Finding Out Who You Are: A Self-exploration View of Education

Sungmin Park[*]

April 15, 2022

Abstract

I study the role of education as self-exploration. Students in my model have different priors about their talents and update their beliefs after receiving noisy signals about themselves. I characterize a socially optimal design of the signal structure. An optimal structure encourages a career in which participating students are on average more confident. I apply the model to students in the United States and estimate the parameters from data. Advanced science classes in high school tend to encourage science majors. Their estimated self-exploration value is a four-percent increase in earnings after graduation.

Keywords: Education, Occupational choice, Value of information, Information design
JEL Codes: I21, J23, D83

1 Introduction

1.1 Motivation and results

People commonly say that education is not just about gaining skills or a diploma but also about finding oneself, about one's interests and talents. Through experimentation and experience, students realize what they like and what they are good at. In primary school, these experiments are often classroom exercises, toys, and games. In secondary school, they are reading, problem-solving, discussion, and writing. In college, they are general education classes before declaring one's major. Even in more specialized, graduate and professional schools, they are classes across different subfields. Likewise, all stages of education involve some degree of self-exploration.

[*]Department of Economics, The Ohio State University. 1945 North High Street, Columbus, Ohio, United States 43210. Email: park.2881@buckeyemail.osu.edu

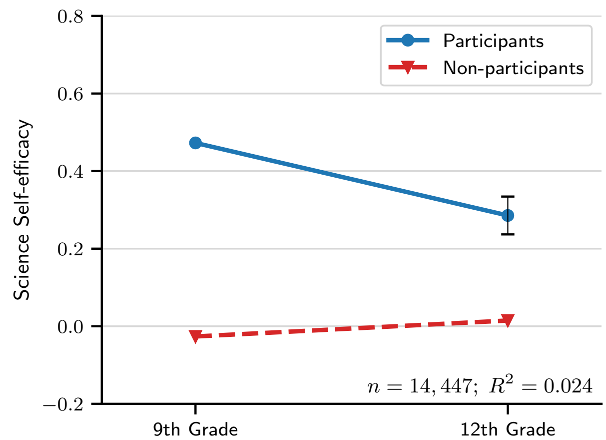

Figure 1: Advanced Science Classes and Changes in Science Self-Efficacy

Note: This figure shows the average degrees of *science self-efficacy* of groups of students who were 9th graders in 2009 from the public-use dataset of the High School Longitudinal Study (HSLS). Science self-efficacy refers to the tendency to agree that the respondent is confident about doing well in exams, understanding textbooks, mastering the skills, and doing the required assignments in the sciences. The *participants* are 2,587 students who took one or more Advanced Placement science classes in high school, whereas the *non-participants* are the remaining 11,860 students who did not take any. The error bar indicates the size of the 95% confidence interval of the estimated difference-in-differences computed with robust standard errors clustered for each student.

Survey data of high school students support this view that education reveals information to students about themselves. The public-use dataset of the High School Longitudinal Study (HSLS) of 2009 contains a measure of students' confidence in their science ability near the beginning and the end of high school. Students taking advanced science classes in high school such as the Advanced Placement (AP) appear to lose significant self-confidence in the sciences (Figure 1). Such belief-updating in school suggests a need for a theory of education beyond the human capital view (that education directly increases one's skills) and the signaling view (that education informs others about one's skills).

Motivated by the commonplace notion and the empirical observation, I study the role of education as self-exploration. Students in my model have different prior beliefs about their talents: whether they are talented *hunters* or talented *gatherers*. With costly effort, they can participate in an educational program that sends noisy signals about their talents. After observing a signal, each student chooses a career in hunting or gathering. Truly talented hunters are better off choosing hunting and truly talented gathers are better off choosing gathering.

In this simple setting, I characterize the socially optimal design of the signal structure of education (hereafter *educational structure*). An educational structure is optimal if and only if (a) it is optimal for the average participant and (b) it induces the set of participants for whom the information is most valuable (Theorem 1). The intuition behind this result is that every participating student's *ex ante* expected payoff is linear in his belief, allowing the optimal structure to depend only on the average belief of the participants. This characterization simplifies the optimization problem of searching over an infinite parameter space into that of searching over a finite collection of potential participant sets.

Building on this result, I prove that an optimal educational structure encourages a career in which its participating students are on average more confident (Theorem 2). Suppose that students' beliefs are such that they are on average more confident in choosing the hunting career than the gathering career. Then an optimal educational structure tends to recommend the hunting career with greater probability. This result arises because the students' average beliefs imply that the marginal benefit of signal accuracy for a talented hunter is greater than that for a talented gatherer.

I apply the model to high school students in the United States and estimate its parameters with a sample of 13,003 students. The sample is from the same public-use dataset of High School Longitudinal Study (HSLS). Assuming that students are either talented in science or not, the estimated parameters imply that students' beliefs of being talented in science are distributed with a mean of 35 percent and a standard deviation of 4 percentage points. The Advanced Placement (AP) science classes tend to encourage students to choose science majors in college. The current educational structure is such that the average student receives a science-encouraging signal with 61 percent probability after taking such a class. The estimated average value of information from taking an advanced science classes is a 3 percent increase in students' earnings after graduation. Under a counterfactual educational structure when the advanced science classes are science-*discouraging*, the estimated value is a larger increase of 12 percent in earnings. The interpretation is that the current AP science class structure over-recommends the sciences to a student population whose confidence in the subject is low.

1.2 Related Literature

The contribution of this study is formalizing the value and design of the self-exploring aspect of education largely overlooked by the existing theories.

The model is an application of the theories of information design and costly information acquisition.

Human capital and signaling On the one hand, the foundational human capital models by Becker (1962, 2009), Ben-Porath (1967), and Mincer (1974) see education as an investment that directly enhances one's productivity. On the other hand, the alternative, signaling models by Arrow (1973), Spence (1978), and Weiss (1995) see education as a socially wasteful but privately valuable means to inform firms about one's productivity. Following these models, the vast empirical literature attributes the return to education to either human capital investment or signaling (Card, 1994, 1999, 2001; Kroch and Sjoblom, 1994; Altonji, 1995; Keane and Wolpin, 1997; Altonji and Pierret, 1998; Chevalier et al., 2004; Psacharopoulos and Patrinos, 2004; Fang, 2006; Lange, 2007; Hussey, 2012; Patrinos, 2016; Aryal, Bhuller, and Lange, 2019; Huntington-Klein, 2021). Both views assume that students are fully informed about themselves. In contrast, students in my model are only partially informed about themselves. The value of education as self-exploration is in reducing that self-uncertainty. This third view fills the gap between the two existing theories by proposing a role of education that is both privately and socially valuable but does not directly increase one's productivity.

Belief-updating Many papers include empirically driven models with students' uncertainty and belief-updating about themselves. In a key early work, Manski (1989) views college education as an experimentation on students' ability to graduate. Despite a possible dropout, attending college can increase students' *ex ante* expected utility if the payoff from graduation is large enough.[1] In Altonji (1993), students are uncertain about their preferences. Specifically, some are indifferent between studying science, studying humanities, and working. Some are indifferent between studying humanities and working, but they dislike studying science. The rest dislike studying altogether. The different beliefs lead to different educational choices and outcomes. In Arcidiacono (2004), Altonji, Blom, and Meghir (2012), and Altonji, Arcidiacono, and Maurel (2016), students are unsure about their multidimensional ability and preferences across different fields of study. Different beliefs lead to varying *ex ante* expected returns to different fields of study. Empirically, Stinebrickner and Stinebrickner (2014) find from survey data that many students enter college with optimistic beliefs about completing science

[1] Similarly, Comay, Melnik, and Pollatschek (1973) examine students facing uncertainty about graduation, although they do not highlight the effect of education on *ex ante* expected payoffs.

degrees, yet relatively few students end up completing them after updating their beliefs. In a randomized controlled trial, Owen (2020) finds that male and female students update their beliefs differently when informed about their relative performance in science classes.

To my knowledge, only one existing paper, Arcidiacono, Aucejo, Maurel, and Ransom (2016), examines the value of an optimal information structure. They estimate a model of belief-updating students choosing education and occupations, and find that college graduates would have 33 percent higher wages under perfect information about their abilities. Compared to their work, mine abstracts from multistage educational choices and multidimensional abilities. Instead, it focuses on a single-stage educational choice when ability is binary. The gain from this simplification is that it provides the first mathematical characterization of a socially optimal educational structure when perfect information is infeasible.

Information design My work is an application of the established ideas of Bayesian persuasion and information design to the problem of educational design. The canonical paper by Kamenica and Gentzkow (2011) examines a Sender's optimal choice of the information structure that affects the Receiver's behavior. Many extensions of their model, as surveyed by Bergemann and Morris (2019) and Kamenica (2019), include incorporating (i) multiple receivers, (ii) multiple senders, and (iii) dynamic elements. My work belongs to the first category, as I consider one Sender (educational designer) and multiple Receivers (students). More specifically, mine belongs to the class of models with multiple Receivers with heterogeneous beliefs whose actions have no consequences on other Receivers' utility. Under this category, Alonso and Câmara (2016), for example, consider the problem of a politician sending a public signal to influence voters with different beliefs. For another example, Arieli and Babichenko (2019) examine the problem of a seller sending private signals to potential buyers to persuade them to purchase a product. Mine differs from these existing works in that there is no single common state of the world. Rather, each Receiver (student) has his own independent state which alone affects his payoff.

Costly information acquisition Unlike most Senders in the Bayesian persuasion literature, my Sender (educational designer) is benevolent and faces an educational resource constraint. In this regard, my work is related to the eminent literature on rational inattention and costly information acquisition such as Matějka and McKay (2015) and Caplin, Dean, and Leahy (2019). However, whereas these studies use cost functions based on the distribution of posterior beliefs, notably the changes in the Shannon entropy, I use a cost that is a direct function of the statistical powers of the information structure. This form of cost function facilitates my analysis of the optimal educational design as it does not depend on students' prior beliefs.

1.3 Outline

Section 2 describes the model of education as an experiment on students' talents. Section 3 analyzes the optimal design of education for students' aggregate welfare. Section 4 examines the effects of changes in beliefs, technology, and the cost of effort. Section 5 extends the baseline model to allow stochastic choice, human capital investment, and imperfectly observed beliefs. Section 6 estimates the educational structure of advanced science classes in U.S. high schools. Section 7 concludes.

2 Model of Education as Self-Exploration

There are students indexed as $i \in I = \{1, 2, \ldots, n\}$. Student i's *state* or *talent* is $\omega^i \in \Omega = \{\omega_g, \omega_h\}$ where ω_g denotes being a talented gatherer and ω_h denotes being a talented hunter. Each student i has a rational and publicly known *prior belief* $p^i \in (0, 1)$ on the event that $\omega^i = \omega_h$. A student's *action* or *career* is $a \in A = \{a_g, a_h\}$ where a_g denotes a gathering career and a_h denotes a hunting career. A student's *productivity* is $u(\omega, a)$ where $u : \Omega \times A \longrightarrow \mathbb{R}_+$. The only restriction on u is that a talented gatherer is better off choosing the gathering career and a talented hunter is better off choosing the hunting career. That is, $u(\omega_g, a_g) > u(\omega_g, a_h)$ and $u(\omega_h, a_h) > u(\omega_h, a_g)$. The values $u_g = u(\omega_g, a_g) - u(\omega_g, a_h)$ and $u_h = u(\omega_h, a_h) - u(\omega_h, a_g)$ are called the *mismatch costs* of a talented hunter and a talented gatherer.

Before deciding on a career, students may *participate* in an educational program by paying a costly effort $\delta \geq 0$ to receive a signal $s \in S = \{s_g, s_h\}$. Signals are realized with conditional probabilities $\Pr(s_g|\omega_g) = 1 - \Pr(s_h|\omega_g) = \gamma$ and $\Pr(s_h|\omega_h) = 1 - \Pr(s_g|\omega_h) = \eta$. The interpretation is that γ represents the accuracy of signals when the true state is ω_g, and η represents the accuracy of signals when the true state is ω_h. The ordered pair $(\gamma, \eta) \in \Theta$ is the *educational structure*, where

$$\Theta = \left\{(\gamma, \eta) \in [0, 1]^2 \;\middle|\; \gamma + \eta \geq 1\right\} \tag{1}$$

is its parameter space. The *posterior belief function* is $Q_{\gamma,\eta} : (0, 1) \times S \longrightarrow (0, 1)$ whose value is the posterior belief q after receiving signal s under the

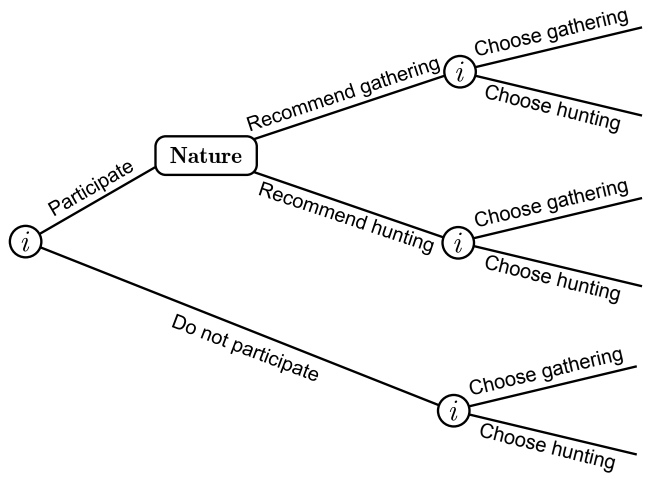

Figure 2: Education and Occupational Choice

Note: This figure illustrates the sequence of moves of every student i given his prior belief $p_i = \Pr(\omega_i = \omega_h)$ and the educational structure (γ, η). If the student participates in the educational program, Nature sends a signal s_g (encourage gathering) or s_h (encourage hunting) with conditional probabilities $\Pr(s_g|\omega_g) = 1 - \Pr(s_h|\omega_g) = \gamma$ and $\Pr(s_h|\omega_h) = 1 - \Pr(s_g|\omega_h) = \eta$.

educational structure (γ, η) when the prior belief is p. The signals s_g and s_h are interpreted as recommendations toward gathering and hunting, respectively.[2] An educational structure $(\gamma, \eta) \in \Theta$ is *uninformative* if $\gamma + \eta = 1$, *informative* if $\gamma + \eta > 1$, and *perfectly informative* if $\gamma = \eta = 1$. Figure 2 illustrates each student's sequential choices.

The *educational designer* is an imaginary agent representing teachers, parents, and policymakers who collectively shape the underlying educational structure. However, their attention is limited, so not every educational structure is feasible. Its choice is restricted to an element of the *feasible set* $\widehat{\Theta} = \{(\gamma, \eta) \in \Theta \mid C(\gamma, \eta) \leq B, B > 0\}$, where the constant B is called the *educational attention budget*, and the *educational attention cost function* $C : \Theta \longrightarrow \mathbb{R}$ satisfies Assumptions 1–4.

Assumption 1 (Smoothness)**.** *C is continuously differentiable.*

Assumption 2 (Curvature)**.** *C is strictly increasing in γ and η on Θ and strictly convex on the interior of Θ.*

Assumption 3 (Symmetry)**.** *$C(\gamma, \eta) = C(\eta, \gamma)$ for any γ and η.*

[2] This interpretation is valid as long as $\gamma + \eta \geq 1$ as required by $(\gamma, \eta) \in \Theta$ and is without loss of generality. If we instead have $\gamma + \eta \leq 1$, we may relabel the probabilities as $\gamma' = 1 - \gamma$ and $\eta' = 1 - \eta$ as well as relabeling the signals as $s'_g = s_h$ and $s'_h = s_g$ to arrive at the same interpretation.

