializing. For instance, consider a scenario in which the experimenting agent succeeds once (with posterior belief $\mu_{(S,F)}$ or $\mu_{(F,S)}$) while the rival who chose to specialize succeeds. Then, the former agent is not hired. Moreover, the chances of the specializing agent succeeding are higher the greater θ_H.

Lemma 2. Mimicking the rival.
If condition (**WFS**) holds, there exists $e^M > 0$ such that the agents competing *within* the Flexible System both choosing $e_F^A, 1 - e_F^A \in (0, e^M]$ is an equilibrium.

Proof: See Appendix C.

6.2 Competition *between* the Flexible and the Rigid Systems

We now consider a scenario where an agent from the Flexible system competes with another agent from the Rigid system. The equilibrium where both agents experiment remains

[12] The statement immediately follows from (**A1**).

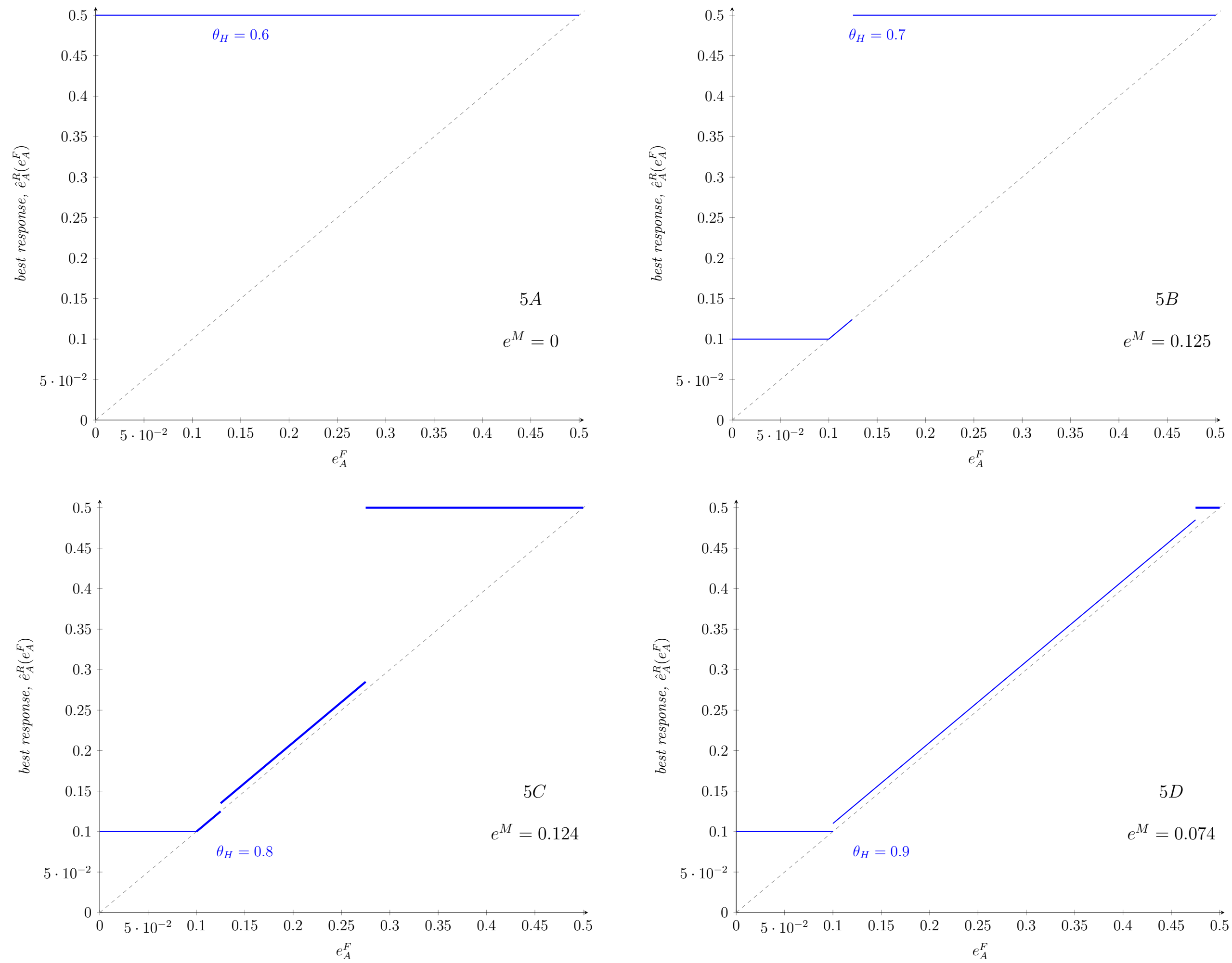

Figure 5. Best response correspondences (Rigid System) if $\overline{e} = 0.1$, $\mu_0 = 0.2$, $\theta_L = 0.2$, $k = 3$, and $\overline{y} = 10$.

intact in this environment as none of the additional constraints in the Rigid system are binding. Equilibrium where the competing agents specialize is not feasible anymore. The main takeaway from this section is two-fold. First, if the Rigid System is sufficiently restrictive ($\overline{e} \geqslant e^M$), the only equilibrium is the one where both agents experiment. Second, if the Rigid System is not very restrictive ($\overline{e} < e^M$), then there is a symmetric equilibrium with both agents choosing effort $e_A \in [\overline{e}, e^M]$:

Lemma 3. Restrictive rigid system and the effect of between competition. For any set of parameters, there exist a high enough value of $\overline{e}_{max}$, such that

- if $\overline{e} > \overline{e}_{max}$, there is a unique equilibrium where both agents ***experiment***.

We provide the intuition and explain the mechanics using Figure 5, where we plot the best response correspondence $\hat{e}_A^R(e_A^F)$ for the agent from the Rigid System for $\overline{e} = 0.1$. As can be seen, for low values of θ_H in Figure 5A, experimentation is a dominant strategy. If the two types are sufficiently close in terms of productivity, the best response to even a specializing rival is to experiment. This is reminiscent of competition within the Flexible System (see Figure 4). In this case, both agents experimenting is the only equilibrium.

For higher values of θ_H, the additional constraint in the Rigid System affects the agent's choices. For example, in 5B, the agent cannot mimic a rival who chooses $e_A^F < 0.1$. As a result, the equilibria possible for these parameters are $e_A^R = e_A^F \in [\bar{e}, e^M]$. In 5D, the Rigid System is very restrictive, and the only possible equilibrium is the one where both agents experiment.

7 Applications

7.1 Education systems

Every year, about six million high schoolers and two million college freshmen in the U.S. struggle with mandatory algebra classes.[13] There is an ongoing debate regarding mandatory classes in U.S. high schools.[14] While some advocate for keeping mandatory classes, others argue that all high school courses should be elective.[15] Why are students subject to this mandatory ordeal? What are the implications of mandatory classes for students trying to discover their productivity and future job market perspectives?

The questions are even more relevant given the differences in high school and university systems across countries. For example, in countries like China, India, and Russia (the "Eastern System"), most classes are mandatory in high schools. In contrast, in Europe, Australia, and the USA (the "Western system"), pupils choose subjects they take.[16] There are similar differences at the college/university level. While students in the Eastern system are admitted to a particular department and, therefore, must choose their specialization in advance, students in the Western system typically do not have to commit to a specialization in their studies at the time of admission.

The question of whether mandatory classes are beneficial for graduates has been extensively addressed in the education literature, albeit with mixed evidence. Levine and Zimmerman (1995) find a positive effect of taking more high-school math on future wages;Görlitz and Gravert (2018) analyze the effect of increasing the instruction time in core STEM subjects in Germany. Similarly, Berggren and Jeppsson (2021) examines the effect of a decrease

[13] This extends beyond algebra to a standard mathematics sequence. See https://www.nytimes.com/2021/11/04/us/california-math-curriculum-guidelines.html

[14] See https://www.nytimes.com/2024/05/22/nyregion/middle-school-math-algebra.html

[15] https://www.washingtonpost.com/news/answer-sheet/wp/2013/01/22/why-all-high-school-courses-should-be-elective/?noredirect=on&utm_term=.99a96781fdfctext

[16] In some countries, pupils choose subjects in addition to a so-called "core" set of classes.

in mandatory mathematics requirements on education outcomes and earnings in Sweden. Goodman (2019) shows that mandatory STEM classes positively affect job returns (up to 10% for some groups). Jia (2021) examines the impact of stricter high school math requirements on the likelihood of completing a degree in STEM fields.[17] At the same time, Betts and Rose (2001) suggest that schools should not make math courses mandatory but rather encourage students to take them. Malamud (2010) studies the trade-off between early and late specialization assuming that talent is field-specific.[18]

Our results highlight the effect of mandatory classes on students' opportunities to explore their productivity and their implications in the labor market. Our model captures the idea that hiring decisions are made after observing information correlated with the applicant's productivity. A firm hires an applicant whose productivity is unknown. It is common knowledge that the applicant can be either productive or not. The firm makes a hiring decision after looking at the academic track record. The wage reflects both the expected output and the bargaining power.

The Flexible and the Rigid systems capture the pivotal difference between the Western and Eastern systems. In the Western system, students might take a broad range of classes (experiment) or devote their effort toward a narrow set of subjects (specialize).[19] This opportunity is captured in our model by the Flexible system environment. In the Eastern system, students must take mandatory classes across a range of disciplines. Thus, a narrow specialization is not feasible. This feature is captured by the Rigid system environment.

We show that students in the Eastern system might be worse due to mandatory classes. In particular, we derive conditions under which specialization (devoting the entire effort endowment to a narrow set of subjects), prohibited in this system, is optimal. For instance, this is the case if the productivity distribution has "fat tails" (the types are sufficiently different in productivity in our model).

