Education, Experimentation, and Entrepreneurship

Tobias Brünner[*], Guido Friebel[†], Richard Holden[‡], Suraj Prasad[§]

April 29, 2016

PRELIMINARY DRAFT

Abstract

Countries approach higher education in very different ways. In Canada and the United States, students are encouraged to *experiment* across fields whereas in countries in Europe like France and Germany, students are forced to *specialize* in a field. We provide an explanation for these differences based on occupational choices available to a worker. In our model there are two phases in a person's career: an educational phase where she learns about her talents, and an occupational phase where she puts her talents to productive use, either as an entrepreneur or an employee. A key implication from our theory is that entrepreneurs experiment more than employees. This is because entrepreneurs, who are residual claimants, are affected only by good draws of their talent whereas employees, who bargain over wages in an imperfect labor market, are also affected by bad draws. Consequently, a system with experimentation is optimal when there are easier opportunities for entrepreneurship.

1 Introduction

Since at least Mincer (1958) and Becker (1964) economists have understood the importance of human capital in determining the economic outcomes of individuals. Becker (1964) emphasized that human capital is something that one can invest in, and the quintessential investment in human capital is education. Viewed from this perspective – where education simply adds to the stock of human capital – one might expect to find that approaches to education are roughly similar around the world. Yet this could not be further from the truth.

Countries like Canada and the United States have an education system that encourages *experimentation*: students are allowed – and even mandated – to specialize

[*]Lincoln Business School, University of Lincoln. tbruenner@lincoln.ac.uk

[†]Department of Management and Microeconomics, Goethe University, Frankfurt. gfriebel@wiwi.uni-frankfurt.de

[‡]UNSW Australia Business School. richard.holden@unsw.edu.au

[§]School of Economics, University of Sydney. suraj.prasad@sydney.edu.au

late and to choose courses across a broad range of fields. At the other end of the spectrum, countries in Europe like France and Germany have education systems that demand *specialization*–they require students to pick a field (which in some cases is vocational) early and stick to it. Most other countries fall somewhere between these two extremes, though closer to the specialization end. In Australia, for example, students choose their field of specialization early but have some flexibility within their degree to try out different areas.[1]

In our paper we see these systems through the lens of an analyst concerned with the optimal design of institutions. We start by asking the following question. Which system, experimentation or specialization, is more efficient? Taking the view that people use education to discover their talents, we show that experimentation is efficient for a surprisingly broad range of cases. Using this result as a benchmark, we then turn our attention to the question of why countries adopt different education systems. The reason that we propose here is that countries have very different opportunities for entrepreneurship. In particular, we show that when it is relatively easier to become an entrepreneur, as is the case in countries like Canada and the U.S, then experimentation is the optimal system. Conversely, when opportunities for entrepreneurship are difficult, like they are in France and Germany, then specialization is the optimal system.

In our model we consider a setting where there are two phases in a person's career: educational and occupational. In the educational phase there are no direct, payoff relevant consequences, but the person can learn about her talents (or her future productivity in an area). In the latter, occupational phase, the person can choose to be an entrepreneur or an employee, and payoffs are realized as a function of the person's talents. This interaction between education and occupational choice – while complicating matters – makes for interesting economics.

A key innovation of our framework is to distinguish, in the occupational phase, between *entrepreneurship* and *employment*. We see these alternatives as stark representations of the two basic alternatives in work. The fundamental economic difference between the two modes of work is who is the residual claimant on output. In our model, an entrepreneur is the residual claimant on output, whereas under employment, the surplus created by the employment relationship is bargained over with the employer in an imperfect labor market. The specific imperfection that we consider is that of a monopsonistic labor market. As Manning (2003) argues persuasively, a monopsony need not be thought of as just as a setting with a single firm. Rather, it can be interpreted more broadly as a labor market where frictions create rents from employment and where firms have bargaining power in setting wages.

To analyze this model, we start with the efficient sampling strategy as a bench-

[1] Singapore and the U.K are other examples of intermediate cases. Though students in these countries mostly specialize early, some universities have started to offer degrees that are closer to the North American system of experimentation. Also within the U.K, there is some variation. Scotland, for example, has a system that is much closer to the North American system.

mark. That is we consider a person who has sampled a talent and ask whether it is efficient to sample the same type of talent again and specialize or to sample a different type of talent and experiment. One might expect that this answer depends on whether the initial signal was good or bad. Surprisingly, this turns out not to be the case. We show that it is efficient for a person to experiment regardless of her signal in the other talent. Intuitively, since a person can choose the area she works in, there is an *upside* effect: good draws of a talent have a benefit because of a higher posterior mean, whereas bad draws entail no cost because the person can work in a different area. With experimentation where a talent is sampled for the first time there is more to learn: more weight is placed on the signal of talent when updating beliefs, and extreme values of talent are more likely. Consequently, the upside effect is larger for experimentation.

We then turn to how education choices interact with occupational choices. Our main result here is that entrepreneurs experiment whereas employees with low bargaining power specialize. The reason entrepreneurs experiment is simple; as residual claimants they have an incentive to make the efficient choice. Employees with low bargaining power, on the other hand, specialize for the following reason. When wages are determined by bargaining and labor markets are imperfect, an employee's wage depends on her reservation wage, which is the lowest of all her talents. This creates a *downside* effect: good draws of a talent have no benefit whereas bad draws lower the reservation wage. Specialization, by muting learning, helps to reduce this downside effect.

Because entrepreneurs experiment more, they have a broader education profile. In this sense, our paper is related to Lazear's analysis of who becomes an entrepreneur (Lazear (2005)). But the implications from our model are distinct. In our model, the breadth in education arises from an entrepreneur sampling *unrelated* areas whereas in Lazear's work, an entrepreneur's varied background is across *related* areas. Our model thus complements his. Another difference is that in our setting, the production function for entrepreneurship and employment are the same. So we run an unbiased horse race between entrepreneurship and employment.

Having analyzed the individual-level model of optimal sampling and occupational choice we are then in a position to discuss the broader institutional design question concerning the structure of educational systems and how they differ across countries, from both a positive and normative perspective. Our main message is that the optimality of an education system is inextricably linked to the opportunities available to individuals in the working phase of their career. In particular, we point our that countries with easier opportunities for entrepreneurship should optimally have educational systems that encourage experimentation. Experimentation is also optimal as a system when labor markets grant workers more bargaining power (say because of increased competition in the labor market), when workers have better options to fall back on in the labor market, and when human capital is more general.

Our paper connects to a number of significant and well-developed literatures, and we will not offer a comprehensive overview of those literatures, but content ourselves with highlighting some of the more salient connections.

As we noted earlier, we have a clear connection to the enormous literature on education as a form of human capital, the formal study of which began with Mincer (1958) and was propelled forward due to pathbreaking contributions by Schultz (1961) and Becker (1964). This literature, with the exception of a few papers (Altonji (1993) and Arcidiacono (2004)), has paid little attention to the details of an education system. We address this issue by taking a different view of education – one where a person discovers her talents.[2] In this respect our paper is similar to Malamud (2011), who looks at how the timing of specialization across the Scottish and English systems, affects the likelihood of switching fields.[3] Also related, is a model by Bordon and Fu (2015) where early specialization leads to positive peer effects through sorting (as students have a similar comparative advantage) but a higher chance of a mismatch. Using Chilean data they conduct a counterfactual experiment and show that moving to a regime with late specialization marginally increases welfare. All of these papers above do not consider the interaction between education and occupation choices.

Similarly, there is a literature that focuses on occupation choice but that does not consider education. In Lucas (1978), entrepreneurial ability affects the productivity of other workers in a firm. These spillovers result in more able individuals becoming entrepreneurs. In Kihlstrom and Laffont (1979), individuals who are less risk averse become entrepreneurs. Also taking an economy wide perspective, Banerjee and Newman (1993) analyze a model in which there is an interaction between the occupational choice of individuals and the distribution of wealth. Strikingly, they show that the initial distribution of wealth can have persistent effects on the occupational structure of an economy.

A key feature of our model is the separation of the education phase where there are no flow payoffs from the occupation phase. This distinguishes our paper from Jovanovic (1979) and Miller (1984) where there is a tradeoff between experimentation and exploitation. Our model is also tied to a literature on learning and job assignments (Meyer (1994) and Ortega (2001)). These papers, however assume that a firm can commit to a learning policy so that decisions do not depend on interim signals. Finally, the idea that workers get rewarded in a labor market that infers ability is close to the career concerns literature (Dewatripont, Jewitt, and Tirole (1999) and (Holmstrom (1982))). The difference in our paper is that ability has

[2]Theodore Schultz while analyzing higher education points out that "the much neglected activity is that of discovering talent." (Schultz (1968)). He also mentions that "there are many signs that indicate that one of the strongest features of the U.S. higher education is in discovering talent." Our result regarding the efficiency of experimentation confirms this conjecture.

[3]Malamud's model like ours has normally distributed talents and signals. A key difference though is that we allow sampling choices to depend on realizations of the signal whereas he does not.

multiple dimensions.

From an even broader perspective, there is now a large literature emphasizing the importance of a variety of institutions for development.[4] This literature is quite explicit about the role of institutions that help develop human capital and the links to growth.

Finally, when thinking about how to structure an educational system to provide appropriate incentives for occupational choice–as in our model–many of the considerations that arise are similar to those that firms face in structuring their organization. Chandler (1962), for instance, made the classic observation that a firms strategy and organizational structure are inextricably linked.[5] One can naturally think of an educational system as an organization that provides incentives for students to pursue a particular strategy. Like the organizational economics literature, our model has the property that it is difficult to balance two objective (experimentation and specialization) and that they system should be optimally designed to deliver on one objective.

The paper proceeds as follows. In section 2 we outline the model. Section 3 considers the first-best or efficient benchmark. Section 4, which is the heart of the paper, analyzes optimal sampling and occupational choice. Section 5 explores a number of natural and important extensions, while section 6 contains some concluding remarks.

2 The model

2.1 Environment and Production

There are two sectors: sector A and sector B. Associated with each sector is one risk neutral firm; this assumption builds labor market imperfections into our model. There is one risk neutral agent who chooses to work in either sector A or sector B. There are two ways an agent can work in a sector: she can either employ herself and be an *entrepreneur* or she can be an *employee* at a firm in the sector.

The agent's talent in sector A (B) is given by $\eta^A(\eta^B)$. This talent is unknown and is distributed normally with mean 0 and variance $\sigma_\eta^2 > 0$. Talents across sectors are independent of one another. Production depends on the talent of the agent in the sector. An agent who works in sector A (B) produces an output $\eta^A(\eta^B)$.