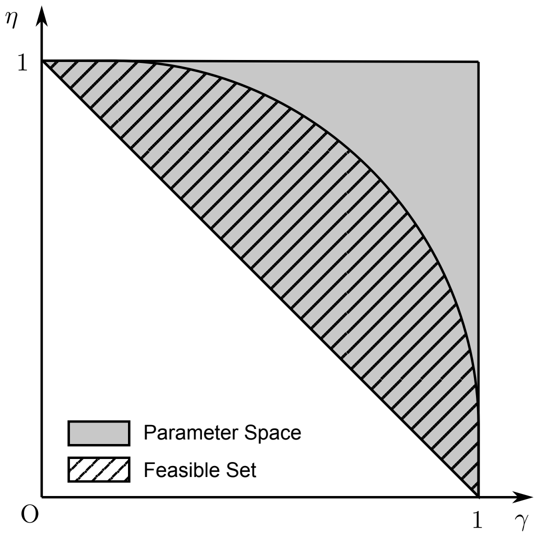

Figure 3: Feasible Education

Assumption 4 (Boundary regularity)**.**

(a) $C(\gamma, \eta) = 0$ *if* $\gamma + \eta = 1$*,*

(b) $C(\gamma, \eta) > B$ *if* $\gamma = \eta = 1$*, and*

(c) $\lim_{\gamma \to 1} \frac{\partial}{\partial \gamma} C(\gamma, \eta) = \lim_{\eta \to 1} \frac{\partial}{\partial \eta} C(\gamma, \eta) = \infty$.

Assumption 4(a) means that uninformative education is always feasible. Assumption 4(b) means that perfectly informative education is infeasible. Assumption 4(c) means that a signal accuracy (γ or η) is infinitely costly on the margin as it approaches 1. Figure 3 illustrates the feasible set of educational structures.

The *ex post* payoff of a student i with state $\omega^i \in \Omega$ and career $a^i \in A$ is $u(\omega^i, a^i) - \delta d^i$ where $d^i = 1$ for a participant and $d^i = 0$ for a non-participant. Every student updates his belief using Bayes' rule and chooses an action that maximizes his expected payoff at each stage. As a tie-breaking rule, students that are indifferent between participating and not participating do not participate. Students that are indifferent between the hunting and gathering careers choose hunting. These tie-breaking rules, however, apply only in knife-edge cases and affect neither the value nor the optimal design of education.

3 Optimal Design of Education

3.1 Characterization

Let $V : (0,1) \times \Theta \longrightarrow \mathbb{R}$ map a student's prior belief $p \in (0,1)$ and an educational structure $(\gamma, \eta) \in \Theta$ to the student's *ex ante* expected payoff under (γ, η) conditional on p. Then

$$V(p, \gamma, \eta) = \max \left\{ V_0(p), V_1(p, \gamma, \eta) - \delta \right\},$$

where $V_0(p)$ is a non-participating student's expected productivity and $V_1(p, \gamma, \eta)$ is a participating student's expected productivity before receiving a signal from the educational structure (γ, η). Let

$$U_g(p) = \sum_{\omega \in \Omega} \Pr(\omega) u(a_g, \omega) = (1-p) u(a_g, \omega_g) + p u(a_g, \omega_h), \text{ and} \tag{2}$$

$$U_h(p) = \sum_{\omega \in \Omega} \Pr(\omega) u(a_h, \omega) = (1-p) u(a_h, \omega_g) + p u(a_h, \omega_h). \tag{3}$$

The two functions $U_g(p)$ and $U_h(p)$ represent the expected productivities of choosing gathering and hunting, respectively, when one's belief is p. Then the functions V_0 and V_1 take values

$$V_0(p) = \max\{U_g(p), U_h(p)\}, \text{ and} \tag{4}$$

$$V_1(p, \gamma, \eta) = \mathbb{E}\left[V_0\big(Q_{\gamma,\eta}(p, s)\big) \mid p, \gamma, \eta \right]. \tag{5}$$

Let the *student welfare function* $W : \Theta \longrightarrow \mathbb{R}$ be defined as

$$W(\gamma, \eta) = \sum_{i \in I} V(p^i, \gamma, \eta). \tag{6}$$

An educational structure $(\gamma^*, \eta^*) \in \widehat{\Theta}$ is *(socially) optimal* if $W(\gamma^*, \eta^*) \geq W(\gamma, \eta)$ for all $(\gamma, \eta) \in \widehat{\Theta}$. An educational structure (γ, η) is *nontrivial* if at least one student $i \in I$ participates in it. Let $\overline{p}_D = \frac{1}{|D|} \sum_{i \in D} p^i$ for all nonempty subsets $D \subset I$. Let $F : (0,1) \longrightarrow \widehat{\Theta}$ map p to the solution (γ, η) to the system of equations

$$\frac{\frac{\partial}{\partial \gamma} C(\gamma, \eta)}{\frac{\partial}{\partial \eta} C(\gamma, \eta)} = \frac{1-p}{p} \frac{u_g}{u_h} \quad \text{and} \quad C(\gamma, \eta) = B. \tag{7}$$

To verify that F is well-defined, observe that the system of equations (7) is

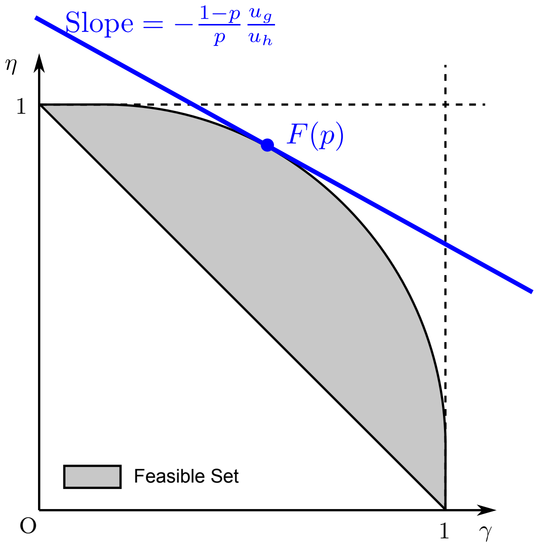

Figure 4: Definition of F

the first order condition to the problem of maximizing $(1-p)u_g\gamma + pu_h\eta$ subject to $C(\gamma, \eta) = B$. The maximizer exists because the maximand is continuous in (γ, η) and the constraint set is closed and bounded. By Assumption 4(c), the solution is an interior point of Θ. It is unique because C is strictly convex. Figure 4 illustrates the definition of F.

Theorem 1. *Suppose $(\gamma^*, \eta^*) \in \widehat{\Theta}$ is nontrivial. Then (γ^*, η^*) is optimal if and only if there exists $D^* \subset I$ that satisfies $(\gamma^*, \eta^*) = F(\bar{p}_{D^*})$ and*

$$D^* \in \underset{D \in \mathcal{P}(I)}{\operatorname{argmax}} W\big(F(\bar{p}_D)\big). \tag{8}$$

This result means that finding an optimal educational structure is equivalent to (i) targeting the average student of a group, and (ii) finding the best group of students to target. In other words, a benevolent educational designer trying to maximize students' welfare does not lose anything by taking a two-step procedure: first finding an optimal subset of students and then applying an optimal design as a function of only the average belief of that subset. This characterization of optimal education simplifies the designer's problem of searching for (γ^*, η^*) over an infinite set Θ to that of searching for D^* over a finite collection $\mathcal{P}(I)$.

Define Θ_D as the set of (γ, η) under which the set of participants is D. For every $(\gamma, \eta) \in \Theta$, such participant set uniquely exists because every student with belief p^i under (γ, η) either participates or not.

Let us prove several lemmas that are part of the proof the the theorem.

Lemma 1. *The student welfare function $W(\gamma, \eta)$ is*
(a) *continuous on Θ,*
(b) *strictly increasing in γ and η on the nontrivial subset of Θ,*
(c) *affine on Θ_D for every $D \subset I$, and*
(d) *convex.*

Proof. Observe that

$$W(\gamma, \eta) = \sum_{i \in I} V(p^i, \gamma, \eta) = \sum_{i \in I} \max \left\{ V_0(p^i), V_1(p^i, \gamma, \eta) - \delta \right\}. \tag{9}$$

Observe that a participant's expected productivity is

$$V_1(p, \gamma, \eta) = \sum_{\omega \in \Omega} \sum_{s \in S} \sum_{a \in A} \Pr(\omega) \Pr(s \mid \omega) \Pr(a \mid s) u(\omega, a). \tag{10}$$

I claim that $\Pr(a_g | s_g) = \Pr(a_h | s_h) = 1$ for any participant. Suppose that this statement is false. Then (i) $\Pr(a_g | s_g) = \Pr(a_g | s_h) = 1$, (ii) $\Pr(a_h | s_g) = \Pr(a_h | s_h) = 1$, or (iii) $\Pr(a_h | s_g) = \Pr(a_g | s_h) = 1$. In all of the three cases, $V_0(p) \geq V_1(p, \gamma, \eta)$, meaning that the student does not participate, a contradiction.

Then

$$V_1(p, \gamma, \eta) = (1-p) \left[\gamma u(\omega_g, a_g) + (1-\gamma) u(\omega_g, a_h) \right] + p \left[(1-\eta) u(\omega_h, a_g) + \eta u(\omega_h, a_h) \right], \tag{11}$$

so $V_1(p, \gamma, \eta)$ is linear in (γ, η) and strictly increasing in γ and η.

Observe from equation (9) that we can write

$$W(\gamma, \eta) = \max_{D \in \mathcal{P}(I)} \left(\sum_{i \in I \setminus D} V_0(p^i) + \sum_{i \in D} [V_1(p^i, \gamma, \eta) - \delta] \right). \tag{12}$$

Let $W(\gamma, \eta | D)$ denote the maximand in the above. Then $W(\gamma, \eta | D)$ is affine in (γ, η). It is strictly increasing in γ and η as long as D is nonempty. Since $W(\gamma, \eta)$ is a maximum of an affine function, it is continuous and convex. Because $W(\gamma, \eta)$ is a maximum of strictly increasing function on the nontrivial subset of Θ, it is strictly increasing in γ and η on the nontrivial subset of Θ. Finally, observe that $W(\gamma, \eta) = W(\gamma, \eta | D)$ wherever $(\gamma, \eta) \in \Theta_D$, for all $D \subset I$. So $W(\gamma, \eta)$ is affine on Θ_D for all $D \subset I$. ■

Lemma 2. *An optimal educational structure exists.*

Proof. By Assumptions 1–4, the feasible set $\widehat{\Theta}$ is closed and bounded. By

Lemma 1, the student welfare function W is continuous. So by Weierstrass's Extreme Value Theorem, there exists $(\gamma^*, \eta^*) \in \widehat{\Theta}$ such that $W(\gamma^*, \eta^*) = \sup_{(\gamma,\eta)\in\widehat{\Theta}} W(\gamma, \eta)$. ■

Lemma 3. *Every nontrivial optimal educational structure is an interior point of* Θ.

Proof. Suppose that $(\gamma^*, \eta^*) \in \Theta$ is nontrivial and optimal. Because at least one student participates in (γ^*, η^*), $V_0(p^i) < V_1(p^i, \gamma^*, \eta^*)$ for some $i \in I$. This result is impossible if $\gamma^* + \eta^* = 1$, so $\gamma^* + \eta^* > 1$. Because the perfectly informative education $(1, 1)$ is infeasible and the marginal costs of γ and η approach infinity as each goes to 1, $\gamma^* < 1$ and $\eta^* < 1$. Thus $(\gamma^*, \eta^*) \in \text{Int } \Theta$. ■

Lemma 4. *Suppose* $(\gamma^*, \eta^*) \in \Theta_{D^*}$ *is optimal and* D^* *is nonempty. Then* $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$.

Proof. By Lemma 3, (γ^*, η^*) is in the interior of Θ. Then (γ^*, η^*) is a solution to the problem

$$\max_{(\gamma,\eta)\,\in\,\text{Int }\Theta\,\cap\,\Theta_{D^*}} W(\gamma, \eta) \quad \text{subject to} \quad C(\gamma, \eta) \leq B. \tag{13}$$

First, suppose that (γ^*, η^*) is in the interior of Θ_{D^*}. Since W is affine on Θ_{D^*} (Lemma 1) and C is strictly convex (Assumption 2), (γ^*, η^*) is the only interior solution in Θ_{D^*}. Then the first order conditions to the above maximization problem are

$$\frac{\frac{\partial}{\partial\gamma}C(\gamma^*, \eta^*)}{\frac{\partial}{\partial\eta}C(\gamma^*, \eta^*)} = \frac{\frac{\partial}{\partial\gamma}W(\gamma^*, \eta^*)}{\frac{\partial}{\partial\eta}W(\gamma^*, \eta^*)} \quad \text{and} \quad C(\gamma^*, \eta^*) = B. \tag{14}$$

Observe that

$$\frac{\partial}{\partial\gamma}W(\gamma^*, \eta^*) = \frac{\partial}{\partial\gamma}W(\gamma^*, \eta^*|D^*) = \frac{\partial}{\partial\gamma}\sum_{i\in D} V_1(p^i, \gamma^*, \eta^*) = \sum_{i\in D^*}(1 - p^i)u_g, \text{ and} \tag{15}$$

$$\frac{\partial}{\partial\eta}W(\gamma^*, \eta^*) = \frac{\partial}{\partial\eta}W(\gamma^*, \eta^*|D^*) = \frac{\partial}{\partial\eta}\sum_{i\in D} V_1(p^i, \gamma^*, \eta^*) = \sum_{i\in D^*} p^i u_h. \tag{16}$$

The above two equations together imply that

$$\frac{\frac{\partial}{\partial\gamma}W(\gamma^*, \eta^*)}{\frac{\partial}{\partial\eta}W(\gamma^*, \eta^*)} = \frac{1 - \overline{p}_{D^*}}{\overline{p}_{D^*}}\frac{u_g}{u_h}, \tag{17}$$

thus $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$.

Second, suppose that (γ^*, η^*) is on the boundary of Θ_{D^*}. I claim that every nonempty Θ_D for some $D \subset I$ is a convex polygon. A student $i \in I$ participates under (γ, η) if and only if $V_1(p^i, \gamma, \eta) - \delta > V_0(p^i)$. Since V_1 is affine in (γ, η), $i \in I$ participates under (γ, η) if and only if (γ, η) is in the intersection of Θ and some halfspace in $\mathbb{R}^2$. Then $D \subset I$ is the set of participants under (γ, η) if and only if (γ, η) is in the intersection of Θ and n halfspaces in $\mathbb{R}^2$. So a nonempty Θ_D is a convex polygon.

Define

$$\Phi^+ = \{(\gamma, \eta) \in \Theta \mid \eta > \eta^*, W(\gamma, \eta) = W(\gamma^*, \eta^*)\}, \text{ and} \tag{18}$$
$$\Phi^- = \{(\gamma, \eta) \in \Theta \mid \eta < \eta^*, W(\gamma, \eta) = W(\gamma^*, \eta^*)\}. \tag{19}$$

Because W is continuous and strictly increasing on the nontrivial subset of Θ (Lemma 1), Φ^+ and Φ^- are curves. For every $\varepsilon > 0$, let $B_\varepsilon(\gamma^*, \eta^*)$ denote the open ball around (γ^*, η^*) with radius ε. Let D^+ and D^- denote the participants such that Θ_{D^+} and Θ_{D^-} are convex polygons that contain $\Phi^+ \cap B_\varepsilon(\gamma^*, \eta^*)$ and $\Phi^- \cap B_\varepsilon(\gamma^*, \eta^*)$ for all $\varepsilon \in (0, \bar{\varepsilon})$ for some $\bar{\varepsilon} > 0$.

Suppose $\overline{p}_{D^+} \neq \overline{p}_{D^-}$. Then $\overline{p}_{D^+} > \overline{p}_{D^-}$ by the convexity of W (Lemma 1). Then because C is smooth (Assumption 1), Φ^+ or Φ^- intersects with the interior of the feasible set $\widehat{\Theta}$. Then there exists some $(\gamma', \eta') \in \widehat{\Theta}$ such that $W(\gamma', \eta') > W(\gamma^*, \eta^*)$. Then $(\gamma^* \eta^*)$ cannot be optimal.