We also study the effect of market structure on the students' incentives to experiment/specialize. An environment where the agent has sufficiently strong bargaining power makes experimentation less likely. The reason is that, with strong bargaining power, the applicant captures a higher portion of his expected productivity through the wage. If specialization is socially efficient, it remains optimal for a student with strong bargaining power. If the student captures only a small portion of the generated surplus, his incentives are not

[17] Rose and Betts (2004) estimate the effect of high school math courses on earnings a decade after graduation. See also Altonji et al. (2012) for a survey of the literature on the demand for high school and post-secondary education by various fields.

[18] Dur et al. (2022) discuss the effect of students choosing college-major pairs jointly in college admissions. See also Malamud (2011).

[19] Another example of experimentation is the "double major." See, e.g., Del Rossi and Hersch (2008) and Zafar (2012).

aligned well with the social ones. Then, it is optimal to specialize less and experiment more.

Finally, we analyze the effect of competition between students from the Eastern and the Western systems in the international market. Consider a company (or a Ph.D. program) accepting applicants from different countries. Applicants from the Flexible system anticipate competition from candidates from the Rigid System. Our model predicts that the former choose to experiment (allocate their effort across multiple courses/disciplines) even if it would not be optimal without competition. Thus, we highlight a negative externality that the Rigid system imposes on applicants from the Flexible system.

The education literature has long acknowledged the connection across education systems. For example, Kimhi et al. (2019) argue that higher education in the US has obtained features of an arms race.[20] The reason for the arms race is the need to gain an edge in a competitive job market. Our paper addresses one aspect of this arms race - "*excessive experimentation.*" Our model predicts that students in the Western system choose to experiment more often due to potential competition from abroad. For instance, one of the features of "excessive experimentation" can be found in the increase of the double majors in the USA.[21] The number of college graduates with more than one undergraduate major has been steadily increasing in the U.S., where the number of double majors is forty percent in many schools.[22] From the market point of view, students who manage to double major are good at getting work done in general. Thus, one of the explanations for the rise in double majors is that students see it as a way to add one more credential to their resumes. Our results indicate that part of the "credential arms race" is due to the risk of competition from more restrictive education systems. Indeed, Hanks et al. (2024) estimate that double majors lower sensitivity to earnings shocks.

From a normative perspective, practitioners argue that the extent of experimentation should be limited. For example, David Coleman, the president of the College Board (a non-profit organization that designed the SAT and Advanced Placement exams in the USA), states:[23]

"... Applications for college have as many as 10 spaces for students to fill out with activities outside of class. How about three? Let's say to students: Share 1 to 3 things you

[20]See also Leuven and Oosterbeek (2011) for a survey of the economics literature on over-education.

[21]Del Rossi and Hersch (2008) report that, in 2006, about twenty-five percent of college graduates had double major (see also Del Rossi and Hersch (2016) for more recent data).

[22]According to The Chronicle of Higher Education, forty percent of students at Vanderbilt University choose to double major; the number of double majors has risen fifty percent in five years at the University of California at Davis, and double majors doubled since 1993 at MIT. See https://www.chronicle.com/blogs/next/the-worrying-rise-of-double-majors

[23]See https://www.nytimes.com/2018/10/24/opinion/higher-education-double-major-extracurricular-activities.html

are devoted to outside your classwork. If you want to do more than three things outside of class, that's great, but not to get into college."

Our results highlight a drastic difference in limiting the extent of experimentation depending on the nature of the economy. In a closed economy (i.e., in a local market), limiting the extent of experimentation might improve efficiency. In an open economy, students from flexible systems might be worse off with a limited extent of experimentation. The optimal educational system (the amount of mandatory classes) depends on the degree and origin of labor market competition.

When studying how the differences in the distribution of talents across countries explain the trade patterns, Grossman and Maggi (2000) make two points regarding the educational systems. First, if talent discovery affects a country's comparative trade advantage, the optimal education policy depends on the policies adopted by trading partners. Second, given the heterogeneity of trading patterns, countries optimally adopt different educational approaches. Our analysis explicitly addresses these two points. Regarding the first point, we show that if the Eastern System is sufficiently restrictive (measured by the time students must devote to mandatory classes), there is a unique equilibrium in which all agents experiment. In addition, if the Eastern System is not very restrictive, then there is a symmetric equilibrium with both agents allocating some effort toward different classes. Regarding the second point, our analysis allows for asymmetric equilibria, such as the ones where competing agents specialize in different subjects.

7.2 Organizational Economics

Baron and Kreps (1999) present three possible classifications of jobs. "Star" jobs are occupations in which the low type is not that bad, but the high type brings much revenue for a company; "guardian" jobs are occupations in which the high type is satisfactory but the low type results in significant losses for the company; "foot-soldier" jobs, in which the high and the low types are not that different. Relating our results to the three types of jobs, specialization is efficient for the star jobs. In contrast, experimentation is efficient for the guardians and foot soldiers (see Proposition 1 for the necessary and sufficient conditions). Intuitively, in a star job, such as producing knowledge and innovation, firms are looking for groundbreaking results. For example, a pharmaceutical company benefits tremendously from a successful vaccine that captures the market. Experimentation generates more dispersed posterior beliefs about the expected productivity. As a result, experimentation is riskier from the company's point of view than specialization. In the case of a star job, the company would like to lower the chances of hiring the low type. Therefore, specialization is efficient. In the case of guardian jobs, such as commercial airline pilots, the low type might lead to a disaster outcome, such as an airplane crash. Experimentation is riskier for the applicant since he has a chance of failing more than once. Such an outcome, however, allows the company to rule out the low type with a higher probability. Therefore, experimentation is efficient. In the case of foot soldiers, the expected productivity after a failure does not decrease significantly. As a result, experimentation, although riskier, increases the expected productivity.

Consider the domestic market, where all the job applicants are from the same system. According to Corollary 1, experimentation (specialization) is less (more) likely in the environment with higher applicant bargaining power. Therefore, stars *over*-experiment unless they have all the bargaining power. In contrast, the guardians and foot soldiers *under*-experiment unless they have all the bargaining power. Similarly, competition within the domestic market leads to inefficiencies since specialization is more likely (see Lemma 1). Consequently, stars *over*-specialize, and the guardians and foot soldiers *under*-experiment due to domestic competition.

Our results indicate that competition in the international market might rule out specialization. The reason is that the externality from the Rigid system is so strong that the only equilibrium is one where all applicants experiment. Therefore, stars might over-experiment to hedge themselves against international competition. Table 1 below summarizes the key distortions.

Job	**Efficient to**	**↓ Applicant's Bargaining Power leads to**	**Domestic Competition** *leads to*	**International Competition** *leads to*
star	specialize	***under*** specialization	***over*** specialization	***under*** specialization
guardian, foot soldier	experiment	***over*** experimentation	***under*** experimentation	***over*** experimentation

Table 1: Distortions due to bargaining power, domestic and international competition.

Given the distortion in the efficient effort allocation, applicants either do not specialize enough or specialize too often. Hiring companies would like to alleviate these inefficiencies, for instance, such companies might perform additional training or set up a probation period for the applicants. Specifically, in many companies, employees devote some time to duties not related to their future tasks during the probation period. For example, some airline companies in the U.S. require all newly hired flight attendants to complete community service during the probation period.[24] In private banks, it is common for those on probation period to be assigned tasks not explicitly related to future duties (for example, a salesperson might spend some time in the accounting department).[25] At the same time, there are jobs where the probation period is devoted to tasks perfectly correlated with future duties. Examples are residency for future doctors and the Comprehensive Judicial Orientation Program for newly appointed judges in the US.

We propose two ways to interpret these differences in the probation period approach. First, suppose the probation period is the main channel for companies to learn about applicant productivity. For example, this is the case for low-skilled jobs where applicants' experience and education are irrelevant. The environment where job applicants must devote some effort to auxiliary tasks during the probation period is captured by the Rigid system in our model. Our results suggest that workers might be damaged by this constraint. For instance, this is the case when workers differ sufficiently in their productivity. Moreover, this inefficiency is exacerbated if workers have sufficient bargaining power vis-á-vis the employer.

Second, the probation period might be an attempt by the employer to alleviate inefficiencies summarized in Table 1. For example, consider the airline companies hiring flight attendants and pilots (i.e., guardians) in the domestic market. For these jobs, experimentation is efficient. However, our model predicts that domestic competition leads to under-experimentation (see the fourth column in Table 1). Thus, the optimal probation policy is to force the applicants to experiment. Thus, our economic intuition is consistent with some airline companies asking workers to perform tasks, such as community work, not explicitly related to their future duties during the probation period.

Similarly, a probation period in a private bank when a salesperson (i.e., a foot soldier) hired domestically performs accounting tasks is a way to force experimentation. In contrast, for highly skilled surgeons (i.e., guardians) hired internationally, it is optimal to specialize during the probation period. We summarize the optimal probation training for domestic and international markets in Table 2.

7.3 Additional Applications

Financing Investment Projects. Our model could be also seen as representing a financing relationship between an inventor or entrepreneur who has a project of uncertain quality. Financiers are willing to finance only good projects yielding positive NPV. Inventors can run a set of independent tests to provide certificates of their projects' quality to financiers who

[24] See `https://liveandletsfly.com/flight-attendants-community-service/` for a discussion.

[25] See also `https://www.wsj.com/articles/SB10001424052970204571404577257533620536076` for the discussion of job rotation during trainee programs in companies like General Electric.

Job	Optimal Probation (***Domestic*** *Market)*	Optimal Probation (***International*** *Market)*
star	specialization	specialization
guardian, **foot soldier**	experimentation	specialization

Table 2: The optimal probation policies.

may ask for all possible tests to be run (rigid system) or leave the choice of how many tests to run to the inventor himself (flexible system). If the project is financed and developed, the inventor and investors share the surplus it generates; such expected surplus would drive inventors' effort choice.

Drug Development. Experimentation is a crucial aspect of drug development processes. Our model can be applied to this framework too. Consider the case of a pharmaceutical company developing a new drug. Drug authorities may be more or less flexible across countries relative to the required types and amounts of tests to be run before the drug can be sold. Certain pharmaceutical companies may find it optimal to undergo as many tests as possible, while others prefer to focus on a certain type of tests only. Heterogeneous institutional environments may imply different choices by pharmaceutical companies.