[4] See North and Thomas (1973) and North (1981) for important early contributions; and Acemoglu, Johnson, and Robinson (2002) who more recently pioneered causal understandings of the impact of institutions on growth. Relatedly, Laporta, de Silanes, Shleifer, and Vishny (1998) and Laporta, de Silanes, Shleifer, and Vishny (2000) demonstrate, *inter alia* that legal origins of institutions can have extremely long-run effects.

[5] Porter (1985) and Roberts (2004) develop this line of reasoning further, concluding that different tasks require different organizational designs.

2.2 Sampling Talents

Prior to working, the agent can *sample* (or learn about) her talents in a sector over two periods. An agent who samples sector i, $i = A, B$, in period t, $t = 1, 2$, draws an informative signal $s_t^i = \eta^i + \epsilon_t^i$ at the end of the period, where ϵ_t^i is an idiosyncratic error term, which is normally distributed with mean 0 and variance $\sigma_\epsilon^2 > 0$. The error terms are independent across periods.

The key constraint that the agent faces is that she can only sample one type of talent per period. If the agent samples the same type of talent over both periods, we say that she *specializes*. On the other hand, if the agent samples different types of talents over both periods, we say that she *experiments*.

For future reference, note that the posterior mean of the agent's talent in sector A at the end of period 1 is given by

$$E(\eta^A | s_1^A) = \lambda_1 s_1^A$$

where $\lambda_1 = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\epsilon^2}$.

For period 2, the posterior mean of the agent's talent in sector A if she specializes is given by

$$E(\eta^A | s_1^A, \hat{s}_2^A) = \lambda_1 s_1^A + \lambda_2 \hat{s}_2^A$$

where $\lambda_2 = \frac{\sigma_\eta^2}{2\sigma_\eta^2 + \sigma_\epsilon^2}$ and where $\hat{s}_2^A = s_2^A - \lambda_1 s_1^A$.

If the agent experiments on the other hand by sampling sector B (for the first time) in period 2, the posterior mean of the agent's talent in sector B is given by

$$E(\eta^B | s_2^B) = \lambda_1 s_2^B$$

Also for future reference, let F^A denote the distribution function for $\hat{s}_2^A$ given s_1^A and let F^B denote the distribution function for s_2^B.

2.3 Occupation Choice

The agent has to choose between being an entrepreneur or an employee. If she chooses entrepreneurship, she has to incur a fixed cost $K \geq 0$. The most natural way to interpret K is to think of it as the cost of starting a business. To start a business, an entrepreneur must raise capital, register her business, comply with regulations, and as in Lazear (2005) acquire skills needed to run a business. More broadly we can think of K as a proxy for the entrepreneurial environment in a country.

The main difference in our model between entrepreneurship and employment is in how returns from output accrue to the agent. In the case of entrepreneurship, the agent is the residual claimant: she keeps the entire surplus generated from her talent for herself. By contrast, in the employment case, the agent bargains over the surplus

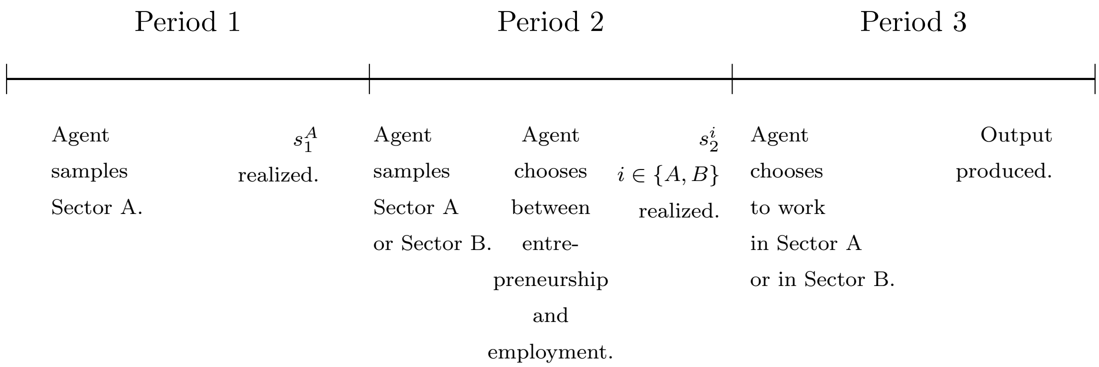


Figure 1: Timeline.

with her employer in a monopsonistic labor market. For this case, we assume that wages are determined by Nash Bargaining where $\mu \in [0, 1)$ is the agent's bargaining weight.

Throughout our paper, we assume that the agent chooses entrepreneurship if she is indifferent between being an entrepreneur and an employee.

2.4 Timing and Information Structure

There are three periods in the model: two sampling periods followed by a working period. We assume without loss of generality that the agent samples sector A in the first period. Thus for all of our analysis, we treat the realized signal in period 1 for talent A, s_1^A, as an exogenous parameter. The timing and information structure of the model then is as follows.

The agent samples sector A at the start of period 1. At the end of this period, she draws a publicly observable signal s_1^A. Conditional on this realized signal the agent decides which sector to sample at the start of the second period. In the middle of the second period, the agent chooses whether to be an entrepreneur or an employee.[6] And at the end of the second period, the signal s_2^i, where $i \in \{A, B\}$, is realized. At the start of period 3, the agent decides which sector to work in. Finally, at the end of period 3 production takes place. Figure 1 depicts the timing of the model.

3 Efficiency

There are two factors that affect efficiency. First, it is efficient for production to take place in a firm so that the agent does not have to incur the fixed cost K of starting

[6]This assumption is made to keep the model tractable. In our extensions, we allow this choice to be made at the start of the working phase and show that our results are qualitatively robust.

a business. Second, efficiency requires that the agent's sampling choice maximizes expected output. In this section, we compare the expected surplus (output) from specializing versus experimenting, given the realization of the first period signal s_1^A. We first sketch the total surplus functions associated with specialization and experimentation. We then compare the expected surplus across these two sampling strategies.

Consider the surplus functions associated with specialization and experimentation. Consider specialization first. Because the agent can pick which sector to work in after sampling talents, the surplus from specialization is given by

$$\begin{aligned} TS(A,A) &= \max\{E(\eta_A|s_1^A,\hat{s}_2^A), E(\eta_B|s_1^A,\hat{s}_2^A)\} \\ &= \max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}. \end{aligned}$$

Similarly, the surplus from experimentation is given by

$$\begin{aligned} TS(A,B) &= \max\{E(\eta_A|s_1^A,s_2^B), E(\eta_B|s_1^A,s_2^B)\} \\ &= \max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}. \end{aligned}$$

These surplus functions are plotted in Figure 2 (for a positive first period signal) and Figure 3 (for a negative first period signal): the surplus function for specialization (the blue graph) as a function of $\hat{s}_2^A$, and the surplus function for experimentation (the red graph) as a function of s_2^B.

The expected surplus from specialization, V_S is then given by

$$V_S = E_{\hat{s}_2^A}[TS(A,A)|s_1^A] = E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A]$$

and the expected surplus from experimentation V_E is given by

$$V_E = E_{s_2^B}[TS(A,B)] = E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}]$$

Looking at Figures 2 and 3, it is not clear which of the two sampling strategies yields a higher expected surplus. Notice that the surplus functions overlap. Also expectations are taken with respect to different random variables. Our main result in this section is that experimentation always yields a higher expected surplus relative to specialization.

But first we state a useful lemma.

Lemma 1 *Let x be a normally distributed random variable with mean 0. Let a be a positive real number and let c and d be real numbers. Then* $E_x[\max\{ax+c,d\}] = E_x[\max\{ax+d,c\}]$.

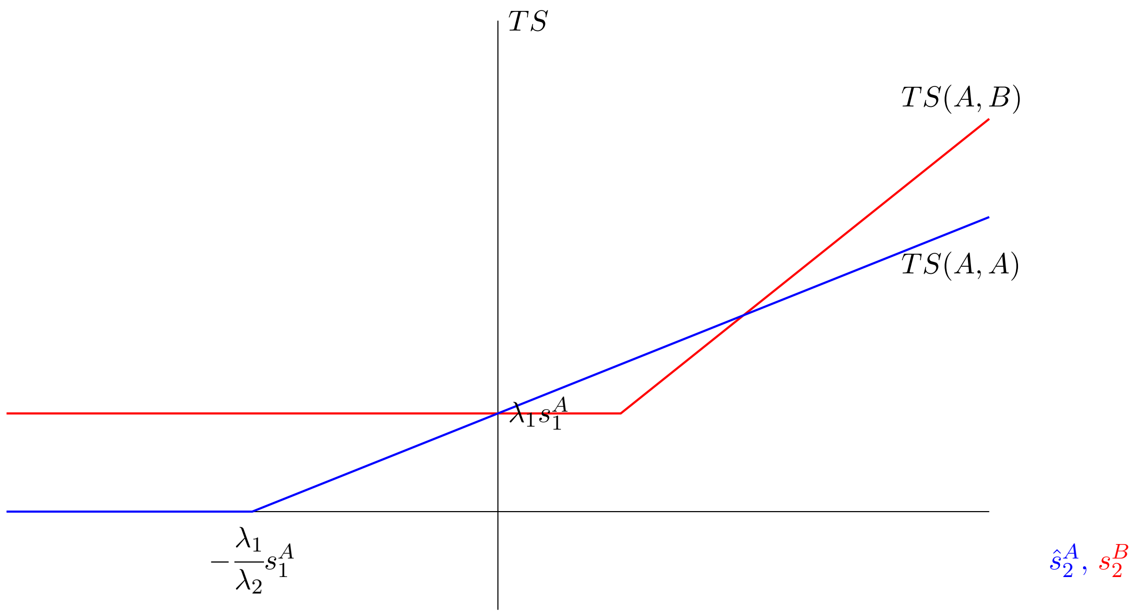


Figure 2: Total Surplus Functions when $s_1^A > 0$

The lemma above says that when a random variable is normally distributed with a mean of zero, then interchanging intercepts across components of the max function does not change the expected value of the max function. It is worth pointing out that the lemma above holds not just for a normal distribution but for any symmetric distribution with mean 0.

We now turn to our main result in this section.

Proposition 1 *Experimentation, where the agent samples different sectors in each period, is efficient.*

To understand the intuition for this result it helps to take a closer look at the surplus functions in the figures above. In particular, notice that there is an *upside* effect: a high signal in the second period increases the posterior mean of the sampled talent and thus increases surplus whereas a low signal entails no cost because the agent can switch to the non-sampled sector. It turns out that the upside effect is stronger in the case of experimentation for the following two reasons.

First, since the agent's talent in Sector B is sampled for the first time in the case of experimentation, the weight placed on this signal is larger relative to the weight placed on the signal in the specialization case ($\lambda_1 > \lambda_2$). This is because a signal drawn for the first time is more informative about talent.