So it must be that $\overline{p}_{D^+} = \overline{p}_{D^-}$. Then $\overline{p}_{D^+} = \overline{p}_D = \overline{p}_{D^-}$, for every D such that Θ_D is adjacent to (γ^*, η^*) because $D = D^+$ or $D = D^-$ or $\overline{p}_{D^+} \geq \overline{p}_D \geq \overline{p}_{D^-}$ by the convexity of W. Then W is differentiable at (γ^*, η^*). Then the first order conditions of the maximization problem yield $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$. ■

Proof of Theorem 1 First, I show the "only if" part of the statement. Suppose $(\gamma^*, \eta^*) \in \widehat{\Theta}$ is nontrivial. Suppose (γ^*, η^*) is optimal. I need to show that there exists $D^* \subset I$ that satisfies $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$ and solves

$$\max_{D \in \mathcal{P}(I)} W\big(F(\overline{p}_D)\big). \tag{20}$$

Let D^* denote the participant set under γ^*, η^*. By Lemma 4, $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$. Since (γ^*, η^*) is optimal, $W(\gamma^*, \eta^*) \geq W(\gamma, \eta)$ for all $(\gamma, \eta) \in \Theta$. Then for all $D \subset I$, we have

$$W(F(\overline{p}_{D^*})) = W(\gamma^*, \eta^*) \geq W(F(\overline{p}_D)). \tag{21}$$

So D^* solves the maximization problem (20).

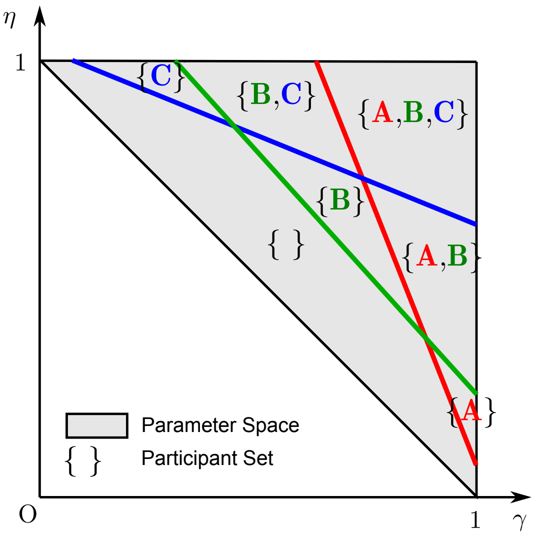

Figure 5: Every nonempty Θ_D is a convex polygon: an example with three students

Second, I show the "if" part of the statment. Suppose $(\gamma^*, \eta^*) \in \widehat{\Theta}$. Suppose there exists $D^* \subset I$ that satisfies $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$ and solves the maximization problem (20). I need to show that (γ^*, η^*) is optimal. Suppose it is not. Then by Lemma 2, some other optimal (γ', η') exists. Let D' denote the participant set under (γ', η'). By Lemma 4, $(\gamma', \eta') = F(\overline{p}_{D'})$. Because (γ', η') is optimal and (γ^*, η^*) is not, $W\big(F(\overline{p}'_D)\big) > W\big(F(\overline{p}_{D^*})\big)$. Then D^* does not solve (20), a contradiction. So (γ^*, η^*) is optimal. ■

A key step in the proof of Theorem 1 is that a participant's *ex ante* expected productivity $V_1(p, \gamma, \eta)$ is affine in p and affine in (γ, η) separately. This result arises because every participating student necessarily follows the action recommended by the signal he receives—otherwise, the signal is not useful to him and he would not participate in the first place. This fact makes every nonempty the set Θ_D of educational structures that induce a participant set D a convex polygon, as illustrated in Figure 5. It also makes the student welfare function W affine on each Θ_D, so that each iso-welfare curves are kinked lines as in Figure 6. Then an optimal education must occur at the tangency of its iso-welfare curve and the feasible set, leading to the condition $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$.

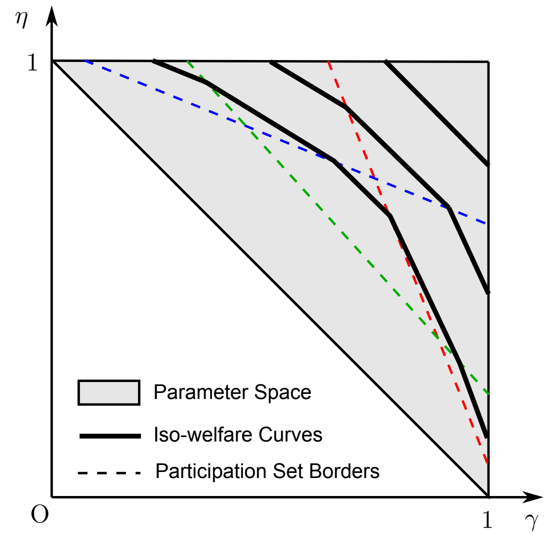

Figure 6: The student welfare function W is affine on Θ_D

3.2 Direction of Encouragement

By building on Theorem 1, we get the next result about the direction of encouragement under an optimal educational structure.

Theorem 2. *Suppose $(\gamma^*, \eta^*) \in \Theta_{D^*}$ is nontrivial and optimal. Then $\overline{p}_{D^*} = \frac{u_g}{u_g+u_h}$ implies $\gamma^* = \eta^*$, $\overline{p}_{D^*} < \frac{u_g}{u_g+u_h}$ implies $\gamma^* > \eta^*$, and $\overline{p}_{D^*} > \frac{u_g}{u_g+u_h}$ implies $\gamma^* < \eta^*$.*

This result means that an optimal education encourages a career in which the average participant is more confident. By being *more confident* in gathering or hunting, I mean that the belief is smaller or larger than the threshold $\frac{u_g}{u_g+u_h}$. This interpretation is valid because

$$U_g(p) \gtreqless U_h(p) \quad \text{if and only if} \quad p \lesseqgtr \frac{u_g}{u_g + u_h}, \tag{22}$$

where $U_g(p)$ and $U_h(p)$ are defined as in (2) and (3). Furthermore, by *encouraging* the gathering career and the hunting career, I mean the conditions $\gamma > \eta$ and $\gamma < \eta$, respectively. To see why, observe that the probability of receiving a gathering-recommending signal s_g for a student with belief p is

$$\Pr(s_g|p) = \sum_{\omega \in \Omega} \Pr(\omega) \Pr(s_g|\omega) = (1-p)\gamma + p(1-\eta), \tag{23}$$

which is strictly increasing in γ and strictly decreasing in η. Since $\Pr(s_h|p) = 1 - \Pr(s_g|p)$, the opposite is true for $\Pr(s_h|p)$, the same student's probability of receiving a hunting-recommending signal. So $\gamma > \eta$ means that the educational structure tends to recommend the gathering career with a greater probability. In particular, for a student with belief $p = 0.5$, $\Pr(s_g|p) < 0.5 < \Pr(s_h|p)$. Similarly, $\gamma < \eta$ means that the educational structure tends to recommend the hunting career with a greater probability. The condition $\gamma = \eta$ means that the educational structure is balanced.

Definition. Let F_g and F_h be the two component functions of the vector-valued function F. That is, let $F_g : (0,1) \longrightarrow [0,1]$ and $F_h : (0,1) \longrightarrow [0,1]$ such that $F(p) = (F_g(p), F_h(p))$.

The following lemma is used in the proof of the theorem.

Lemma 5. *F_g is strictly decreasing in p and F_h is strictly increasing in p.*

Proof. Suppose $0 < p < p' < 1$ and let $(\gamma, \eta) = F(p)$ and $(\gamma', \eta') = F(p')$. Then

$$\frac{\frac{\partial}{\partial\gamma}C(\gamma,\eta)}{\frac{\partial}{\partial\eta}C(\gamma,\eta)} = \frac{1-p}{p}\frac{u_g}{u_h} > \frac{1-p'}{p'}\frac{u_g}{u_h} = \frac{\frac{\partial}{\partial\gamma}C(\gamma',\eta')}{\frac{\partial}{\partial\eta}C(\gamma',\eta')}. \tag{24}$$

Since C is strictly increasing in both variables and is strictly convex, the above inequality implies $\gamma > \gamma'$ and $\eta < \eta'$. ∎

Proof of Theorem 2 By Theorem 1, $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$. Suppose $\overline{p}_{D^*} = \frac{u_g}{u_g+u_h}$. Then

$$\frac{\frac{\partial}{\partial\gamma}C(\gamma^*,\eta^*)}{\frac{\partial}{\partial\eta}C(\gamma^*,\eta^*)} = \frac{1-\overline{p}_{D^*}}{\overline{p}_{D^*}}\frac{u_g}{u_h} = 1. \tag{25}$$

Then by the strict convexity (Assumption 2) and symmetry (Assumption 3) of C, $\gamma^* = F_g(\overline{p}_{D^*}) = F_h(\overline{p}_{D^*}) = \eta^*$. Next, suppose $\overline{p}_{D^*} < \frac{u_g}{u_g+u_h}$. By Lemma 5, $\gamma^* = F_g(\overline{p}_{D^*}) > F_g\left(\frac{u_g}{u_g+u_h}\right) = F_h\left(\frac{u_g}{u_g+u_h}\right) > F_h(\overline{p}_{D^*}) = \eta^*$. Finally, suppose $\bar{t} > t^*$. Then $\gamma^* = F_g(\overline{p}_{D^*}) < F_g\left(\frac{u_g}{u_g+u_h}\right) = F_h\left(\frac{u_g}{u_g+u_h}\right) < F_h(\overline{p}_{D^*}) = \eta^*$. ∎

Figure 7 illustrates the proof of Theorem 2. The function F maps a belief p to the point on the boundary of the feasible set whose slope is $-\frac{1-p}{p}\frac{u_g}{u_h}$. When the average belief of the participating students is equal to the threshold $\frac{u_g}{u_g+u_h}$, this slope is -1, so (γ^*, η^*) is at the midpoint on the boundary. When the average belief is greater than the threshold, this slope is flatter, so (γ^*, η^*) is closer to the upper-left corner of the parameter space Θ. When the average belief is less than the threshold, this slope is steeper, so (γ^*, η^*) is closer to the bottom-right corner of the parameter space Θ.

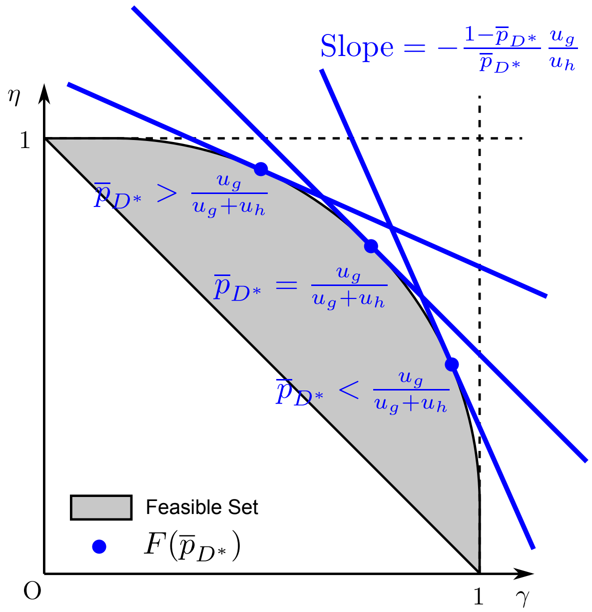

Figure 7: Optimal educational structure and the direction of encouragement

3.3 Participant Sets

Theorems 1 and 2 tell us to target an optimal set of participants and toward which direction, without telling us what such a set may resemble. The next result characterizes the set of participants under any educational structure.

Theorem 3. *A nonempty subset $D \subset I$ is a participant set under some $(\gamma, \eta) \in \Theta$ if and only if there exists an interval $(b, c) \subset (0, 1)$ that satisfies*
(a) (Separation). *$i \in D$ if and only if $p^i \in (b, c)$, and*
(b) (Regularity).

$$\delta(c - b) \le \min\Big\{(1-c)\big[U_g(b) - U_h(b)\big],\ b\big[U_h(c) - U_g(c)\big]\Big\}. \tag{26}$$

The separation condition means that a student participates if and only if he is sufficiently uncertain about his talent: that is, his belief is within an interval (b, c). Put differently, if two students with different beliefs participate, any student with an intermediate belief between the two also participates. The regularity condition requires that such interval does not tilt too much toward either direction. When $\delta = 0$, the condition is equivalent to only requiring that the threshold $\frac{u_g}{u_g+u_h}$ belongs to the interval. The condition becomes stronger as δ increases. Taken together, the theorem means that any such participant set characterized by an interval of beliefs can be induced by some educational structure.

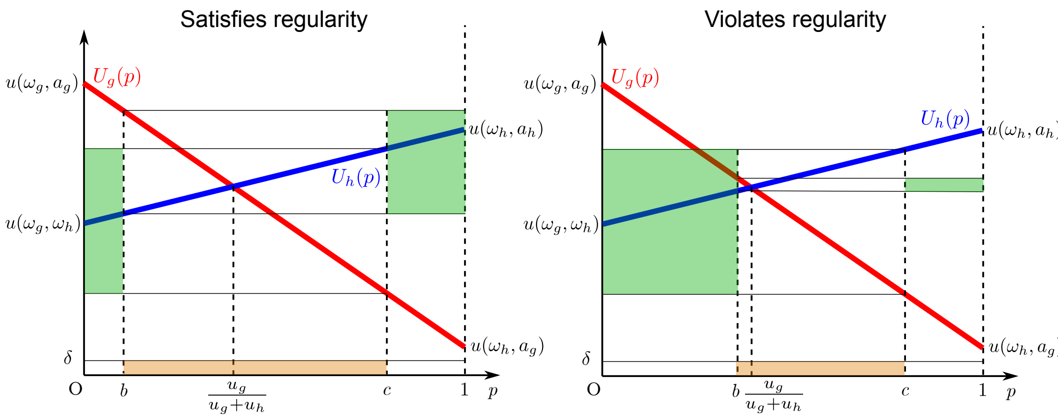

Figure 8: Interval regularity: satisfied (left) and violated (right)

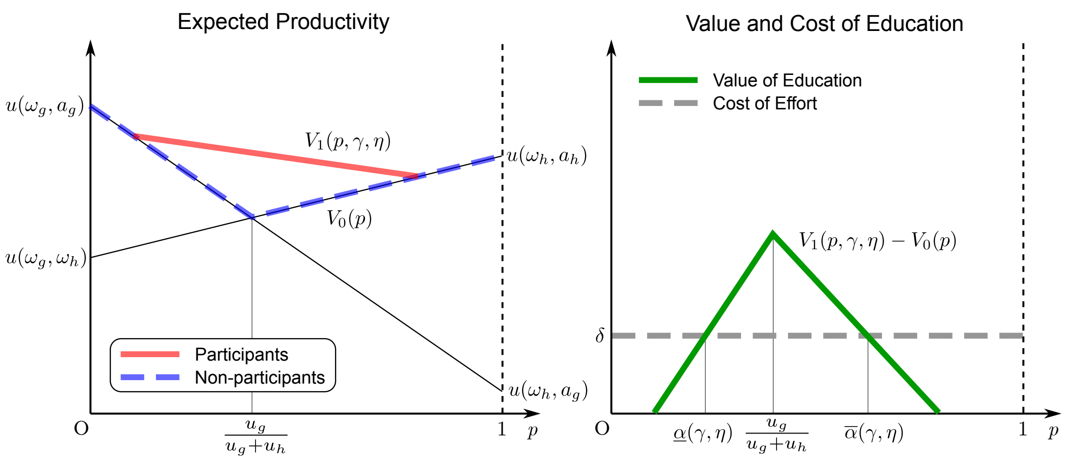

Figure 9: Any participation set is separated by an interval

Figure 8 shows examples in which the regularity condition is satisfied (left) and violated (right). Each panel indicates an interval (b, c). The area of the shaded rectangle at the bottom represents the left-hand side of (26). The areas of the top left and top right shaded triangles represent each of the two terms being minimized over on the right-hand side of (26). The regularity is violated if the bottom rectangle's area is larger than any of the two top rectangles' areas.

