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

Sungmin Park*

October 24, 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 its parameters from data. Advanced science classes in high school tend to encourage science majors in college. Their estimated self-exploration value is a 5-percent average increase in future earnings.

Keywords: Education, Occupational choice, Value of information, Information design
JEL Codes: I21, J24, 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

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

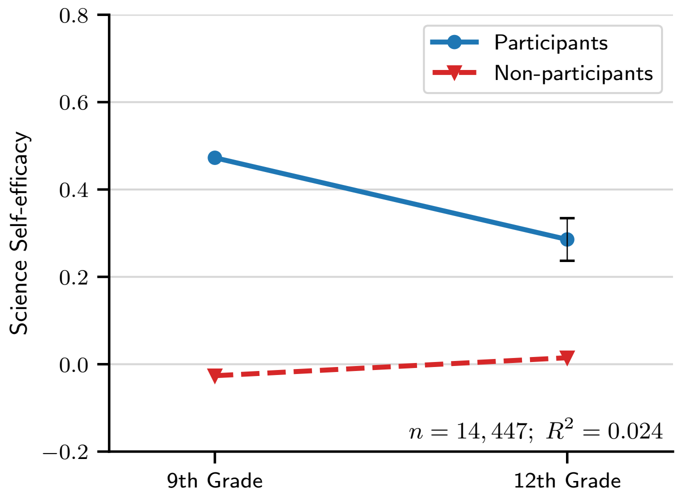

Figure 1: Advanced Science Classes and Changes in Science Self-Efficacy

Note: This figure shows the average degrees of *science self-efficacy* of total 14,447 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 coefficient computed with robust standard errors clustered for each student. See Table 1 in Subsection 5.3 for all estimated coefficients.

professional schools, there are classes across different subfields. Likewise, all stages of education involve some degree of self-exploration.

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 of 2009 (HSLS) 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 that value the information the most (Theorem 1). The intuition behind this result is that every participating student's *ex ante* expected payoff is linear in his prior 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 a socially optimal educational structure encourages a career in which its participating students are on average more confident (Theorem 2). If the average participant is believed to have greater comparative advantage in gathering, the educational structure should be gathering-encouraging. That is, the signal accuracy for the truly talented gatherer should be greater than that for the truly talented hunter. I say that such a signal structure is gathering-encouraging, in that it has a relatively higher chance of misleading students with no talent in gathering into the gathering career. Similarly, if the average participant is believed to have comparative advantage in hunting, the educational structure should be hunting-encouraging.

An important extension I consider is a model in which education allows students to not only learn about themselves but also to accumulate human capital, increasing future productivity directly. I show that the results from the benchmark model continue to hold with some modification. When we allow human capital accumulation, students participate in education for different reasons. Some are sufficiently certain about their talents and have nothing to gain from additional information, yet still participate to raise their *ex post* productivity. Others are less certain about their talents and gain from both. I show that an optimal information structure of education encourages the career in which this latter group of students are on average more confident.

To illustrate an application, I adapt the model for advanced science education in the United States and estimate its parameters using a sample of 6,638 high school students who later attended 4-year colleges. The estimated parameters imply that most students' confidence in the sciences is low, ranging from about 15 to 31 percent. Despite this comparative disadvantage, advanced science classes tends to encourage students to pursue science majors in college. Under the current structure, students without science talent receive accurate signals only 74 percent of the time, whereas students with science talent receive accurate signals 91 percent of the time. As a result, the estimated average value of providing these classes is a 5-percent increase on students' future earnings. However, under the opposite structure of signals, the estimated value would be 12 percent. The interpretation is that the current advanced science classes in U.S. high schools are overly recommending 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 between fields and learn them after first year of college. 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 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. Using university administrative data and a regression discontinuity design, Li and Xia (2022) estimate that higher grades caused students more likely to major in those subjects, and argue that grades let students to learn their comparative advantage.

To my knowledge, only one existing paper, Arcidiacono, Aucejo, Maurel, and Ransom

[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.

(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; more specifically, it 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) 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 each Receiver has his own independent state rather than a common state of the world.

Costly information acquisition My work is also related to the eminent literature on rational inattention and costly information acquisition such as Matějka and McKay (2015) and Caplin, Dean, and Leahy (2019). These studies use cost functions based on the distribution of posterior beliefs, notably the changes in the Shannon entropy. In comparison, I use a cost that is a direct function of the parameters of the information structure. In the terminology of Denti, Marinacci, and Rustichini (2022), the cost function is experiment-based rather than posterior-based.

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 benchmark model to allow stochastic choice, human capital investment, imperfectly observed beliefs, and more than two talents and careers. 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

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

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 educational structure (γ, η) when the prior belief is p. Specifically, by Bayes' rule, the posterior belief function is

$$Q_{\gamma,\eta}(p, s) = \begin{cases} \dfrac{(1-\eta)p}{\gamma(1-p) + (1-\eta)p} & \text{if } s = s_g, \text{ and} \\ \dfrac{\eta p}{(1-\gamma)(1-p) + \eta p} & \text{if } s = s_h. \end{cases}$$

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* $\widehat{\Theta} = \{(\gamma, \eta) \in \Theta \mid C(\gamma, \eta) \leq B\}$, where $B > 0$ is a constant called the *educational attention budget*, and the *educational attention cost function* $C : \Theta \longrightarrow \mathbb{R}$ satisfies Assumptions 1–3.

[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.

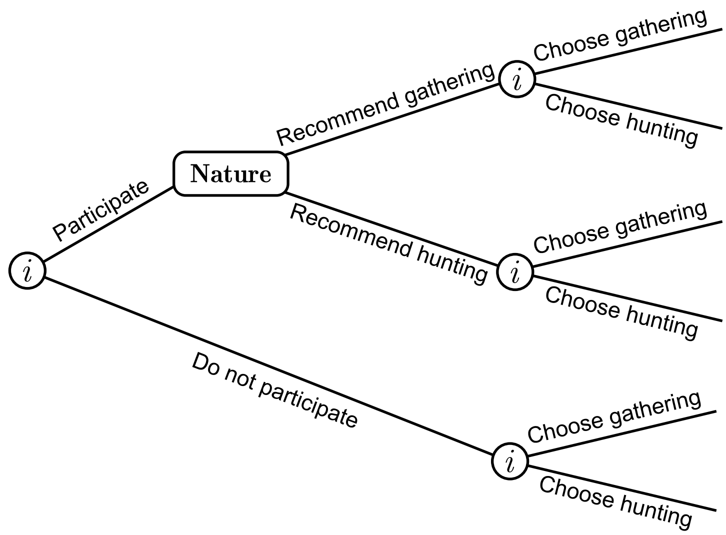

Figure 2: Education and Occupational Choice

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 $\Pr(s_h|\omega_h) = 1 - \Pr(s_g|\omega_h) = \eta$.

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 (Boundary regularity). *C satisfies*
(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 3(a) means that uninformative education is always feasible. Assumption 3(b) means that perfectly informative education is infeasible. Assumption 3(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.

Example 1 (Expected reduction in entropy from a fixed prior). Let $\lambda > 0$ and $p_0 \in (0, 1)$ be given. Let $H : (0, 1) \longrightarrow \mathbb{R}$ be defined as $H(p) = (1 - p)\log(1 - p) + p \log p$. Let the educational cost function be $C(\gamma, \eta) = \lambda \left(H(p_0) - \mathbb{E}\left[H(Q_{\gamma,\eta}(p_0, s))\right]\right)$, where the expectation is taken with respect to the possible realizations of the signal s. Then C satisfies Assumptions 1–3.

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

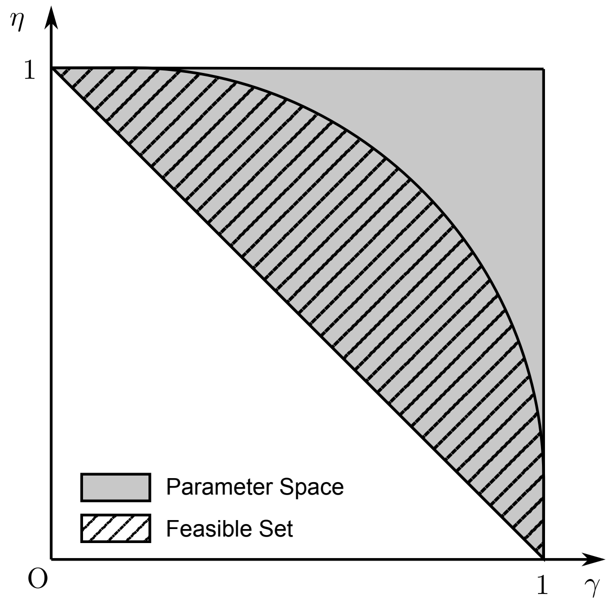

Figure 3: Feasible Education

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

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

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

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

$$U_h(p) = \sum_{\omega \in \Omega} \Pr(\omega) u(a_h, \omega) = (1-p) u(a_h, \omega_g) + p u(a_h, \omega_h). \tag{2}$$

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

V_1 take values

$$V_0(p) = \max\{U_g(p), U_h(p)\}, \text{ and}$$
$$V_1(p,\gamma,\eta) = \mathbb{E}\left[V_0\big(Q_{\gamma,\eta}(p,s)\big) \mid p,\gamma,\eta\right],$$

where the expectation is taken with respect to the possible realizations of the signal s.

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

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

An educational structure (γ^*,η^*) in the feasible set $\widehat{\Theta}$ is *(socially) optimal* if $W(\gamma^*,\eta^*) \geq W(\gamma,\eta)$ for all $(\gamma,\eta) \in \widehat{\Theta}$. The *participant set* of an educational structure (γ,η) is the set of students $i \in I$ that participates in (γ,η). An educational structure (γ,η) is *nontrivial* if its participant set is nonempty. Let $\overline{p}_D = \frac{1}{|D|}\sum_{i\in D} p^i$ for all nonempty subsets $D \subset I$. Let $\mathcal{P}(I)$ denote the collection of all subsets of I.