Second, both the signals $\hat{s}_2^A$ and s_2^B have the same mean of 0, but the signal in

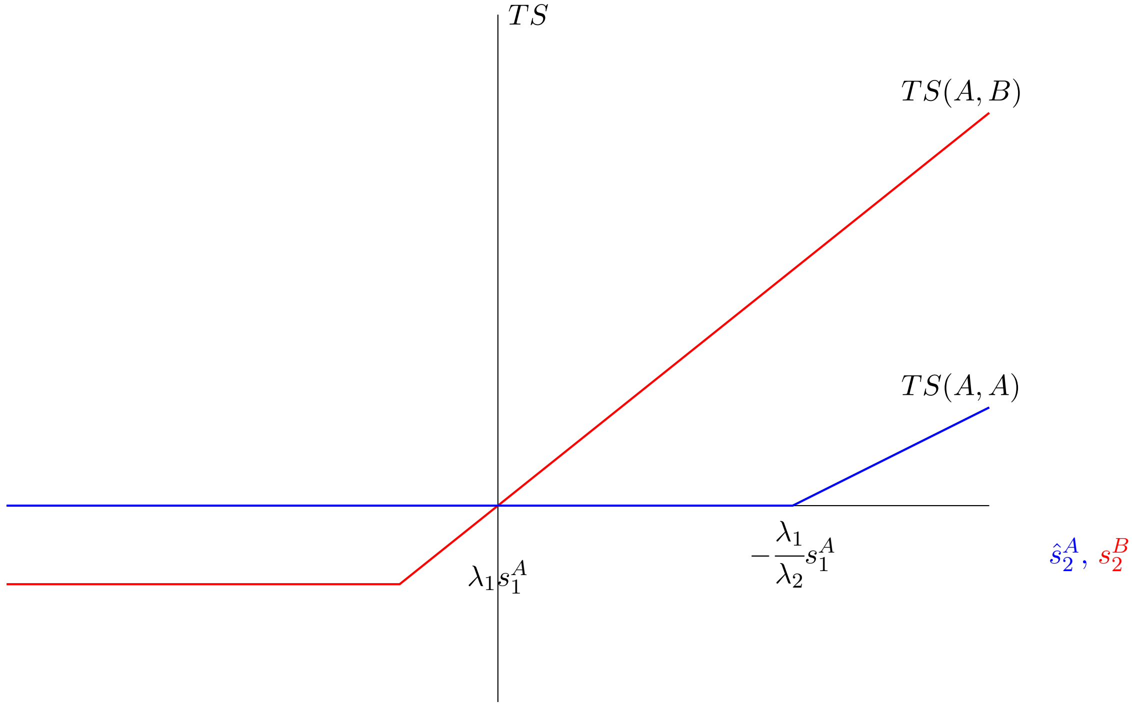


Figure 3: Total Surplus Functions when $s_1^A < 0$

sector B which is drawn for the first time has larger variance.[7] Or put differently the signal $\hat{s}_2^A$ second order stochastically dominates the signal s_2^B. This is because less is known about a talent which is sampled for the first time.

To summarize the intuition for the proposition, there is *more to learn* from experimentation: the weight placed on signal B when updating beliefs is stronger ($\lambda_1 > \lambda_2$) and extreme values of signal B are more likely ($\hat{s}_2^A$ second order stochastically dominates s_2^B). As a result, the upside effect is larger for experimentation. This larger upside effect combined with the symmetry of the normal distribution ensures that experimentation yields a higher expected surplus relative to specialization.

Given that experimentation does better than specialization always, we now look at how the difference in the expected surplus across both of these cases varies as we vary parameters in our model.

Proposition 2 *$V_E - V_S$ is*

i strictly increasing in σ_η^2

ii strictly decreasing in $|s_1^A|$.

The intuition for the first part of Proposition 2 is simple. As the variance of talents gets larger, there is more to learn from experimentation which makes it more valuable relative to specialization. To see why the second part of Proposition 2 holds, notice from Lemma 1 that we can rewrite $E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] = E_{\hat{s}_2^A}[\max\{\lambda_2 \hat{s}_2^A, \lambda_1 s_1^A\}|s_1^A]$. Also observe that the signals $\hat{s}_2^A$ and s_2^B have the same conditional mean. Thus, the only reason the surplus functions differ is because of the floors associated with the surplus function. When the floor is too low, the floor plays little or no role in the surplus function and since both signals have the same mean the expected surpluses are close. On the other hand, when the floor is too high, it is all that matters for surplus and once again expected surpluses across both cases are close to one another.

4 Optimal Sampling and Occupation Choice

We now turn to how the agent samples her talents and chooses her occupation in equilibrium. As mentioned earlier, there are two dimensions which affect efficiency. The first is the fixed cost K associated with starting a new business. The second dimension is the expected output from experimenting or specializing. The problem, however, is that each occupation can improve efficiency only along one of these dimensions. In this section, we explore how this tradeoff across both dimensions affects optimal occupation choice.

Given s_1^A, the agent has four choices in the second period: specialize and become an entrepreneur, experiment and become and entrepreneur, specialize and become

[7] The signal $\hat{s}_2^A \sim N(0, (1-\lambda_1^2)(\sigma_\eta^2 + \sigma_\epsilon^2))$, whereas the signal $s_2^B \sim N(0, \sigma_\eta^2 + \sigma_\epsilon^2)$.

an employee, and experiment and become an employee. The respective expected utilities associated with each of these options is given below.

$$EU^{ent}(AA) = E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] - K.$$

$$EU^{ent}(AB) = E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] - K.$$

$$EU^{emp}(AA) = E_{\hat{s}_2^A}[\mu(\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}) + (1-\mu)(\min\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\})|s_1^A].$$

$$EU^{emp}(AB) = E_{s_2^B}[\mu(\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}) + (1-\mu)(\min\{\lambda_1 s_1^A, \lambda_1 s_2^B\})].$$

The proposition below shows the agent's optimal sampling strategy when she is an entrepreneur and when she is an employee.

Proposition 3 *i An entrepreneur optimally experiments.*

ii An employee optimally specializes when $\mu < \frac{1}{2}$, and experiments otherwise.

Proposition 3 says that an entrepreneur is more likely to experiment with her education and thus will have a broader education profile when compared to an employee. This result is consistent with a recent theory of entrepreneurship put forward by Edward Lazear (Lazear (2004) and Lazear (2005)). But the source of breadth in our model is different. In our setting, entrepreneurs have breadth across unrelated areas whereas in Lazear's work, the breadth is across related areas.[8] This yields a distinct testable implication: those who choose entrepreneurship are more likely to have experimented across very different areas. A simple way to test this implication is to go back to Lazear's paper which looks at the graduate level education profile of Stanford business students and examine the undergraduate profile of this same group. The prediction from our paper would be that entrepreneurs also have a broader undergraduate education profile across unrelated fields.

The intuition for the first part of Proposition 3 is simple. It is efficient to experiment and since an entrepreneur is the residual claimant (she gets all of the surplus from her talent), it is optimal for her to choose experimentation

As for the second part of the proposition, to see why an employee with low bargaining power specializes, consider the extreme case where the agent has no bargaining power ($\mu = 0$). In this case, her wage equals her reservation wage, which is

[8] By unrelated we mean a setting where talents are independent of each other and where they do not interact in the entrepreneurs payoff function. More generally, our results continue to hold even if talents are correlated, though the gains from experimentation are smaller.

the lesser of her two talents. This leads to a *downside* effect: high signals in the second period yield no benefit (as wages are capped above by the non-sampled sector) while low signals entail a cost in terms of a lower reservation wage. Specialization in this case helps reduce this downside effect and is thus optimal.

The next proposition looks at the determinants of occupation choice.

Proposition 4 *In equilibrium, the agent is more likely to choose entrepreneurship when*

i K *is low.*

ii μ *is low.*

iii σ_η^2 *is high.*

Proposition 4 says that entrepreneurship is more likely when i) the cost of starting a business is low, ii) when labor markets offer the agent a larger share of the surplus (say because of competition in the labor market) so that being an employee is a closer substitute to that of being an entrepreneur and when iii) the gains from entrepreneurship are larger. The intuition itself for each of these points is simple. The value of the Proposition, however, lies in combining it with Proposition 3. Put together, these two propositions tell us that experimentation is optimal when there are easier opportunities for entrepreneurship. Or viewed more from an economy wide perspective, they tell us that countries with better entrepreneurial environments should have higher rates of entrepreneurship and an education system with experimentation.

The following scatter plot in Figure 4, which we construct using information from Ardagna and Lusardi (2010), suggests a pattern consistent with the results above. On the horizontal axis is a regulatory index ranging from zero to one, called ENTER which measures the barriers and costs entrepreneurs face when deciding to open a business. These barriers and costs include the average number of procedures officially required to start a business and the time taken for these procedures. The data for this index are drawn from the Doing Business Database (the World Bank) and the Index of Economic Freedom (the Heritage Foundation) across years in the late 1990's. On the vertical axis is a measure of entrepreneurial activity which is drawn from the micro survey data set, the Global Entrepreneurship Monitor (GEM) for the years 2001 and 2002. This data set collects cross national data on entrepreneurship by sampling at least 2000 randomly selected individuals (Adult Population Survey) and an average of 35 national experts per country. The measure of entrepreneurial activity called TEAOPP (Total Opportunity Entrepreneurial Activity) gives the percentage of individuals surveyed in a country, who are starting a business or are owner managers of young firms to take advantage of a business opportunity.