8 Conclusions

In economic environments characterized by uncertainty about agents' productivity, it may be efficient for them to engage in several experimentation activities to convey signals about their productivity to their counterparts. However, in some scenarios, such experimentation may be constrained by institutions. For instance, pupils may be forced to attend several classes in different disciplines instead of focusing on a single topic, or companies' interns may be asked to perform several tasks to signal their match with the firm.

We present a simple model of experimentation in which agents can signal their type to prospective principals by exerting costly effort on two tasks. They can choose whether to *specialize* (i.e., devoting all of their efforts to one task only) or to *experiment* (i.e., devoting some effort to both tasks). We study this decision in two different institutional setups: a rigid system in which agents must exert some positive effort on both tasks and a flexible one, in which they have full discretion on how much effort to devote to each task.

Absent competition between agents, in a flexible system, specialization is more likely when the worker has sufficient bargaining power vis-$\acute{a}-$vis the principal, as she is entitled to a larger share of the surplus generated, which is then larger in expectation, with specialization. Indeed, when the agent devotes all of her effort to a single task, the posterior beliefs about her ability are more volatile (entailing higher variance, as well as higher expected value). In a rigid system, instead, specialization is not feasible, which, per se, constrains agents' optimal strategies, as well as social efficiency.

If several agents compete against each other while signaling their ability, more inefficiencies may arise. When competing agents are subject to the same institutional constraints (i.e., they are from the same system) specialization is less likely, as experimentation reduces volatility in the expected payoffs for agents as it increases the probabilities of producing better posterior beliefs than opponents. In a framework where agents are homogeneous, this induces symmetric equilibria. A new equilibrium arises in this setup: agents may mimic each other's effort. However, when agents are subject to different systems, i.e., they come respectively from a rigid and a flexible institutional system, then the mimicking equilibrium is less likely to be realized as the pressure of competition induces the agent from the flexible system to adapt to the choices made by the agent from the rigid system, namely, to specialize less and experiment more often, thus implying a further source of inefficiencies in the economy.

The main prediction is that inefficient institutional rigidities may hamper also flexible systems, in open economic setups in which agents with heterogeneous institutional backgrounds compete against each other for scarce resources.

References

Ali, S. N., N. Haghpanah, X. Lin, and R. Siegel (2022). How to sell hard information. *The Quarterly Journal of Economics 137*(1), 619–678.

Altonji, J. G., E. Blom, and C. Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu. Rev. Econ. 4*(1), 185–223.

Bar-Isaac, H. and R. Lévy (2022). Motivating employees through career paths. *Journal of Labor Economics 40*(1), 95–131.

Baron, J. and D. Kreps (1999). *Strategic Human Resources: Frameworks for General Managers*. John Wiley.

Berggren, A. and L. Jeppsson (2021). The impact of upper secondary school flexibility on sorting and educational outcomes. *Economics of Education Review 81*, 102080.

Betts, J. R. and H. Rose (2001). Math matters: The links between high school curriculum, college graduation, and earnings.

Bonatti, A. and J. Hörner (2017). Career concerns with exponential learning. *Theoretical Economics 12*(1), 425–475.

Brunner, T., G. Friebel, R. Holden, and S. Prasad (2021). Incentives to discover talent. *The Journal of Law, Economics, and Organization 38*(2), 309–344.

Canidio, A. and P. Legros (2023). Task discretion, labor-market frictions, and entrepreneurship. *The Journal of Law, Economics, and Organization 39*(2), 420–455.

Del Rossi, A. F. and J. Hersch (2008). Double your major, double your return? *Economics of Education Review 27*(4), 375–386.

Del Rossi, A. F. and J. Hersch (2016). The private and social benefits of double majors. *Journal of Benefit-Cost Analysis 7*(2), 292–325.

DeMarzo, P. M., I. Kremer, and A. Skrzypacz (2019). Test design and minimum standards. *American Economic Review 109*(6), 2173–2207.

Dur, U., Y.-C. Kao, and S. Paiement (2022). Market design of college-major switches for reducing student-major mismatch.

Gibbons, R. and M. Waldman (1999a). Careers in organizations: Theory and evidence. *Handbook of labor economics 3*, 2373–2437.

Gibbons, R. and M. Waldman (1999b). A theory of wage and promotion dynamics inside firms. *The Quarterly Journal of Economics 114*(4), 1321–1358.

Goodman, J. (2019). The labor of division: Returns to compulsory high school math coursework. *Journal of Labor Economics 37*(4), 1141–1182.

Görlitz, K. and C. Gravert (2018). The effects of a high school curriculum reform on university enrollment and the choice of college major. *Education Economics 26*(3), 321–336.

Grossman, G. M. and G. Maggi (2000). Diversity and trade. *American Economic Review 90*(5), 1255–1275.

Hanks, A. S., S. Jiang, X. Qian, B. Wang, and B. A. Weinberg (2024). Do double majors face less risk? an analysis of human capital diversification. Technical report, National Bureau of Economic Research.

Hermalin, B. E. (1993). Managerial preferences concerning risky projects. *The Journal of Law, Economics, and Organization 9*(1), 127–135.

Hirshleifer, D. and A. V. Thakor (1992). Managerial conservatism, project choice, and debt. *The Review of Financial Studies 5*(3), 437–470.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *The review of Economic studies 66*(1), 169–182.

Jia, N. (2021). Do stricter high school math requirements raise college stem attainment? *Economics of Education Review 83*, 102140.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Khalil, F., J. Lawarree, and A. Rodivilov (2020). Learning from failures: Optimal contracts for experimentation and production. *Journal of Economic Theory 190*, 105107.

Kimhi, O., B. Shahar, and T. Harel (2019). Higher education: Too much of a (potentially) good thing. *Wake Forest L. Rev. 54*, 761.

Kurino, M. and Y. Kurokawa (2023). Job rotation or specialization? a dynamic matching model analysis. *Review of Economic Design*, 1–31.

Leuven, E. and H. Oosterbeek (2011). Overeducation and mismatch in the labor market. *Handbook of the Economics of Education 4*, 283–326.

Levine, P. B. and D. J. Zimmerman (1995). The benefit of additional high-school math and science classes for young men and women. *Journal of Business & Economic Statistics 13*(2), 137–149.

Lizzeri, A. (1999). Information revelation and certification intermediaries. *The RAND Journal of Economics*, 214–231.

Malamud, O. (2010). Breadth versus depth: the timing of specialization in higher education. *Labour 24*(4), 359–390.

Malamud, O. (2011). Discovering one's talent: learning from academic specialization. *ILR Review 64*(2), 375–405.

Milbourn, T. T., R. L. Shockley, and A. V. Thakor (2001). Managerial career concerns and investments in information. *RAND Journal of Economics*, 334–351.

Pagano, M. and L. Picariello (2023). Talent discovery, layoff risk and unemployment insurance. *European Economic Review 154*, 104406.

Picariello, L. (2020). Organizational design with portable skills. *Available at SSRN 2901951*.

Rodina, D. and J. Farragut (2016). Inducing effort through grades. Technical report, Tech. rep., Working paper.

Rodivilov, A., D. Shin, and X. Zhao (2022). Overachieving and obsessive behavior as signaling devices under career concern. *Journal of Institutional and Theoretical Economics (JITE) 178*(4), 311331.

Rose, H. and J. R. Betts (2004). The effect of high school courses on earnings. *Review of economics and statistics 86*(2), 497–513.

Zafar, B. (2012). Double majors: One for me, one for the parents? *Economic Inquiry 50*(2), 287–308.

9 Appendix A. Proof of Proposition 1.

The social planner solves the following problem:

$$\max_{e_A, e_B} \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m)\Big[\mu_m f(\theta_H) + (1-\mu_m) f(\theta_L)\Big] - C(e_A, e_B)$$
$$\text{subject to } e_A, e_B \geqslant 0 \text{ and } e_A + e_B = 1.$$

To simplify notation, we denote $e_A = e$ and $e_B = 1 - e$, respectively.

The social planner's optimization problem then can be rewritten as

$$\max_{0 \leqslant e \leqslant 1} \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m)\Big[\mu_m f(\theta_H) + (1-\mu_m) f(\theta_L)\Big] - C(e, 1-e). \tag{24}$$

To simplify the objective function, we rewrite the expected output as

$$\begin{aligned} &\sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m)\Big[\mu_m f(\theta_H) + (1-\mu_m) f(\theta_L)\Big] = \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m)\big[\mu_m(\overline{y} + k) - k\big] \\ =&(\overline{y} + k) \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m)\mu_m - k \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m). \end{aligned} \tag{25}$$

Next, given that the expected posterior is equal to the prior (the Bayes' rule),

$$\sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m)\mu_m = \mu_0 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m)\mu_m, \tag{26}$$

and the probabilities of obtaining signals sum up to one:

$$\sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) = 1 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m). \tag{27}$$

Using (25) the objective function can be rewritten as

$$(\overline{y} + k)\Big[\mu_0 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m)\mu_m\Big] - k\Big[1 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m)\Big] - C(e, 1-e). \tag{28}$$

To find the optimal solution, we proceed in two steps. First, we assume that the solution is interior, i.e., $e \in (0,1)$. In this case, the agent cannot generate signal $\emptyset$ on any task. Therefore, the only beliefs μ_m that are lower than the lower bound $\hat{\mu}$ are $\mu_{(F,F)}$. This simplifies the agent's objective function. We then derive conditions that guarantee that the objective function is concave in the effort level e and, consequently, the interior solution is a global maximum. Second, we derive conditions under which the objective function is convex in e and, consequently, the interior solution is not a global maximum. In this case, the optimal solution is a corner one, i.e., $e \in \{0,1\}$.