Figure 9 illustrates why any participant set is separated by an interval. For a fixed educational structure (γ, η), the left panel shows the expected productivities of participants and non-participants depending on their prior beliefs. Students who are sufficiently confident in either careers do not change their career choice regardless of signals, so they do not not find education valuable. Those who are sufficiently uncertain base their actions on the recommendations they receive, so they find education *ex ante* valuable. The value of education shown on the right panel represents this difference between participants' and non-participants' expected productivities. Then the participation set is the set of students whose value of education is greater than the cost of effort, with beliefs anywhere between the two intersections.

Definition. Let $\underline{\alpha}, \overline{\alpha} : \Theta \longrightarrow \mathbb{R}$ such that

$$\underline{\alpha}(\gamma, \eta) = \frac{(1-\gamma)u_g + \delta}{(1-\gamma)u_g + \eta u_h} \quad \text{and} \quad \overline{\alpha}(\gamma, \eta) = \frac{\gamma u_g - \delta}{\gamma u_g + (1-\eta)u_h}. \tag{27}$$

Let us prove the following two lemmas for the proof of the theorem.

Lemma 6. *A student with belief p participates under $(\gamma, \eta) \in \Theta$ if and only if $\underline{\alpha}(\gamma, \eta) < p < \overline{\alpha}(\gamma, \eta)$.*

Proof. A student with belief p participates under $(\gamma, \eta) \in \Theta$ if and only if

$$V_1(p, \gamma, \eta) - V_0(p) > \delta. \tag{28}$$

We know $V_1(p, \gamma, \eta)$ from equation (11). Observe also that $V_0(p) = U_g(p)$ if $p \le \frac{u_g}{u_g+u_h}$, and $V_0(p) = U_h(p)$ if $p \ge \frac{u_g}{u_g+u_h}$. Then

$$V_1(p, \gamma, \eta) - V_0(p) = \begin{cases} -(1-p)(1-\gamma)u_g + p\eta u_h & \text{if } p \le \frac{u_g}{u_g+u_h}, \\ (1-p)\gamma u_g - p(1-\eta)u_h & \text{if } p \ge \frac{u_g}{u_g+u_h}. \end{cases} \tag{29}$$

This value is greater than δ if and only if $\frac{(1-\gamma)u_g+\delta}{(1-\gamma)u_g+\eta u_h} < p < \frac{\gamma u_g - \delta}{\gamma u_g + (1-\eta)u_h}$. ■

Lemma 7. *For all $(\gamma, \eta) \in \Theta$, $\frac{u_g}{u_g+u_h}$ attains*

$$\max_{p \in (0,1)} \left[V_1(p, \gamma, \eta) - V_0(p) \right] = (\gamma + \eta - 1)\frac{u_g u_h}{u_g + u_h}. \tag{30}$$

Proof. Equation (29) implies

$$V_1\left(\frac{u_g u_h}{u_g + u_h}, \gamma, \eta\right) - V_0\left(\frac{u_g u_h}{u_g + u_h}\right) \ge V_1(p, \gamma, \eta) - V_0(p), \tag{31}$$

for all $p \in (0, 1)$. ■

Proof of Theorem 3 First, I show the "only if" part of the theorem. Suppose D is a nonempty participant set under (γ, η). Then $V_1(p^i, \gamma, \eta) - V_0(p^i) > \delta$ for some $i \in D$. Then $(\gamma + \eta - 1)\frac{u_g u_h}{u_g+u_h} > \delta$ by Lemma 7. Let $b = \underline{\alpha}(\gamma, \eta)$ and $c = \overline{\alpha}(\gamma, \eta)$. Then by Lemma 6, $i \in D$ if and only if (b, c). So condition (a) is satisfied. Furthermore, observe that

$$(1-c)\big[U_g(b)-U_h(b)\big]-\delta(c-b)=\frac{(1-\eta)u_h\cdot[(\gamma+\eta-1)u_gu_h-\delta(u_g+u_h)]}{[(1-\gamma)u_g+\eta u_h]\cdot[\gamma u_g+(1-\eta)u_h]}, \tag{32}$$

$$b\big[U_h(c)-U_g(c)\big]-\delta(c-b)=\frac{[(1-\gamma)u_h+2\delta]\cdot[(\gamma+\eta-1)u_gu_h-\delta(u_g+u_h)]}{[(1-\gamma)u_g+\eta u_h]\cdot[\gamma u_g+(1-\eta)u_h]}. \tag{33}$$

Both (32) and (33) are positive because $(\gamma+\eta-1)\frac{u_gu_h}{u_g+u_h}>0$. So condition (b) is satisfied.

Second, I show the "if" part of the theorem. Let a nonempty subset $D\subset I$ be given. Suppose there exists an interval $(b,c)\subset(0,1)$ that satisfies conditions (a) and (b). I need to show that there exists $(\gamma,\eta)\in\Theta$ that induces D. Observe that $\frac{u_g}{u_g+u_h}\in(b,c)$. Consider the system of linear equations

$$V_1(b,\gamma,\eta)-V_0(p)=\delta,\ \text{and} \tag{34}$$

$$V_1(c,\gamma,\eta)-V_0(p)=\delta. \tag{35}$$

Any solution $(\gamma,\eta)\in\Theta$ to this system of equations induces D.

Let $\gamma',\gamma'',\eta',\eta''\in(0,1)$ be defined so that

$$V_1(b,\gamma',1)-V_0(p)=\delta, \tag{36}$$

$$V_1(c,\gamma'',1)-V_0(p)=\delta, \tag{37}$$

$$V_1(b,1,\eta')-V_0(p)=\delta,\ \text{and} \tag{38}$$

$$V_1(c,1,\eta'')-V_0(p)=\delta. \tag{39}$$

Observe that there exists a solution $(\gamma,\eta)\in\Theta$ to the system of equations (34)–(35) if $\gamma'\geq\gamma''$ and $\eta'\leq\eta''$. Condition (b) implies that $\gamma'\geq\gamma''$ and $\eta'\leq\eta''$. ■

4 Comparative Statics

4.1 Changes in beliefs

We now examine the optimal design of education and educational participation when students' beliefs change. Let $\mathbf{p}$ denote the vector of beliefs $(p^1,p^2,\ldots,p^n)$. Let $\mathbf{p}'$ denote the vector of beliefs $(p'^1,p'^2\ldots,p'^n)$. For any $D,D'\subset I$, let $\overline{p}_D=\frac{1}{|D|}\sum_{i\in D}p^i$ and $\overline{p}'_{D'}=\frac{1}{|D'|}\sum_{i\in D'}p'^i$.

Corollary 1. *Suppose* $(\gamma^*,\eta^*)\in\Theta_{D^*}$ *is optimal for students with beliefs*

$\mathbf{p}$. *Suppose* $(\gamma', \eta') \in \Theta_{D'}$ *is optimal for students with beliefs* $\mathbf{p}'$. *Then the following statements are equivalent:*
(a) $\overline{p}_{D^*} < \overline{p}'_{D'}$,
(b) $\gamma^* > \gamma'$, *and*
(c) $\eta^* < \eta'$.

Proof. Suppose (a) holds. Then Theorem 2 implies (b) and (c). Suppose (b) or (c) holds. Then Lemma 5 implies (a). ■

This corollary means that when beliefs change, education should become more hunting-encouraging if and only if the average belief of the resulting participant set is more hunting-confident than that of the previous participant set. In the special case when the participant set remains the same as $D' = D^*$, the interpretation is simply that education should become more hunting-encouraging if and only if the current participants become more hunting-confident.

There are two cases when it is reasonable to assume $D' = D^*$. First, an obvious sufficient condition is that the educational attention budget B is large enough to always induce everyone to participate.

Corollary 2. *Suppose* $\{(\gamma, \eta) \in \Theta \mid C(\gamma, \eta) = B\} \subset \Theta_I$. *Then* $D' = D^* = I$.

Proof. Because $C(\gamma, \eta) = B$ implies $(\gamma, \eta) \in \Theta_I$, every optimal educational structure is in Θ_I. ■

Second, another sufficient condition is that the changes in beliefs are sufficiently small. Write the student welfare function as

$$W(\gamma, \eta, \mathbf{p}) = \sum_{i \in I} \max \left\{ V_0(p^i), V_1(p^i, \gamma, \eta) - \delta \right\}. \tag{40}$$

Corollary 3. *Suppose* (γ^*, η^*) *is the unique maximizer of* $W(\gamma, \eta, \mathbf{p})$ *subject to* $(\gamma, \eta) \in \widehat{\Theta}$. *Suppose* (γ^*, η^*) *is in the interior of* Θ_{D^*}. *Then there exists* $\varepsilon > 0$ *such that for any belief vector* $\mathbf{p}'$, $|\mathbf{p}' - \mathbf{p}| < \varepsilon$ *implies*

$$\underset{(\gamma,\eta)\in\widehat{\Theta}}{\operatorname{argmax}}\, W(\gamma, \eta, \mathbf{p}') \subset \Theta_{D^*} \tag{41}$$

Proof. By the Maximum Theorem, the maximizer of $W(\gamma, \eta, \mathbf{p})$ with respect to $(\gamma, \eta) \in \widehat{\Theta}$ is upper-hemicontinuous at $\mathbf{p}$. Since the maximizer is a singleton at $\mathbf{p}$, it is continuous at $\mathbf{p}$. ■

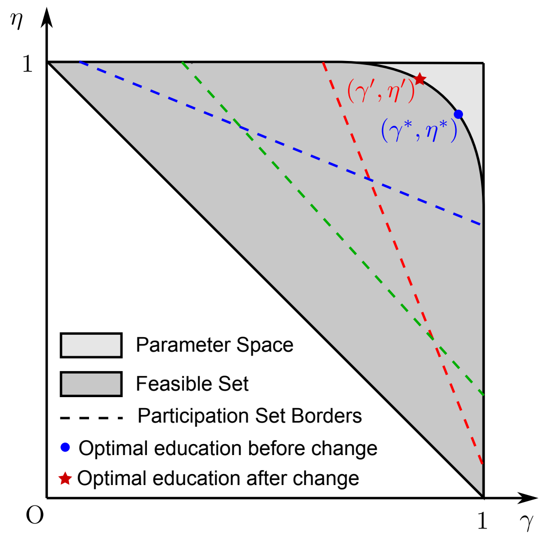

Figure 10: Every student participates regardless of changes in beliefs when the educational attention budget is large

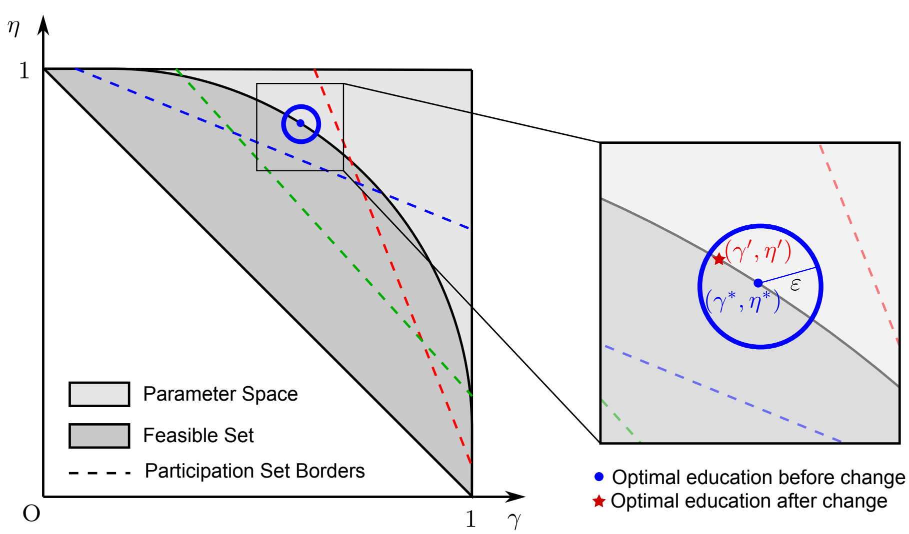

Figure 11: Participation set remains the same when changes in beliefs are small

Figures 10 and 11 illustrate the two sufficient conditions for the participant set to remain the same after changes in beliefs. When the educational attention budget is large (Figure 10, the upper-right boundary of the feasible set is contained in the polygon Θ_I, the set of educational structure that induce everyone to participate. When the changes in beliefs are small (Figures 11), the change in the optimal educational structure, too, is small. The new structure remains in the original polygon Θ_D.

Next, we take the educational structure as given and look at how students' participation and welfare change as students become more confident.

Definition. A belief vector $\mathbf{p}'$ is more *dispersed* than $\mathbf{p}$ if, for all $i \in I$, $p^i \leq \frac{u_g}{u_g+u_h}$ implies $p'^i \leq p^i$ and $p^i \geq \frac{u_g}{u_g+u_h}$ implies $p'^i \geq p^i$.

Corollary 4. *Let $(\gamma, \eta) \in \Theta$ be given. Suppose D and D' are the participant sets under (γ, η) when students' beliefs are $\mathbf{p}$ and $\mathbf{p}'$, respectively. Suppose $\mathbf{p}'$ is more dispersed than $\mathbf{p}$. Then*
(a) *$D' \subset D$, and*
(b) $W(\gamma, \eta, \mathbf{p}') - W(0.5, 0.5, \mathbf{p}') < W(\gamma, \eta, \mathbf{p}) - W(0.5, 0.5, \mathbf{p})$.

Proof. For all $i \in I$, $p'^i \leq p^i \leq \frac{u_g}{u_g+u_h}$ or $\frac{u_g}{u_g+u_h} \leq p^i \leq p'^i$. Suppose $i \in D'$. Then by Lemma 6, $i \in D$. Then $D' \subset D$.

From equation (29), every $j \in I$ satisfies

$$V_1(\gamma, \eta, p'^j) - V_0(p'^j) \leq V_1(\gamma, \eta, p^j) - V_0(p^j). \tag{42}$$

Then

$$\begin{aligned} W(\gamma, \eta, \mathbf{p}') - W(0.5, 0.5, \mathbf{p}') &= \sum_{j \in D} \max \left\{ V_1(\gamma, \eta, p'^j) - V_0(p'^j), 0 \right\} && (43) \\ &\leq \sum_{j \in D} \max \left\{ V_1(\gamma, \eta, p^j) - V_0(p^j), 0 \right\} && (44) \\ &= W(\gamma, \eta, \mathbf{p}) - W(0.5, 0.5, \mathbf{p}), && (45) \end{aligned}$$

as desired. ■

In other words, as students become more confident, the set of participating students srhinks and the welfare gain from education decreases.

4.2 Changes in technology

Consider changes in technology reflected through changes in the *ex post* productivities $u(\omega, a)$.

Corollary 5. *Let the belief vector* $\mathbf{p}$ *be given. Suppose* $(\gamma^*, \eta^*) \in \Theta_{D^*}$ *is optimal when the mismatch costs are* u_g *and* u_h*. Suppose* $(\gamma', \eta') \in \Theta_{D^*}$ *is optimal when the mismatch costs are* u'_g *and* u'_h*. Then the following statements are equivalent:*
(a) $\dfrac{u_g}{u_h} > \dfrac{u'_g}{u'_h}$,
(b) $\gamma^* > \gamma'$, *and*
(c) $\eta^* < \eta'$.

Proof. Theorem 2 implies that $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$ and $(\gamma', \eta') = F\left(\overline{p}_{D^*} \frac{u'_g}{u'_h} \frac{u_h}{u_g}\right)$. Then by Lemma 5, (a), (b), and (c) are equivalent. ■

This corollary means that the education should become more hunting-encouraging if and only if the technological change widens the mismatch cost of talented hunters. Consider a technological innovation that improves the productivity of talented hunters (ω_h) choosing the hunting career (a_h) but not others: for example, an introduction of a sophisticated weapon. Then the education should be more hunting-encouraging.