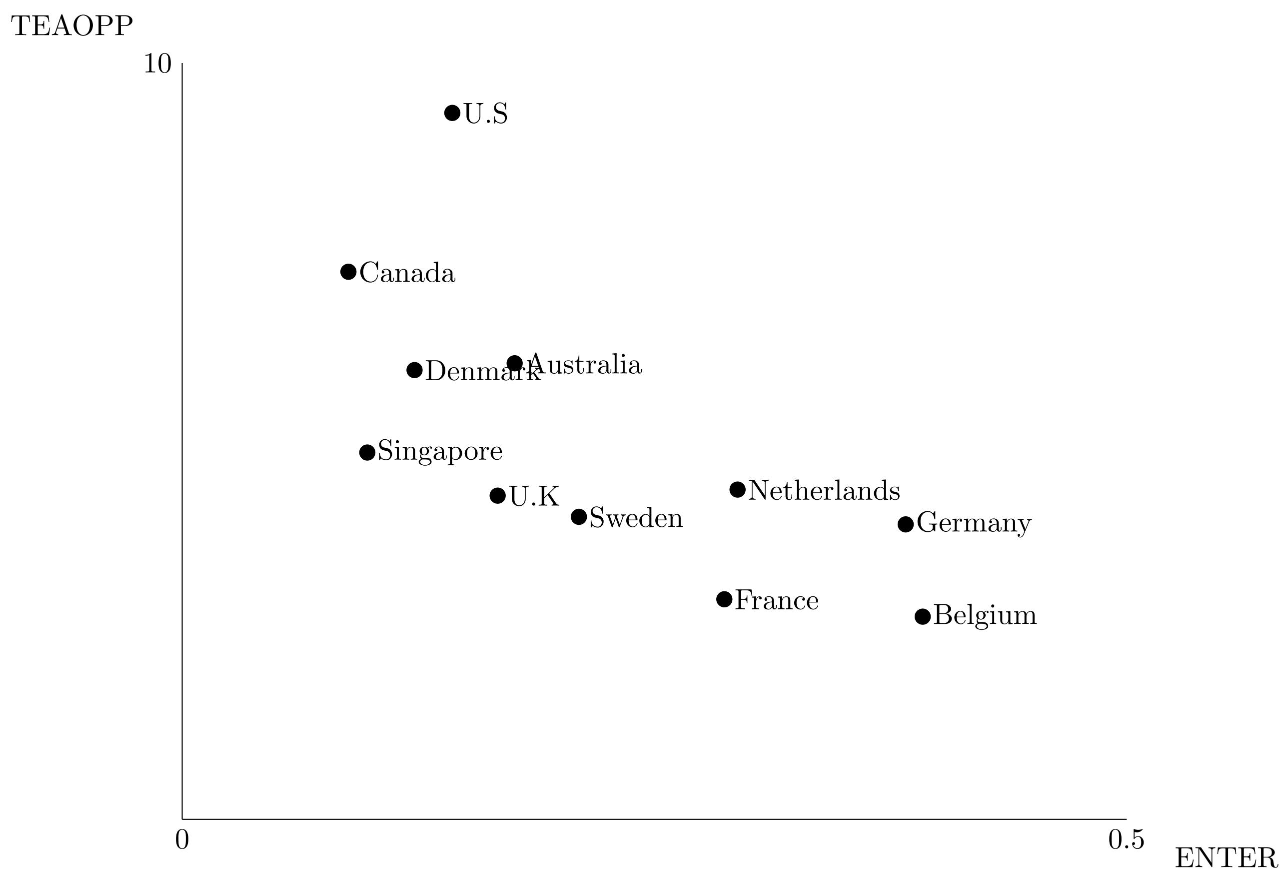


Figure 4: Regulatory constraints on starting a business and entrepreneurship rates across countries.

The scatter plot reflects a clear pattern. The U.S and Canada are both located on the upper left corner of the plot with low regulation, high entrepreneurship rates and a system with experimentation. France and Germany, on the other hand, along with some other countries in Europe like Belgium and the Netherlands, are located in the lower right corner with high regulation, low entrepreneurship rates and systems with specialization. Other countries like Australia, the U.K., Singapore and some Scandinavian countries lie in-between: these countries have fewer regulations but do not have the appropriate education systems to stimulate experimentation and entrepreneurship.

5 Extensions

5.1 Human Capital

We now introduce human capital into our analysis. When an agent samples a sector, she does not just get a signal of her talent she also acquires human capital $H > 0$. Output in each sector is the agent's talent plus her human capital. When human capital is general, we can rewrite the surplus functions as

$$\begin{aligned} TS^{General}(A,A) &= \max\{E(\eta_A|s_1^A, \hat{s}_2^A) + 2H, E(\eta_B|s_1^A, \hat{s}_2^A) + 2H\} \\ &= \max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A + 2H, 2H\}. \end{aligned}$$

$$\begin{aligned} TS^{General}(A,B) &= \max\{E(\eta_A|s_1^A, s_2^B) + 2H, E(\eta_B|s_1^A, s_2^B) + 2H\} \\ &= \max\{\lambda_1 s_1^A + 2H, \lambda_1 s_2^B + 2H\}. \end{aligned}$$

When human capital is specific to a sector, on the other hand, the surplus functions become

$$\begin{aligned} TS^{Specific}(A,A) &= \max\{E(\eta_A|s_1^A, \hat{s}_2^A) + 2H, E(\eta_B|s_1^A, \hat{s}_2^A)\} \\ &= \max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A + 2H, 0\}. \end{aligned}$$

$$\begin{aligned} TS^{Specific}(A,B) &= \max\{E(\eta_A|s_1^A, s_2^B) + H, E(\eta_B|s_1^A, s_2^B) + H\} \\ &= \max\{\lambda_1 s_1^A + H, \lambda_1 s_2^B + H\}. \end{aligned}$$

The following proposition characterizes the efficient sampling strategy with human capital.

Proposition 5 *i When human capital is general across sectors, experimentation is efficient.*

ii When human capital is specific to a sector, specialization is efficient for a sufficiently large first period signal s_1^A, and experimentation is efficient for a sufficiently small first period signal s_1^A.

With general human capital, nothing changes in our analysis: experimentation is still efficient regardless of the first period signal. But when human capital is specific to a sector, our main result in Proposition 1 changes. With specific human capital, when an agent gets a really good draw in sector A, then it is efficient for her to sample the same sector again. And when she gets a really bad draw in sector A, efficiency dictates that she should experiment instead. The intuition for the result is the following. Because human capital is sector specific, it is lost if the agent ends up working for a different sector. For a large first period signal in Sector A, the agent is more likely to work in Sector A and thus prefers to specialize. For a low first period signal on the other hand, the agent is more likely to work in Sector B and thus experiments instead.

5.2 Outside Option in the Labor Market

We now consider a setting where the agent has an outside option in the labor market where she is guaranteed a utility of $\underline{u}$ if she chooses not to work in either Sector A or Sector B. We assume that $\underline{u}$ is less than the average (ex ante expected) productivity in either sector, i.e., $\underline{u} < 0$.

In this subsection we see how the existence of this outside option changes incentives to sample talents and to choose occupations. To keep things simple, we work with the case where the agent has no bargaining power with a firm ($\mu = 0$).

With the outside option in the labor market, it is once again efficient for the agent to experiment. The outside option does not bind in the case of specialization and when it does bind for the case with experimentation (a low first period signal in sector A) it increases the expected surplus.

The following proposition looks at the optimal sampling strategy of an entrepreneur and an employee.

Proposition 6 *Consider the case where the agent has an outside option in the labor market with $\underline{u} < 0$ and let $\mu = 0$.*

i An entrepreneur optimally experiments.

ii An employee optimally experiments when $s_1^A > \dfrac{|\underline{u}|}{\lambda_1}$ and specializes when $s_1^A < \dfrac{\underline{u}}{\lambda_1}$.

With an outside option, an entrepreneur still has incentives to experiment. The difference from Proposition 3 is that for high realizations of the first period signal an employee also has incentives to experiment. To see why this is the case, note that the outside option mitigates the stronger downside effect from experimentation. This, combined with fact that a high first period signal raises the wage ceiling ensures that employees experiment.

For a low first period signal, however, an employee will still have incentives to inefficiently specialize. So overall, the existence of an outside option in the labor market leads to a more efficient outcome with respect to experimentation.

5.3 Asymmetric Model

So far in our model, sectors are symmetric: talents in both sectors have the same mean and the same variance. In this section, we allow for asymmetries across sectors.

Let $\eta^A \sim N(0, \sigma_\eta^2)$ and $\eta^B \sim N(b, v\sigma_\eta^2)$ where $v > 0$ and where b is any real number. Also let $\epsilon_t^A \sim N(0, \sigma_\epsilon^2)$ and $\epsilon_t^B \sim N(0, w\sigma_\epsilon^2)$ for $t = 1, 2$ with $w > 0$. Using this information structure we have that $E(\eta^B | s_2^B) = (1 - \lambda_1^B)b + \lambda_1^B s_2^B$, where $\lambda_1^B = \frac{v\sigma_\eta^2}{v\sigma_\eta^2 + w\sigma_\epsilon^2}$. The unconditional distribution of the signal s_2^B is normal with mean b and variance $v\sigma_\eta^2 + w\sigma_\epsilon^2$.

For the following proposition we assume that the agent samples Sector A first, and we give sufficient conditions under which experimentation is efficient.

Proposition 7 *Let $1 - \lambda_1^2 < v\lambda_1 + w(1 - \lambda_1) < v(1 + \lambda_1)$. Then, experimentation, where the worker samples different sectors in each period, is efficient.*

Notice first in Proposition 7, that the parameter b plays no role in the sufficient conditions. Thus our result that experimentation is efficient, holds regardless of the difference in means across talents. What our result does depend on is the variances of talents and the variances of the error terms. The first condition in the proposition, $1 - \lambda_1^2 < v\lambda_1 + w(1 - \lambda_1)$, ensures that signal B has larger variance than signal A. The second condition, $v\lambda_1 + w(1-\lambda_1) < v(1+\lambda_1)$, ensures that the weight placed on signal B while updating the mean is higher. These two conditions, combined with the symmetry of the normal distribution ensure that experimentation, once again, is efficient.

Next, we relax the assumption that the agent exogenously samples Sector A in the first period. We restrict our attention to the case where $v > 1$ and $w = 1$. Thus, talent has a larger prior variance in Sector B and signals are equally noisy across sectors. For this case we have a very simple condition under which experimentation is efficient.

Proposition 8 *Let $v > 1$, $w = 1$, and let the agent choose which sector to sample in period 1. Then experimentation is efficient if and only $\frac{\sigma_\eta^2}{\sigma_\epsilon^2} \geq \frac{v-1}{v}$.*

The proposition above offers a simple condition that is both necessary and sufficient for experimentation to be efficient. The left hand side of the condition $\frac{\sigma_\eta^2}{\sigma_\epsilon^2}$ is simply the signal to noise ratio, whereas the right hand side of the condition, $\frac{v-1}{v}$ measures the degree to which variances across sectors are asymmetric. The proposition then says that as long as the signal to noise ratio is at least as large as the degree of asymmetry in the variances, then experimentation is efficient, regardless of the signal drawn in the first period. Notice, that a sufficient condition for experimentation is that the signal to noise ratio is at least as large as one.

5.4 Alternative Timing

Now let us assume that the agent can take the decision to become an entrepreneur or an employee after having seen the realization of the second signal. Figure 4 depicts the alternative timing of the model.

