The Knowledge Trap: Human Capital and Development Reconsidered*

Benjamin F. Jones†

July 2014

Abstract

This paper presents a model where human capital differences may explain several central phenomena in the world economy. Human capital differences emerge via the division of labor, which underpins the quality and quantity of skilled workers. Low quality occurs when skilled workers fail, collectively, to embody advanced knowledge. Traditional human capital accounting is shown to underestimate resulting skill differences between rich and poor nations. The theory may explain price, wage and income differences across countries and further suggests novel interpretations of immigrant outcomes, poverty traps, and the brain drain, among other applications.

Keywords: human capital, education, technology, TFP, relative prices, wages, cross-country income differences, immigration, international trade, multinationals, poverty traps, skill-bias

*I wish to thank Daron Acemoglu, Pol Antras, Mark Bils, Andrew Hertzberg, Chad Jones, Sam Kortum, Jin Li, Pablo Montagnes, Ben Olken, Scott Stern, Kane Sweeney and participants in presentations at Northwestern, Washington University, Penn State, the LSE, Yale, Wisconsin, the Philadelphia Fed, Chicago, BYU, BREAD, and the NBER Summer Institute for helpful comments. All errors are my own.

†Kellogg School of Management and NBER. Contact: 2001 Sheridan Road, Room 609, Evanston, IL 60208. Email: bjones@kellogg.northwestern.edu.

1 Introduction

To explain several central phenomena in economics, from the wealth and poverty of nations to patterns of world trade, standard economic frameworks require large, residual productivity differences. That is, explanations rely on some critical factor of production that is distinct from the contributions of physical and human capital. This paper presents an alternative view, showing how one may put "ideas" back into people, presenting a model where human capital differences can play an expanded role in the world economy and may help explain many stylized facts.

The starting point for this paper is a viewpoint where advanced productive knowledge is too great for one person to know, so that implementing advanced ideas will rely on a division of labor (Jones 2009). Skilled workers are seen as vessels of ideas. Human capital investment is seen as the embodiment of ideas in people, where the division of labor is needed to aggregate advanced knowledge. Productivity advantages emerge in the *collective* productivity of skilled workers, where specialists working in teams bring greater knowledge into production.

The theory thus builds on Adam Smith's foundational observation that the division of labor can bring high productivity as well as both classic and modern theories of economic geography that emphasize specialized skills (Smith 1776, Marshall 1920, Saxenian 1994). In this paper, labor division is explicitly motivated as above: it is necessary for employing the modern economy's advanced ideas in production, whether engineering jet turbines, performing thoracic surgery, or managing bond issuances. However, the acquisition of advanced knowledge is also challenging to achieve. Three challenges are emphasized. First, deep expertise may be hard to acquire locally (e.g. university quality is low). Second, coordination costs in production may be especially high.[1] Third, there may be strategic complementarities in decisions to specialize, creating persistent poverty when the initial supply of specialists is low. For any (or all) of these reasons, a low-productivity equilibrium may persist. I call such outcomes a "knowledge trap" because the unspecialized equilibrium

[1] The idea that coordination costs of teamwork limit the gains from specialization follows Becker & Murphy (1992). More broadly, the limits to specialization considered in this paper are based on local frictions, rather than on the extent of the market as in Smith (1776).

features shallower collective knowledge.[2]

In the model, differences in labor division among skilled workers underpin productivity differences across economies. At the same time, the macroeconomic implications of these productivity differences depend on a second form of labor division: the division between skilled and unskilled workers. That is, the equilibrium depends on both the *quality* and *quantity* of labor types. Quality depends on the collective capacity of skilled workers to obtain and aggregate advanced knowledge. Quantity depends on workers' collective decisions to obtain higher education.

When quality differences exist among skilled workers, the quantity of skilled labor naturally adjusts. For example, if substantial wage advantages emerge for skilled workers, more workers may naturally become skilled so that the relative price (and hence relative wages) of skilled versus unskilled services falls. This quantity adjustment pins down the model. In equilibrium, labor allocations adjust so that the wage returns become decoupled from the quality of skilled labor, with real income effects shared equally by skilled and unskilled workers alike.

This tandem of quality and quantity is crucial to understanding the macroeconomic effects. Among other implications, it poses significant challenges to traditional human capital accounting methods. The traditional approach infers cross-country skill differences from within-country returns to schooling, but in the model quantity adjustments mean that the entire wage distribution shifts, so that within-country wage equilibria say little about cross-country skill differences. Estimation approaches based on immigrant behavior face similar challenges. The wage gains experienced by unskilled workers who immigrate from poor to rich countries need not be explained by technology residuals; in this model, unskilled wage gains follow simply because unskilled workers, working as farm hands or taxi drivers, gain by moving to a place where they are relatively scarce.