Definition. Let $F : (0,1) \longrightarrow \widehat{\Theta}$ map p to the solution (γ,η) of the system of equations

$$\frac{\frac{\partial}{\partial\gamma}C(\gamma,\eta)}{\frac{\partial}{\partial\eta}C(\gamma,\eta)} = \frac{1-p}{p}\frac{u_g}{u_h} \quad \text{and} \quad C(\gamma,\eta) = B. \tag{3}$$

To verify that F is well-defined, observe that the system of equations (3) is 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 3(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 its participant set D^* satisfies $(\gamma^*,\eta^*) = F(\bar{p}_{D^*})$ and*

$$D^* \in \operatorname*{argmax}_{D\in\mathcal{P}(I)} W\big(F(\bar{p}_D)\big). \tag{4}$$

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

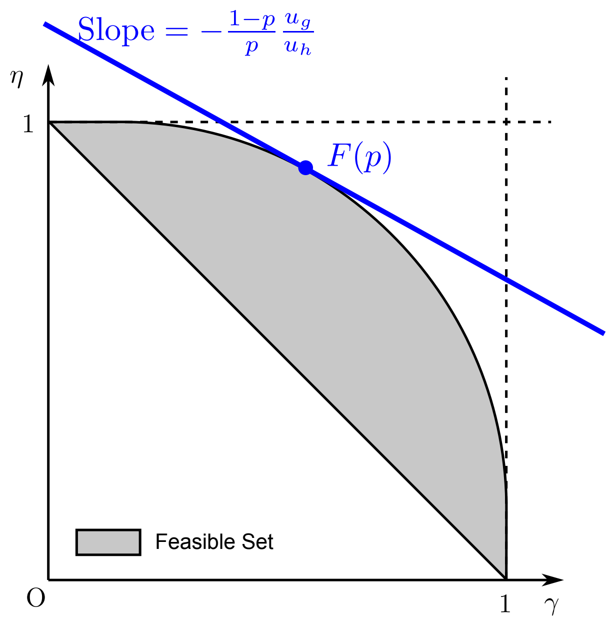

Figure 4: Definition of F

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

$$W(\gamma, \eta) = \sum_{i \in I} V(p^i, \gamma, \eta) = \sum_{i \in I} \max \left\{ V_0(p^i), V_1(p^i, \gamma, \eta) - \delta \right\}. \tag{5}$$

Observe that a participant's expected productivity is

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

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

$$V_1(p, \gamma, \eta) = (1-p)\left[\gamma u(\omega_g, a_g) + (1-\gamma) u(\omega_g, a_h)\right] + p\left[(1-\eta) u(\omega_h, a_g) + \eta u(\omega_h, a_h)\right], \tag{6}$$

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

Observe from equation (5) that we can write

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

Let $W(\gamma,\eta|D)$ denote the maximand in the above equation. 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 1–3, 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

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

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

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

Observe that

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

The above two equations together imply that

$$\frac{\frac{\partial}{\partial\gamma}W(\gamma^*,\eta^*)}{\frac{\partial}{\partial\eta}W(\gamma^*,\eta^*)} = \frac{1-\overline{p}_{D^*}}{\overline{p}_{D^*}}\frac{u_g}{u_h}, \tag{7}$$

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

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

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. By Lemma 4, the participant set D^* of (γ^*,η^*) satisfies $(\gamma^*,\eta^*) = F(\overline{p}_{D^*})$. Since (γ^*,η^*) is optimal, $W(\gamma^*,\eta^*) \geq W(\gamma,\eta)$

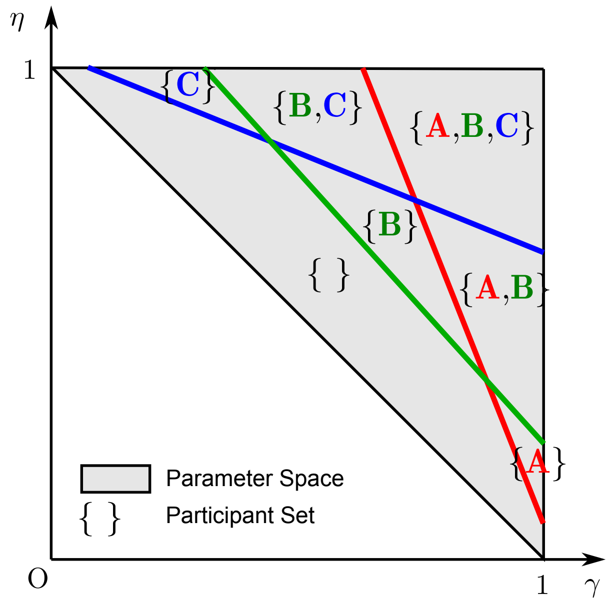

Figure 5: Every nonempty Θ_D is a convex polygon: an example with three students

for all $(\gamma, \eta) \in \Theta$. Then for all $D \subset I$, we have

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

So D^* satisfies (4).

Second, I show the "if" part of the statment. Suppose $(\gamma^*, \eta^*) \in \widehat{\Theta}$. Suppose its participant set D^* satisfies $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$ and (4). 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 satisfy (4), 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 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 $(\gamma^*, \eta^*) = F(\overline{p}_{D^*})$.

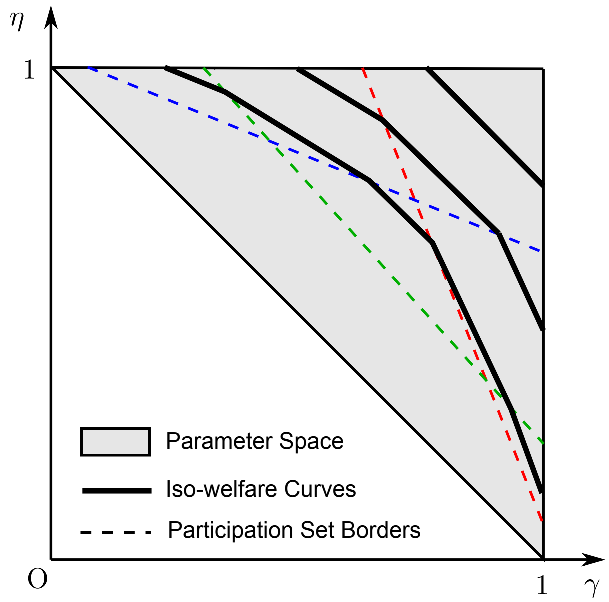

Figure 6: The student welfare function W is affine on Θ_D

3.2 Direction of encouragement

Let us place an additional restriction on the cost function C for the remainder of the paper.

Assumption 4 (Symmetry)**.** *$C(\gamma, \eta) = C(\eta, \gamma)$ for any γ and η.*

This assumption means that, for any budget $B > 0$, if an educational structure (γ, η) is feasible, the opposite structure (η, γ) is feasible as well. Although this assumption is not crucial, it simplifies the analysis and interpretation of the results that follow. Building on this assumption and 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 $\frac{u_g}{u_g+u_h}$. In other words, a student is believed to have comparative advantage in gathering if $p < \frac{u_g}{u_g+u_h}$ and in hunting if $p > \frac{u_g}{u_g+u_h}$. This interpretation is valid because

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

where $U_g(p)$ and $U_h(p)$ are defined as in (1) and (2). Furthermore, by *encouraging* the gathering career and the hunting career, I mean the conditions $\gamma > \eta$ and $\gamma <$

η, respectively. To see why, observe that the probability of receiving a gathering-recommending signal s_g for a student with belief p is

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

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.

Recall that $F : (0,1) \longrightarrow \widehat{\Theta}$ is defined as the vector-valued function that maps a belief p to the solution (γ, η) of the system of equations (3). Let F_g and F_h be the two component functions of 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

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

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

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

Then by the strict convexity (Assumption 2) and symmetry (Assumption 4) 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 $\overline{p}_{D^*} > \frac{u_g}{u_g+u_h}$. 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 on the midpoint of the boundary. When the average belief is greater than

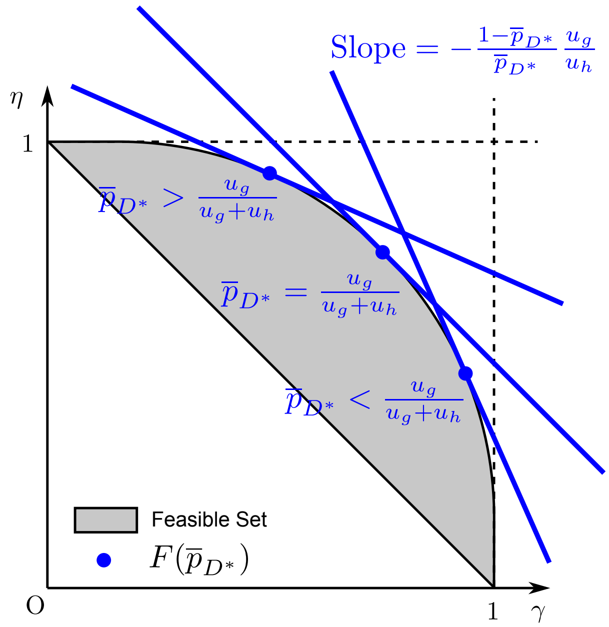

Figure 7: Optimal educational structure and the direction of encouragement

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.

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 look like. 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).

$$\delta(c-b) \leq \min\Big\{(1-c)\big[U_g(b) - U_h(b)\big],\ b\big[U_h(c) - U_g(c)\big]\Big\}. \tag{8}$$

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 stricter as δ increases. Taken together, the theorem means that any participant set characterized by a regular interval of beliefs can be induced by some