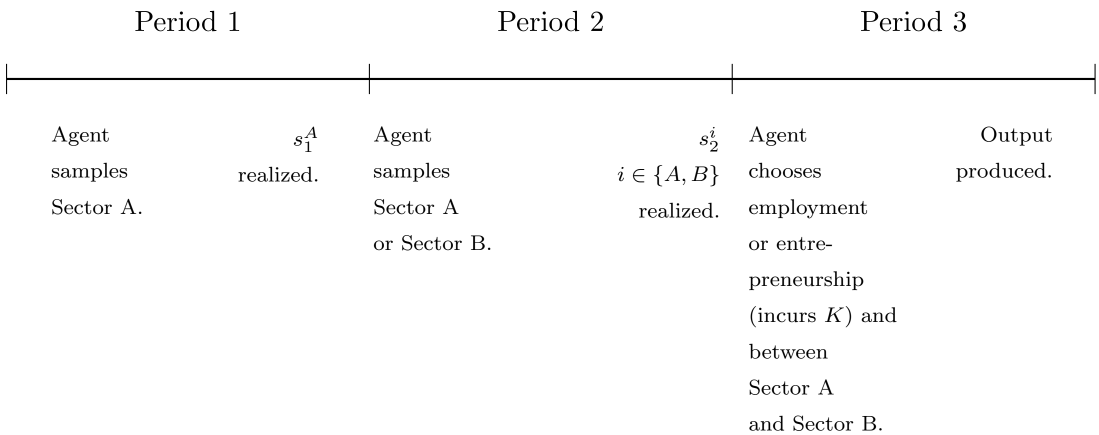


Figure 5: Alternative Timeline.

If the agent chooses to specialize, her utility (ex post) is

$$U(AA) = \begin{cases} \mu \max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\} + (1-\mu) \min\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\} & \text{if } K > \overline{K}_{spec}, \\ \max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\} - K & \text{otherwise,} \end{cases} \tag{1}$$

where $\overline{K}_{spec} = (1-\mu)[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\} - \min\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}]$.

If the agent chooses to experiment, her utility (ex post) is

$$U(AB) = \begin{cases} \mu \max\{\lambda_1 s_1^A, \lambda_1 s_2^B\} + (1-\mu) \min\{\lambda_1 s_1^A, \lambda_1 s_2^B\} & \text{if } K > \overline{K}_{exp}, \\ \max\{\lambda_1 s_1^A, \lambda_1 s_2^B\} - K & \text{otherwise,} \end{cases} \tag{2}$$

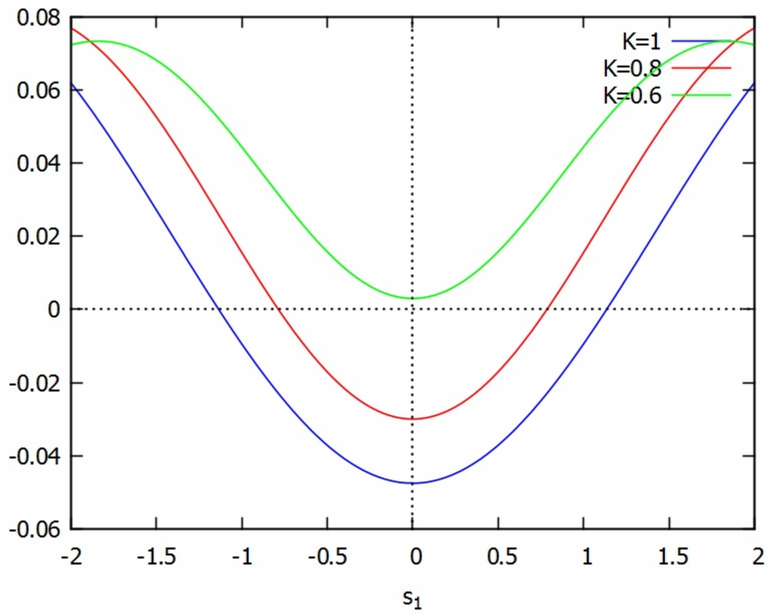
where $\overline{K}_{exp} = (1-\mu)[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\} - \min\{\lambda_1 s_1^A, \lambda_1 s_2^B\}]$.

By comparing the expected utilities $\int U(AA) dF^A$, and $\int U(AB) dF^B$ we see whether the agent will specialize or experiment. If K is very low, the agent will choose to become an entrepreneur and will get the entire surplus. Thus, by Proposition 1 it is best for the agent to experiment. Similarly, when $\mu \to 1$, the agent gets the entire surplus even when employed, and it is therefore optimal to experiment.

However, when the agent's bargaining weight is low, specialization can be optimal for the agent when the costs of becoming an entrepreneur are high. This is illustrated in Figure 6. Figure 6 shows the expected utilities for experimentation and specialization assuming that $\lambda_1 = 0.5, \lambda_2 = 0.33, \mu = 0.25$. The different curves correspond to different levels of the cost of entrepreneurship ($K = 1, K = 0.8$ and $K = 0.6$).

When the agent's first signal is close to zero, her talents across the two sectors will not differ by much in expected terms. As a result, the agent does not gain much from choosing entrepreneurship. If the cost of becoming an entrepreneur is large

Figure 6: The difference in expected utilities from experimentation over specialization for various levels of K.



such an agent expects to choose employment. In this case the agent is better off when she specializes. However, the interval of first period signals for which specialization is optimal shrinks as K decreases from 1 to 0.8. For $K = 0.6$, specialization is never optimal.

5.5 How Sensitive is the Result to the Normal-Normal Model?

The fact that the result that experimentation is more efficient than specialization is independent of the realization of the signal drawn in the first period is surprising. We conjecture that the independence on the first period signal is specific to the normal-normal model and, more specifically, to the property that the variance of the updated normal distribution is independent of the first period signal.

To explore this conjecture we analyze a slightly more general information structure. We assume that the agent's talent, η^i, in sector $i = \{A, B\}$ follows a Student t-distribution with $\nu > 2$ degrees of freedom, a mean of zero and scale parameter of $\frac{\nu-2}{\nu}\sigma_\eta^2$, i.e.,

$$\eta^i \sim t_\nu(0, \tfrac{\nu-2}{\nu}\sigma_\eta^2).$$

As before, conditional on η^i, signals are normally distributed with mean η^i and variance σ_ϵ^2. When the prior distributions for the agent's talents follow a t-distribution and signals are normally distributed, the posterior distributions for the worker's talents are also t-distributions DeGroot (1970). The posterior means of the agent's talents are the same as for the normal-normal model.

Similarly, the unconditional distribution of the first signal and the conditional distribution of the second signal given the first signal follow Student t-distributions. More specifically,

$$\begin{aligned} s_2^B &\sim t_\nu(0, \tfrac{\nu-2}{\nu}(\sigma_\eta^2 + \sigma_\epsilon^2)) \\ &\text{and} \\ \hat{s}_2^A | s_1^A &\sim t_{\nu+1}(0, (\tfrac{\nu-2}{\nu} + \tfrac{1}{\nu+1}\tfrac{(s_1^A)^2}{\sigma_\eta^2+\sigma_\epsilon^2})(1-\lambda_1^2)(\sigma_\eta^2 + \sigma_\epsilon^2)) \end{aligned}$$

These posterior distributions are very similar to the ones obtained in the normal-normal model. Posterior means are identical and as $\nu \to \infty$, the variances and distributions converge to the normal-normal model.

The crucial difference to the normal-normal model is that the posterior variance of the second signal in sector A depends on the first signal. The greater the magnitude of the first signal, the greater the posterior variance of the second signal. If s_1^A is very high or very low, the posterior variance of the second signal from sector A can get larger than the unconditional variance of the signal from sector B. In this case it can be efficient to sample from sector A again and, thus, specialize.

This is illustrated in Figure 7, which shows the difference in expected surplus from experimentation and specialization. Here, $\sigma_\eta^2 = \sigma_\epsilon^2 = 1, \lambda_1 = 0.5, \lambda = 0.33$ and $\nu = 3$.

Figure 7: The difference in expected surplus from experimentation over specialization $(V_E - V_S)$

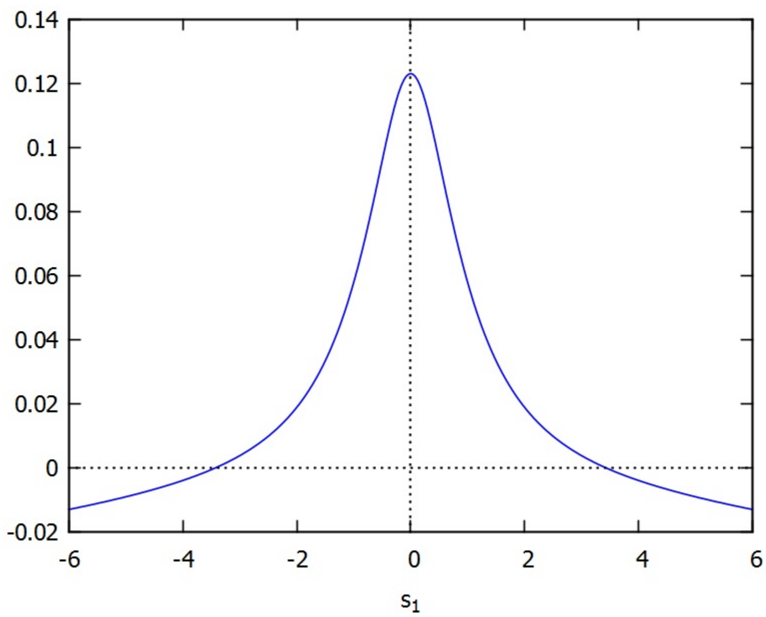


Figure 7 confirms our conjecture: The result that experimentation is more efficient than specialization *for all* realizations of the first signal relies on the normal-normal model. In Figure 7 specialization is more efficient when the first signal is below -3.4 or above 3.4. Note that for the parameter values used in Figure 7, the signal s_1^A follows a t-distribution with 3 degrees of freedom, mean zero and standard deviation of $\sqrt{2}$. Given this distribution, the probability of choosing a signal below -3.4 or above 3.4 is less than 10%. Thus, for the vast majority of realized signals, experimentation is still efficient.

5.6 Option not to Sample Talent

So far in our analysis, we have assumed that the agent has to sample talents in both periods. In this subsection we relax this assumption and give the agent the option of not sampling her talent in the second period. We once again focus on the case where firms have all of the bargaining power with $\mu = 0$.

If the agent chooses entrepreneurship, her expected utility from experimenting, $V_E - K$, exceeds her expected utility from specialization, $V_S - K$, (from Proposition 1) and her expected utility from not sampling in the second period which is $\max\{\lambda_1 s_1^A, 0\} - K$ (since the max function is convex). If the agent chooses employment, on the other hand, her expected utility from not sampling which is $\min\{\lambda_1 s_1^A, 0\}$ dominates her expected utility from specialization, $E_{\hat{s}_2^A}[\min\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A]$, and her expected utility from experimentation, $E_{s_2^B}[\min\{\lambda_1 s_1^A, \lambda_1 s_2^B\}]$.

Thus the agent will sample her talent in the second period if and only if

$$V_E - K \geq \min\{\lambda_1 s_1^A, 0\}$$

Since $V_E > \min\{\lambda_1 s_1^A, 0\}$ there is a threshold level of K below which the agent chooses to sample her talent.