Interior solution $e \in (0,1)$**.** We now assume that the social planner is choosing an effort $e \in (0,1)$ and, therefore, the objective function can be rewritten as

$$(\overline{y}+k)\big[\mu_0 - \tau(\mu_{(F,F)}]\mu_{(F,F)}\big] - k\big[1-\tau(\mu_{(F,F)})\big] - C(e,1-e), \tag{29}$$

where we used the fact that the only beliefs that are lower than the lower bound $\hat{\mu}$ are $\mu_{(F,F)}$.

Given that

$$\tau(\mu_{(F,F)}) = g((F,F)|\theta_H)\mu_0 + g((F,F)|\theta_L)(1-\mu_0), \tag{30}$$

and

$$\tau(\mu_{(F,F)})\mu_{(F,F)} = \tau(\mu_{(F,F)})\frac{g((F,F)|\theta_H)\mu_0}{\tau(\mu_{(F,F)})} = g((F,F)|\theta_H)\mu_0, \tag{31}$$

the objective function can be rewritten as

$$(\overline{y}+k)\Big[\mu_0 - g((F,F)|\theta_H)\mu_0\Big] - k\Big[1 - g((F,F)|\theta_H)\mu_0 - g((F,F)|\theta_L)(1-\mu_0)\Big] - C(e,1-e). \tag{32}$$

Finally, given that

$$g((F,F)|\theta) = \big[1-\theta e\big]\big[1-\theta(1-e)\big] \text{ for } \theta = \{\theta_L, \theta_H\}, \tag{33}$$

the objective function simplifies to

$$\begin{aligned}
&(\overline{y}+k)\Big[\mu_0 - g((F,F)|\theta_H)\mu_0\Big] - k\Big[1 - g((F,F)|\theta_H)\mu_0 - g((F,F)|\theta_L)(1-\mu_0)\Big] - C(e,1-e)\\
=&(\overline{y}+k)\mu_0\Big[1 - \big[1-\theta_H e\big]\big[1-\theta_H(1-e)\big]\Big]\\
&- k\Big[1 - \mu_0\big[1-\theta_H e\big]\big[1-\theta_H(1-e)\big] - (1-\mu_0)\big[1-\theta_L e\big]\big[1-\theta_L(1-e)\big]\Big] - C(e,1-e).
\end{aligned} \tag{34}$$

Taking the first-order derivative with respect to the effort level e we obtain:

$$
\begin{aligned}
& -\mu_0(\overline{y}+k)\Big[-\theta_H(1-\theta_H(1-e))+(1-\theta_H e)\theta_H\Big] \\
& -k\Big[-\mu_0\big[(1-\theta_H e)\theta_H-\theta_H(1-\theta_H(1-e))\big]-(1-\mu_0)\big[(1-\theta_L e)\theta_L-\theta_L(1-\theta_L(1-e))\big]\Big] \\
& -(2e-1) \\
= & -\mu_0(\overline{y}+k)\theta_H^2(1-2e)-k\Big[-\mu_0\theta_H^2(1-2e)-(1-\mu_0)\theta_L^2(1-2e)\Big]-(2e-1) \\
= & (1-2e)\Big[-\mu_0(\overline{y}+k)\theta_H^2+k\mu_0\theta_H^2+k(1-\mu_0)\theta_L^2+1\Big] \\
= & (1-2e)\Big[k(1-\mu_0)\theta_L^2-\overline{y}\mu_0\theta_H^2+1\Big]=0.
\end{aligned}
$$

Thus, the First Order Condition is

$$(1-2e)\Big[1-\big[\overline{y}\mu_0\theta_H^2-k(1-\mu_0)\theta_L^2\big]\Big]=0. \tag{35}$$

Global maximum. To guarantee that the solution given by the (35) above is a global maximum, the following Second Order Condition must be satisfied as well:

$$-2\Big[1-\big[\overline{y}\mu_0\theta_H^2-k(1-\mu_0)\theta_L^2\big]\Big]<0. \tag{36}$$

Therefore, if $\overline{y}\mu_0\theta_H^2-k(1-\mu_0)\theta_L^2<1$, then we have an interior solution

$$e^*=\frac{1}{2}. \tag{37}$$

However, if $\overline{y}\mu_0\theta_H^2-k(1-\mu_0)\theta_L^2\geqslant 1$, then we have a corner solution

$$e^*\in\{0,1\}. \tag{38}$$

This completes the proof of Proposition 1. *Q.E.D.*

10 Appendix B. Proof of Proposition 2.

The agent solves the following program:

$$
\begin{aligned}
[P^F] \quad & \max_{e_A,e_B}\textstyle\sum_{\mu_m\geqslant\hat{\mu}}\tau(\mu_m)w(\mu_m)-C(e_A,e_B) \\
& \text{subject to } e_A,e_B\geqslant 0 \text{ and } e_A+e_B=1.
\end{aligned}
$$

To simplify notation, we denote $e_A = e$ and $e_B = 1 - e$, respectively.

The agent's optimization problem then can be rewritten as

$$\max_{0 \leqslant e \leqslant 1} \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) w(\mu_m) - C(e, 1-e). \tag{39}$$

To simplify the objective function, we rewrite the expected wage using (14) as

$$\begin{aligned} \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) w(\mu_m) &= \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) \gamma \big[\mu_m (\overline{y} + k) - k \big] \\ &= \gamma \Bigg[(\overline{y} + k) \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) \mu_m - k \sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) \Bigg]. \end{aligned} \tag{40}$$

Next, given that the expected posterior is equal to the prior (the Bayes' rule):

$$\sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) \mu_m = \mu_0 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m) \mu_m, \tag{41}$$

and the probabilities of obtaining all the possible signals sum up to one:

$$\sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) = 1 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m), \tag{42}$$

the expression in (40) can be rewritten as

$$\sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) w(\mu_m) = \gamma \Bigg[(\overline{y} + k) \Bigg(\mu_0 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m) \mu_m \Bigg) - k \Bigg(1 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m) \Bigg) \Bigg]. \tag{43}$$

The objective function then can be rewritten as

$$\begin{aligned} &\textstyle\sum_{\mu_m \geqslant \hat{\mu}} \tau(\mu_m) w(\mu_m) - C(e, 1-e) \\ &= \gamma \Bigg[(\overline{y} + k) \Bigg(\mu_0 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m) \mu_m \Bigg) - k \Bigg(1 - \sum_{\mu_m < \hat{\mu}} \tau(\mu_m) \Bigg) \Bigg] - C(e, 1-e). \end{aligned} \tag{44}$$

To find the optimal solution, we proceed in two steps. First, we assume that the solution is interior, i.e., $e \in (0, 1)$. In this case, the agent cannot generate signal $\emptyset$ on any task. Therefore, the only beliefs μ_m that are lower than the lower bound $\hat{\mu}$ are $\mu_{(F,F)}$. This allows for the simplification of the agent's objective function. Then, we derive conditions that guarantee the objective function is concave in e and, consequently, the interior solution is a global maximum. Second, we derive conditions under which the objective function is convex in e and, consequently, the interior solution is not a global maximum. In this case, the optimal solution is a corner one, i.e., $e \in \{0, 1\}$.

Interior solution $e \in (0, 1)$. We now assume that the agent is choosing an effort $e \in (0, 1)$ and, therefore, the objective function can be rewritten as

$$\gamma\Big[(\overline{y}+k)\big[\mu_0 - \tau(\mu_{(F,F)})\mu_{(F,F)}\big] - k\big[1-\tau(\mu_{(F,F)})\big]\Big] - C(e, 1-e). \tag{45}$$

Given that

$$\tau(\mu_{(F,F)}) = g((F,F)|\theta_H)\mu_0 + g((F,F)|\theta_L)(1-\mu_0), \tag{46}$$

and

$$\tau(\mu_{(F,F)})\mu_{(F,F)} = \tau(\mu_{(F,F)})\frac{g((F,F)|\theta_H)\mu_0}{\tau(\mu_{(F,F)})} = g((F,F)|\theta_H)\mu_0, \tag{47}$$

the objective function becomes

$$\gamma\Big[(\overline{y}+k)\big[\mu_0 - g((F,F)|\theta_H)\mu_0\big] - k\big[1 - g((F,F)|\theta_H)\mu_0 - g((F,F)|\theta_L)(1-\mu_0)\big]\Big] - C(e, 1-e). \tag{48}$$

Finally, given that

$$g((F,F)|\theta) = \big[1-\theta e\big]\big[1-\theta(1-e)\big] \text{ for } \theta = \theta_L, \theta_H, \tag{49}$$

the objective function simplifies to

$$\gamma(\overline{y}+k)\mu_0\Big[1 - \big[1-\theta_H e\big]\big[1-\theta_H(1-e)\big]\Big]$$
$$-\gamma k\Big[1 - \mu_0\big[1-\theta_H e\big]\big[1-\theta_H(1-e)\big] - (1-\mu_0)\big[1-\theta_L e\big]\big[1-\theta_L(1-e)\big]\Big] - C(e, 1-e).$$

The first-order condition with respect to the effort level e is as follows:

$$\begin{aligned}
&-\gamma\mu_0(\overline{y}+k)\Big[-\theta_H(1-\theta_H(1-e)) + (1-\theta_H e)\theta_H\Big] \\
&\gamma k\Big[\mu_0\big[-\theta_H(1-\theta_H(1-e)) + (1-\theta_H e)\theta_H\big](1-\mu_0)\big[-\theta_L(1-\theta_L(1-e)) + (1-\theta_L e)\theta_L\big]\Big] \\
=&-\gamma\mu_0(\overline{y}+k)\theta_H^2(1-2e) - \gamma k\Big(-\mu_0\theta_H^2(1-2e) - (1-\mu_0)\theta_L^2(1-2e)\Big) \qquad VM \\
=&\gamma(1-2e)\Big(-\mu_0(\overline{y}+k)\theta_H^2 + k\mu_0\theta_H^2 + k(1-\mu_0)\theta_L^2\Big) \\
=&\gamma(1-2e)\Big(k(1-\mu_0)\theta_L^2 - \overline{y}\mu_0\theta_H^2\Big) = 2e-1.
\end{aligned}$$

The First Order Condition can be rewritten as

$$(1-2e)\Big[1-\gamma\big[\overline{y}\mu_0\theta_H^2-k(1-\mu_0)\theta_L^2\big]\Big]=0. \tag{50}$$

Global maximum. To guarantee that the solution given by the (50) above is a global maximum, the following Second Order Condition must be satisfied as well:

$$-2\Big[1-\gamma\big[\overline{y}\mu_0\theta_H^2-k(1-\mu_0)\theta_L^2\big]\Big]<0. \tag{51}$$

Therefore, if $\gamma\big[\overline{y}\mu_0\theta_H^2-k(1-\mu_0)\theta_L^2\big]<1$, then we have an interior solution

$$e^*=\frac{1}{2}. \tag{52}$$

However, if $\gamma\big[\overline{y}\mu_0\theta_H^2-k(1-\mu_0)\theta_L^2\big]\geqslant 1$, then we have a corner solution

$$e^*\in\{0,1\}. \tag{53}$$

This completes the proof of Proposition 2. *Q.E.D.*

11 Appendix C. Within Competition Analysis.

Outline of the proof: We characterize the equilibrium in four steps. In Step 1, we characterize and simplify the agent's objective function. In Step 2, we characterize four relevant cases (depending on the value of the agent's objective function) for the analysis. In Step 3, we prove Lemma 1. In Step 4, we prove Lemma 2.