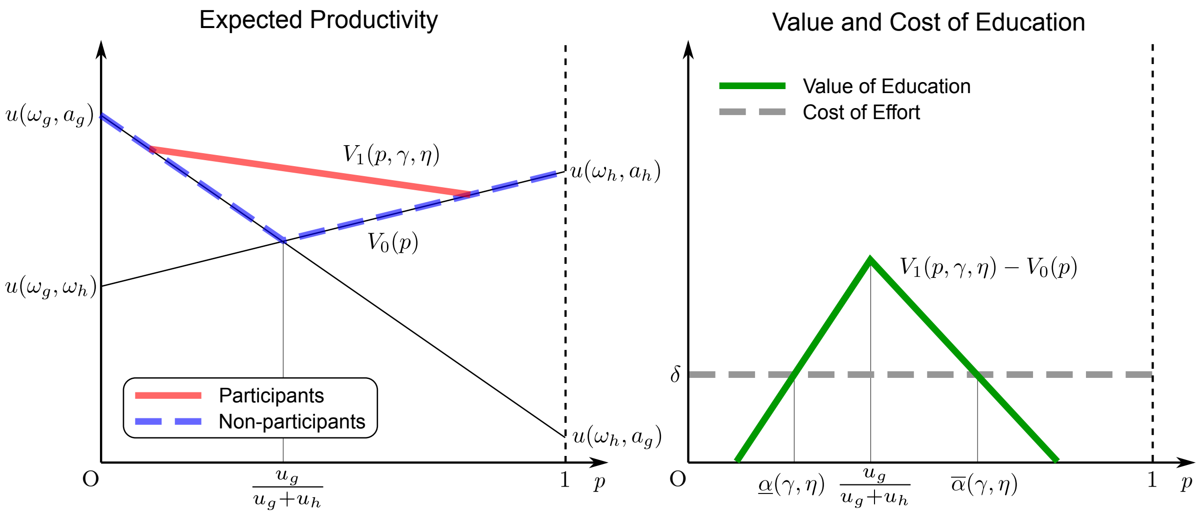

Figure 8: Any participation set is separated by an interval

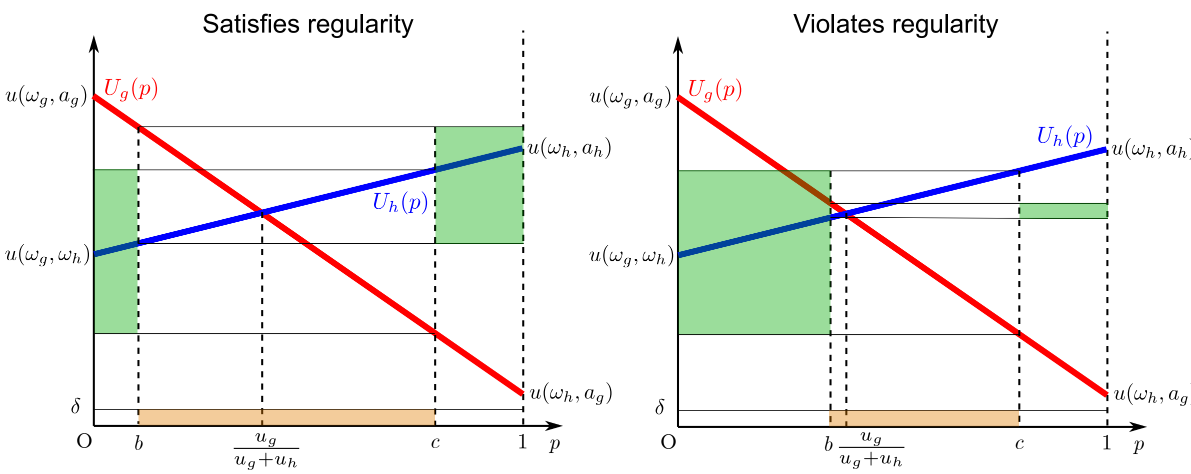

Figure 9: Interval regularity: satisfied (left) and violated (right)

educational structure.

Figure 8 illustrates why any participant set is separated by an interval. For a fixed educational structure (γ, η), the left panel shows the expected productivity 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 productivity. 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.

Figure 9 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 (8). 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 (8). The regularity condition is violated if the bottom rectangle's area is larger than any of the two top rectangles' areas.

Let us make the following definition and establish two lemmas for the proof of the theorem.

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

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

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

$$V_1(p,\gamma,\eta) - V_0(p) > \delta.$$

We know $V_1(p,\gamma,\eta)$ from equation (6). 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

$$V_1(p,\gamma,\eta) - V_0(p) = \begin{cases} -(1-p)(1-\gamma)u_g + p\eta u_h & \text{if } p \le \frac{u_g}{u_g+u_h}, \\ (1-p)\gamma u_g - p(1-\eta)u_h & \text{if } p \ge \frac{u_g}{u_g+u_h}. \end{cases} \tag{9}$$

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 $(\gamma,\eta) \in \Theta$, $\frac{u_g}{u_g+u_h}$ is a maximizer of the problem*

$$\max_{p\in(0,1)} \left[V_1(p,\gamma,\eta) - V_0(p)\right]. \tag{10}$$

The maximized value of (10) is $(\gamma+\eta-1)\frac{u_g u_h}{u_g+u_h}$.

Proof. Equation (9) implies

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

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

$$(1-c)\big[U_g(b)-U_h(b)\big]-\delta(c-b)=\frac{(1-\eta)u_h\cdot[(\gamma+\eta-1)u_gu_h-\delta(u_g+u_h)]}{[(1-\gamma)u_g+\eta u_h]\cdot[\gamma u_g+(1-\eta)u_h]}, \quad (11)$$

$$b\big[U_h(c)-U_g(c)\big]-\delta(c-b)=\frac{[(1-\gamma)u_h+2\delta]\cdot[(\gamma+\eta-1)u_gu_h-\delta(u_g+u_h)]}{[(1-\gamma)u_g+\eta u_h]\cdot[\gamma u_g+(1-\eta)u_h]}. \quad (12)$$

Both (11) and (12) 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

$$V_1(b,\gamma,\eta)-V_0(b)=\delta,\ \text{and} \quad (13)$$
$$V_1(c,\gamma,\eta)-V_0(c)=\delta. \quad (14)$$

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

$$\begin{aligned}V_1(b,\gamma',1)-V_0(b)&=\delta,\\ V_1(b,1,\eta')-V_0(b)&=\delta,\\ V_1(c,\gamma'',1)-V_0(c)&=\delta,\ \text{and}\\ V_1(c,1,\eta'')-V_0(c)&=\delta.\end{aligned}$$

The line segment between the two points $(\gamma',1)$ and $(1,\eta')$ is the set of $(\gamma,\eta)\in\Theta$ that satisfy equation (13). Similarly, the line segment between the two points $(\gamma'',1)$ and $(1,\eta'')$ is the set of $(\gamma,\eta)\in\Theta$ that satisfy equation (14). Then there exists a solution to the system of equations (13)–(14) if (and only if) the two line segments have a nonempty intersection, i.e. either (i) $\gamma'\geq\gamma''$ and $\eta'\leq\eta''$ or (ii) $\gamma'\leq\gamma''$ and $\eta'\geq\eta''$. 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 $(p^1,p^2,\dots,p^n)$. Let $\mathbf{p}'$ denote the vector of beliefs $(p'^1,p'^2\dots,p'^n)$. 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

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

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* $\mathbf{p}'$, $||\mathbf{p}' - \mathbf{p}||_2 < \varepsilon$ *implies*

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

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 $\mathbf{p}$. ■

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

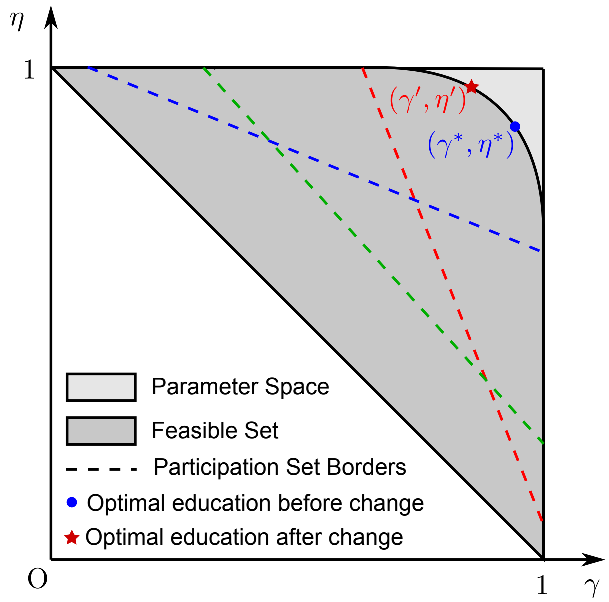

Figure 10: Every student participates regardless of changes in beliefs when the educational attention budget is large

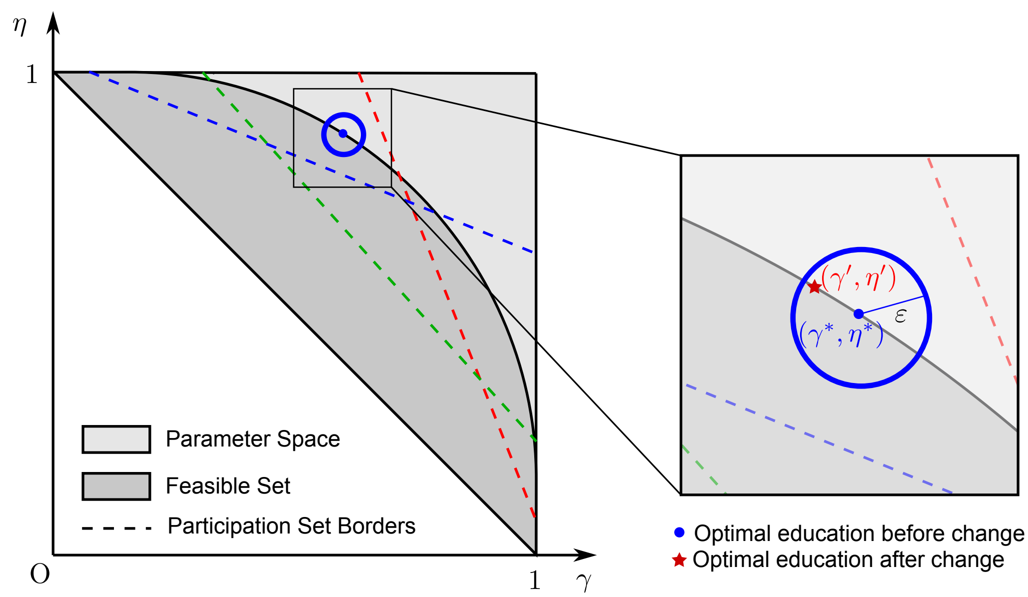

Figure 11: Participation set remains the same when changes in beliefs are small

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 $i \in I$, $p^i \le \frac{u_g}{u_g+u_h}$ implies $p'^i \le p^i$ and $p^i \ge \frac{u_g}{u_g+u_h}$ implies $p'^i \ge p^i$.

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

From equation (9), every $j \in I$ satisfies

$$V_1(\gamma,\eta,p'^j) - V_0(p'^j) \le V_1(\gamma,\eta,p^j) - V_0(p^j).$$

Then

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

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* productivity $u(\omega,a)$.

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

$$W(\gamma, \eta, \mathbf{p}, \delta) = \sum_{i \in I} \max\left\{V_0(p^i), V_1(p^i, \gamma, \eta) - \delta\right\}. \tag{15}$$

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 (15) implies (b). ■

5 Extension

The benchmark 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 in the benchmark model are deterministic and do not allow errors. Second, students in the benchmark model do not learn anything that directly improves their productivity. Third, their prior beliefs are perfectly observable to an econometrician. Fourth, their 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 allow students' choices to be 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

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

where ν_0^i and ν_1^i are 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* productivity 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

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

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 independently and identically distributed as Gumbel$(0, \beta_d)$, the cdf G is the logit function:

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

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

$$U_g(q^i) + \nu_g^i \leq U_h(q^i) + \nu_h^i,$$

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

$$\begin{aligned} U_g(q) &= (1-q)u(\omega_g, a_g) + qu(\omega_h, a_g), \text{ and} \\ U_h(q) &= (1-q)u(\omega_g, a_h) + qu(\omega_h, a_h). \end{aligned}$$

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

$$\begin{aligned} \Pr(a^i = a_h | q^i) &= \Pr\left[\nu_g^i - \nu_h^i \leq U_h(q^i) - U_g(q^i)\right] \\ &= H\left[U_h(q^i) - U_g(q^i)\right], \end{aligned}$$

where H denotes the cumulative distribution function of the random variable $\nu_g^i - \nu_h^i$.