6 Conclusion

We develop a model of education and occupation choice. In the education phase, a person learns about her talents and in the occupation phase she uses these talents in production as an entrepreneur or an employee. We suggest a new way to distinguish between an entrepreneur and an employee: the entrepreneur is a residual claimant whereas an employee bargains over her wage. This simple structure yields the following insights. First, experimentation is efficient for a surprisingly broad range of parameters. Second, our theory presents a view of an entrepreneur as an experimenter. Though this implies that an entrepreneur has a broader education profile, the reasons for this breadth are different from the existing literature. Finally, we show that the optimality of experimentation as an education system depends on opportunities for entrepreneurship in the economy. This possibly explains why the

U.S. and Canada have high rates of entrepreneurship along with a system with experimentation whereas Germany and France have lower rates of entrepreneurship along with a system of specialization.

Appendix

Proof of Lemma 1:

Let F denote the distribution function of x. Since the normal distribution is symmetrical around zero, $F(x) = 1 - F(-x)$. Then

$$\begin{aligned} E_x[\max\{ax+c,d\}] - E_x[\max\{ax+d,c\}] &= F(\tfrac{d-c}{a})\, d + a \int_{\frac{d-c}{a}}^{\infty} x dF + (1 - F(\tfrac{d-c}{a}))\, c \\ &- F(\tfrac{c-d}{a}) c - a \int_{\frac{c-d}{a}}^{\infty} x dF - (1 - F(\tfrac{c-d}{a}))d \\ &= a \int_{\frac{d-c}{a}}^{\infty} x dF - a \int_{\frac{c-d}{a}}^{\infty} x dF \\ &= 0\,, \end{aligned}$$

where the last step again follows from the symmetry of the normal distribution.■

Proof of Proposition 1: We split the proof into three claims.

Claim 1: $E_{\hat{s}_2^A}[TS(A,A)|s_1^A] = E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] \leq E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 s_2^B, 0\}]$.

Proof The distribution of signal $\hat{s}_2^A$ given s_1^A is $N(0,(1-\lambda_1^2)(\sigma_\eta^2 + \sigma_\epsilon^2))$. The distribution of signal s_2^B is $N(0,\sigma_\eta^2 + \sigma_\epsilon^2)$. Therefore the two random variables $\hat{s}_2^A$ and s_2^B have the same mean but the former has smaller variance than the latter. Thus $\hat{s}_2^A$ second order stochastically dominates s_2^B. Since the max function is convex, $E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}] \leq E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 s_2^B, 0\}]$.■

Claim 2: $E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 s_2^B, 0\}] < E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_1 s_2^B, 0\}]$.

Proof Consider two possible cases.

First, suppose $s_1^A \leq 0$. Then $\max\{\lambda_1 s_1^A + \lambda_2 s_2^B, 0\} \leq \max\{\lambda_1 s_1^A + \lambda_1 s_2^B, 0\}$ with the inequality strict for s_2^B sufficiently large. Thus $E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 s_2^B, 0\}] < E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_1 s_2^B, 0\}]$.

Second, suppose $s_1^A > 0$. Then $\max\{\lambda_1 s_1^A, \lambda_2 s_2^B\} \leq \max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}$ with the inequality strict for s_2^B sufficiently large. From Lemma 1, it follows that $E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 s_2^B, 0\}] = E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_2 s_2^B\}] < E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] = E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_1 s_2^B, 0\}]$.■

Claim 3: $E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_1 s_2^B, 0\}] = E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] = E_{s_2^B}[TS(A,B)]$.

Proof This claim follows from Lemma 1.■

Taking all three claims together, the result holds.■

Proof of Proposition 2: We first prove that

$$V_E - V_S = \int_{\frac{|s_1^A|}{\sigma_B}}^{\frac{\lambda_1|s_1^A|}{\lambda_2\sigma_A}} (\lambda_1\sigma_B z - \lambda_1|s_1^A|) f_z dz + \int_{\frac{\lambda_1|s_1^A|}{\lambda_2\sigma_A}}^{\infty} (\lambda_1\sigma_B z - \lambda_2\sigma_A z) f_z dz \,,$$

where z is distributed normally with mean 0 and variance 1, σ_A is the standard deviation of the random variable $\hat{s}_2^A$, and σ_B is the standard deviation of the random variable s_2^B.

Consider two cases. Suppose $s_1^A \geq 0$. Then

$$\begin{aligned}
V_E - V_S &= E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] - E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] \\
&= E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] - E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A, \lambda_2 \hat{s}_2^A\}|s_1^A] \\
&= E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\} - \lambda_1 s_1^A] - E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A, \lambda_2 \hat{s}_2^A\} - \lambda_1 s_1^A|s_1^A] \\
&= \int_{s_1^A}^{\frac{\lambda_1 s_1^A}{\lambda_2}} (\lambda_1 s_2^B - \lambda_1 s_1^A) dF^B + \int_{\frac{\lambda_1 s_1^A}{\lambda_2}}^{\infty} (\lambda_1 s_2^B - \lambda_1 s_1^A) dF^B - \int_{\frac{\lambda_1 s_1^A}{\lambda_2}}^{\infty} (\lambda_2 \hat{s}_2^A - \lambda_1 s_1^A) dF^A \\
&= \int_{\frac{|s_1^A|}{\sigma_B}}^{\frac{\lambda_1|s_1^A|}{\lambda_2\sigma_A}} (\lambda_1\sigma_B z - \lambda_1|s_1^A|) f_z dz + \int_{\frac{\lambda_1|s_1^A|}{\lambda_2\sigma_A}}^{\infty} (\lambda_1\sigma_B z - \lambda_2\sigma_A z) f_z dz \,.
\end{aligned}$$

Next, suppose $s_1^A < 0$.

$$\begin{aligned}
V_E - V_S &= E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] - E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] \\
&= E_{s_2^B}[\max\{\lambda_1 s_1^A + \lambda_1 s_2^B, 0\}] - E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] \\
&= \int_{-s_1^A}^{-\frac{\lambda_1 s_1^A}{\lambda_2}} (\lambda_1 s_2^B + \lambda_1 s_1^A) dF^B + \int_{-\frac{\lambda_1 s_1^A}{\lambda_2}}^{\infty} (\lambda_1 s_2^B + \lambda_1 s_1^A) dF^B - \int_{-\frac{\lambda_1 s_1^A}{\lambda_2}}^{\infty} (\lambda_2 \hat{s}_2^A + \lambda_1 s_1^A) dF^A \\
&= \int_{\frac{|s_1^A|}{\sigma_B}}^{\frac{\lambda_1|s_1^A|}{\lambda_2\sigma_A}} (\lambda_1\sigma_B z - \lambda_1|s_1^A|) f_z dz + \int_{\frac{\lambda_1|s_1^A|}{\lambda_2\sigma_A}}^{\infty} (\lambda_1\sigma_B z - \lambda_2\sigma_A z) f_z dz \,.
\end{aligned}$$

Next consider the comparative static results with respect to σ_η^2 and $|s_1^A|$ respectively.

i

$$\begin{aligned}\frac{\partial V_E - V_S}{\partial \sigma_\eta^2} &= \frac{\partial \frac{\lambda_1 |s_1^A|}{\lambda_2 \sigma_A}}{\partial \sigma_\eta^2}\left(\frac{\lambda_1 \sigma_B - \lambda_2 \sigma_A}{\lambda_2 \sigma_A}\right) - \frac{\partial \frac{|s_1^A|}{\sigma_B}}{\partial \sigma_\eta^2}(0) \\ &- \frac{\partial \frac{\lambda_1 |s_1^A|}{\lambda_2 \sigma_A}}{\partial \sigma_\eta^2}\left(\frac{\lambda_1 \sigma_B - \lambda_2 \sigma_A}{\lambda_2 \sigma_A}\right) \\ &+ \int_{\frac{\lambda_1 |s_1^A|}{\lambda_2 \sigma_A}}^{\infty} \left(\lambda_1 \frac{\partial \sigma_B}{\partial \sigma_\eta^2} + \frac{\partial \lambda_1}{\partial \sigma_\eta^2}\sigma_B - \lambda_2 \frac{\partial \sigma_A}{\partial \sigma_\eta^2} - \frac{\partial \lambda_2}{\partial \sigma_\eta^2}\sigma_A\right) z f_z dz\,.\end{aligned}$$

Notice that the first two lines in the expression above cancel each other out. Also, since $\lambda_1 > \lambda_2$, $\sigma_B > \sigma_A$, $\frac{\partial \lambda_1}{\partial \sigma_\eta^2} > \frac{\partial \lambda_2}{\partial \sigma_\eta^2}$ and $\frac{\partial \sigma_B}{\partial \sigma_\eta^2} > \frac{\partial \sigma_A}{\partial \sigma_\eta^2}$, it follows that $\frac{\partial V_E - V_S}{\partial \sigma_\eta^2} > 0$.

ii Note that we can rewrite

$$V_E - V_S = (\lambda_1 \sigma_A - \lambda_2 \sigma_B)\left(\frac{1}{\sqrt{2\pi}} - \int_0^{\frac{|s_1^A|}{\sigma_B}} z f_z dz\right) - \int_{\frac{|s_1^A|}{\sigma_B}}^{\frac{\lambda_1 |s_1^A|}{\lambda_2 \sigma_A}} (\lambda_2 \sigma_A z - \lambda_1 |s_1^A|) f_z dz.$$

$$\frac{\partial V_E - V_S}{\partial |s_1^A|} = -(\lambda_1 \sigma_A - \lambda_2 \sigma_B)\frac{|s_1^A|}{\sigma_B^2} + |s_1^A|\left(\frac{\lambda_2 \sigma_A}{\sigma_B} - \lambda_1\right) - \int_{\frac{|s_1^A|}{\sigma_B}}^{\frac{\lambda_1 |s_1^A|}{\lambda_2 \sigma_A}} \lambda_1 f_z dz.$$

Notice that $\dfrac{\lambda_2 \sigma_A}{\sigma_B} < \dfrac{\lambda_2 \sigma_A}{\sigma_A} < \lambda_1$. Thus $\dfrac{\partial V_E - V_S}{\partial |s_1^A|} < 0$.

Proof of Proposition 3:

i In the entrepreneurship case, $EU^{ent}(AB) - EU^{ent}(AA) = V_E - V_S$, which from Proposition 1 is strictly positive.■

ii We first prove that

$$E_{\hat{s}_2^A}[\min\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\} | s_1^A] = -V_S(-s_1^A) \tag{1}$$

and

$$E_{s_2^B}[\min\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] = -V_E(-s_1^A) \tag{2}$$

where $V_S(-s_1^A)$ is the expected surplus from specialization for a first period signal of $-s_1^A$, and where $V_E(-s_1^A)$ is the expected surplus from experimentation for a first period signal of $-s_1^A$.