Note that this corollary assumes that the optimal educational structure remains the same before and after the technological change. As in the case of changes in beliefs, this assumption is reasonable if (a) the technological change is sufficiently small or (b) the educational attention budget (B) is sufficiently large.

4.3 Change in the cost of effort

An increase in the students' cost of effort shrinks the participation set and the welfare gain from education. Write the student welfare function as

$$W(\gamma, \eta, \mathbf{p}, \delta) = \sum_{i \in I} \max\left\{V_0(p^i), V_1(p^i, \gamma, \eta) - \delta\right\}. \tag{46}$$

Corollary 6. *Let the belief vector* $\mathbf{p}$ *be given. Let* $(\gamma, \eta) \in \Theta$ *be given. Suppose* D *and* D' *are the participant sets when the costs of participation are* δ *and* δ'*, respectively. Suppose* $\delta' > \delta$*. Then*
(a) $D' \subset D$, *and*
(b) $W(\gamma, \eta, \mathbf{p}, \delta') - W(0.5, 0.5, \mathbf{p}, \delta') < W(\gamma, \eta, \mathbf{p}, \delta) - W(0.5, 0.5, \mathbf{p}, \delta)$.

Proof. Lemma 6 implies (a). Equation (46) implies (b). ■

5 Extension

The baseline model examined so far yields readily interpretable properties. However, its assumptions may be too strong in practice, for example, in an empirical application. First, students' educational participation and career decisions are deterministic and do not allow errors. Second, students do not learn anything that directly improves their productivity. Third, students' prior beliefs are perfectly observable to an econometrician. Fourth, students' ability and career choices are binary. This section considers extensions of the model in these four directions.

5.1 Stochastic choice

Consider the following modifications to make students' choices stochastic. In the educational choice stage, a student $i \in I$ with prior belief p^i participates under educational structure $(\gamma, \eta) \in \Theta$ if and only if

$$V_1(p^i, \gamma, \eta) - \delta + \nu_1^i > V_0(p^i) + \nu_0^i, \tag{47}$$

where ν_0^i and ν_1^i are independently and identically distributed random variables. The interpretation is that these random variables represent mistakes in students' choices or factors other than their beliefs and the educational structure. Following the earlier definitions, the functions V_1 and V_0 are the *ex ante* productivities when a student with belief p chooses to participate under (γ, η) and does not, respectively. Then a student i's probability of participation given a prior belief p^i is

$$\begin{aligned} \Pr(d^i = 1|p^i) &= 1 - \Pr\left[\nu_1^i - \nu_0^i \leq V_0(p^i) - V_1(p^i, \gamma, \eta) + \delta\right] &(48)\\ &= 1 - G\left[V_0(p^i) - V_1(p^i, \gamma, \eta) + \delta\right], &(49)\end{aligned}$$

where G is the cumulative distribution function of the random variable $\nu_1^i - \nu_0^i$. For example, if both ν_1^i and ν_0^i are from Gumbel$(0, \beta)$, the cdf G is the logit function:

$$G(v) = \frac{e^{v/\beta}}{1 + e^{v/\beta}}. \tag{50}$$

In the career choice stage, A student $i \in I$ with a posterior belief q^i chooses the hunting career if and only if

$$U_g(q^i) + \nu_g^i \leq U_h(q^i) + \nu_h^i, \tag{51}$$

where ν_g^i and ν_h^i are independently and identically distributed random variables. The interpretation is the same as those for ν_0^i and ν_1^i. As in the baseline model, the functions U_g and U_h are the expected productivity of choosing gathering and hunting careers given posterior belief q:

$$U_g(q) = (1-q)u(\omega_g, a_g) + qu(\omega_h, a_g), \text{ and} \tag{52}$$

$$U_h(q) = (1-q)u(\omega_g, a_h) + qu(\omega_h, a_h). \tag{53}$$

Then a student i's probability of choosing the hunting career given a posterior belief q^i is

$$\Pr(a^i = a_h | q^i) = \Pr\left[\nu_g^i - \nu_h^i \leq U_h(q^i) - U_g(q^i)\right] \tag{54}$$

$$= H\left[U_h(q^i) - U_g(q^i)\right], \tag{55}$$

where H denotes the cumulative distribution function of the random varible $\nu_g^i - \nu_h^i$. For example, if ν_g^i and ν_h^i are from Gumbel$(0, \beta)$, $H(v) = \frac{e^{v/\beta}}{1+e^{v/\beta}}$.

With these choice probabilities, the function V_0 takes values

$$V_0(p) = [1 - \Pr(a_h|p)] \cdot U_g(p) + \Pr(a_h|p) \cdot U_h(p). \tag{56}$$

Recall that $Q_{\gamma,\eta}(p, s)$ is the posterior belief function whose value is the posterior belief after receiving signal s when the prior belief is p. The function V_1 takes values

$$V_1(p,\gamma,\eta) = \Pr(s_g|p,\gamma,\eta)V_0\big(Q_{\gamma,\eta}(p,s_g)\big) + \Pr(s_h|p,\gamma,\eta)V_0\big(Q_{\gamma,\eta}(p,s_h)\big). \tag{57}$$

The *ex ante* productivity of a student with belief p before making the participation decision is

$$V(p,\gamma,\eta) = [1 - \Pr(d=1|p,\gamma,\eta)]V_0(p) + \Pr(d=1|p,\gamma,\eta)V_1(p,\gamma,\eta). \tag{58}$$

The *student welfare function* is

$$W(\gamma,\eta) = \sum_{i \in I} V(p^i, \gamma, \eta). \tag{59}$$

An educational structure (γ^*, η^*) is *optimal* if it maximizes $W(\gamma,\eta)$ subject to $(\gamma,\eta) \in \widehat{\Theta}$.

In general, it is difficult to characterize the optimal design of education in this stochastic choice setting. However, there is a special case in which the optimum is the same as in the baseline model.

Definition. The extended model with stochastic choice features *fixed proba-*

bilities of mistakes if, for some for some $\alpha, \beta \in (0, 0.5)$

$$G(v) = \begin{cases} \alpha & \text{if } v \leq 0, \\ 1 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad H(v) = \begin{cases} \beta & \text{if } v \leq 0, \\ 1 & \text{otherwise.} \end{cases}$$

A set $D \subset I$ is an *expected participant set* under $(\gamma, \eta) \in \Theta)$ if

$$D = \left\{ i \in I | V_1(p^i, \gamma, \eta) - \delta > V_0(p^i) \right\}. \tag{60}$$

Let Θ_D denote the set of educational structures whose expected participant set is D. An educational structure $(\gamma, \eta) \in \Theta_D$ is *nontrivial* if D is nonempty.

Theorem 4. *Theorem 1 holds in the extended model with fixed probabilities of mistakes. That is, suppose $(\gamma^*, \eta^*) \in \widehat{\Theta}$ is nontrivial. Then (γ^*, η^*) is optimal if and only if there exists $D^* \subset I$ that satisfies $(\gamma^*, \eta^*) = F(\bar{p}_{D^*})$ and solves*

$$\max_{D \in \mathcal{P}(I)} W\big(F(\bar{p}_D)\big). \tag{61}$$

Proof. Fix a nontrivial educational structure $(\gamma, \eta) \in \Theta$. Observe that

$$V(p, \gamma, \eta) = \max \Big\{ (1 - \alpha)[V_1(p, \gamma, \eta) - \delta] + \alpha V_0(p), \tag{62}$$

$$\alpha[V_1(p, \gamma, \eta) - \delta] + (1 - \alpha)V_0(p) \Big\}. \tag{63}$$

We have

$$V_1(p, \gamma, \eta) = \sum_{\omega \in \Omega} \sum_{s \in S} \sum_{a \in A} \Pr(\omega) \Pr(s \,|\, \omega) \Pr(a \,|\, s) u(\omega, a). \tag{64}$$

Let D denote the expected participant set under (γ, η). For all $i \in D$, $\Pr(a_g | s_g) = \Pr(a_h | s_h) = 1 - \beta$ and $\Pr(a_h | s_g) = \Pr(a_g | s_h) = \beta$, so

$$V_1(p^i, \gamma, \eta) = p^i \left[\gamma\big((1 - \beta)u_{gg} + \beta u_{gh}\big) + (1 - \gamma)\big(\beta u_{gg} + (1 - \beta)u_{gh}\big) \right] \tag{65}$$

$$+ (1 - p^i) \left[(1 - \eta)\big((1 - \beta)u_{hg} + \beta u_{hh}\big) + \eta\big(\beta u_{hg} + (1 - \beta)u_{hh}\big) \right], \tag{66}$$

where $u_{gg} = u(\omega_g, a_g)$, $u_{gh} = u(\omega_g, a_h)$, $u_{hg} = u(\omega_h, a_g)$, and $u_{hh} = u(\omega_h, a_h)$. Then for all $i \in D$,

$$\frac{\partial}{\partial \gamma} V_1(p^i, \gamma, \eta) = (1 - 2\beta)(1 - p)u_g, \tag{67}$$

$$\frac{\partial}{\partial \eta} V_1(p^i, \gamma, \eta) = (1 - 2\beta)p u_h. \tag{68}$$

For those not in the expected participant set under (γ, η), either $\Pr(a_g|s_g) = \Pr(a_g|s_h) = 1 - \beta$ or $\Pr(a_h|s_g) = \Pr(a_h|s_h) = 1 - \beta$, so

$$V_1(p^i, \gamma, \eta) = V_0(p^i). \tag{69}$$

Then

$$V(p^i, \gamma, \eta) = (1 - 2\alpha) \cdot \max\left\{V_1(p^i, \gamma, \eta) - V_0(p^i) - \delta, 0\right\} + V_0(p^i), \tag{70}$$

thus

$$W(\gamma, \eta) = (1 - 2\alpha) \cdot \sum_{i \in D} [V_1(p, \gamma, \eta) - V_0(p) - \delta] + \sum_{i \in I} V_0(p^i). \tag{71}$$

Lemmas 1–3 hold in the extended model from the above equation. Therefore, wherever W is differentiable, we have

$$\frac{\frac{\partial}{\partial \gamma} W(\gamma, \eta)}{\frac{\partial}{\partial \eta} W(\gamma, \eta)} = \frac{\sum_{i \in D} \frac{\partial}{\partial \gamma} V_1(\gamma, \eta)}{\sum_{i \in D} \frac{\partial}{\partial \eta} V_1(\gamma, \eta)} = \frac{\sum_{i \in D} (1 - p^i) u_g}{\sum_{i \in D} p^i u_h} = \frac{1 - \overline{p}_D}{\overline{p}_D} \frac{u_g}{u_h}. \tag{72}$$

So Lemma 4 holds in the extended model. The rest follows from the same proof as in Theorem 1. ■

Corollary 7. *Theorem 2 holds in the extended model with a fixed probability ε of mistakes. That is, suppose $(\gamma^*, \eta^*) \in \Theta_{D^*}$ is nontrivial and optimal. Then $\overline{p}_{D^*} = \frac{u_g}{u_g + u_h}$ implies $\gamma^* = \eta^*$, $\overline{p}_{D^*} < \frac{u_g}{u_g + u_h}$ implies $\gamma^* > \eta^*$, and $\overline{p}_{D^*} > \frac{u_g}{u_g + u_h}$ implies $\gamma^* < \eta^*$.*

Proof. From Theorem 4, $\overline{p}_{D^*} = \frac{u_g}{u_g + u_h}$ implies $\gamma^* = \eta^*$. Observe that the first component of F is decreasing in p and the second component of $F(p)$ is increasing in p. Then $\overline{p}_{D^*} < \frac{u_g}{u_g + u_h}$ implies $\gamma^* > \eta^*$, and $\overline{p}_{D^*} > \frac{u_g}{u_g + u_h}$ implies $\gamma^* < \eta^*$. ■

This corollary mean that, with a fixed probability of mistakes, an optimal educational structure encourages a career in which only the expected participants are more confident on average. By expected participants, we mean those who intend to participate, not those who end up participating unintentionnaly by mistake. The reason is that the unintended participants do not value the information from educational signals in the first place. Furthermore, the same belief threhold $u_g u_g + u_h$ determines whether a set of expected participants are confident in the gathering or the hunting career. The result arises because the fixed probability of mistakes shrinks the expected mismatch costs proportionally. Figure 12 illustrates this point.

Figure 12: Fixed probability of mistakes shrinks expected mismatch costs proportionally

5.2 Human capital accumulation

Next, we consider an extended model with *human capital accumulation*. That is, education has not only imforms students about their talents but also directly increases students' *ex post* productivity.

Define the *human capital accumulation function* as $\psi : \Omega \times A \longrightarrow \mathbb{R}$ that maps (ω, a) to the increase in the *ex post* productivity for a participating student with talent ω and career a. Then a student's *ex post* payoff is

$$u(\omega, a) + [\psi(\omega, a) - \delta]d, \tag{73}$$

where d is the indicator of participation. Let $\psi_g = \psi(\omega_g, a_g) - \psi(\omega_g, a_h) > -u_g$ and $\psi_h = \psi(\omega_h, a_h) - \psi(\omega_h, a_g) > -u_h$. Let $F_\psi : (0,1) \longrightarrow \widehat{\Theta}$ map p to the solution (γ, η) to the system of equations

$$\frac{\frac{\partial}{\partial\gamma}C(\gamma,\eta)}{\frac{\partial}{\partial\eta}C(\gamma,\eta)} = \frac{1-p}{p}\frac{u_g+\psi_g}{u_h+\psi_h} \quad \text{and} \quad C(\gamma,\eta) = B. \tag{74}$$

A participating student is a *complier* if $\Pr(a_g|s_g) = \Pr(a_h|s_h) = 1$. A student is an *always-hunter* if $\Pr(a_h|s_g) = \Pr(a_h|s_h) = 1$. A student is a *always-gatherer* if $\Pr(a_g|s_g) = \Pr(a_g|s_h) = 1$. Let Θ_D denote the set of educational structures under which the set of participating compliers is D. An educational structure $(\gamma, \eta) \in \Theta_D$ is *nontrivial* if D is nonempty.