Step 1. **Deriving the expected utility function.** Given that agents spend their entire unit of effort ($e_A^s+e_B^s=1$) in equilibrium, to simplify notation, we denote $e_A^s=e^s$ and $e_B^s=1-e^s$ for $s=\{F,R\}$.

Since the agent is hired with necessity only if he succeeds at least in one task, three outcomes are relevant for determining the agent's expected utility. First, with probability $\tau\big(\mu^F_{(S,S)}\big)$, the agent succeeds in both tasks. He then receives $\frac{1}{2}w(\mu^F_{(S,S)})$ if his rival produces the same signal (with probability $\tau(\mu^R_{(S,S)})$) and receives $w(\mu^F_{(S,S)})$ if his rival produces any other signal than (S,S) (with probability $1-\tau(\mu^R_{(S,S)})$). Second, in case the agent succeeds in one task only, he is paid $\frac{1}{2}w(\mu^F_{(S,F)})$ if his rival generates identical posterior which happens in two cases: either the rival produces the same signal (S,F) with exactly the same effort choice

$(e^R = e^F)$ or the rival succeeds on the other task (F,S) with symmetric effort allocation $(e^R = 1-e^F)$. In addition, when the agent succeeds only once $((S,F))$, he is paid $w(\mu^F_{(S,F)})$ if the second agent generates lower posterior beliefs regarding his type, which happens with probability $1 - \tau\big(\mu^R_{(S,S)}\big) - \tau\big(\mu^R_{(S,F)}\big)\mathbb{1}_{e^R<e^F} - \tau\big(\mu^R_{(F,S)}\big)\mathbb{1}_{e^R>1-e^F}$. Third, similarly to the previous case, the agent can succeed in task B only and generate signal (F,S).

The agent's objective function is[26]

$$\begin{aligned}
&\tau\big(\mu^F_{(S,S)}\big)\left[\frac{1}{2}\tau(\mu^R_{(S,S)})w(\mu^F_{(S,S)}) + (1-\tau(\mu^R_{(S,S)}))w(\mu^F_{(S,S)})\right]\\
&+\tau\big(\mu^F_{(S,F)}\big)w(\mu^F_{(S,F)})\left[1-\tau\big(\mu^R_{(S,S)}\big) - \tau\big(\mu^R_{(S,F)}\big)\mathbb{1}_{e^R<e^F} - \tau\big(\mu^R_{(F,S)}\big)\mathbb{1}_{e^R>1-e^F}\right]\\
&+\frac{1}{2}\tau\big(\mu^F_{(S,F)}\big)w(\mu^F_{(S,F)})\left[\tau\big(\mu^R_{(S,F)}\big)\mathbb{1}_{e^R=e^F} + \tau\big(\mu^R_{(F,S)}\big)\mathbb{1}_{e^R=1-e^F}\right]\\
&+\tau\big(\mu^F_{(F,S)}\big)w(\mu^F_{(F,S)})\left[1-\tau\big(\mu^R_{(S,S)}\big) - \tau\big(\mu^R_{(F,S)}\big)\mathbb{1}_{e^R>e^F} - \tau\big(\mu^R_{(S,F)}\big)\mathbb{1}_{e^R<1-e^F}\right]\\
&+\frac{1}{2}\tau\big(\mu^F_{(F,S)}\big)w(\mu^F_{(F,S)})\left[\tau\big(\mu^R_{(F,S)}\big)\mathbb{1}_{e^R=e^F} + \tau\big(\mu^R_{(S,F)}\big)\mathbb{1}_{e^R=1-e^F}\right]\\
&-C(e^F, 1-e^F).
\end{aligned} \tag{54}$$

To simplify the objective function in (54), we rewrite the distributions over the posterior beliefs τ (for a given e^j) as:

$$\begin{aligned}
&\tau(\mu^j_{(S,S)})\\
&\quad= g((S,S)|\theta_H)\mu_0 + g((S,S)|\theta_L)(1-\mu_0) = \theta_H e^j \theta_H(1-e^j)\mu_0 + \theta_L e^j \theta_L(1-e^j)(1-\mu_0)\\
&\quad= e^j(1-e^j)\big(\theta_H^2\mu_0 + \theta_L^2(1-\mu_0)\big) = e^j(1-e^j)\mathbb{E}[\theta^2];
\end{aligned} \tag{55}$$

$$\begin{aligned}
&\tau(\mu^j_{(S,F)})\\
&\quad= g((S,F)|\theta_H)\mu_0 + g((S,F)|\theta_L)(1-\mu_0)\\
&\quad= \theta_H e^j[1-\theta_H(1-e^j)]\mu_0 + \theta_L e^j[1-\theta_L(1-e^j)](1-\mu_0)\\
&\quad= e^j\big(\theta_H[1-\theta_H(1-e^j)]\mu_0 + \theta_L[1-\theta_L(1-e^j)](1-\mu_0)\big) = e^j\mathbb{E}[\theta(1-\theta(1-e^j))];
\end{aligned} \tag{56}$$

[26] A characteristic function $\mathbb{1}_{t\in\mathbb{T}}$ is defined as $\mathbb{1}_{t\in\mathbb{T}} = \begin{cases} 1, & t\in\mathbb{T} \\ 0, & t\notin\mathbb{T}\end{cases}$.

$$
\begin{aligned}
&\tau(\mu^j_{(F,S)}) \\
&\qquad =g((F,S)|\theta_H)\mu_0 + g((F,S)|\theta_L)(1-\mu_0) \\
&\qquad =[1-\theta_H e^j]\theta_H(1-e^j)\mu_0 + [1-\theta_L e^j]\theta_L(1-e^j)(1-\mu_0) \\
&\qquad =(1-e^j)\big([1-\theta_H e^j]\theta_H\mu_0 + [1-\theta_L e^j]\theta_L(1-\mu_0)\big) = (1-e^j)\mathbb{E}[\theta(1-\theta e^j)].
\end{aligned} \tag{57}
$$

Next, we rewrite the wage functions $w(\mu^j_m)$ as:

$$
\begin{aligned}
&w(\mu^j_{(S,S)}) \\
&\qquad =\gamma\left[\mu^j_{(S,S)}(\overline{y}+k)-k\right] = \gamma\left[\frac{g((S,S)|\theta_H)\mu_0}{\tau(\mu^j_{(S,S)})}(\overline{y}+k)-k\right] \\
&\qquad =\gamma\left[\frac{\mu_0\theta_H^2 e^j(1-e^j)}{e^j(1-e^j)\mathbb{E}[\theta^2]}(\overline{y}+k)-k\right] = \gamma\left[\frac{\mu_0\theta_H^2}{\mathbb{E}[\theta^2]}(\overline{y}+k)-k\right];
\end{aligned} \tag{58}
$$

$$
\begin{aligned}
&w(\mu^j_{(S,F)}) \\
&\qquad =\gamma\left[\mu^j_{(S,F)}(\overline{y}+k)-k\right] = \gamma\left[\frac{g((S,F)|\theta_H)\mu_0}{\tau(\mu^j_{(S,F)})}(\overline{y}+k)-k\right] \\
&\qquad =\gamma\left[\frac{\theta_H e^j[1-\theta_H(1-e^j)]\mu_0}{e^j\mathbb{E}[\theta(1-\theta(1-e^j))]}(\overline{y}+k)-k\right] = \gamma\left[\frac{\theta_H[1-\theta_H(1-e^j)]\mu_0}{\mathbb{E}[\theta(1-\theta(1-e^j))]}(\overline{y}+k)-k\right];
\end{aligned} \tag{59}
$$

$$
\begin{aligned}
&w(\mu^j_{(F,S)}) \\
&\qquad =\gamma\left[\mu^j_{(F,S)}(\overline{y}+k)-k\right] = \gamma\left[\frac{g((F,S)|\theta_H)\mu_0}{\tau(\mu^j_{(F,S)})}(\overline{y}+k)-k\right] \\
&\qquad =\gamma\left[\frac{[1-\theta_H e^j]\theta_H(1-e^j)\mu_0}{(1-e^j)\mathbb{E}[\theta(1-\theta e^j)]}(\overline{y}+k)-k\right] = \gamma\left[\frac{[1-\theta_H e^j]\theta_H\mu_0}{\mathbb{E}[\theta(1-\theta e^j)]}(\overline{y}+k)-k\right].
\end{aligned} \tag{60}
$$