For example, if ν_g^i and ν_h^i are independently and identically distributed as Gumbel$(0, \beta_a)$, $H(v) = \frac{e^{v/\beta_a}}{1+e^{v/\beta_a}}$.

With these choice probabilities, the function V_0 takes values

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

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

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

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

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

The *student welfare function* is

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

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 benchmark model.

Definition. The extended model with stochastic choice features *fixed probabilities of mistakes* if, for some for some $\alpha, \beta \in (0, 0.5)$

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

where M is a constant greater than $\frac{u_g u_h}{u_g + u_h}$. A set $D \subset I$ is an *expected participant set* under $(\gamma, \eta) \in \Theta$ if

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

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 its expected participant set D^* satisfies $(\gamma^*, \eta^*) = F(\bar{p}_{D^*})$ and*

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

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

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

We have

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

Let D denote the expected participant set under (γ, η). By Lemma 7, $M > V_1(p, \gamma, \eta) - V_0(p) - \delta$ for all $p \in (0, 1)$ and for all $(\gamma, \eta) \in \Theta$. Then 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

$$\begin{aligned} V_1(p^i, \gamma, \eta) = p^i & \left[\gamma\big((1-\beta)u_{gg} + \beta u_{gh}\big) + (1-\gamma)\big(\beta u_{gg} + (1-\beta)u_{gh}\big)\right] \\ & + (1-p^i)\left[(1-\eta)\big((1-\beta)u_{hg} + \beta u_{hh}\big) + \eta\big(\beta u_{hg} + (1-\beta)u_{hh}\big)\right], \end{aligned}$$

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 $i \in D$,

$$\begin{aligned} \frac{\partial}{\partial \gamma} V_1(p^i, \gamma, \eta) &= (1-2\beta)(1-p)u_g, \\ \frac{\partial}{\partial \eta} V_1(p^i, \gamma, \eta) &= (1-2\beta)p u_h. \end{aligned}$$

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

$$V_1(p^i, \gamma, \eta) = V_0(p^i).$$

Then

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

thus

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

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

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

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

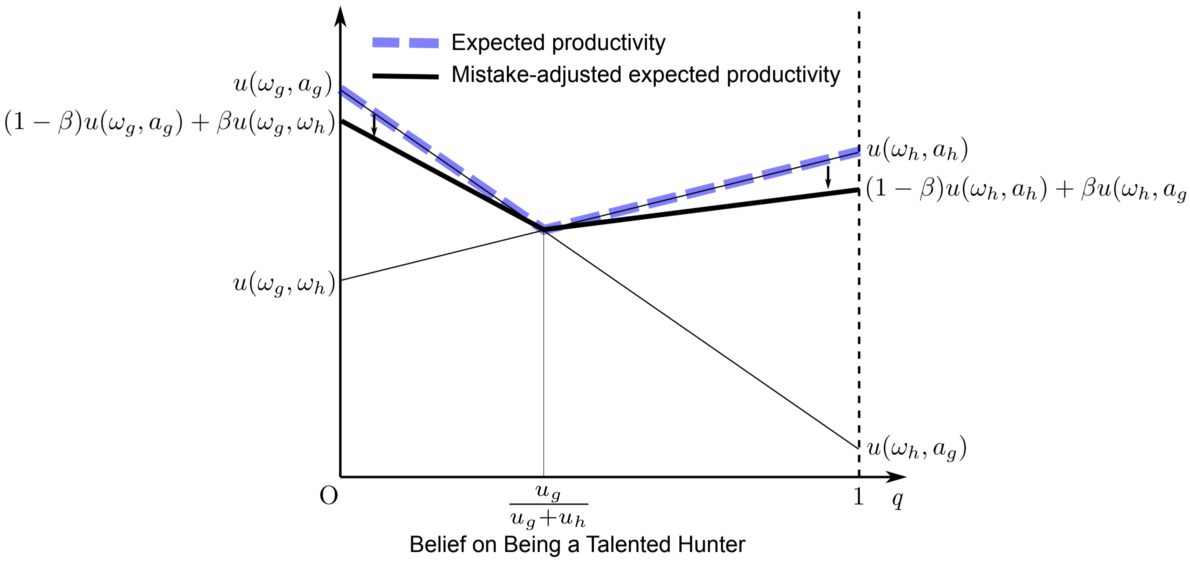

Figure 12: Fixed probability of mistakes shrinks expected mismatch costs proportionally

Corollary 7. *Theorem 2 holds in the extended model with fixed probabilities 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 by mistake. The reason is that the non-expected participants do not value the information from educational signals in the first place. Furthermore, the same belief threshold $\frac{u_g}{u_g+u_h}$ 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.

5.2 Human capital accumulation

Next, we consider an extended model with *human capital accumulation.* That is, education not only informs 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

$$u(\omega, a) + [\psi(\omega, a) - \delta]d,$$

where d is the indicator of participation. Let ψ_g denote $\psi(\omega_g, a_g) - \psi(\omega_g, a_h)$ and let ψ_h denote $\psi(\omega_h, a_h) - \psi(\omega_h, a_g)$. Let us maintain our earlier assumption that the mismatch costs are positive: $u_g > 0$ and $u_h > 0$. Assume also that $u_g + \psi_g > 0$ and $u_h + \psi_h > 0$, ensuring that the costs of mismatches between talents and careers remain positive after students gain additional human capital. Let $F_\psi : (0, 1) \longrightarrow \widehat{\Theta}$ map p to the solution (γ, η) to the system of equations

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

Definition. A participating student is a *complier* if $\Pr(a_g|s_g) = \Pr(a_h|s_h) = 1$. A participating student is a *always-gatherer* if $\Pr(a_g|s_g) = \Pr(a_g|s_h) = 1$. A participating student is an *always-hunter* if $\Pr(a_h|s_g) = \Pr(a_h|s_h) = 1$.

In the benchmark model without human capital accumulation, every participant was a complier. However, with human capital accumulation, distinguishing compliers, always-gatherers, and always-hunters is important because students participate in education for different reasons. Some may value the signals about their talents and follow the recommended careers (thus called *compliers*). Others may be certain that they will pursue gathering or hunting regardless of the signals they receive (thus called *always-gatherers* or *always-hunters*, respectively), yet still be willing to participate to increase their future productivity.

Say that the *compliant participant set* of an educational structure (γ, η) is the set of participants who are compliers under (γ, η). Let Θ_D denote the set of educational structures under which the compliant participant set 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 its compliant participant set D^* satisfies $(\gamma^*, \eta^*) = F_\psi(\bar{p}_{D^*})$ and*

$$D^* \in \operatorname*{argmax}_{D \in \mathcal{P}(I)} W\big(F_\psi(\bar{p}_D)\big).$$

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

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

We have

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

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

$$\begin{aligned} V_1(p, \gamma, \eta) = (1-p)\big[\gamma(u_{gg} + \psi_{gg}) + (1-\gamma)(u_{gh} + \psi_{gh})\big] \\ + p\big[(1-\eta)(u_{hg} + \psi_{hg}) + \eta(u_{hh} + \psi_{hh})\big], \end{aligned}$$

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

$$\begin{aligned} \frac{\partial}{\partial\gamma} V_1(p^i, \gamma, \eta) &= (1-p)(u_g + \psi_g), \\ \frac{\partial}{\partial\eta} V_1(p^i, \gamma, \eta) &= p(u_h + \psi_h). \end{aligned}$$

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

$$\begin{aligned} V_1(p, \gamma, \eta) &= V_0(p) + \Psi_g(p) \quad \text{for all } i \in D_g, \text{ and} \\ V_1(p, \gamma, \eta) &= V_0(p) + \Psi_h(p) \quad \text{for all } i \in D_h. \end{aligned}$$

Then

$$\begin{aligned} W(\gamma, \eta) = \sum_{i \in D} V_1(p^i, \gamma, \eta) + \sum_{i \in D_g} [V_0(p) + \Psi_g(p)] + \sum_{i \in D_h} [V_0(p) + \Psi_h(p)] \\ + \sum_{i \in I \setminus (D \cup D_g \cup D_h)} V_0(p). \end{aligned}$$

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

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

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 human capital accumulation 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 $\gamma^* < \eta^*$. ■

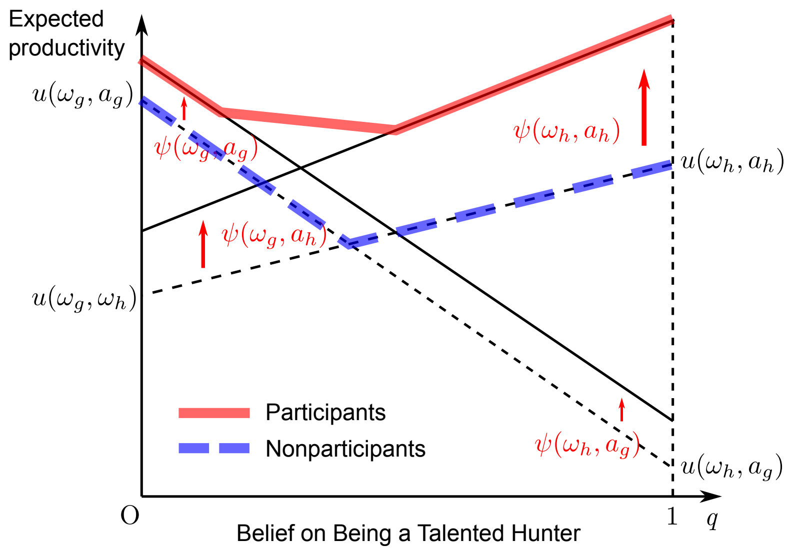

Figure 13: Human capital accumulation changes the mismatch costs for participants

This corollary means that, with human capital accumulation, an optimal educational structure encourages a career in which only the *compliant* participants are confident on average. This result is due to the fact that always-gatherers and always-hunters would 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 productivity change the mismatch costs for all participants. Figure 13 illustrates this point. The three linear pieces in the participants' expected productivity (in red) represent the expected productivity of always-gatherers, compliers, and always-hunters.