Theorem 5. *Theorem 1 holds in the extended model with human capital accumulation if F is replaced with F_ψ. That is, suppose $(\gamma^*, \eta^*) \in \widehat{\Theta}$ is* *nontrivial. Then* (γ^*, η^*) *is optimal if and only if there exists* $D^* \subset I$ *that satisfies* $(\gamma^*, \eta^*) = F_\psi(\bar{p}_{D^*})$ *and solves*

$$\max_{D \in \mathcal{P}(I)} W\big(F_\psi(\bar{p}_D)\big). \tag{75}$$

Proof. Let $(\gamma, \eta) \in \Theta_D$ be given. Let D_g denote the set of participating always-gatherers under (γ, η). Let D_h denote the set of participating always-hunters under (γ, η). We have

$$V(p, \gamma, \eta) = \max\{V_1(p, \gamma, \eta) - \delta, V_0(p)\}. \tag{76}$$

We have

$$V_1(p, \gamma, \eta) = \sum_{\omega \in \Omega} \sum_{s \in S} \sum_{a \in A} \Pr(\omega) \Pr(s \,|\, \omega) \Pr(a \,|\, s) u(\omega, a). \tag{77}$$

For all $i \in D$, $\Pr(a_g|s_g) = \Pr(a_h|s_h) = 1$. Then for all $i \in D$,

$$V_1(p, \gamma, \eta) = (1-p)\big[\gamma(u_{gg} + \psi_{gg}) + (1-\gamma)(u_{gh} + \psi_{gh})\big] \tag{78}$$

$$+ p\big[(1-\eta)(u_{hg} + \psi_{hg}) + \eta(u_{hh} + \psi_{hh})\big], \tag{79}$$

denoting $u_{gh} = u(\omega_g, a_h)$, $\psi_{gh} = \psi(\omega_g, a_h)$, and so on. Then for all $i \in D$,

$$\frac{\partial}{\partial \gamma} V_1(p^i, \gamma, \eta) = (1-p)(u_g + \psi_g), \tag{80}$$

$$\frac{\partial}{\partial \eta} V_1(p^i, \gamma, \eta) = p(u_h + \psi_h). \tag{81}$$

In contrast, $\Pr(a_g|s_g) = \Pr(a_g|s_h) = 1$ for all $i \in D_g$ and $\Pr(a_h|s_g) = \Pr(a_h|s_h) = 1$ for all $i \in D_h$. Writing $\Psi_g(p) = (1-p)\psi(\omega_g, a_g) + p\psi(\omega_h, a_g)$ and $\Psi_h(p) = (1-p)\psi(\omega_g, a_h) + p\psi(\omega_h, a_h)$, we have

$$V_1(p, \gamma, \eta) = V_0(p) + \Psi_g(p) \quad \text{for all } i \in D_g, \text{ and} \tag{82}$$

$$V_1(p, \gamma, \eta) = V_0(p) + \Psi_h(p) \quad \text{for all } i \in D_h. \tag{83}$$

Then

$$W(\gamma, \eta) = \sum_{i \in D} V_1(p^i, \gamma, \eta) + \sum_{i \in D_g} [V_0(p) + \Psi_g(p)] + \sum_{i \in D_h} [V_0(p) + \Psi_h(p)] \tag{84}$$

$$+ \sum_{i \in I \setminus (D \cup D_g \cup D_h)} V_0(p). \tag{85}$$

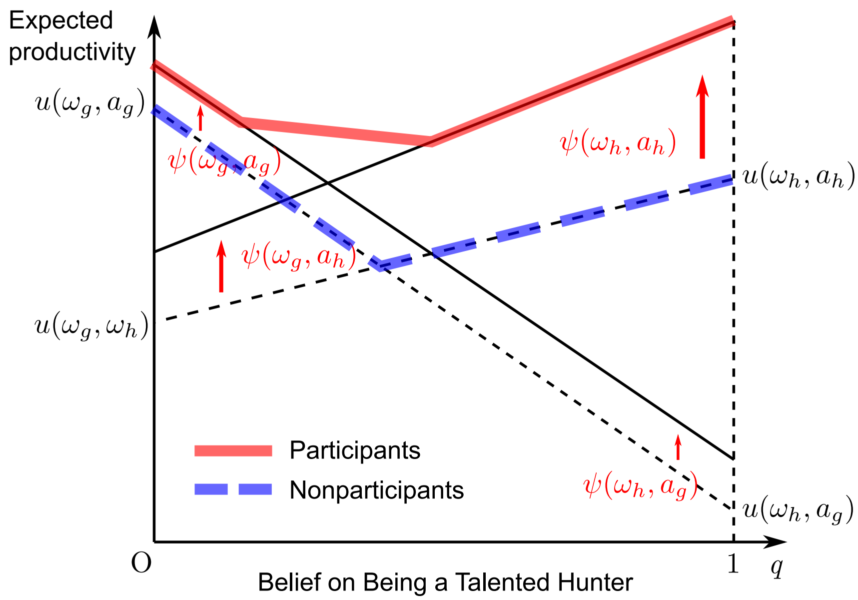

Figure 13: Human capital accumulation changes the mismatch costs for participants

Lemmas 1–3 hold in the extended model from the above equation. Therefore, wherever W is differentiable, we have

$$\frac{\frac{\partial}{\partial \gamma}W(\gamma,\eta)}{\frac{\partial}{\partial \eta}W(\gamma,\eta)} = \frac{\sum_{i\in D}\frac{\partial}{\partial \gamma}V_1(\gamma,\eta)}{\sum_{i\in D}\frac{\partial}{\partial \eta}V_1(\gamma,\eta)} = \frac{\sum_{i\in D}(1-p^i)(u_g+\psi_g)}{\sum_{i\in D}p^i(u_h+\psi_h)} = \frac{1-\overline{p}_D}{\overline{p}_D}\frac{u_g+\psi_g}{u_h+\psi_h}. \tag{86}$$

So Lemma 4 holds in the extended model. The rest follows from the same proof as in Theorem 1. ■

Corollary 8. *Theorem 2 holds in the extended model with a fixed probability* ε *of mistakes if* $\frac{u_g}{u_g+u_h}$ *is replaced with* $\frac{u_g+\psi_g}{u_g+\psi_g+u_h+\psi_h}$*. That is, suppose* $(\gamma^*,\eta^*)\in\Theta_{D^*}$ *is nontrivial and optimal. Then* $\overline{p}_{D^*}=\frac{u_g+\psi_g}{u_g+\psi_g+u_h+\psi_h}$ *implies* $\gamma^*=\eta^*$*,* $\overline{p}_{D^*}<\frac{u_g+\psi_g}{u_g+\psi_g+u_h+\psi_h}$ *implies* $\gamma^*>\eta^*$*, and* $\overline{p}_{D^*}>\frac{u_g+\psi_g}{u_g+\psi_g+u_h+\psi_h}$ *implies* $\gamma^*<\eta^*$*.*

Proof. From Theorem 5, $\overline{p}_{D^*}=\frac{u_g+\psi_g}{u_g+\psi_g+u_h+\psi_h}$ implies $\gamma^*=\eta^*$. Observe that the first component of $F_\psi(p)$ is decreasing in p and the second component of $F_\psi(p)$ is increasing in p. Then $\overline{p}_{D^*}<\frac{u_g+\psi_g}{u_g+\psi_g+u_h+\psi_h}$ implies $\gamma^*>\eta^*$, and $\overline{p}_{D^*}>\frac{u_g+\psi_g}{u_g+\psi_g+u_h+\psi_h}$ implies $\gamma^*<\eta^*$. ■

This corollary means that, with human capital accumulation, an optimal educational structure encourages a career in which only the complying participants are confident on average. This result is due to the fact that participating always-gatherers and always-hunters participate in education only because of its human capital value, not its information value. Meanwhile, the threshold belief level adjusts to $\frac{u_g+\psi_g}{u_g+\psi_g+u_h+\psi_h}$ because the direct changes to ex post productivities change the mismatch costs for all participants. Figure 13 illustrates this point.

5.3 Multidimensional talent and many careers

This extension allows students to have multidimensional ability and to choose from more than two careers.

Let the space of talents (states) Ω be any set. For example, let $\Omega \subset \mathbb{R}^m$ for some positive integer m. Let the set of careers (actions) A be any finite set, i.e. $A = \{a_1, a_2, \ldots, a_k\}$. Each student i has a publicly known rational belief $p^i \in \Delta(\Omega)$. The *productivity function* is $u : \Omega \times A \longrightarrow \mathbb{R}$. The *human capital function* is $u : \Omega \times A \longrightarrow \mathbb{R}$. An individual i's *ex post payoff* is

$$u(\omega^i, a^i) + [\psi(\omega^i, a^i) - \delta]d^i, \tag{87}$$

where $\delta \geq 0$ is the cost of participation and d is the indicator of a student's participation. An individual i's *outcome* is

$$y_i = u(\omega^i, a^i) + \psi(\omega^i, a^i)d^i + \varepsilon_i, \tag{88}$$

where ε_i is an independently and identically distributed error. The *educational signal space* is $S = \{s_1, s_2, \ldots, s_k\}$. *Education* is a signal from a conditional probability $\pi : \Omega \longrightarrow \Delta(S)$. *Educational structure* refers to the conditional probability π.

The expected productivity of choosing a career a without human capital accumulation is

$$U_a(p) = \int_{\omega \in \Omega} u(\omega, a) dp(\omega). \tag{89}$$

The expected human capital accumulation of choosing a career a is

$$\Psi_a(p) = \int_{\omega \in \Omega} \psi(\omega, a) dp(\omega). \tag{90}$$

For all $a \in A$, let ν_a be an with an independently and identically distributed random variable. A student with a participation decision d and a posterior belief q^i chooses a career a if

$$U_a(q) + \Psi_a(q)d + \nu_a \geq U_{a'}(q) + \Psi_a(q)d + \nu_{a'}. \tag{91}$$

Let $\mu_a(q)$ denote the probability of the event (91) when $d = 0$ and let $\mu_a^{\psi}(q)$ denote the probability of the event (91) when $d = 1$. A non-participant's expected productivity is

$$V_0(p) = \sum_{a \in A} \mu_a(p) U_a(p). \tag{92}$$

A participant's expected productivity given a posterior belief $q \in \Delta(\Omega)$ is

$$V_0^{\psi}(q) = \sum_{a \in A} \mu_a^{\psi}(q) \left[U_a(q) + \Psi_a(q) \right]. \tag{93}$$

Let $Q_\pi(p, s)$ denote the posterior belief for a student with belief p who receives a signal $s \in S$ from an educational structure π. A participant's expected productivity is

$$V_1^{\psi}(p, \pi) = \mathbb{E}\left[V_0^{\psi}\big(Q_\pi(p, s)\big) \,\Big|\, p \right] \tag{94}$$

The individual *treatment effect* for student i is

$$\tau^i = \mathbb{E}[y_i | i, d^i = 1] - \mathbb{E}[y_i | i, d^i = 0]. \tag{95}$$

The individual *conditional treatment effect* for student i given a career a is

$$\tau_a^i = \mathbb{E}[y_i | i, d^i = 1, a^i = a] - \mathbb{E}[y_i | i, d^i = 0, a^i = a]. \tag{96}$$

Theorem 6. *The individual treatment effect satisfies*

$$\tau^i = \sum_{a \in A} \mu_a(p^i) \tau_a^i + V_1^{\psi}(p^i, \pi) - V_0^{\psi}(p^i). \tag{97}$$

Proof. We have

$$\begin{aligned} \mathbb{E}[y_i | i, d^i = 1] &= \sum_{a \in A} \sum_{s \in S} \Pr(a, s | i, d^i = 1) \mathbb{E}[y_i | i, d^i = 1, a^i = a, s^i = s] && (98) \\ &= \sum_{s \in S} \Pr(s | i, d^i = 1) V_0^{\psi}\left(Q_\pi(p^i, s)\right) && (99) \\ &= V_1^{\psi}\left(p^i, \pi\right), && (100) \end{aligned}$$

where the second equality uses the fact that $\Pr(a, s | i, d^i = 1) = \Pr(a | i, d^i = 1, s^i = s) \Pr(s | i, d^i = 1)$ and the definition of V_0^{ψ}. Also,

$$\mathbb{E}[y_i | i, d^i = 0] = \sum_{a \in A} \Pr(a | i, d^i = 0) \mathbb{E}[y_i | i, d^i = 0, a^i = a]. \tag{101}$$

Then

$$\tau^i = V_1^\psi(p^i, \pi) - \sum_{a \in A} \mu_a(p^i) \mathbb{E}[y_i | i, d^i = 0, a^i = a] \tag{102}$$

$$= V_1^\psi(p^i, \pi) - \sum_{a \in A} \mu_a(p^i) \mathbb{E}[y_i | i, d^i = 1, a^i = a] \tag{103}$$

$$+ \sum_{a \in A} \mu_a(p^i) \mathbb{E}[y_i | i, d^i = 1, a^i = a] - \sum_{a \in A} \mu_a(p^i) \mathbb{E}[y_i | i, d^i = 0, a^i = a] \tag{104}$$

$$= V_1^\psi(p^i, \pi) - V_0^\psi(p^i) + \sum_{a \in A} \mu_a(p^i) \tau_a^i, \tag{105}$$

as desired. ■

This theorem means that the individual treatment effect is decomposed into the sum of individual conditional treatment effects across all careers and the value of information from the educational signal. The first component represents the human capital accumulation from education, whereas the second component represents the self-exploration value of education.

Define the (local) *average treatment effect* for a set $D \subset I$ of students as

$$ATE_D = \mathbb{E}[\tau^i \,|\, i \in D]. \tag{106}$$

Define the (local) *conditional average treatment effect* for a subset D of choosing a career a as

$$CATE_D(a) = \mathbb{E}[\tau_a^i \,|\, i \in D]. \tag{107}$$

Corollary 9. *Suppose that $\mu_a(p^i)$ and τ_a^i are uncorrelated for a set D of students. Then*

$$ATE_D = \sum_{a \in A} \mathbb{E}[\mu_a(p^i)] \cdot CATE_D(a) + \mathbb{E}\left[V_1^\psi(p^i, \pi) - V_0^\psi(p^i) \,\middle|\, i \in D\right]. \tag{108}$$

This corollary suggests that the (local) average value of self-exploration from education can be estimated as a residual. It is the difference between the aggregate return to education (the average treatment effect) and the return to education that controls for occupational choices (an weighted average of the conditional average treatment effects). Therefore, one can test the significance of the self-exploration value of education by estimating the size of this residual.

For example, Lemieux (2014) uses Canadian survey and census data of about 11,000 respondents to estimate the return to college education (a Bachelor's degree) on earnings with (*i.e.* CATE) and without (*i.e.* ATE) occupational controls. Estimating that the ATE is about 55 percent and the weighted average of CATE's across 24 occupational categories is about 31 percent, he concludes that the difference of about 24 percentage points is due to the better occupational choices rather than direct increase in productivity.

5.4 Imperfectly observed beliefs

The next extension addresses a practical concern that an econometrician typicaly does not observe students' beliefs as probabilities but rather as a measure on some arbitrary scale. For example, many survey data report respondents answers as an index or standardized scores.

Suppose students $i \in I = \{1, \ldots, n\}$ are randomly sampled from a population. Let us maintain that the prior belief of each student $i \in I$ is $p^i \in (0, 1)$ and his participation decision is $d_i \in \{0, 1\}$. If he does not participate, he receives no signal. If he participates, he receives a signal $s_i \in \{s_g, s_h\}$ with conditional probabilities $P(s_g|\omega_g) = \gamma$ and $P(s_h|\omega_h) = \eta$. The posterior belief satisfies

$$q^i = \begin{cases} p^i, & \text{if } d_i = 0 \\ Q_{\gamma,\eta}(p^i, s) & \text{if } d_i = 1 \end{cases} \tag{109}$$

From Bayes rule, the posterior belief function Q takes values

$$Q_{\gamma,\eta}(p^i, s_g) = \frac{(1-\eta)p^i}{\gamma(1-p^i) + (1-\eta)p^i}, \text{ and} \tag{110}$$

$$Q_{\gamma,\eta}(p^i, s_h) = \frac{\eta p^i}{(1-\gamma)(1-p^i) + \eta p^i}. \tag{111}$$

Suppose an econometrician does not observe these beliefs but only observes their standardized values. That is, the econometrician observes the *prior belief scores* $x_1, x_2, \ldots, x_n$ and the *posterior belief scores* $z_1, z_2, \ldots, z_n$, where, for every student $i \in I$,

$$x_i = \frac{p^i - \mu_p}{\sigma_p}, \tag{112}$$

$$z_i = \frac{q_i - \mu_q}{\sigma_q}. \tag{113}$$

The constants μ_p and μ_q are the means of prior and posterior beliefs of all students and σ_p and σ_q are their standard deviations.

Definition. The *difference-in-differences* in belief scores between participants and non-participants is

$$\phi = \Big(\mathbb{E}[z_i \,|\, d_i = 1] - \mathbb{E}[x_i \,|\, d_i = 1]\Big) - \Big(\mathbb{E}[z_i \,|\, d_i = 0] - \mathbb{E}[x_i \,|\, d_i = 0]\Big). \quad (114)$$

Theorem 7. *Suppose $\mathbb{E}[x_i \,|\, d_i = 1] > \mathbb{E}[x_i \,|\, d_i = 0]$. Then the educational program is informative if and only if $\phi < 0$.*

Proof. Observe that $q_i = p^i + \epsilon_i$ where

$$\epsilon_i = \begin{cases} 0, & \text{if } d_i = 0 \\ \dfrac{(1-\eta)p^i}{\gamma(1-p^i)+(1-\eta)p^i} - p^i, & \text{if } d_i = 1 \text{ and } s_i = s_g, \text{ and} \\ \dfrac{\eta p^i}{(1-\gamma)(1-p^i)+\eta p^i} - p^i, & \text{if } d_i = 1 \text{ and } s = s_h. \end{cases} \quad (115)$$

Then $\mathbb{E}[\epsilon_i|p^i] = 0$, thus $\mu_q = \mu_p$ and $\sigma_q^2 = \sigma_p^2 + \mathrm{Var}(\epsilon)$. Also, $\mathrm{Var}(\epsilon) > 0$ if and only if $\gamma + \eta > 1$. So the educational program is informative if and only if $\sigma_q > \sigma_p$.