Plugging the expression for τ and w into (54), the agent's objective function becomes:

$$
\begin{aligned}
& e^F(1-e^F)\gamma\Big(\mu_0\theta_H^2(\overline{y}+k)-k\mathbb{E}[\theta^2]\Big)\Big[1-\frac{e^R(1-e^R)\mathbb{E}[\theta^2]}{2}\Big] \\
+& e^F\gamma\Big(\theta_H[1-\theta_H(1-e^F)]\mu_0(\overline{y}+k)-k\mathbb{E}\big[\theta(1-\theta(1-e^F))\big]\Big)\times \\
& \Big[1-e^R(1-e^R)\mathbb{E}[\theta^2]-e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\mathbb{1}_{e^R<e^F}-(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]\mathbb{1}_{e^R>1-e^F} \\
& +\frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2}\mathbb{1}_{e^R=e^F}+\frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2}\mathbb{1}_{e^R=1-e^F}\Big] \\
+& (1-e^F)\gamma\Big((1-\theta_He^F)\theta_H\mu_0(\overline{y}+k)-k\mathbb{E}[\theta(1-\theta e^F)]\Big)\times \\
& \Big[1-e^R(1-e^R)\mathbb{E}[\theta^2]-(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]\mathbb{1}_{e^R>e^F}-e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\mathbb{1}_{e^R<1-e^F} \\
& +\frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2}\mathbb{1}_{e^R=e^F}+\frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2}\mathbb{1}_{e^R=1-e^F}\Big]-C(e^F,1-e^F) \\
=& e^F(1-e^F)\gamma\Big(\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\Big)\Big[1-\frac{e^R(1-e^R)\mathbb{E}[\theta^2]}{2}\Big] \\
+& e^F\gamma\Big(\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L+(1-e^F)\big(k(1-\mu_0)\theta_L^2-\mu_0\overline{y}\theta_H^2\big)\Big)\times \\
& \Big[1-e^R(1-e^R)\mathbb{E}[\theta^2]-e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\mathbb{1}_{e^R<e^F}-(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]\mathbb{1}_{e^R>1-e^F} \\
& +\frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2}\mathbb{1}_{e^R=e^F}+\frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2}\mathbb{1}_{e^R=1-e^F}\Big] \\
+& (1-e^F)\gamma\Big(\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L+e^F\big(k(1-\mu_0)\theta_L^2-\mu_0\overline{y}\theta_H^2\big)\Big)\times \\
& \Big[1-e^R(1-e^R)\mathbb{E}[\theta^2]-(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]\mathbb{1}_{e^R>e^F}-e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\mathbb{1}_{e^R<1-e^F} \\
& +\frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2}\mathbb{1}_{e^R=e^F}+\frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2}\mathbb{1}_{e^R=1-e^F}\Big]-C(e^F,1-e^F).
\end{aligned}
\tag{61}
$$

To simplify the objective function further, we define the following three functions

$$
\begin{aligned}
J_1 &= 1 - \frac{e^R(1-e^R)\mathbb{E}[\theta^2]}{2};\\
J_2 &= 1 - e^R(1-e^R)\mathbb{E}[\theta^2] - e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\mathbb{1}_{e^R<e^F} - (1-e^R)\mathbb{E}[\theta(1-\theta e^R)]\mathbb{1}_{e^R>1-e^F}\\
&\quad + \frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2}\mathbb{1}_{e^R=e^F} + \frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2}\mathbb{1}_{e^R=1-e^F};\\
J_3 &= 1 - e^R(1-e^R)\mathbb{E}[\theta^2] - (1-e^R)\mathbb{E}[\theta(1-\theta e^R)]\mathbb{1}_{e^R>e^F} - e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\mathbb{1}_{e^R<1-e^F}\\
&\quad + \frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2}\mathbb{1}_{e^R=e^F} + \frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2}\mathbb{1}_{e^R=1-e^F}.
\end{aligned}
\tag{62}
$$

Using (62), the agent's objective function becomes:

$$
\begin{aligned}
U(e^F, e^R) :=& e^F(1-e^F)\gamma\Big(\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\Big)J_1\\
&+ e^F\gamma\Big(\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L + (1-e^F)\big(k(1-\mu_0)\theta_L^2 - \mu_0\overline{y}\theta_H^2\big)\Big)J_2\\
&+ (1-e^F)\gamma\Big(\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L + e^F\big(k(1-\mu_0)\theta_L^2 - \mu_0\overline{y}\theta_H^2\big)\Big)J_3\\
&- C(e^F, 1-e^F)\\
=& e^F(1-e^F)\gamma\Big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\Big]\Big(J_1 - J_2 - J_3\Big)\\
&+ \gamma\Big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\Big]\Big(J_2e^F + J_3(1-e^F)\Big) - C(e^F, 1-e^F).
\end{aligned}
\tag{63}
$$

Step 2. **The best response correspondences.** To characterize all possible equilibria, we need to determine the best response correspondence, $e^F(e^R)$. We first note that the agent's utility function satisfies the following property:

$$
U(e^F, e^R) = U(e^F, 1-e^R) = U(1-e^F, e^R) = U(1-e^F, 1-e^R). \tag{64}
$$

This property significantly simplifies analysis as it allows us to determine the best response correspondence $e^F(e^R) \in [0, \frac{1}{2}]$ for $0 \leqslant e^R \leqslant \frac{1}{2}$ only.

We next determine the values of $J_1 - J_2 - J_3$ and $J_2e^F + J_3(1-e^F)$ for all possible combinations of the effort levels $e^F \in [0, \frac{1}{2}]$ and $e^R \in [0, \frac{1}{2}]$.

1) Case 1: $e^R < e^F \leqslant \frac{1}{2}$:

$J_1 - J_2 - J_3$

$$\begin{aligned} &= 1 - \frac{e^R(1-e^R)\mathbb{E}[\theta^2]}{2} \\ &-1 + e^R(1-e^R)\mathbb{E}[\theta^2] + e^R\mathbb{E}[\theta(1-\theta(1-e^R))] \\ &-1 + e^R(1-e^R)\mathbb{E}[\theta^2] + e^R\mathbb{E}[\theta(1-\theta(1-e^R))] \\ &= 2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1. \end{aligned} \tag{65}$$

$J_2 e^F + J_3(1 - e^F)$

$$\begin{aligned} &= \Big(1 + e^R(1-e^R)\mathbb{E}[\theta^2] + e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\Big)e^F \\ &+ \Big(1 + e^R(1-e^R)\mathbb{E}[\theta^2] + e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\Big)(1-e^F) \\ &= 1 + e^R(1-e^R)\mathbb{E}[\theta^2] + e^R\mathbb{E}[\theta(1-\theta(1-e^R))] = 1 + e^R\mathbb{E}[\theta]. \end{aligned} \tag{66}$$

Substituting (65) and (66) in (63), the agent's objective function becomes:
$U(e^F, e^R)$

$$\begin{aligned} &= e^F(1-e^F)\gamma\Big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\Big]\Big(2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\Big) \\ &+ \gamma\Big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\Big]\Big(1 + e^R\mathbb{E}[\theta]\Big) - C(e^F, 1-e^F). \end{aligned} \tag{67}$$

Consider the value of $U(e^F, e^R)$ in (67):

$$\begin{aligned} U(e^F, e^R) = &e^F(1-e^F)\gamma\Big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\Big]\Big(2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\Big) \\ &+ \gamma\Big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\Big]\Big(1 + e^R\mathbb{E}[\theta]\Big) - \Big(\frac{(e^F)^2}{2} + \frac{(1-e^F)^2}{2}\Big) \end{aligned} \tag{68}$$

$$\begin{aligned} &= \gamma\Big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\Big]\Big(1 + e^R\mathbb{E}[\theta]\Big) + \frac{1}{2} \\ &+ e^F\Big(\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\big) + 1\Big) \\ &- (e^F)^2\Big(\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\big) + 1\Big), \end{aligned} \tag{69}$$

that is a monotonic function of e^F. To see the reason for monotonicity, note that the first-order derivative with respect to $e^F \in [0, \frac{1}{2}]$ is:

$$(1-2e^F)\Big(\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\big(2e^R\mathbb{E}[\theta]-\frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2]-1\big)+1\Big). \tag{70}$$

Thus, if

$$\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\big(2e^R\mathbb{E}[\theta]-\frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2]-1\big)+1>0, \tag{71}$$

then $U(e^F, e^R)$ in Case 1 is increasing in $e^F \in [0, \frac{1}{2}]$, and it is decreasing otherwise. Therefore, $U(e^F, e^R)$ in Case 1 achieves maximum at $e^F = \frac{1}{2}$ if $\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\big(2e^R\mathbb{E}[\theta]-\frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2]-1\big)+1>0$, and achieves maximum at $e^F = e^R + \varepsilon$ for small $\varepsilon > 0$ otherwise.