5.3 Imperfectly observed beliefs

The next extension addresses a practical concern that an econometrician typically does not observe students' beliefs as probabilities but rather observe them 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

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

From Bayes rule, the posterior belief function Q takes values

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

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

$$x_i = \frac{p^i - \mu_p}{\sigma_p}, \tag{16}$$

$$z_i = \frac{q^i - \mu_q}{\sigma_q}. \tag{17}$$

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.

Let $\mathbb{E}[x_i \,|\, d_i = 1]$ and $\mathbb{E}[z_i \,|\, d_i = 1]$ denote the averages of x_i and z_i among participants. Similarly, let $\mathbb{E}[x_i \,|\, d_i = 0]$ and $\mathbb{E}[z_i \,|\, d_i = 0]$ denote the averages of x_i and z_i among non-participants.

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

$$\phi = \Big(\mathbb{E}[z_i \,|\, d_i = 1] - \mathbb{E}[x_i \,|\, d_i = 1]\Big) - \Big(\mathbb{E}[z_i \,|\, d_i = 0] - \mathbb{E}[x_i \,|\, d_i = 0]\Big). \tag{18}$$

Theorem 6. *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* $\phi < 0$. *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

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

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 (16)–(17) and the fact that $\mu_q = \mu_p$, we have

$$z_i = \frac{\sigma_p}{\sigma_q} x_i + \frac{\epsilon_i}{\sigma_q}.$$

By substituting for z_i into (18), we obtain

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

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

$$y_{it} = \beta_0 + \beta_1\, d_i + \beta_2\, post_t + \phi\, d_i \times post_t + \varepsilon_{it}, \tag{19}$$

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.

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 squares (OLS) estimates of the difference-in-differences regression (19) 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.

5.4 Many talents and many careers

The final extension in this subsection allows more than two talents and more than two careers. It is difficult to characterize the optimal *design* of education with many talents and many careers. However, we can continue to analyze the *value* of education: in particular, how the value of education can be separated into *ex ante* and *ex post* components. This extension aligns the model more closely with the empirical literature on the returns to multiple educational field and occupational choices, for example, Altonji, Arcidiacono, and Maurel (2016) and Arcidiacono et al. (2020).

Let the space of talents (states) Ω be any finite set $\{\omega_1, \omega_2, \ldots, \omega_m\}$. Let the set of careers (actions) A be any finite set $\{a_1, a_2, \ldots, a_k\}$. Each student i has a publicly known rational belief $p^i \in \Delta(\Omega)$. The *ex post productivity function* is $u : \Omega \times A \longrightarrow \mathbb{R}$. The *human capital accumulation function* is $\psi : \Omega \times A \longrightarrow \mathbb{R}$. An individual i's *ex post payoff* is

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

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

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

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 with 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

$$U_a(p) = \sum_{\omega \in \Omega} p(\omega) u(\omega, a).$$

The expected human capital accumulation of choosing a career a is

$$\Psi_a(p) = \sum_{\omega \in \Omega} p(\omega) \psi(\omega, a).$$

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

$$U_a(q) + \Psi_a(q)d + \nu_a \geq U_{a'}(q) + \Psi_a(q)d + \nu_{a'}. \tag{20}$$

Let $\mu_a(q)$ denote the probability of the event (20) conditional on q and $d = 0$. Let $\mu_a^{\psi}(q)$ denote the probability of the event (20) conditional on q and $d = 1$. A non-participant's expected productivity is

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

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

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

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

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

where the expectation is taken with respect to the possible realizations of the signal s.

The individual *treatment effect* for student i is

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

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

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

Theorem 7. *The individual treatment effect satisfies*

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

Proof. We have

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

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,

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

Then from the definition of τ^i, we have

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

Then by subtracting and adding back the same terms, we can write

$$\begin{aligned}\tau^i = V_1^{\psi}(p^i,\pi) - \sum_{a\in A}\mu_a(p^i)\mathbb{E}[y_i|i,d^i=1,a^i=a] \\ + \sum_{a\in A}\mu_a(p^i)\mathbb{E}[y_i|i,d^i=1,a^i=a] - \sum_{a\in A}\mu_a(p^i)\mathbb{E}[y_i|i,d^i=0,a^i=a],\end{aligned} \tag{21}$$

By the definition of U_a and Ψ_a, $\mathbb{E}[y_i|i,d^i=1,a^i=a] = U_a(p^i)+\Psi_a(p^i)$. Then the second term of (21) is $V_0^{\psi}(p)$. By the definition of τ_a^i, the sum of the third and fourth terms of (21) is $\sum_{a\in A}\mu_a(p^i)\tau_a^i$. Thus, we have

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

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 (the *ex post* gains), whereas the second component represents the self-exploration value of education (the *ex ante* gains).

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

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

a simple average of the treatment effect τ^i over all $i\in D$. Define the (local) *conditional average treatment effect* for a subset D of choosing a career a as

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

a simple average of the conditional treatment effect τ_a^i given the career a over all $i\in D$.

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

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

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 (a 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.

6 Empirical Application: Advanced Science Education

To illustrate an application of the self-exploration model of education, I estimate the value and design of advanced science classes in U.S. high schools.

Advanced Placement (AP) classes—or International Baccalaureate (IB), their international equivalent—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), math (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 also an education policy tool because of its potential role in encouraging students into sciences. For example, former President George 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 high school students' confidence in the sciences and the educational structure of the AP science classes suggest that the average value of providing these classes are about a 5-percent increase in students' earnings 8 years after their high school graduation. However, the estimates suggest that the classes may be too science-encouraging relative to the students' comparative advantage. With the opposite, science-discouraging design, the value of these classes would increase by 7 percentage points.

6.1 Data

I use the public-use datasets of the High School Longitudinal Study (HSLS) and the Education Longitudinal Study (ELS). First, the HSLS contains survey and transcript data following a nationally representative sample of 23,503 students who were 9th graders in 2009, from 944 high schools across 10 states in the United States. The survey 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, and their decisions a_i to pursue a STEM (science, technology, engineering, and math, including computer science; hereafter *science*) major in college. 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. I only use the data of students who have attended some 4-year colleges by the year 2016. There are 8,294 such students, which constitute about 35% of all respondents. After dropping observations with missing initial self-efficacy scores, there are 6,638 observations of students in the baseline sample.

Second, I use both the HSLS and the ELS[3] to obtain the predicted values of students' annual log earnings y_{it} as of 8 years after their graduation from high school.[4] The predicted log earnings correspond to students' *ex post* productivity in the model. I compute the predicted values on two estimated regressions: (a) a Logit regression on the binary outcome of being employed (having any earnings) and (b) an OLS regression on the actual earnings. The explanatory variables are the students' gender, race, parental education, college selectivity, completed college major, and college grades. Therefore, the predicted values from these regressions reflect the predictions based on the information at the time of completing college degrees. Appendix Table A.1 shows the estimated coefficients of these regressions. The pseudo-R^2 and the adjusted R^2 for the two regressions are about 5 percent and 10 percent, respectively. I compute the predicted earnings y_i for my sample as

$$\begin{aligned} y_i = &\Pr(\,employed_i \mid controls_i\,) \cdot \mathbb{E}[\,\log(earnings_i) \mid controls_i\,] \\ &+ \big(1 - \Pr(\,employed_i \mid controls_i\,)\big) \cdot \log(100), \end{aligned} \tag{22}$$

where i represents a respondent from the HSLS sample, $\Pr(employed_i|controls_i)$ is the predicted probability of i having some reported earnings, and $\mathbb{E}[\log(earnings_i)|controls_i]$ is i's predicted log earnings conditional on being employed. The predicted probability of i having no reported is $1 - \Pr(employed_i|controls_i)$. The term with $\log(100)$ means that I assume that the respondents with no reported earnings had an annual earnings of 100 U.S. dollars.[5] Thus, the resulting predicted earnings measure y_i does not condition on i being employed; large variations in y_i reflect the probabilities that respondents are not matched in the labor market.

Table 2 shows the summary statistics of these data. The 6,638 students in the baseline sample—all of the 4-year-college-attending respondents—had an average initial self-efficacy in sciences around 0.2. Since the variable is standardized in the total population

[3] Similar to HSLS but conducted earlier, the Education Longitudinal Study (ELS) contains survey and transcript data following a nationally representative sample of 15,362 high school students who were 10th graders in 2002.

[4] The reason for this step is that the most recent data of the HSLS are from year 2016. The earnings data, in particular, would represent the respondents' income in year 2015, only 3 years after the respondents' high school graduation. This means that most students' earnings would be from part-time jobs and would not necessarily represent their long-run productivity.

[5] The 100 dollars is the observed minimum of all annual earnings reported in the Education Longitudinal Study (ELS).

Table 2: Summary of beliefs, choices, and outcomes

		Initial Self-Efficacy in Science (x_i)		Predicted Log Earnings (y_i)	
Sample	Obs	Mean	Std Dev	Mean	Std Dev
All	6,638	0.206	(0.960)	9.106	(0.899)
Non-participant, non-science major	3,892	0.030	(0.949)	9.116	(0.841)
Non-participant, science major	857	0.355	(0.928)	9.081	(1.028)
Participant, non-science major	1,104	0.376	(0.903)	9.172	(0.853)
Participant, science major	785	0.676	(0.896)	8.995	(1.065)
Female	3,681	0.068	(0.959)	9.061	(0.877)
Non-participant, non-science major	2,296	-0.083	(0.947)	9.102	(0.825)
Non-participant, science major	336	0.224	(0.899)	8.761	(1.019)
Participant, non-science major	700	0.265	(0.927)	9.201	(0.814)
Participant, science major	349	0.516	(0.926)	8.797	(1.049)
Male	2,957	0.377	(0.934)	9.163	(0.923)
Non-participant, non-science major	1,596	0.192	(0.930)	9.135	(0.862)
Non-participant, science major	521	0.439	(0.938)	9.288	(0.981)
Participant, non-science major	404	0.568	(0.827)	9.123	(0.915)
Participant, science major	436	0.805	(0.850)	9.153	(1.052)

Note: The summary statistics are from a baseline sample of 6,638 high school students who were 9th graders in 2009 and enrolled in 4-year colleges during the period 2012-2016. The sample uses the public-use dataset of High School Longitudinal Study (HSLS). *Participants* are students who have taken at least one Advanced Placement (AP) science classes during high school (that is, $d_i = 1$). *Science majors* are students who considered majoring in a STEM field most seriously when first entering college (that is, $a_i = 1$). *Initial self-efficacy in science* (x_i) measures 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. The measure is standardized to have zero mean and unit standard deviation for the total population of students. *Predicted log earnings* (y_i) are predicted values of log earnings 8 years after high school graduation. They are based on an estimated regression of log earnings on gender, race, parental education, college major, college selectivity, and college grade point average (GPA) using the public-use data of the Education Longitudinal Study (ELS).

of students, this statistic means that the 9th grade students who later attended 4-year colleges were on average more confident in the sciences than others. Their mean predicted log earnings is 9.1, about 8,955 U.S. dollars. In other words, an average student in the sample faced an expected earnings at this level, taking into account the probability that they may not be employed.