Moreover, from equations (112)–(113) and the fact that $\mu_q = \mu_p$, we have

$$z_i = \frac{\sigma_p}{\sigma_q} x_i + \frac{\epsilon_i}{\sigma_q}. \quad (116)$$

By substituting for z_i into (114), we obtain

$$\phi = \left(\frac{\sigma_p}{\sigma_q} - 1\right)\Big(\mathbb{E}[x_i \,|\, d_i = 1] - \mathbb{E}[x_i \,|\, d_i = 0]\Big). \quad (117)$$

Since $\mathbb{E}[x_i \,|\, d_i = 1] > \mathbb{E}[x_i \,|\, d_i = 0]$, the condition $\sigma_q > \sigma_p$ is equivalent to the condition $\phi < 0$. Therefore, the educational program is informative if and only if $\phi < 0$. ■

This theorem means that the difference-in-differences in belief scores between participants and non-participants only tells us how the educational structure is informative about students' talents. It does not tell us whether students become more confident in one direction or not. The reason is that participants' beliefs become more dispersed if and only if the educational structure is informative. The participants' average posterior belief remains at the average prior belief, by Bayes plausibility. Therefore, if the educational structure is informative, the greater dispersion shrinks the participants' posterior belief scores toward zero. If the educational structure is uninformative, the participants' posterior belief scores remain at the prior belief scores.

With a dataset of belief scores x_i and z_i, define variables $y_{i0} = x_i$, $y_{i1} = z_i$, $post_0 = 0$, and $post_1 = 1$ for all $i \in I$. We can estimate the coefficients of the linear regression

$$y_{it} = \beta_0 + \beta_1 \, d_i + \beta_2 \, post_t + \phi \, d_i \times post_t + \varepsilon_{it}, \tag{118}$$

where the error term ε satisfies $\mathbb{E}[\varepsilon_{it} | d_i, post_t] = 0$. Then the coefficient ϕ from this equation satisfies the definition of the difference-in-differences, and is consistently estimated with the ordinary least squares (OLS) estimator.

Table 1 shows an example of such difference-in-difference linear regression estimates. The High School Longitudinal Study (HSLS) of 2009 contains data on students' *self-efficacy* in sciences and maths in standard scores in their 9th and 12th grades. Self-efficacy refers to the tendency to agree that the respondent is confident about doing well in exams, understanding textbooks, mastering the skills, and doing the required assignments in the subject. The dataset also contains information on students' coursework in advanced sciecne or math classes during high school. The estimate of the difference-in-differences in science self-efficacy scores is significantly negative, implying that advanced science classes in high school are informative about students' own science ability. In contrast, the same estimate in math self-efficacy score does not significantly differ from zero, implying that advanced math classes in high school are not informative about students' own math ability. Figure 1 from the Introduction is a graphical representation of the estimated regression of science self-efficacy.

6 Empirical Application

This section presents an empirical application of the self-exploration model of education. I estimate how science-encouraging Advanced Placement science classes are in high schools in the United States.

Advanced Placement (AP) classes are an important part of American high school education. They are college-level classes available to participating high schools in various subjects such as sciences (physics, chemistry, biology, and environmental science), maths (calculus and statistics), social studies, and English. The program is popular because it lets advanced students pursue a subject more deeply and earn college course credit. 38 percent of U.S. public high school graduates in the class of 2020 have participated in at least one AP class. (College Board, 2020).

AP science classes are especially significant in education policy because of its potential role in encouraging students into sciences. For example, former

Table 1: Advanced math and science classes and students' self-efficacy scores: difference-in-differences regressions

	Dependent variable: Standard score in	
Independent variables	Science Self-efficacy	Math Self-efficacy
Intercept	-0.026***	-0.005
	(0.009)	(0.009)
Advanced science participation	0.499***	
	(0.020)	
Advanced math participation		0.458***
		(0.019)
Post	0.041***	-0.018*
	(0.011)	(0.010)
Advanced science participation × Post	-0.228***	
	(0.025)	
Advanced math participation × Post		0.013
		(0.022)
R-squared	0.024	0.031
Observations per student	2	2
Number of participants	2,587	2,702
Number of non-participants	11,860	13,154

Note: This table shows the ordinary least square (OLS) estimates of the difference-in-differences regression (118) using the public-use dataset of the High School Longitudinal Study (HSLS) of 2009. *Science self-efficacy* and *math self-efficacy* refers to the tendency to agree that the respondent is confident about doing well in exams, understanding textbooks, mastering the skills, and doing the required assignments in the sciences and maths, respectively. *Advanced science participation* equals 1 if a student participated in at least one Advanced Placement science class and equals 0 otherwise. *Advanced science participation* equals 1 if a student participated in at least one Advanced Placement science math and equals 0 otherwise. *Post* equals 0 if the student was in 9th grade and equals 1 if the student was in 12th grade. Numbers in paranthese are robust standard errors clustered for each student. Stars *, **, and *** indicate statistical significance at 10, 5, and 1 percent levels.

U.S. President Geoarge W. Bush (2006) supported expanding the AP science courses for this reason:

> Third, we need to encourage children to take more math and science, and to make sure those courses are rigorous enough to compete with other nations. [...] Tonight I propose to train 70,000 high school teachers to lead Advanced Placement courses in math and science, bring 30,000 math and science professionals to teach in classrooms, and give early help to students who struggle with math, so they have a better chance at good, high-wage jobs. [...]

My estimates of the educational structure of the AP science classes suggest that they are indeed informative about students' science ability and are science-encouraging.

6.1 Data

I use the public-use dataset of High School Longitudinal Study (HSLS). It contains survey and transcript data following a nationally representative sample of about 23,503 students who were 9th graders in 2009, from 944 high schools across 10 states in the United States. The dataset contains observations from three periods: years 2009, 2012, and 2016. I use the data on the students' initial self-efficacy standard scores x_i in science , their participation decisions d_i into any AP science classes, their decisions a_i to pursue a science (STEM) major in college, and their log earnings[3] y_i three years after high school graduation.[4] The variable d_i equals 1 if the student participated in at least one AP science class, and 0 otherwise. The variable a_i equals 1 if the student indicated at the end of high school that he will pursue science major in college. After dropping observations with missing data, I retain 13,003 observations of students for the baseline sample.

Table 2 shows the summary statistics of these variables as well as the proportion of female students in the sample. Although the prior belief scores are standard scores, students in the sample have an average belief slightly greater than zero after dropping missing observations. About 18 percent of the students in the sample participated in at least one advanced science class during their high school years. About 20 percent of the students in the sample indicated that they would pursue a science major in college.

[3]The earnings data are recorded in 13 different intervals: 0, [1,1000], [1001, 2500], [2501, 5000], [5001,10001], $\ldots$, and [55001,75000]. I merge the first two intervals and let each y_i be the log of the midpoint of the interval that it belongs to.

[4]This earnings measure has two shortcomings in this application. First, we need students' lifetime earnings rather than earnings in a given year. Second, students in science majors typically complete 4-year college degrees, so most of their earnings are from summer or part-time jobs. However, these shortcomings should not bias the estimation of the educational structure as long as it does not distort the relative size of the mismatch costs.

Table 2: Summary statistics

Variable	Symbol	Obs	Mean	Std Dev
Prior belief score in science	x	13,003	0.066	0.988
Advanced Science (AP) science participation	d	13,003	0.183	0.387
Decision to pursue science major in college	a	13,003	0.195	0.397
Log earnings 3 years after graduation	y	13,003	1.577	1.371
Female	none	13,003	0.518	0.500

Note: The summary statistics are from a baseline sample of 13,003 high school students who were 9th graders in 2009. The sample uses the public-use dataset of High School Longitudinal Study (HSLS). Log earnings are in thousands of U.S. dollars.

6.2 Estimation

I infer the parameters of the extended model using maximum likelihood estimation. The baseline model analyzed in Sections 2–3 are not adequate for estimation because its predictions are deterministic, even though students' decisions in the data are noisy. The extended model for estimation in this section includes all of the three additional components introduced in Section 5: stochastic choice, human capital accumulation, and imperfectly observed beliefs.

Consider a student $i \in I$ and let us suppress the index i for now. As a convention, interpret the hunting talent (ω_h) and hunting career (a_h) as the science talent and science career. Similarly, interpret the gathering talent (ω_g) and gathering career (a_g) as the non-science talent and non-science career. Maintain that the prior belief score x in science satisfies

$$x = \frac{p - \mu_p}{\sigma_p}, \tag{119}$$

where p is the student's prior belief in science that is unobservable to the econometrician. The student's participation decision follows

$$d = \begin{cases} 1, & \text{if } V_1(p, \gamma, \eta) - \delta + \nu_1 > V_0(p) + \nu_0 \\ 0, & \text{otherwise.} \end{cases} \tag{120}$$

The error terms ν_0 and ν_1 are independently and identically distributed as Gumbel$(0, \beta)$. The student's decision to pursue a science major in college follows

$$a = \begin{cases} 0, & \text{if } U_g(q) + \nu_g > U_h(q) + \nu_h \\ 1, & \text{otherwise,} \end{cases} \tag{121}$$

by letting $a_g = 0$ and $a_h = 1$. The error terms ν_g and ν_h are independently and identically distributed as Gumbel$(0, \beta)$. Finally, a person's ex-post productivity y is given as

$$y = u(\omega, a) + [\psi(\omega, a) - \delta]d + \varepsilon, \tag{122}$$

where the error term ε is independently and identically distributed as $N(0, \sigma_\varepsilon)$. Impose that

$$\psi(\omega_g, a_g) = \psi(\omega_h, a_g) = 0, \text{and} \tag{123}$$

$$\psi(\omega_g, a_h) = \psi(\omega_h, a_h) = \overline{\psi}. \tag{124}$$

The interpretation is that taking an advanced science class improves one's science human capital only, by $\overline{\psi}$.

Denote the parameter vector as

$$\Lambda = \left(\mu_p, \sigma_p, \gamma, \eta, \delta, \beta, u_{gg}, u_{gh}, u_{hg}, u_{hh}, \sigma_\varepsilon, \overline{\psi}\right). \tag{125}$$

Let $\mathbf{x} = (x, d, a, y)$ denote the data vector for the student i. The likelihood function for this student is

$$\mathcal{L}^i(\Lambda|\mathbf{x}) = f_{d,a,y|\Lambda,x}(d, a, y \,|\, \Lambda, x) = \sum_{\omega \in \Omega} \sum_{q \in (0,1)} f_{d,q,a,\omega,y|\Lambda,x}(d, q, a, \omega, y|\Lambda, x), \tag{126}$$

where $f_{d,a,y|\Lambda,x}$ denotes the joint probability density function of observed variables (d, a, y). This joint probability density function is computed by summing $f_{d,q,a,\omega,y|\Lambda,x}$ over the unobservables ω and q. We have

$$\begin{aligned} & f_{d,q,a,\omega,y}(d, q, a, \omega, y|\Lambda, x) \\ & = \Pr(d|\Lambda, p) \cdot \Pr(q|\Lambda, p, d) \cdot \Pr(a|\Lambda, q) \cdot \Pr(\omega|\Lambda, q) \cdot f(y|\Lambda, a, \omega). \end{aligned} \tag{127}$$

The first probability term is given as

$$\Pr(d = 1|\Lambda, x) = \frac{\exp\left\{\frac{1}{\beta}\left[v(\mu_p + \sigma_p x, \gamma, \eta) - \delta\right]\right\}}{1 + \exp\left\{\frac{1}{\beta}\left[v(\mu_p + \sigma_p x, \gamma, \eta) - \delta\right]\right\}}, \text{ and} \tag{128}$$

$$\Pr(d = 0|\Lambda, x) = 1 - \Pr(d = 1|\Lambda, x). \tag{129}$$

The second probability term, $\Pr(q|\Lambda, p, d)$ is given as

$$\Pr(q|\lambda, p, d) = \begin{cases} 1 & \text{if } d = 0 \text{ and } q = p \\ (1-p)\gamma + p(1-\eta) & \text{if } d = 1 \text{ and } q = Q_{\gamma,\eta}(p, s_g) \\ (1-p)(1-\gamma) + p\eta & \text{if } d = 1 \text{ and } q = Q_{\gamma,\eta}(p, s_h) \\ 0 & \text{otherwise.} \end{cases} \tag{130}$$

The third probability term is given as

$$\Pr(a = a_g|\Lambda, q) = \frac{\exp\left\{\frac{1}{\beta}\left[U_g(q) - U_h(q) - \overline{\psi}\right]\right\}}{1 + \exp\left\{\frac{1}{\beta}\left[U_g(q) - U_h(q) - \overline{\psi}\right]\right\}}, \text{ and} \tag{131}$$

$$\Pr(a = a_h|\Lambda, q) = 1 - \Pr(a = a_g|\Lambda, q). \tag{132}$$

The fourth probability term is simply $\Pr(\omega|\Lambda, q) = q$. The last term is given as

$$f(y|\Lambda, a, \omega) = \phi\left(\frac{y - u(\omega, a) - \overline{\psi}a}{\sigma_\epsilon}\right), \tag{133}$$

where ϕ is the probability density function of the Standard Normal distribution.

The maximum likelihood estimator is the vector $\widehat{\Lambda}$ in the set of possible parameters that maximize the sum of log likelihood $\sum_{i \in I} \log \mathcal{L}^i(\Lambda|\mathbf{x}^i)$. The estimated variance-covariance matrix of the estimator is the inverse of the minus Hessian of the sum of log likelihood with respect to the parameter vector.

6.3 Results: how science-encouraging are AP science classes?

Table 3 shows the benchmark estimates of the model parameters. This benchmark estimation imposes that there is no human capital accumulation from AP science classes, that is, $\overline{\psi} = 0$. Estimates of the mean and standard deviation of students' prior beliefs are 0.354 and 0.038. These estimates imply that most students' prior beliefs lie between being around 25 to 45 percent confident about their science ability. Observe that the estimated belief threshold for being under-confident or over-confident in science is

$$\frac{\widehat{u}_g}{\widehat{u}_g + \widehat{u}_h} = \frac{2.367 - 0.072}{2.367 - 0.072 + 2.243 - (-0.220)} = 0.482, \tag{134}$$

which is greater than students' belief range. Therefore, the estimates suggest that students are largely underconfident in the sciences.

Table 3: Estimated Parameters

Symbol	Description	Estimate	Std Err
Beliefs, education, and choices			
μ_p	Mean of prior beliefs on science talent	0.354	0.005
σ_p	Standard deviation of prior beliefs on science talent	0.038	0.002
γ	Signal accuracy for a non-science-talented student	0.548	0.026
η	Signal accuracy for a science-talented student	0.904	0.030
δ	Cost of participation in AP science classes	0.650	0.032
β	Stochastic choice error scaling factor	0.314	0.016
Ex post productivities			
$u(\omega_g, a_g)$	Non-science-talented student choosing non-science	2.367	0.010
$u(\omega_g, a_h)$	Non-science-talented student choosing science	0.072	0.040
$u(\omega_h, a_g)$	Science-talented student choosing non-science	-0.220	0.018
$u(\omega_h, a_h)$	Science-talented student choosing science	2.243	0.024
Other parameter			
σ_ε	Standard deviation of log earnings noise	0.748	0.006

Note: This table shows the benchmar estimates of the extended model parameters using maximum likelihood estimation. The estimates are based on a sample of 13,003 students. The human capital accumulation from Advanced Placement science classes is imposed to be zero.