Consider equation (1) first. Notice that

$$\begin{aligned} E_{\hat{s}_2^A}[\min\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] &= \int_{-\infty}^{-\frac{\lambda_1}{\lambda_2}s_1^A} (\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A) dF^A + (1 - F^A(-\frac{\lambda_1}{\lambda_2}s_1^A))\, 0 \\ &= -\int_{-\frac{\lambda_1}{\lambda_2}(-s_1^A)}^{\infty} (\lambda_1(-s_1^A) + \lambda_2 \hat{s}_2^A) dF^A \\ &= -E_{\hat{s}_2^A}[\max\{\lambda_1(-s_1^A) + \lambda_2 \hat{s}_2^A, 0\}|(-s_1^A)] \\ &= -V_S(-s_1^A)\,, \end{aligned}$$

where the second line follows from the fact that $\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A = -(\lambda_1(-s_1^A) + \lambda_2(-\hat{s_2^A}))$ and from the symmetry of the normal distribution.

Similarly, for equation (2),

$$\begin{aligned} E_{s_2^B}[\min\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] &= \int_{-\infty}^{s_1^A} \lambda_1 s_2^B dF^B + (1 - F^B(s_1^A))\, \lambda_1 s_1^A \\ &= -\int_{(-s_1^A)}^{\infty} \lambda_1 s_2^B dF^B - F^B(-s_1^A)\lambda_1(-s_1^A) \\ &= -E_{s_2^B}[\max\{\lambda_1(-s_1^A), \lambda_1 s_2^B\}] \\ &= -V_E(-s_1^A)\,, \end{aligned}$$

where the second line follows from the fact that $\lambda_1 s_2^B = -\lambda_1(-s_2^B)$ and from the symmetry of the normal distribution.

Using (1) and (2), we can rewrite

$$\begin{aligned} EU^{emp}(AB) - EU^{emp}(AA) &= \mu(E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] - E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A]) \\ &+ (1-\mu)(E_{s_2^B}[\min\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] - E_{\hat{s}_2^A}[\min\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A]) \\ &= \mu(V_E - V_S) - (1-\mu)(V_E(-s_1^A) - V_S(-s_1^A)) \\ &= (2\mu - 1)(V_E - V_S)\,, \end{aligned}$$

where the last line follows from the fact that $V_E - V_S$ only depends on the absolute value of the first period signal $|s_1^A|$ (from Proposition 2).

Since $V_E - V_S > 0$ (from Proposition 1), it follows that $EU^{emp}(AB) - EU^{emp}(AA) \geq 0$ if and only if $\mu \geq \frac{1}{2}$. ■

Proof of Proposition 4: Consider two cases. First, suppose $\mu \geq \frac{1}{2}$. Then the agent chooses entrepreneurship over employment if and only if

$$E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] - K \geq E_{s_2^B}\mu[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] + (1-\mu)E_{s_2^B}[\min\{\lambda_1 s_1^A, \lambda_1 s_2^B\}]$$

Rearranging the inequality above, we get

$$\lambda_1 \left[E_{s_2^B}[\max\{s_1^A, s_2^B\}] - E_{s_2^B}[\min\{s_1^A, s_2^B\}] \right] \geq \frac{K}{(1-\mu)}. \tag{3}$$

Observe that the right hand side of (3) is strictly increasing in μ and strictly increasing in K. Thus entrepreneurship is more likely when μ and K are low.

To see how the parameter σ_η^2 effects the entrepreneurship decision, notice that the left-hand side of (3) is the product of two terms, λ_1 and the term in square brackets. Clearly, λ_1 is increasing in σ_η^2. By second order stochastic dominance, the expectation of the max function increases in the variance of the random variable and the expectation of the min function decreases. Thus, the term in square brackets is increasing in σ_η^2.

It follows that the left hand side is strictly increasing in σ_η^2 and so entrepreneurship is more likely when σ_η^2 is high.

Next, consider the second case where $\mu < \frac{1}{2}$. Using the fact that $E_{\hat{s}_2^A}[\min\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] = -V_S(-s_1^A)$ (from equation (1) in the proof of Proposition 3), it follows that the agent chooses entrepreneurship over employment if and only if

$$(V_E(s_1^A) - V_S(s_1^A)) + (1-\mu)(V_S(s_1^A) + V_S(-s_1^A)) \geq K \tag{4}$$

Observe that the left hand side of (4) is strictly decreasing in μ and the right hand side of (4) is strictly increasing in K. Thus entrepreneurship is more likely when μ and K are low.

Since $V_E - V_S$ is strictly increasing in σ_η^2 (from Proposition 2) and since V_S is strictly increasing in σ_η^2 (by second order stochastic dominance) it follows that the left hand side of the inequality above is strictly increasing in σ_η^2. Thus entrepreneurship is more likely when σ_η^2 is high.

■

Proof of Proposition 5:

i Suppose human capital is general. Then

$$E_{s_2^B}[TS^{General}(A,B)] - E_{\hat{s}_2^A}[TS^{General}(A,A)|s_1^A] = V_E - V_S.$$

From Proposition 1, it follows that experimentation is efficient.

ii Suppose human capital is specific to a sector.

We can write $E_{s_2^B}[TS^{Specific}(A,B)] = V_E + H$. Similarly, we can write

$$E_{\hat{s}_2^A}[TS^{Specific}(A,A)|s_1^A] = V_S + H + g(s_1^A)\,,$$

where $g(s_1^A) = \int_{\frac{-(\lambda_1 s_1^A+2H)}{\lambda_2}}^{0} 2H dF^A + \int_{\frac{-(\lambda_1 s_1^A+2H)}{\lambda_2}}^{\frac{-\lambda_1 s_1^A}{\lambda_2}} (\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A) dF^A$.

Thus the expected gain in surplus from experimenting over specializing is given by $V_E - V_S - g(s_1^A)$. In the limit as s_1^A tends to infinity, $V_E - V_S - g(s_1^A)$ tends to $-H$ and as s_1^A tends to minus infinity, $V_E - V_S - g(s_1^A)$ tends to H. Furthermore,

$$g'(s_1^A) = \frac{2H\lambda_1}{\lambda_2} f^A\left(\frac{-(\lambda_1 s_1^A+2H)}{\lambda_2}\right) + \lambda_1 \int_{\frac{-(\lambda_1 s_1^A+2H)}{\lambda_2}}^{\frac{-\lambda_1 s_1^A}{\lambda_2}} dF^A - \frac{2H\lambda_1}{\lambda_2} f^A\left(\frac{-(\lambda s_1^A+2H)}{\lambda_2}\right) > 0\,.$$

Thus there is a threshold level of the first period signal, above which it is efficient to specialize. And since $V_E - V_S > 0$, there is a threshold level of the first period signal below which it is efficient to experiment.

Proof of Proposition 6:

i

$$\begin{aligned} EU_{\underline{u}}^{ent}(AA) &= E_{\hat{s}_2^A}[\max\{\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}, \underline{u}\}|s_1^A] \\ &= E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, 0\}|s_1^A] = EU^{ent}(AA). \end{aligned}$$

$$\begin{aligned} EU_{\underline{u}}^{ent}(AB) &= E_{s_2^B}[\max\{\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}, \underline{u}\}|s_1^A] \\ &\geq \max\{E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}], \underline{u}\} \\ &\geq E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 s_2^B\}] = EU^{ent}(AB). \end{aligned}$$

Thus it is optimal for the agent to experiment when she is an entrepreneur.

ii When the agent takes up employment and specializes in the case with a wage floor, her payoff for any realization of $\hat{s}_2^A$ is bounded below by the outside option $\underline{u}$ and is bounded above by the mean of the non-sampled sector 0. Also for $\hat{s}_2^A$ sufficiently small, the agent's payoff is $\underline{u}$ and for $\hat{s}_2^A$ sufficiently large, the agent's payoff is 0. Thus $EU_{\underline{u}}^{emp}(AA) \in (\underline{u}, 0)$.

On the other hand, when the agent takes up employment and experiments, her expected utility when $s_1^A > \dfrac{|\underline{u}|}{\lambda_1}$, is

$$
\begin{aligned}
EU_{\underline{u}}^{emp}(AB) &= \int_{-\infty}^{\frac{\underline{u}}{\lambda_1}} \underline{u} dF^B + \int_{\frac{\underline{u}}{\lambda_1}}^{\frac{|\underline{u}|}{\lambda_1}} \lambda_1 s_2^B dF^B + \int_{\frac{|\underline{u}|}{\lambda_1}}^{\infty} \min\{\lambda_1 s_1^A, \lambda_1 s_2^B\} dF^B \\
&= \int_{-\infty}^{\frac{\underline{u}}{\lambda_1}} \underline{u} dF^B + \int_{\frac{\underline{u}}{\lambda_1}}^{\frac{|\underline{u}|}{\lambda_1}} \lambda_1 s_2^B dF^B + \int_{\frac{|\underline{u}|}{\lambda_1}}^{\infty} |\underline{u}| dF^B \\
&+ \int_{\frac{|\underline{u}|}{\lambda_1}}^{\infty} (\min\{\lambda_1 s_1^A, \lambda_1 s_2^B\} - |\underline{u}|) dF^B \\
&= \int_{\frac{|\underline{u}|}{\lambda_1}}^{\infty} (\min\{\lambda_1 s_1^A, \lambda_1 s_2^B\} - |\underline{u}|) dF^B > 0\,.
\end{aligned}
$$

Thus $EU_{\underline{u}}^{emp}(AB) > EU_{\underline{u}}^{emp}(AA)$ when $s_1^A > \dfrac{|\underline{u}|}{\lambda_1}$.

When $s_1^A < \dfrac{\underline{u}}{\lambda_1}$ the agent's payoff when she takes up employment and experiments is $\underline{u}$. Thus $EU_{\underline{u}}^{emp}(AB) < EU_{\underline{u}}^{emp}(AA)$ when $s_1^A < \dfrac{\underline{u}}{\lambda_1}$. ■

Proof of Proposition 7: Let $\hat{s}_2^B = s_2^B - b$. We split the proof into three claims.

Claim 1: $E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, b\}] \leq E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^B, b\}]$.