Each of the 6,638 students in the sample belongs to one of four groups depending on their participation in AP science classes and their major choice: (1) a non-participant and a non-science major, (2) a non-participant and a science major, (3) a participant and a non-science major, and (4) a participant and a science major. About 28% participated in some Advanced Placement science classes, and about 71% of the participants eventually chose to pursue a science major when they first entered college. In contrast, among the non-participants, much smaller percentage, 22%, chose to pursue a science major. The (3) participant-and-non-science-major group had the highest predicted log earnings, followed by (1), (2), and (4).

We see a roughly similar distribution of choices in the female and male subsamples: AP science participants had much higher percentage of students who later chose to pursue science majors. However, we also see remarkable differences. The female subsample had much lower confidence in the sciences than the male subsample for all of the four groups. The female subsample also had lower predicted earnings for all of the four groups, and did so especially for the two groups that chose to pursue science majors.

6.2 Estimation

I infer the parameters of the extended model using maximum likelihood estimation. The benchmark model analyzed in Sections 2–3 are not adequate for estimation because its predictions are deterministic, whereas students' decisions in the data are noisy. The extended model for estimation includes three additional components introduced in Section 5: stochastic choice, human capital accumulation, and imperfectly observed beliefs. It maintains the assumption of binary states (talent in non-science or science) and binary actions (decisions to pursue a non-science major or a science major).

Consider a student $i \in I$ while suppressing the index. As a convention, let us interpret the hunting talent (ω_h) and the hunting career (a_h) as the science talent and the choice to pursue a science major in college. Similarly, let us interpret the gathering talent (ω_g) and the gathering career (a_g) as the non-science talent and choice to pursue a non-science major. Let us maintain that the prior belief score x in science satisfies

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

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

The student's participation decision follows

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

where V_1^{ψ} is the expected earnings of a participant with belief under the educational structure (γ, η) and the human capital accumulation function $\psi : \Omega \times A \longrightarrow \mathbb{R}$. The error terms ν_0 and ν_1 are independently and identically distributed as Gumbel$(0, \beta_d)$. A higher β_d means that the errors ν_0 and ν_1 have greater variance.

The student's decision to pursue a science major in college follows

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

by letting $a_g = 0$ and $a_h = 1$. The error terms ν_g and ν_h are independently and identically distributed as Gumbel$(0, \beta_a)$.

Finally, a person's *ex post* productivity y, measured in the data as predicted earnings, is

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

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

Let us denote the parameter vector as

$$\Lambda = (\mu_p, \sigma_p, \gamma, \eta, \delta, \beta_d, \beta_a, u_{gg}, u_{gh}, u_{hg}, u_{hh}, \psi_{gg}, \psi_{gh}, \psi_{hg}, \psi_{hh}, \sigma_{\varepsilon}).$$

where we denote $u_{gh} = u(\omega_g, a_h)$, $\psi_{gh} = \psi(\omega_g, a_h)$, and so on. Let $\mathbf{x} = (x, d, a, y)$ denote the data vector for the student i. The likelihood function for this student is

$$\mathcal{L}^i(\Lambda|\mathbf{x}) = f_{d,a,y|\Lambda,x}(d, a, y \mid \Lambda, x) = \sum_{\omega \in \Omega} \sum_{q \in (0,1)} f_{d,q,a,\omega,y|\Lambda,x}(d, q, a, \omega, y|\Lambda, x), \tag{23}$$

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

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

The first probability term is

$$\Pr(d=1|\Lambda,x) = \frac{\exp\left\{\frac{1}{\beta_d}\left[v^{\psi}(\mu_p+\sigma_p x,\gamma,\eta)-\delta\right]\right\}}{1+\exp\left\{\frac{1}{\beta_d}\left[v^{\psi}(\mu_p+\sigma_p x,\gamma,\eta)-\delta\right]\right\}}, \text{ and}$$
$$\Pr(d=0|\Lambda,x) = 1-\Pr(d=1|\Lambda,x),$$

where $v^{\psi}(p,\gamma,\eta)=V_1^{\psi}(p,\gamma,\eta)-V_0(p)$. The second probability term, $\Pr(q|\Lambda,p,d)$ is

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

The third probability term is

$$\Pr(a=0|\Lambda,q,d) = \frac{\exp\left\{\frac{1}{\beta_a}\left[U_g(q)+\Psi_g(q)-U_h(q)-\Psi_h(q)\right]\right\}}{1+\exp\left\{\frac{1}{\beta_a}\left[U_g(q)+\Psi_g(q)-U_h(q)-\Psi_h(q)\right]\right\}}, \text{ and}$$
$$\Pr(a=1|\Lambda,d,q) = 1-\Pr(a=0|\Lambda,q).$$

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

$$f(y|\Lambda,d,a,\omega)=\phi\left(\frac{y-u(\omega,a)-\psi(\omega,a)\cdot d}{\sigma_\varepsilon}\right),$$

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

Table 3 shows the estimates of the model parameters. Under the first category, we have parameters about beliefs, education, and choices. The estimated mean of prior beliefs, μ_p, is 0.26. The estimated standard deviation of prior beliefs, σ_p, is 0.03. Taken together, these two estimates imply that students' beliefs are scattered roughly between 20 and 30 percent. Most 9th grade students have only about a quarter chance of having a science talent.

The central parameter of our interest, the information structure (γ,η) of advanced science classes, is estimated to be $(0.74, 0.91)$. That is, the signal accuracy for a student talented in non-science is 74 percent, whereas that for a student talented in

Table 3: Parameter estimates and the value of advanced science education

Parameter	Description	Estimate	Std Err
Belief, education, and choices			
μ_p	Mean prior belief on science talent	0.256	(0.007)
σ_p	Std dev of prior beliefs on science talent	0.032	(0.003)
γ	Signal accuracy for non-science-talented students	0.736	(0.021)
η	Signal accuracy for science-talented students	0.909	(0.028)
δ	Cost of participation	0.300	(0.032)
β_d	Scale of participation choice errors	0.159	(0.018)
β_a	Scale of major choice errors	0.263	(0.021)
Ex post productivity			
$u(\omega_g, a_g)$	Non-science-talented student choosing a non-science major	9.412	(0.012)
$u(\omega_g, a_h)$	Non-science-talented student choosing a science major	8.165	(0.042)
$u(\omega_h, a_g)$	Science-talented student choosing a non-science major	7.736	(0.040)
$u(\omega_h, a_h)$	Science-talented student choosing a science major	9.823	(0.029)
Human capital accumulation			
$\psi(\omega_g, a_g)$	Non-science-talented student choosing a non-science major	-0.030	(0.023)
$\psi(\omega_g, a_h)$	Non-science-talented student choosing a science major	-0.381	(0.058)
$\psi(\omega_h, a_g)$	Science-talented student choosing a non-science major	-0.399	(0.086)
$\psi(\omega_h, a_h)$	Science-talented student choosing a science major	-0.151	(0.041)
Other parameter			
σ_ε	Std dev of predicted log earnings errors	0.565	(0.006)
Average value of providing advanced science education			
$\widetilde{W}(\gamma, \eta)$	Status quo	0.048	(0.010)
$\widetilde{W}(0.5, 0.5)$	No information	-0.008	(0.002)
$\widetilde{W}(\eta, \gamma)$	Opposite educational structure	0.123	(0.030)
$\widetilde{W}(1, 1)$	Full information	0.442	(0.032)
Average log likelihood		-2.347	–
Pseudo-R^2 on participation choice		0.033	–
Pseudo-R^2 on science major choice		0.066	–
Observations		6,638	–

Note: This table shows the maximum likelihood estimates of the extended model parameters and the value of education using a sample of 6,638 respondents in the High School Longitudinal Study (HSLS) who were 9th graders in 2009. Pseudo-R^2 follows McFadden (1973).

science is 91 percent. This educational structure is highly science-*encouraging*, in the sense that 26 percent of students with no talent in science would receive misguided recommendations to pursue science, whereas few science-talented students would receive misguided recommendations to pursue anything other than science.

The remaining estimates in the category, though not as important, are informative. The estimated cost of participation, 0.3, suggests that taking advanced science classes is significntly painful to students, to a degree that they would avoid it unless they see an expected 35 percent gain in their future earnings. The two scale parameters of choice errors, β_d and β_a estimated as 0.16 and 0.26, indicate that there are significant factors other than students' beliefs about their science talent that influence their decisions.[6]

The next set of parmeters are the values of the *ex post* productivity function u. The estimates imply that students who are talented at non-science subjects (correctly) choosing non-science majors have predicted annual log earnings of 9.4 (about 12,000 USD) 4 years after college graduation. Using the same interpretation, we see that those talented at non-science subjects who (incorrectly) choose science majors have predicted annual log earnings of 8.2 (about 3,500 USD). Those talented at science subjects who (incorrectly) choose non-science majors have predicted annual log earnings of 7.7 (about 2,300 USD). Those talented at science subjects who (correctly) choose science majors have predicted annual log earnings of 9.8 (about 18,400 USD). The seemingly small predicted annual earnings, especially for students with "incorrect" choices, are due to the fact that the predicted log earnings account for the probability of not being employed, as defined in equation (22).

Next, we have the values of the human capital accumulation function ψ. The estimate of the value $\psi(\omega_g, a_g)$ is not significantly different from zero, suggesting that taking an advanced science class in high school does not directly affect the future earnings for a student talented in non-science subjects majoring in a non-science subject. However, the estimates of the three remaining function values $\psi(\omega_g, a_h)$, $\psi(\omega_h, a_g)$, and $\psi(\omega_h, a_h)$, are significantly negative at –0.38, –0.40, and –0.15, meaning that an advanced science class directly lowers the future earnings of the other students. The negative estimates are especially large for the mismatched students: non-science-talented students choosing a science major and science-talented students choosing a non-science major. Although it may seem strange that the parameters capturing the accumulation of human capital are negative, a probable reason behind this result is that our estimated model do not account for different earnings trends over the lifecycle for different student types; it interprets one's predicted future earnings 4 years after college graduation as one's ex post outcome. If taking an advanced science class in high school is somehow associated with smaller temporary earnings but larger eventual earnings, the estimated parameters would capture such association. Therefore, the estimated human capture function ψ should not be interpreted literally but rather as parameters that capture all *ex post* effects.