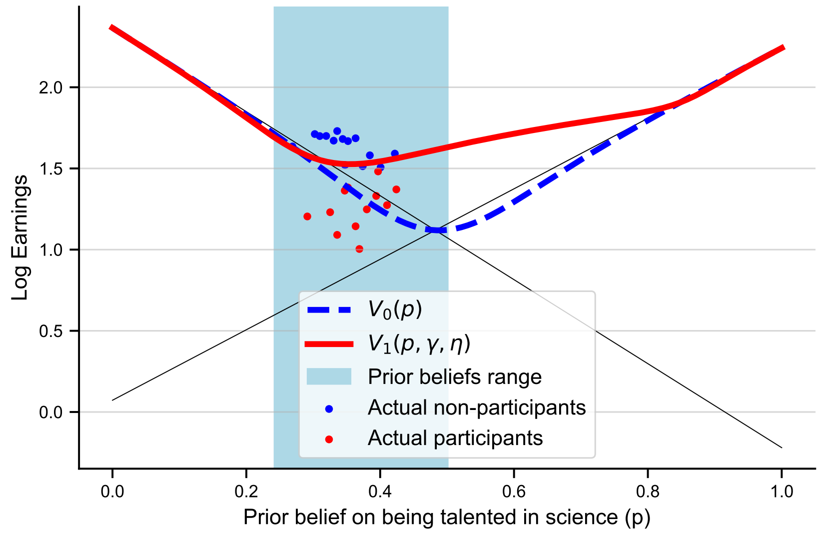

Figure 14: Estimated expected productivity with and without education

Although students are underconfident about sciences, Advanced Placement science classes are largely science-encouraging. The estimated educational structure of AP science classes is $(\widehat{\gamma}, \widehat{\eta}) = (0.548, 0.904)$, with much greater $\widehat{\eta}$ than $\widehat{\gamma}$. This result suggests that the current structure of AP science classes may be suboptimally over-encouraging students into sciences.

Figure 14 illustrates the implications of the estimated parameters. The horizontal axis is a student's prior belief in science talent. The two thin straight lines represent the expected productivities when choosing to pursue a non-science and science majors, respectively. The dashed blue curve represents the expected productivity of a non-participant who optimally chooses between a non-science or a science major givne his belief. The solid red curve represents the expected productivity of a participant who optimally chooses his major after receiving a signal and updating his belief. The difference between the solid and dashed curves represent the *ex ante* gain in productivity, that is, the self-exploration value of education.

The shaded area indicates the estimated range of students' prior beliefs. The figure shows that students who are the most confident in sciences—the ones closer to the threshold—gain the most from the self-exploration aspect of the Advanced Placement science classes. However, the solid red curve is significantly skewed to the right, meaning that it is largely science-encouraging. The figure suggests that making the educational structure skewed to the opposite direction (*i.e.* science-*discouraging*) may increase the self-exploration value of education for a considerable portion of students.

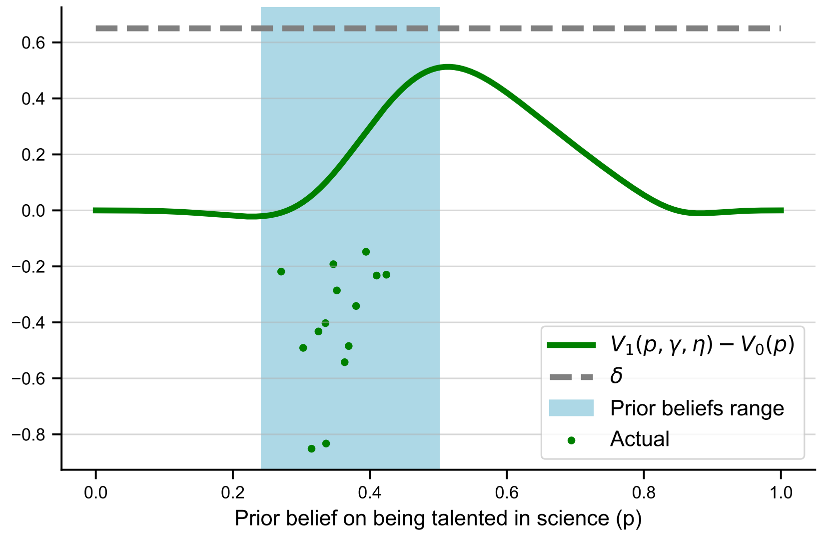

Figure 15: Estimated value of education as self-exploration

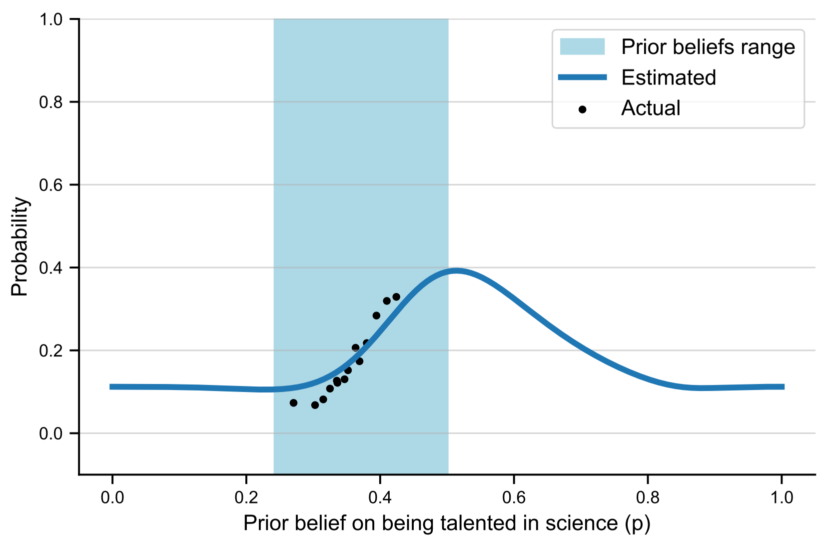

Figure 16: Estimated and actual probability of Advanced Placement (AP) science class participation

Note: Actual probabilities are the percentages of participants in each of twenty quartiles of their prior beliefs.

Table 4: Effects of alternative educational structures

Estimates	Sample (observations)		
	Full (13,003)	Female (6,734)	Male (6,269)
Beliefs and education			
Sample average prior beliefs on science talent	0.356	0.348	0.372
Sample s.d. of prior beliefs on science talent	0.038	0.035	0.041
Signal accuracy for a non-science-talented student ($\widehat{\gamma}$)	0.548	0.653	0.509
Signal accuracy for a science-talented student ($\widehat{\eta}$)	0.903	0.744	1.000
Probability of receiving a science recommendation	0.613	0.485	0.680
Effects of alternative educational structures			
Self-exploration value of education (status quo) $= \frac{1}{n}[W(\widehat{\gamma},\widehat{\eta}) - W(0.5, 0.5)]$	0.036	0.029	0.049
Value of education under the opposite structure $= \frac{1}{n}[W(\widehat{\eta},\widehat{\gamma}) - W(0.5, 0.5)]$	0.122	0.045	0.170
Value of education under perfect information $= \frac{1}{n}[W(1,1) - W(0.5, 0.5)]$	0.689	0.687	0.657

Figures 15 shows the estimated self-exploration value of education in relation to the opportunity cost of participation. This figure shows that taking an Advanced Placement science classes is costly enough that even the most likely participant would take such a class with less than 50 percent probability. Figure 16 shows the actual and estimated probabilities of participation across students with differnet prior beliefs.

Table 4 reports the effects of alternative educational structures in the full sample, female-only subsample, and male-only subsample. With the status quo of estimated educatoinal structure of $(\widehat{\gamma},\widehat{\eta}) = (0.548, 0.903)$, the self-exploration value of education is 0.036. This means that the availability of AP science classes *ex ante* increases students' earnings by 3.6 percent by providing them information that leads to better college major choices. However, restructuring the AP science classes to the opposite direction—making them science-discouraging—would *ex ante* increase students' earnings by 12.2 percent. The intuition behind this contrast is that enouraging science majors to a largely underconfident student population can lead to more severe mismatches.

The next two columns of Table 4 report the parameter estimates when the sample is restricted to female- and male-only. The estimated educational

Table 5: Estimated parameters with human capital accumulation

Parameters	Benchmark (Set $\overline{\psi} = 0$)	$\overline{\psi} = 0.01$	$\overline{\psi} = 0.03$	$\overline{\psi} = 0.05$
μ_p	0.354	0.354	0.354	0.354
σ_p	0.038	0.039	0.039	0.039
γ	0.548	0.549	0.550	0.551
η	0.904	0.901	0.895	0.890
δ	0.650	0.654	0.662	0.671
β	0.314	0.316	0.319	0.323
$u(\omega_g, a_g)$	2.367	2.367	2.367	2.367
$u(\omega_g, a_h)$	0.072	0.069	0.062	0.055
$u(\omega_h, a_g)$	-0.220	-0.220	-0.220	-0.220
$u(\omega_h, a_h)$	2.243	2.240	2.233	2.227
σ_ε	0.738	0.738	0.739	0.739

structure for female students is $(0.653, 0.744)$ while that for male students is $0.509, 1.000$. This result suggests that the current structure of AP science classes is more science-encouraging to male students. Moreover, the self-exploration value of education under both the status quo and the opposite structure is greater for male students than female students. This result suggests that AP science classes are structured to be more informative to male students than female students.

The results examined so far are based on the benchmark case when the human capital accumulation $\overline{\psi}$ is set to zero. Table 5 reports the parameter estimates under slightly different specifications that allow positive human capital accumulation. For example, setting the value $\overline{\psi} = 0.01$ means that taking a AP science class directly improves one's *ex post* earnings by one percent. Even for reasonably large values of the parameter $\overline{\psi}$, the educational structure is estimated to be science-encouraging, around $(0.55, 0.9)$.

7 Conclusion

This paper shows that it is useful to think of education as finding oneself. A benevolent educational designer should encourage a career in which participants are on average more confident. This property of an optimal educational structure extends to settings with stochastic choice and human capital accumulation. An econometrician can estimate the parameters of this model and infer counterfactual outcomes even if students' beliefs are imperfectly observed.

An important caveat to the paper's conclusions is that it assumes that students' prior beliefs are rational—that they represent the true probabilities of their talents. If, for any reason, a group of students' beliefs are biased and differ from their actual probabilities, the optimal educational structure may differ. For example, it may still be desirable to design classes to encourage an underconfident group of students to pursue sciences if their socioeconomic background contributes to that underconfidence.

References

Alonso, Ricardo and Odilon Câmara (2016) "Persuading voters," *American Economic Review*, 106 (11), 3590–3605.

Altonji, Joseph G (1993) "The demand for and return to education when education outcomes are uncertain," *Journal of Labor Economics*, 11 (1, Part 1), 48–83.

——— (1995) "The effects of high school curriculum on education and labor market outcomes," *Journal of Human Resources*, 409–438.

Altonji, Joseph G, Peter Arcidiacono, and Arnaud Maurel (2016) "The analysis of field choice in college and graduate school: Determinants and wage effects," in *Handbook of the Economics of Education*, 5, 305–396: Elsevier.

Altonji, Joseph G, Erica Blom, and Costas Meghir (2012) "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annual Review of Econonomics*, 4 (1), 185–223.

Altonji, Joseph G and Charles R Pierret (1998) "Employer learning and the signalling value of education," in *internal Labour markets, incentives and Employment*, 159–195: Springer.

Arcidiacono, Peter (2004) "Ability sorting and the returns to college major," *Journal of Econometrics*, 121 (1-2), 343–375.

Arcidiacono, Peter, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom (2016) "College attrition and the dynamics of information revelation,"Technical report, National Bureau of Economic Research.

Arieli, Itai and Yakov Babichenko (2019) "Private bayesian persuasion," *Journal of Economic Theory*, 182, 185–217.

Arrow, Kenneth J (1973) "Higher education as a filter," *Journal of Public Economics*, 2 (3), 193–216.

Aryal, Gaurab, Manudeep Bhuller, and Fabian Lange (2019) "Signaling and employer learning with instruments,"Technical report, National Bureau of Economic Research.

Becker, Gary S (1962) "Investment in human capital: A theoretical analysis," *Journal of Political Economy*, 70 (5, Part 2), 9–49.

——— (2009) *Human capital: A theoretical and empirical analysis, with special reference to education*: University of Chicago press.

Ben-Porath, Yoram (1967) "The production of human capital and the life cycle of earnings," *Journal of Political Economy*, 75 (4, Part 1), 352–365.

Bergemann, Dirk and Stephen Morris (2019) "Information design: A unified perspective," *Journal of Economic Literature*, 57 (1), 44–95.

Bush, George W (2006) "Address before a joint session of the Congress on the state of the Union. 31 January 2006," *https://www.presidency.ucsb.edu/documents/address-before-joint-session-the-congress-the-state-the-union-13*.

Caplin, Andrew, Mark Dean, and John Leahy (2019) "Rational inattention, optimal consideration sets, and stochastic choice," *The Review of Economic Studies*, 86 (3), 1061–1094.

Card, David (1994) "Earnings, schooling, and ability revisited."

——— (1999) "The causal effect of education on earnings," *Handbook of labor economics*, 3, 1801–1863.

——— (2001) "Estimating the return to schooling: Progress on some persistent econometric problems," *Econometrica*, 69 (5), 1127–1160.

Chevalier, Arnaud, Colm Harmon, Ian Walker, and Yu Zhu (2004) "Does education raise productivity, or just reflect it?" *The Economic Journal*, 114 (499), F499–F517.

College Board (2020) "AP Cohort Data Report: Graduating Class of 2020."

Comay, Yochanan, Arie Melnik, and Moshe A Pollatschek (1973) "The option value of education and the optimal path for investment in human capital," *International Economic Review*, 421–435.

Fang, Hanming (2006) "Disentangling the college wage premium: Estimating a model with endogenous education choices," *International Economic Review*, 47 (4), 1151–1185.

Huntington-Klein, Nick (2021) "Human capital versus signaling is empirically unresolvable," *Empirical Economics*, 60, 2499–2531.

Hussey, Andrew (2012) "Human capital augmentation versus the signaling value of MBA education," *Economics of Education Review*, 31 (4), 442–451.

Kamenica, Emir (2019) "Bayesian persuasion and information design," *Annual Review of Economics*, 11, 249–272.

Kamenica, Emir and Matthew Gentzkow (2011) "Bayesian persuasion," *American Economic Review*, 101 (6), 2590–2615.

Keane, Michael P and Kenneth I Wolpin (1997) "The career decisions of young men," *Journal of political Economy*, 105 (3), 473–522.

Kroch, Eugene A and Kriss Sjoblom (1994) "Schooling as human capital or a signal: Some evidence," *Journal of Human Resources*, 156–180.

Lange, Fabian (2007) "The speed of employer learning," *Journal of Labor Economics*, 25 (1), 1–35.

Lemieux, Thomas (2014) "Occupations, fields of study and returns to education," *Canadian Journal of Economics/Revue canadienne d'économique*, 47 (4), 1047–1077.

Manski, Charles F (1989) "Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon," *Economics of Education Review*, 8 (4), 305–312.

Matějka, Filip and Alisdair McKay (2015) "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105 (1), 272–98.

Mincer, Jacob (1974) "Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.."

Owen, Stephanie (2020) "College Field Specialization and Beliefs about Relative Performance."

Patrinos, Harry Anthony (2016) "Estimating the return to schooling using the Mincer equation," *IZA World of Labor*.

Psacharopoulos, George and Harry Anthony Patrinos (2004) "Returns to investment in education: a further update," *Education economics*, 12 (2), 111–134.

Spence, Michael (1978) "Job market signaling," in *Uncertainty in Economics*, 281–306: Elsevier.

Stinebrickner, Ralph and Todd Stinebrickner (2014) "Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model," *Journal of Labor Economics*, 32 (3), 601–644.

Weiss, Andrew (1995) "Human capital vs. signalling explanations of wages," *Journal of Economic Perspectives*, 9 (4), 133–154.