2) Case 2: $e^F < e^R \leqslant \frac{1}{2}$:

$$\begin{aligned}
&J_1 - J_2 - J_3\\
&=1-\frac{e^R(1-e^R)\mathbb{E}[\theta^2]}{2}\\
&-1+e^R(1-e^R)\mathbb{E}[\theta^2]\\
&-1+e^R(1-e^R)\mathbb{E}[\theta^2]+(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]+e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\\
&=\mathbb{E}[\theta]-\frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2]-1.
\end{aligned} \tag{72}$$

$$\begin{aligned}
&J_2e^F + J_3(1-e^F)\\
=&\Big(1-e^R(1-e^R)\mathbb{E}[\theta^2]\Big)e^F\\
+&\Big(1-e^R(1-e^R)\mathbb{E}[\theta^2]-(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]-e^R\mathbb{E}[\theta(1-\theta(1-e^R))]\Big)(1-e^F)\\
=&1-e^R(1-e^R)\mathbb{E}[\theta^2]-\Big(\mathbb{E}[\theta]-2e^R(1-e^R)\mathbb{E}[\theta^2]\Big)(1-e^F).
\end{aligned} \tag{73}$$

Substituting (72) and (73) in (63), the agent's objective function becomes:

$$\begin{aligned}
&U(e^F, e^R)\\
=&e^F(1-e^F)\gamma\Big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\Big]\Big(\mathbb{E}[\theta]-\frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2]-1\Big)\\
+&\gamma\Big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\Big]\Big(1-e^R(1-e^R)\mathbb{E}[\theta^2]-\big(\mathbb{E}[\theta]-2e^R(1-e^R)\mathbb{E}[\theta^2]\big)(1-e^F)\Big)\\
-&C(e^F, 1-e^F).
\end{aligned} \tag{74}$$

Consider the value of $U(e^F, e^R)$ in (74):

$$
\begin{aligned}
U(e^F, e^R) =& e^F(1-e^F)\gamma\left[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\right]\left(\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\right) \\
&+\gamma\left[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\right]\left(1 - e^R(1-e^R)\mathbb{E}[\theta^2] - \big(\mathbb{E}[\theta] - 2e^R(1-e^R)\mathbb{E}[\theta^2]\big)(1-e^F)\right) \\
&-\left(\frac{(e^F)^2}{2} + \frac{(1-e^F)^2}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&=\gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big]\left(1 - e^R(1-e^R)\mathbb{E}[\theta^2] - \big(\mathbb{E}[\theta] - 2e^R(1-e^R)\mathbb{E}[\theta^2]\big)\right) \\
&+e^F\left(\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\big) + \gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big]\big(\mathbb{E}[\theta] - 2e^R(1-e^R) \right. \\
&-(e^F)^2\left(\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\big) + 1\right),
\end{aligned}
\tag{75}
$$

that is an increasing function of e^F. To see the reason for monotonicity, note that the first-order derivative with respect to $e^F \in [0, \frac{1}{2}]$ is:

$$
\begin{aligned}
&\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\big) \\
&+\gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big]\big(\mathbb{E}[\theta] - 2e^R(1-e^R)\mathbb{E}[\theta^2]\big) + 1 \\
&-2e^F\left(\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\big) + 1\right).
\end{aligned}
\tag{76}
$$

Given that $\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2 < \mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L$, the derivative is greater than:

$$
\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(\mathbb{E}[\theta] - 2e^R(1-e^R)\mathbb{E}[\theta^2]\big) + 1 - 2e^F > 0, \tag{77}
$$

which is positive since $\mathbb{E}[\theta] - 2e^R(1-e^R)\mathbb{E}[\theta^2]$ for $e^R \in [0, \frac{1}{2}]$ and $1 - 2e^F > 0$ for $e^F \in [0, \frac{1}{2}]$.

Therefore, $U(e^F, e^R)$ in Case 2 achieves maximum at $e^F = e^R - \varepsilon$.

3) Case 3: $e^R = e^F \neq \frac{1}{2}$:

$J_1 - J_2 - J_3$

$$\begin{aligned}
=&1 - \frac{e^R(1-e^R)\mathbb{E}[\theta^2]}{2} \\
&-1 + e^R(1-e^R)\mathbb{E}[\theta^2] - \frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2} \\
&-1 + e^R(1-e^R)\mathbb{E}[\theta^2] - \frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2} \\
=&\frac{5}{2}e^R(1-e^R)\mathbb{E}[\theta^2] - \frac{\mathbb{E}[\theta]}{2} - 1.
\end{aligned} \tag{78}$$

$J_2 e^F + J_3(1-e^F)$

$$\begin{aligned}
=&\left(1 - e^R(1-e^R)\mathbb{E}[\theta^2] + \frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2}\right)e^F \\
&+\left(1 - e^R(1-e^R)\mathbb{E}[\theta^2] + \frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2}\right)(1-e^F) \\
=&1 + \frac{(1-e^R)}{2}\mathbb{E}[\theta] - \frac{3e^R(1-e^R)}{2}\mathbb{E}[\theta^2] + \frac{e^F}{2}(2e^R-1)\mathbb{E}[\theta].
\end{aligned} \tag{79}$$

Substituting (78) and (79) in (63), the agent's objective function becomes:
$U(e^F, e^R)$

$$\begin{aligned}
=&e^F(1-e^F)\gamma\left[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\right]\left(\frac{5}{2}e^R(1-e^R)\mathbb{E}[\theta^2] - \frac{\mathbb{E}[\theta]}{2} - 1\right) \\
&+\gamma\left[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\right]\left(1 + \frac{(1-e^R)}{2}\mathbb{E}[\theta] - \frac{3e^R(1-e^R)}{2}\mathbb{E}[\theta^2] + \frac{e^F}{2}(2e^R-1)\mathbb{E}[\theta]\right) \\
&-C(e^F, 1-e^F).
\end{aligned} \tag{80}$$

4) Case 4: $e^R = e^F = \frac{1}{2}$:
$J_1 - J_2 - J_3$

$$\begin{aligned}
=&1 - \frac{e^R(1-e^R)\mathbb{E}[\theta^2]}{2} \\
&-1 + e^R(1-e^R)\mathbb{E}[\theta^2] - \frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2} - \frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2} \\
&-1 + e^R(1-e^R)\mathbb{E}[\theta^2] - \frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2} - \frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2} \\
=&\frac{7}{2}e^R(1-e^R)\mathbb{E}[\theta^2] - \mathbb{E}[\theta] - 1.
\end{aligned} \tag{81}$$

$$
\begin{aligned}
&J_2e^F + J_3(1-e^F)\\
=&\left(1 - e^R(1-e^R)\mathbb{E}[\theta^2] + \frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2} + \frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2}\right)e^F\\
+&\left(1 - e^R(1-e^R)\mathbb{E}[\theta^2] + \frac{(1-e^R)\mathbb{E}[\theta(1-\theta e^R)]}{2} + \frac{e^R\mathbb{E}[\theta(1-\theta(1-e^R))]}{2}\right)(1-e^F) \quad (82)\\
=&1 + \frac{\mathbb{E}[\theta]}{2} - 2e^R(1-e^R)\mathbb{E}[\theta^2].
\end{aligned}
$$

Substituting $e^R = e^F = \frac{1}{2}$, (81) and (82) in (63), the agent's objective function becomes: $U(e^F, e^R)$

$$
\begin{aligned}
=&\frac{1}{4}\gamma\left[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\right]\left(\frac{7}{8}\mathbb{E}[\theta^2] - \mathbb{E}[\theta] - 1\right)\\
+&\gamma\left[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\right]\left(1 + \frac{\mathbb{E}[\theta]}{2} - \frac{1}{2}\mathbb{E}[\theta^2]\right) - \frac{1}{4}.
\end{aligned} \quad (83)
$$

To summarize the four cases, for a given $e^R \in [0, \frac{1}{2}]$ chosen by the rival, the agent is choosing $e^F \in [0, \frac{1}{2}]$ to maximize the following objective function:

$$
\begin{aligned}
U(e^F, e^R) = &\gamma e^F(1-e^F)\left[\mu_0\overline{y}\theta_H^2 - (1-\mu_0)k\theta_L^2\right]\left(J_1 - J_2 - J_3\right)\\
&+ \gamma\left[\mu_0\overline{y}\theta_H - (1-\mu_0)k\theta_L\right]\left(J_2e^F + J_3(1-e^F)\right) - C(e^F, 1-e^F),
\end{aligned} \quad (84)
$$

where the values of $J_2 + J_3 - J_1$ and $J_3 - J_2$ are as follows:

$$
J_2 + J_3 - J_1 = \begin{cases}
1 - 2e^R\mathbb{E}[\theta] + \frac{1}{2}e^R(1-e^R)\mathbb{E}[\theta^2], & e^R < e^F < 1-e^R\\
1 - \mathbb{E}[\theta] + \frac{1}{2}e^R(1-e^R)\mathbb{E}[\theta^2], & 1-e^F < e^R < e^F\\
1 - \mathbb{E}[\theta] + \frac{1}{2}e^R(1-e^R)\mathbb{E}[\theta^2], & e^F < e^R < 1-e^F\\
1 - 2(1-e^R)\mathbb{E}[\theta] + \frac{1}{2}e^R(1-e^R)\mathbb{E}[\theta^2], & 1-e^R < e^F < e^R\\
1 + \frac{1}{2}\mathbb{E}[\theta] - \frac{5}{2}e^R(1-e^R)\mathbb{E}[\theta^2], & e^R = e^F \neq \frac{1}{2}; e^R = 1-e^F \neq \frac{1}{2}\\
1 + \mathbb{E}[\theta] - \frac{7}{8}\mathbb{E}[\theta^2], & e^R = e^F = 1-e^F = \frac{1}{2}
\end{cases}.
$$

$$J_3 - J_2 = \begin{cases} 0, & e^R < e^F < 1 - e^R \\ \mathbb{E}[\theta] - 2e^R(1-e^R)\mathbb{E}[\theta^2], & 1 - e^F < e^R < e^F \\ -\mathbb{E}[\theta] + 2e^R(1-2e^R)\mathbb{E}[\theta^2], & e^F < e^R < 1 - e^F \\ 0, & 1 - e^R < e^F < e^R \\ \frac{1}{2}(1-2e^R)\mathbb{E}[\theta], & e^R = e^F \neq \frac{1}{2} \\ \frac{1}{2}(2e^R - 1)\mathbb{E}[\theta], & e^R = 1 - e^F \neq \frac{1}{2} \\ 0, & e^R = e^F = 1 - e^F = \frac{1}{2} \end{cases}.$$

Step 3. **Proof of Lemma 1.** To characterize necessary and sufficient conditions for the specialization equilibrium, we determine the best response function to $e^R = 0$. Thus, we compare the value of $U(e^F, 0)$ in Case 1 given by (67) and the value of $U(0,0)$ in Case 3 given by (80). Recalling analysis of Case 1 at Step 2, $U(e^F, e^R)$ in Case 1 achieves maximum at $e^F = \frac{1}{2}$ if $\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\big) + 1 > 0$ (see condition (71)), and achieves maximum at $e^F = e^R + \varepsilon$ for small $\varepsilon > 0$ otherwise.