Proof The distribution of signal $\hat{s}_2^A$ given s_1^A is $N(0, (1-\lambda_1^2)(\sigma_\eta^2 + \sigma_\epsilon^2))$. The distribution of signal $\hat{s}_2^B$ is $N(0, v\sigma_\eta^2 + w\sigma_s^2)$. When $1 - \lambda_1^2 < v\lambda_1 + w(1-\lambda_1)$ the two random variables $\hat{s}_2^A$ and $\hat{s}_2^B$ have the same mean but the former has smaller variance than the latter. Thus $\hat{s}_2^A$ second order stochastically dominates $\hat{s}_2^B$. Since the max function is convex, $E_{\hat{s}_2^A}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^A, b\}] \leq E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 s_2^B, b\}]$.■

Claim 2: $E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^B, b\}] < E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_1^B \hat{s}_2^B, b\}]$.

Proof Notice that $\lambda_1^B > \lambda_2$ when $v\lambda_1 + w(1-\lambda_1) < v(1+\lambda_1)$.

Consider two possible cases.

First, suppose $\lambda_1 s_1^A \leq b$. Then $\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^B, b\} \leq \max\{\lambda_1 s_1^A + \lambda_1^B \hat{s}_2^B, b\}$ with the inequality strict for $\hat{s}_2^B$ sufficiently large. Thus $E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^B, b\}] < E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_1^B \hat{s}_2^B, b\}]$.

Second, suppose $\lambda_1 s_1^A \geq b$. Then $\max\{\lambda_1 s_1^A, \lambda_2 \hat{s}_2^B + b\} \leq \max\{\lambda_1 s_1^A, \lambda_1^B \hat{s}_2^B + b\}$ with the inequality strict for $\hat{s}_2^B$ sufficiently large. From Lemma 1 it follows that

$E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_2 \hat{s}_2^B, b\}] = E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A, \lambda_2 \hat{s}_2^B + b\}] < E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A, \lambda_1 \hat{s}_2^B + b\}] = E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_1^B \hat{s}_2^B, b\}]$.■

Claim 3: $E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_1^B \hat{s}_2^B, b\}] = E_{s_2^B}[\max\{\lambda_1 s_1^A, (1-\lambda_1^B)b + \lambda_1^B s_2^B\}]$.

Proof

$$\begin{aligned}
E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A + \lambda_1^B \hat{s}_2^B, b\}] &= E_{\hat{s}_2^B}[\max\{\lambda_1 s_1^A, \lambda_1^B \hat{s}_2^B + b\}] \\
&= E_{s_2^B}[\max\{\lambda_1 s_1^A, \lambda_1^B (s_2^B - b) + b\}] \\
&= E_{s_2^B}[\max\{\lambda_1 s_1^A, (1-\lambda_1^B)b + \lambda_1^B s_2^B\}]
\end{aligned}$$

where the equality in the first line follows from Lemma 1.
■

Taking all three claims together, the result holds.■

Proof of Proposition 8: Consider two possible cases. First, suppose the agent samples Sector A in the first period. Then since $v > 1$ and $w = 1$ both the inequalities in Proposition 7 hold and it is optimal for the agent to sample Sector B in the second period.

For the second case, suppose that the agent samples Sector B in the first period. There are then two subcases to consider. When $\dfrac{\sigma_\eta^2}{\sigma_\epsilon^2} < \dfrac{v-1}{v}$, the posterior variance of talent in Sector B at the start of period 2 which is given by $\dfrac{v\sigma_\eta^2\sigma_\epsilon^2}{v\sigma_\eta^2 + \sigma_\epsilon^2}$ is larger than the prior variance of talent in Sector A σ_η^2. Thus, as in the proof of Proposition 7, the following three claims hold.

Claim 1: $E_{s_2^A}[\max\{\lambda_1 s_2^A, (1-\lambda_1^B)b + \lambda_1^B s_1^B\}] \leq E_{\hat{s}_2^{B'}}[\max\{\lambda_1 \hat{s}_2^{B'}, (1-\lambda_1^B)b + \lambda_1^B s_1^B\}]$, where $\hat{s}_2^{B'} = s_2^B - ((1-\lambda_1^B)b + \lambda_1^B s_1^B)$. This claim holds because the signals s_2^A and $\hat{s}_2^{B'}$ have the same mean, but the former has smaller variance than the latter.

Claim 2: $E_{\hat{s}_2^{B'}}[\max\{\lambda_1 \hat{s}_2^{B'}, (1-\lambda_1^B)b + \lambda_1^B s_1^B\}] < E_{\hat{s}_2^{B'}}[\max\{\lambda_2^B \hat{s}_2^{B'}, (1-\lambda_1^B)b + \lambda_1^B s_1^B\}]$, where $\lambda_2^B = \frac{v\sigma_\eta^2}{2v\sigma_\eta^2 + \sigma_\epsilon^2}$.

Claim 3: $E_{\hat{s}_2^{B'}}[\max\{\lambda_2^B \hat{s}_2^{B'}, (1-\lambda_1^B)b + \lambda_1^B s_1^B\}] = E_{\hat{s}_2^B}[\max\{0, \lambda_2^B s_2^B + (1-\lambda_2^B)((1-\lambda_1^B)b + \lambda_1^B s_1^B)\}]$. This claim follows from Lemma 1.

Thus for this subcase, specializing in Sector B is optimal. Also since the expected surplus from experimenting does not depend on the order of sampling talents, it follows that specializing in Sector B also does better than sampling Sector A in the first period and Sector B in the second period.

For the second subcase let $\dfrac{\sigma_\eta^2}{\sigma_\epsilon^2} \geq \dfrac{v-1}{v}$. For this subcase, the posterior variance of talent in Sector B at the start of period 2 which is given by $\dfrac{v\sigma_\eta^2\sigma_\epsilon^2}{v\sigma_\eta^2+\sigma_\epsilon^2}$ is less than or equal to the prior variance of talent in Sector A σ_η^2. Thus, as in the proof of Proposition 7, the following three claims hold.

Claim 1: $E_{\hat{s}_2^{B'}}[\max\{0, \lambda_2^B \hat{s}_2^{B'} + (1-\lambda_1^B)b + \lambda_1^B s_1^B\}] \leq E_{s_2^A}[\max\{0, \lambda_2^B s_2^A + (1-\lambda_1^B)b + \lambda_1^B s_1^B\}]$, where $\hat{s}_2^{B'} = s_2^B - ((1-\lambda_1^B)b + \lambda_1^B s_1^B)$. This claim holds because the signals $\hat{s}_2^{B'}$ and s_2^A have the same mean, but the former has smaller variance than the latter.

Claim 2: $E_{s_2^A}[\max\{0, \lambda_2^B s_2^A + (1-\lambda_1^B)b + \lambda_1^B s_1^B\}] < E_{s_2^A}[\max\{0, \lambda_1 s_2^A + (1-\lambda_1^B)b + \lambda_1^B s_1^B\}]$, where $\lambda_2^B = \frac{v\sigma_\eta^2}{2v\sigma_\eta^2+\sigma_\epsilon^2}$.

Claim 3: $E_{s_2^A}[\max\{0, \lambda_1 s_2^A + (1-\lambda_1^B)b + \lambda_1^B s_1^B\} = E_{s_2^A}[\max\{\lambda_1 s_2^A, (1-\lambda_1^B)b + \lambda_1^B s_1^B\}$. This claim follows from Lemma 1.

Thus for this subcase, experimenting is optimal. ■

References

Acemoglu, D., S. Johnson, and J. Robinson (2002): "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 117(4), 1231–1294.

Altonji, J. (1993): "The Demand for and Return to Education When Education Outcomes are Uncertain," *Journal of Labor Economics*, 11(1), 48–83.

Arcidiacono, P. (2004): "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, 121(1-2), 343–375.

Ardagna, S., and A. Lusardi (2010): *Lerner, J., Shoar, A. (Eds.),NBER Volume on International Differences in Entrepreneurship*.chap. Explaining International Differences in Entrepreneurship: The Role of Individual Characteristics and Regulatory Constraints. University of Chicago Press, Chicago.

Banerjee, A., and A. Newman (1993): "Occupational Choice and the Process of Development.," *Journal of Political Economy*, 101, 274–298.

Becker, G. (1964): *Human Capital: a Theoretical and Empirical Analysis, with Special Reference to Schooling.* University of Chicago Press, Chicago.

Bordon, P., and C. Fu (2015): "College-Major Choice to College-Then-Major Choice," *Review of Economic Studies*, 82, 1247–1288.

Chandler, A. (1962): *Strategy and Structure: Chapters in the History of the Industrial Enterprise.* MIT Press, Cambridge.

DeGroot, M. (1970): *Optimal Statistical Decisions.* John Wiley and Sons, New Jersey.

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies," *Review of Economic Studies*, 66, 199–217.

Holmstrom, B. (1982): *Essays in Economics and Management in Honor of Lars Wahlbeck*chap. Managerial Incentive Problems- A Dynamic Perspective. Sweedish School of Economics, Helsinki.

Jovanovic, B. (1979): "Job matching and the theory of turnover.," *Journal of Political Economy*, 87(6), 972–990.

Kihlstrom, R., and J.-J. Laffont (1979): "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion.," *Journal of Political Economy*, 87(4), 719–748.

LAPORTA, R., F. L. DE SILANES, A. SHLEIFER, AND R. W. VISHNY (1998): "Law and Finance.," *Journal of Political Economy*, 106(6), 1113–1155.

——— (2000): "Investor protection and corporate governance.," *Journal of Financial Economics*, 58, 3–27.

LAZEAR, E. (2004): "Balanced Skills and Entrepreneurship," *American Economic Review*, 94(2), 208–211.

——— (2005): "Entrepreneurship," *Journal of Labor Economics*, 23(4), 649–680.

LUCAS, R. (1978): "On the Size Distribution of Business Firms.," *Bell Journal of Economics*, 9(2), 508–523.

MALAMUD, O. (2011): "Discovering One's Talent: Learning from Academic Specialization," *Industrial and Labor Relations Review*, 62(2), 372–405.

MANNING, A. (2003): *Monopsony in Motion.* Princeton University Press, Princeton.

MEYER, M. (1994): "The dynamics of Learning with Team Production: Implications for Task Assignment.," *Quarterly Journal of Economics*, 109(4), 1157–1184.

MILLER, R. (1984): "Job matching and Occupational Choice.," *Journal of Political Economy*, 92(6), 1086–1122.

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution.," *Journal of Political Economy*, 66(4), 281–302.

NORTH, D. (1981): *Structure and change in economic history.* Norton, New York.

NORTH, D., AND R. P. THOMAS (1973): *The rise of the western world: A new economic history.* Cambridge University Press, Cambridge.

ORTEGA, J. (2001): "Job Rotation as a Learning Mechanism.," *Management Science*, 47(10), 1361–1370.

ROBERTS, J. (2004): *The Modern Firm.* Oxford University Press, Oxford.

SCHULTZ, T. (1961): "Investment in human capital.," *American Economic Review*, 51, 1–17.

——— (1968): "Resources for Higher Education: An Economists View.," *Journal of Political Economy*, 67(3), 327–347.