[6] With the error terms having a Gumbel distribution, their estimated standard deviations are $\pi/\sqrt{6} \times 0.16 \approx 0.21$ and $\pi/\sqrt{6} \times 0.26 \approx 0.33$.

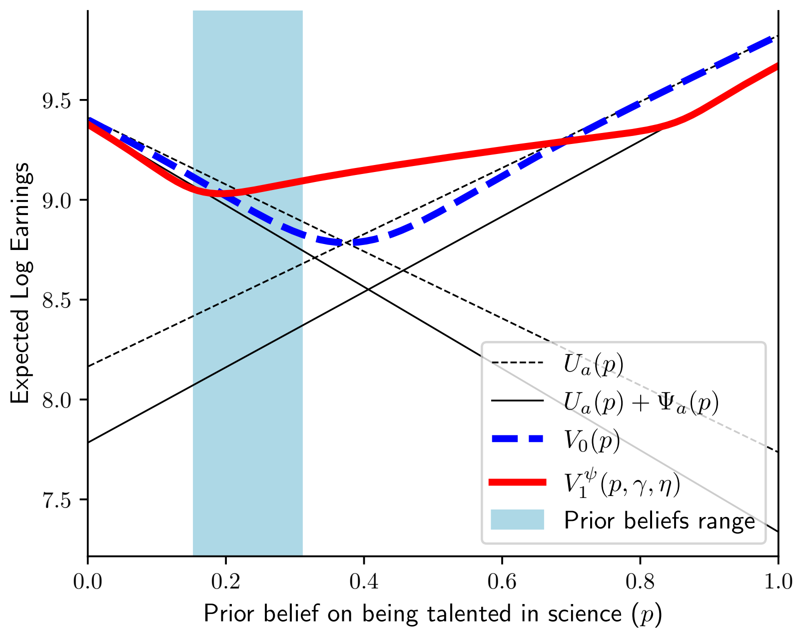

Figure 14: Estimated expected productivity with and without advanced science education

The last parameter, σ_ε, is the standard deviation of error term ε on the log of future earnings, estimated as 0.57. Since this estimate is smaller than the standard deviation of all predicted log earnings (about 0.9, from the first row of Table 2), it indicates that the model explains at least some of the variation in the predicted earnings. The bottom part of Table 3 shows a few other measures of fit. The average of the log likelihood defined in equation (23) is –2.347, meaning that the (geometric) average of the likelihood of the data (d, a, y) given the estimated parameters is $\exp(-2.347) \approx 10\%$. Following McFadden (1973), I also compute the pseudo-R^2, for which 0% means that the model's predictions are no better than independent and identical draws from a bernoulli distribution, and 100% means that the predictions are perfectly accurate. The pseudo-R^2s on the participation choice (d) and the science major choice (a) are 3% and 7%, indicating the model's moderate fit to the data.

Figure 14 shows these parameters in the estimated version of the earlier Figure 13. The horizontal axis is a student's prior belief in science talent. The two thin dashed lines represent a *non-participant*'s expected log earnings $U_g(p)$ and $U_h(p)$ when choosing to pursue a non-science and science majors, respectively. The dashed blue curve is the "smoothed" upper envelope of these two lines, representing the expected log earnings $V_0(p)$ of a non-participant who optimally chooses between a non-science or a science major given his belief p. The smoothing is due to the stochastic choice errors in the major decision. Similarly, the two thin solid lines represent a *participant*'s expected productivity $U_g(p) + \Psi_g(p)$ and $U_h(p) + \Psi_h(p)$ when choosing to pursue a non-science and science majors, respectively. The solid red curve represents the smoothed upper envelope of these two lines, representing the expected log earnings $V_1^\psi(p)$ 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 log earnings, the value of taking an advanced science class in high school.

We can further see from Figure 14 the range of students' prior beliefs and the structure of advanced science education. The figure's shaded area indicates the estimated range of the student beliefs, from about 15 to 31 percent. Notably, all students have prior beliefs significantly below the threshold levels where the thin straight lines cross, around 40 percent, at which they would be indifferent between choosing a non-science major or a science major. Thus, in 9th grade, students are believed to have comparative disadvantage in the sciences; for these students, the optimal educational structure would be non-science-encouraging according to our theory (Theorem 2 and Corollary 8). However, the estimated educational structure is science-encouraging. The estimated expected log earnings for participants (thick solid curve) is such that the resulting self-exploration value is greater on the other side of the threshold. This shape of the curve suggests that the student welfare would rise if the advanced science classes had the opposite structure of the status quo.

I run such thought experiments on four alternative educational structures: (1) the status quo (γ, η), (2) no information $(0.5, 0.5)$, (3) the opposite structure of the status quo (η, γ), and (4) full information $(1, 1)$. Under the estimated status quo $(0.74, 0.91)$, the signal accuracies for the non-science-talented and the science-talented are 74 and 91 percent. Under the no information strcutrue, the signal accuracies are both 50 percent. Under under the opposite structure, the signal accuracies are 91 and 74 percent. Under the full information structure, the signal accuracies are 100 percent. To compare these four structures, I use a modified measure of student welfare $\widetilde{W}$:

$$\widetilde{W}(\gamma, \eta) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\Pr(d_i = 1 \mid p_i)}_{\substack{\text{participation} \\ \text{probability}}} \cdot \underbrace{\left[V_1^\psi(p_i, \gamma, \eta) - V_0(p_i) \right]}_{\text{value of education}}.$$

That is, I take each student's value of education (the expected increase in future earnings) and take an weighted average of this value across all students, with their participation probabilities as their weights. The straightforward interpretation is that this measure represents the average value of providing the advanced science education to students. In other words, it captures the average *intent-to-treat* effect, where the treatment is taking an advanced science class in high school.

The remaining part of Table 3 shows the results of these thought experiments. First, under the status quo, the estimated average value of providing advanced science education in high school is about 5 percent. The interpretation is as follows. Students with higher beliefs (closer to 31%) have higher participation probability and larger value of advanced science education. Students with lower beliefs (closer to 15%) have lower participation probability and smaller value of advanced science education. On average,

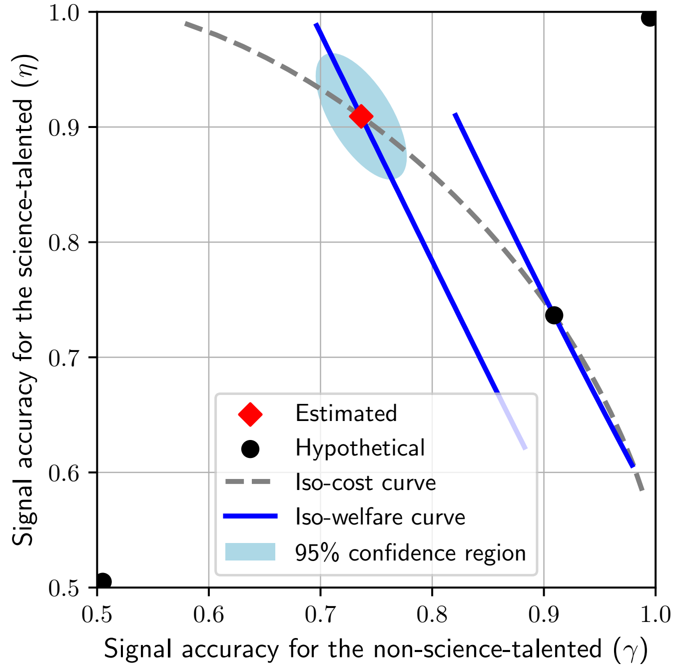

Figure 15: The value and design of advanced science education under estimated and hypothetical educational structures

after adjusting for the participation probabilities, the value of advanced science education is a 5-percent increase in students' predicted earnings 4 years after college.

Second, under the no-information structure, the estimated value of providing advanced science education is –0.8 percent. Since there is no information gain from this structure, this estimate represents the human capital component of the total value of providing this education. Thus, the negative (although small) estimate suggests that participation in an advanced science class is negatively associated with *ex post* earnings.

Third, under the opposite structure, the estimated value of providing the advanced science education is 12 percent. The intuition behind this gain compared to the status quo is that most students have low confidence in the sciences. Thus, it is better to make the structure science-*discouraging*: signals to non-science-talented students should be more accurate than those to science-talented students. The higher signal accuracy for the non-science-talented students would mean that fewer non-science-talented students would receive misguided recommendations to pursue science.

Fourth, under the full information, the estimated value of providing the advanced science education is 44 percent. The full-information structure means that everyone who takes an advanced science class receives a perfectly accurate signal about oneself. Receiving such a signal means that one can choose the correct major and get higher future earnings with certainty. It is thus not surprising that the estimated value of the full information is so large.

Figure 15 illustrates the four educational structures on the parameter space and their welfare implications. The red diamond marker represents the estimated educational structure of the status quo, whereas the black circle markers in the counter-clockwise order represent the educational structures under no information, opposite information, and full information. The gray dotted curve is the *iso-cost curve*, the set of points with the same value of information cost $C(\gamma, \eta)$ whose functional form is the expected reduction in entropy from a fixed prior $p_0 = 0.5$, from Example 1. The blue solid curve is the *iso-welfare curve*, the set of points with the same welfare $\widehat{W}(\gamma, \eta)$. Recall from equation (7) that the slope of an iso-welfare curve is inversely related to the average belief $\overline{p}$. Because the students' estimated average belief is low, the iso-welfare curve passing through the status quo has a steep slope. We see that the iso-welfare curve passing through the hypothetical opposite structure is much closer to the upper-right corner, nearly tangent to the iso-cost curve and thus closer to being a feasible optimum. The figure also shows that the no-information structure would be feasible but not optimal; the full-information structure would be infeasible.

Lastly, parameter estimates from female and male subsamples reveal that there may be important differences in the educational structure faced by students of different genders. Appendix Table A.2 shows the estimated parameters and their welfare implications using female and male subsamples. An immediately noticeable difference is that female students have a significantly lower average belief (20 percent) in science talent than their male counterparts (39 percent). However, female students are estimated to face more informative educational structure, as both of their signal accuracy parameters, γ and η, are greater than those of male students. As a result, the estimated welfare gain from advanced science classes for female students is 4.2 percent in future earnings, whereas that for male students is statistically indistinguishable from zero. Welfare estimates under the hypothetical educational structures suggest that female students would benefit significantly from a more non-science-encouraging educational structure, whereas male students would gain little. Appendix Figure A.1 illustrates this result.

7 Conclusion

This paper has shown 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 under-confident group of students to pursue sciences if their socioeconomic background contributes to such a bias.

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.