Evaluating the condition (71) at $e^R = 0$, $U(e^F, e^R)$ in Case 1 achieves maximum value of

$$\begin{cases} \gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big] - \frac{1}{4}\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] - \frac{1}{4} \text{ if } \gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] < 1; \\ \gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big] - \varepsilon(1-\varepsilon)\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] - (\frac{\varepsilon^2}{2} + \frac{(1-\varepsilon)^2}{2}) \text{ otherwise.} \end{cases}$$

The value of the utility function in Case 3 is $U(0,0) = \gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big]\big(1 + \frac{\mathbb{E}[\theta]}{2}\big) - \frac{1}{2}$. Therefore, there are two possibilities depending on which utility function is greater:

The best response to a specializing agent is to specialize as well, $e^F(0) = 0$, if

(*Scenario 1*) either $\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] < 1$ and

$$\gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big] - \frac{1}{4}\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] - \frac{1}{4} < \gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big]\big(1 + \frac{\mathbb{E}[\theta]}{2}\big) - \frac{1}{2}; \tag{85}$$

(*Scenario 2*) or $\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] \geqslant 1$ and

$$\begin{aligned} \gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big] - \varepsilon(1-\varepsilon)\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] - (\frac{\varepsilon^2}{2} + \frac{(1-\varepsilon)^2}{2}) \\ < \gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big]\big(1 + \frac{\mathbb{E}[\theta]}{2}\big) - \frac{1}{2}. \end{aligned} \tag{86}$$

Note first that condition (85) can be rewritten as

$$1 - 2\mathbb{E}[\theta]\gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big] < \gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]. \tag{87}$$

Thus, combining the two conditions in Scenario 1, the best response to a specializing agent is to specialize as well, $e^F(0) = 0$, if

$$1 - 2\mathbb{E}[\theta]\gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big] < \gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] < 1. \tag{88}$$

Note second that condition (86) is automatically satisfied for small $\varepsilon > 0$. Therefore, in Scenario 2, the only relevant condition is

$$\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big] \geqslant 1. \tag{89}$$

Combining (88) and (89), we conclude that the best response to a specializing agent is to specialize as well, $e^F(0) = 0$, if

$$1 - 2\mathbb{E}[\theta]\gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big] < \gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]. \tag{90}$$

This completes the proof of Lemma 1. *Q.E.D.*

Step 4. **Proof of Lemma 2.** To prove Lemma 2, we compare the value of $U(e^F, e^R)$ in Case 1 given by (67), the value of $U(e^F, e^R)$ in Case 2 given by (74), and the value of $U(e^F, e^R)$ in Case 3 given by (80). We first compare the three values and establish that the value of $U(e^F, e^R)$ in Case 2 is smaller than the value of $U(e^F, e^R)$ in Case 3. Then, we derive conditions for the value of $U(e^F, e^R)$ in Case 3 to be greater than in Case 1.

(*Case 1*) Recalling analysis of Case 1 at Step 2, $U(e^F, e^R)$ in Case 1 achieves maximum at $e^F = \frac{1}{2}$ if $\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\big(2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\big) + 1 > 0$ (see condition (71)),

$$\begin{aligned} U(\frac{1}{2}, e^R) =& \frac{1}{4}\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\left(2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\right) \\ &+\gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big]\left(1 + e^R\mathbb{E}[\theta]\right) - \frac{1}{4}. \end{aligned} \tag{91}$$

and achieves maximum at $e^F = e^R + \varepsilon$ for small $\varepsilon > 0$ otherwise,

$$\begin{aligned} &U(e^R+\varepsilon, e^R) \\ =&(e^R+\varepsilon))\big(1-(e^R+\varepsilon))\big)\gamma\big[\mu_0\theta_H^2\overline{y} - (1-\mu_0)k\theta_L^2\big]\left(2e^R\mathbb{E}[\theta] - \frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2] - 1\right) \\ &+\gamma\big[\mu_0\theta_H\overline{y} - (1-\mu_0)k\theta_L\big]\left(1 + e^R\mathbb{E}[\theta]\right) - \left(\frac{(e^R+\varepsilon)^2}{2} + \frac{(1-e^R-\varepsilon)^2}{2}\right). \end{aligned} \tag{92}$$

(*Case 2*) Recalling analysis of Case 2 at Step 2, $U(e^F, e^R)$ given by (74) achieves maximum at $e^F = e^R - \varepsilon$,

$$
\begin{aligned}
&U(e^R - \varepsilon, e^R)\\
=&(e^R-\varepsilon))\big(1-(e^R-\varepsilon))\big)\gamma\Big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\Big]\left(\mathbb{E}[\theta]-\frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2]-1\right)\\
+&\gamma\Big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\Big]\Big(1-e^R(1-e^R)\mathbb{E}[\theta^2]-\big(\mathbb{E}[\theta]-2e^R(1-e^R)\mathbb{E}[\theta^2]\big)(1-(e^R-\varepsilon))\Big)\\
-&\left(\frac{(e^R-\varepsilon)^2}{2}+\frac{(1-(e^R-\varepsilon)))^2}{2}\right).
\end{aligned}
\tag{93}
$$

(*Case 3*) Evaluating the value of (80) at $e^F = e^R$, the utility function in Case 3 is: $U(e^R, e^R)$

$$
\begin{aligned}
=&e^R(1-e^R)\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\left(\frac{5}{2}e^R(1-e^R)\mathbb{E}[\theta^2]-\frac{\mathbb{E}[\theta]}{2}-1\right)\\
+&\gamma\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]\left(1+\frac{(1-2e^R+2(e^R)^2)}{2}\mathbb{E}[\theta]-\frac{3e^R(1-e^R)}{2}\mathbb{E}[\theta^2]\right)\\
-&\left(\frac{(e^R)^2}{2}+\frac{(1-e^R)^2}{2}\right).
\end{aligned}
\tag{94}
$$

We prove that the value of $U(e^F, e^R)$ in Case 2 is smaller than the value in Case 3. Since $e^R(1-e^R)$ is an increasing function of e^R on $[0, \frac{1}{2}]$, the sufficient condition for the value in (93) to be smaller than the value in (94) can be written as:

$$
\begin{aligned}
&e^R(1-e^R)\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\left(\mathbb{E}[\theta]-\frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2]-1\right)\\
+&\gamma\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]\Big(1-e^R(1-e^R)\mathbb{E}[\theta^2]-\big(\mathbb{E}[\theta]-2e^R(1-e^R)\mathbb{E}[\theta^2]\big)(1-e^R)\Big)\\
<&e^R(1-e^R)\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\left(\frac{5}{2}e^R(1-e^R)\mathbb{E}[\theta^2]-\frac{\mathbb{E}[\theta]}{2}-1\right)\\
+&\gamma\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]\left(1+\frac{(1-2e^R+2(e^R)^2)}{2}\mathbb{E}[\theta]-\frac{3e^R(1-e^R)}{2}\mathbb{E}[\theta^2]\right).
\end{aligned}
\tag{95}
$$

We then simplify the condition (95) above as

$$\begin{aligned}&e^R(1-e^R)\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\frac{3}{2}\Big(\mathbb{E}[\theta]-2e^R(1-e^R)\mathbb{E}[\theta^2]\Big)\\<&\gamma\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]\frac{1}{2}\Big((3-4e^R+2(e^R)^2)\mathbb{E}[\theta]-e^R(1-e^R)(3-2e^R)\mathbb{E}[\theta^2]\Big),\end{aligned}\tag{96}$$

which can be rewritten after substituting $e^R(1-e^R)$ with $\frac{1}{4}$ on the left-hand side (to achieve the highest value) as:

$$\begin{aligned}&\frac{1}{4}\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\frac{3}{2}\Big(\mathbb{E}[\theta]-2e^R(1-e^R)\mathbb{E}[\theta^2]\Big)\\<&\gamma\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]\frac{1}{2}\Big((3-4e^R+2(e^R)^2)\mathbb{E}[\theta]-e^R(1-e^R)(3-2e^R)\mathbb{E}[\theta^2]\Big),\end{aligned}\tag{97}$$

$$\begin{aligned}&\gamma\mathbb{E}[\theta]\left(\frac{\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]}{2}(3-4e^R+2(e^R)^2)-\frac{3\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]}{8}\right)>\\&\gamma\mathbb{E}[\theta^2]e^R(1-e^R)\left(\frac{3\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]}{4}-\frac{\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big](3-2e^R)}{2}\right),\end{aligned}\tag{98}$$

which holds since $\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]>\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]$ and $e^R(1-e^R)\leqslant\frac{1}{4}$.

In Lemma 1, we established that, if (71) holds at $e^R=0$, then utility in Case 3 is grater than in Case 1. Since both the utilities in Case 1 and Case 3 are continuous, there exists a value of e^R called e^M, such that utility in Case 3 is grater than in Case 1 for $e^R\leqslant e^M$:

$$\begin{aligned}&\frac{1}{4}\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\left(2e^R\mathbb{E}[\theta]-\frac{e^R(1-e^R)}{2}\mathbb{E}[\theta^2]-1\right)\\&+\gamma\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]\Big(1+e^R\mathbb{E}[\theta]\Big)-\frac{1}{4}\\<&e^R(1-e^R)\gamma\big[\mu_0\theta_H^2\overline{y}-(1-\mu_0)k\theta_L^2\big]\left(\frac{5}{2}e^R(1-e^R)\mathbb{E}[\theta^2]-\frac{\mathbb{E}[\theta]}{2}-1\right)\\&+\gamma\big[\mu_0\theta_H\overline{y}-(1-\mu_0)k\theta_L\big]\left(1+\frac{(1-2e^R+2(e^R)^2)}{2}\mathbb{E}[\theta]-\frac{3e^R(1-e^R)}{2}\mathbb{E}[\theta^2]\right)\\&-\left(\frac{(e^R)^2}{2}+\frac{(1-e^R)^2}{2}\right).\end{aligned}\tag{99}$$

This completes the proof of Lemma 2. *Q.E.D.*