Arcidiacono, Peter, V Joseph Hotz, Arnaud Maurel, and Teresa Romano (2020) "Ex ante returns and occupational choice," *Journal of Political Economy*, 128 (12), 4475–4522.

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.

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.

Denti, Tommaso, Massimo Marinacci, and Aldo Rustichini (2022) "Experimental cost of information," *American Economic Review*, 112 (9), 3106–23.

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.

Li, Hongyan and Xing Xia (2022) "Grades as Signals of Comparative Advantage: How Letter Grades Affect Major Choices," *Available at SSRN*.

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.

McFadden, Daniel (1973) "Conditional logit analysis of qualitative choice behavior."

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.

Appendix

Table A.1: Estimated regressions of labor market outcomes on demographics and educational outcomes

	Dependent variable (Model):			
	Has any earnings (Logit)		Log earnings (OLS)	
Independent variable	Estimate	Std Err	Estimate	Std Err
Gender				
Female	0.074	(0.072)	-0.193***	(0.026)
Unknown	1.017	(0.951)	0.598	(0.990)
Race				
Amer. Indian/Alaska Native, non-Hispanic	0.808	(0.742)	-0.362*	(0.212)
Asian, Hawaii/Pac. Islander,non-Hispanic	-0.762***	(0.104)	-0.084*	(0.049)
Black or African American, non-Hispanic	-0.105	(0.114)	-0.129***	(0.045)
Hispanic, no race specified	-0.028	(0.194)	-0.018	(0.075)
Hispanic, race specified	-0.319**	(0.150)	-0.089	(0.058)
More than one race, non-Hispanic	-0.408***	(0.154)	-0.023	(0.060)
Unknown	-0.013	(1.092)	-0.647	(0.641)
Parental education				
Less than high school	0.249	(0.214)	-0.207**	(0.093)
Associate's degree	0.220	(0.141)	0.015	(0.047)
Bachelor's degree	0.031	(0.090)	0.001	(0.031)
Master's degree	-0.036	(0.107)	-0.009	(0.037)
Ph.D/M.D/Law/other high lvl prof degree	0.040	(0.127)	-0.053	(0.048)
Unknown	-1.314	(1.092)	-0.218	(1.178)
College selectivity				
Highly selective, 4-year institution	-0.055	(0.087)	0.139***	(0.029)
Inclusive, 4-year institution	-0.082	(0.103)	-0.052	(0.039)
Selectivity not classified, 4-year institution	0.018	(0.111)	-0.060	(0.043)
College major				
Agriculture, Agriculture Operations, and Related Sciences (1)	0.252	(0.421)	-0.104	(0.200)
Natural Resources and Conservation (3)	0.739	(0.538)	-0.009	(0.115)
Architecture and Related Services (4)	2.161**	(1.022)	-0.113	(0.173)
Communication, Journalism, and Related Programs (9)	0.968***	(0.226)	0.173***	(0.057)
Computer and Information Sciences and Support Services (11)	1.283***	(0.435)	0.386***	(0.118)
Education (13)	0.885***	(0.237)	0.164***	(0.058)
Engineering (14)	0.747***	(0.221)	0.583***	(0.056)
Engineering Technologies/Technicians (15)	1.553**	(0.733)	0.362*	(0.215)
Foreign languages, literatures, and Linguistics (16)	0.269	(0.342)	0.085	(0.139)
Family and Consumer Sciences/Human Sciences (19)	1.247**	(0.525)	0.092	(0.109)
English language and literature/letters (23)	0.162	(0.228)	-0.216**	(0.104)
Liberal arts/sci/gen studies/humanities (24)	1.309**	(0.523)	-0.091	(0.136)
Biological and biomedical sciences (26)	-0.578***	(0.151)	-0.320***	(0.087)
Mathematics and statistics (27)	1.118**	(0.477)	0.079	(0.104)
Multi/interdisciplinary studies (30)	0.656**	(0.320)	-0.068	(0.116)
Parks/recreation/leisure/fitness studies (31)	0.197	(0.251)	-0.069	(0.099)
Physical sciences (40)	-0.251	(0.288)	-0.086	(0.153)
Psychology (42)	0.568***	(0.192)	-0.095	(0.075)
Security and protective services (43)	0.745**	(0.343)	0.315***	(0.083)
Public administration/social service (44)	1.204**	(0.471)	0.100	(0.095)
Social sciences (45)	0.758***	(0.168)	0.001	(0.062)
Visual and performing arts (50)	0.826***	(0.210)	-0.185***	(0.069)
Health/related clinical sciences (51)	1.238***	(0.230)	0.447***	(0.061)
Business/management/marketing/related (52)	0.990***	(0.136)	0.435***	(0.037)
History (54)	0.382	(0.284)	-0.201*	(0.115)
Other	1.391***	(0.398)	-0.072	(0.085)
College grades				
College grade-point average (GPA), standardized	0.053	(0.044)	0.121***	(0.018)
Intercept	1.392***	(0.097)	10.080***	(0.036)
Pseudo-R^2	0.053	–	-	-
Adjusted R^2	–	–	0.098	–
Observations	6,908	–	5,808	–

Note: This table shows the estimated regressions of labor market outcomes on demographics and education. The estimates use the public-use dataset of the Education Longitudinal Study (ELS) that follow respondents who were 10th graders in 2002. *Has any earnings* is an indicator variable that equals 1 if a respondent has any earnings as of 2012 and equals 0 otherwise. *Log earnings* is as of 2012 among respondents who have any earnings in that year. *Parental education* is the highest degree earned by either of respondents' parents/guardians. *College major* is the major or field of study on the respondent's most recent bachelor's degree. *College grade-point average* is the respondent's GPA on the transcripts of all known institutions attended. The standard errors for the OLS estimates are heteroskedasticity-robust. Stars *, **, and *** indicate statistical significance at 10, 5, and 1 percent levels.

Table A.2: Parameter estimates and the value of advanced science education: female vs. male subsamples

	Female		Male	
Parameter	Estimate	Std Err	Estimate	Std Err
Belief, education, and choices				
μ_p	0.202	(0.008)	0.393	(0.017)
σ_p	0.025	(0.004)	0.017	(0.006)
γ	0.759	(0.032)	0.743	(0.022)
η	0.973	(0.062)	0.868	(0.022)
δ	0.258	(0.041)	0.192	(0.047)
β_d	0.134	(0.023)	0.051	(0.019)
β_a	0.300	(0.027)	0.155	(0.050)
Ex post productivity				
$u(\omega_g, a_g)$	9.366	(0.014)	9.543	(0.024)
$u(\omega_g, a_h)$	8.102	(0.055)	8.210	(0.080)
$u(\omega_h, a_g)$	7.582	(0.046)	8.214	(0.066)
$u(\omega_h, a_h)$	9.714	(0.053)	9.844	(0.037)
Human capital accumulation				
$\psi(\omega_g, a_g)$	0.010	(0.028)	-0.173	(0.042)
$\psi(\omega_g, a_h)$	-0.317	(0.077)	-0.467	(0.103)
$\psi(\omega_h, a_g)$	-0.272	(0.112)	-0.960	(0.139)
$\psi(\omega_h, a_h)$	-0.151	(0.070)	-0.133	(0.054)
Other parameter				
σ_ε	0.551	(0.008)	0.601	(0.013)
Average value of providing advanced science education				
$\widetilde{W}(\gamma, \eta)$	0.042	(0.013)	0.043	(0.035)
$\widetilde{W}(0.5, 0.5)$	-0.003	(0.004)	0.000	(0.024)
$\widetilde{W}(\eta, \gamma)$	0.204	(0.094)	0.043	(0.035)
$\widetilde{W}(1, 1)$	0.396	(0.042)	0.523	(0.070)
Average log likelihood	-2.213	–	-2.465	–
Pseudo-R^2 on participation choice	0.030	–	0.040	–
Pseudo-R^2 on science major choice	0.067	–	0.061	–
Observations	3,681	–	2,957	–

Note: This table shows the maximum likelihood estimates of the extended model parameters and the value of education using a sample of 3,681 female respondents and 2,957 male respondents in the High School Longitudinal Study (HSLS) who were 9th graders in 2009. See Table 3 for the description of parameters. Pseudo-R^2 follows McFadden (1973).

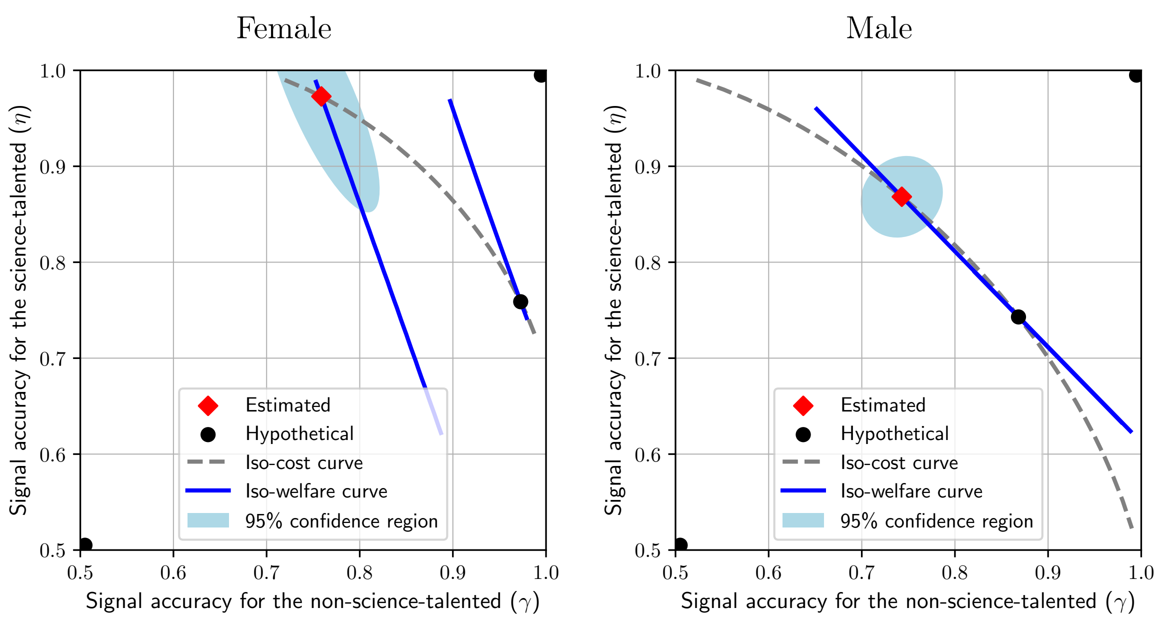

Figure A.1: The value and design of advanced science education: female vs. male subsamples