In sum, human capital is viewed as the embodiment of ideas into people. Rich countries

[2]This perspective has a somewhat different emphasis from classic descriptions of specialization that emphasize the extent of the market (the demand side) or, in more modern literature, coordination costs as limits on specialization (Smith 1776, Becker and Murphy 1992). In those perspectives, specialization is good when it can be achieved. By contrast, in the above perspective, specialization is essential to accessing the stock of advanced ideas in a modern economy. That the division of labor is necessary for employing advanced ideas in production seems inevitable when the stock of productive knowledge is too great for one individual to know. See also Jones (2009).

attain deeper collective knowledge among skilled workers. Resulting adjustments in the quantity of skilled workers mean that the real wages of skilled and unskilled workers rise in equal proportion, even though unskilled workers have no more skill in rich than poor countries. One thus finds a skill-based interpretation of cross-country income differences that can also get wages right, while providing interpretations for many other stylized facts about the world economy.

This paper is organized as follows. Section 2 introduces the core ideas. Section 3 presents a formal model, examining mechanisms for the existence of knowledge traps and their general equilibrium effects. Section 4 discusses several applications and relates them to extant empirical literatures. I show that the model provides an integrated perspective on (i) cross-country income differences, (ii) immigrant labor market outcomes, and (iii) poverty traps, as well as price phenomena, including (iv) why some goods are especially cheap in poor countries and (v) why "Mincerian" wage structures appear in all countries. Section 4 also offers possible insights about (vi) the brain drain and (vii) the role of multinationals in development and then closes by discussing generalizations to inform (viii) international trade patterns, (ix) skill-biased technical change, and (x) income divergence across countries. Section 5 concludes.

Related Literature Many existing papers explore theoretical aspects of the division of labor (e.g. Kim 1989, Becker and Murphy 1992, Garicano 2000). Other papers explore multiple equilibria in human capital (e.g. Kremer 1993, Acemoglu 1996), and still others explore specialization in intermediate goods, i.e. at the firm level, as the source of development failures (e.g. Ciccone and Matsuyama 1996, Rodriguez-Clare 1996, Acemoglu et al. 2006). A key innovation in this paper is to imagine specialization in training as a basis for different organizational forms of labor supply. By emphasizing the division of labor and interactions across workers, the mechanisms here differ substantially from other treatments of human capital quality (e.g. Erosa et al. 2010; Manuelli and Seshadri forthcoming). More precisely, this paper imagines a two-dimensional education decision where both the breadth and duration of education are endogenous choices. There is thus a division of labor among skilled workers (based on breadth), and a division of labor between skilled and unskilled workers (based on duration).

This theoretical approach allows a reinterpretation of several empirical literatures, including the "macro-Mincer" approach in the vast development accounting literature (surveyed in Caselli 2005), which attempts to assess the role of human capital in cross-country income differences. These empirical literatures will be discussed in detail below.

2 The Core Ideas

This section introduces the core ideas in this paper. First, the quality of skilled workers is considered. Second, the quantity of skilled workers is considered. As one application, the tandem of quality and quantity differences is shown to disrupt traditional macroeconomic accounting methods, leading to an understatement of cross-country skill differences. Section 3 integrates these ideas into a formal model before discussing a broader set of applications.

2.1 The Quality of Skilled Labor

Modern production in rich countries appears to involve an enormous variety of expert knowledge, from microprocessors to jet propulsion, from polymer synthesis to optical switches, from radiation oncology to accounting consistent with the GAAP. As measures of this differentiation, consider that the U.S. Census recognizes over 31,000 different occupational titles, the U.S. Patent and Trademark Office recognizes 475 different primary technology classes, the ISI Web of Science organizes more than 15,000 research journals into 252 different fields, and the American Board of Medical Specialities recognizes physician certifications in 145 different areas.[3]

It seems infeasible for one individual to know more than a fraction of a modern economy's advanced knowledge. A basic challenge is then how – and whether – economies load this advanced knowledge into people's minds. If the set of productive knowledge is greater than what one person can acquire, then the acquisition of advanced knowledge becomes a collective enterprise - it depends on a division of labor.

To fix ideas, imagine there are two tasks, A and B, which are complementary in the

[3] See http://www.uspto.gov for patent classifications, Wuchty et al. (2007) for Web of Science field classifications, and http://www.abms.org for medical specialties.

production of a good. For example, the ultimate output could be a microprocessor, a gas turbine, or heart surgery, each of which builds on knowledge across complementary tasks.[4] Now imagine individuals must train to acquire skill. One might train as a "generalist", developing skill at both tasks. Alternatively, one might focus training on one task, becoming especially adept at that task. For simplicity, let training as a generalist produce a skill level 1 at both tasks, while training as a specialist produces a skill level $m > 1$ at one task and 0 at the other.

As an example, let production be $Y = \sqrt{H^A H^B}$ when working alone and cY when pairing with another worker. This Cobb-Douglas production function captures the complementarity between skills, and the term $c < 1$ represents a coordination penalty from working in a team. Output is per unit of clock-time, and the amount of skill applied to a particular task, e.g. H^A, is the summation of skill applied per unit of clock-time.

In this setting, a generalist working alone does best by dividing his time equally between tasks and earning $Y = \frac{1}{2}$. A pairing of complementary specialists optimally applies each worker to their specialty, producing $Y = mc$ for every unit of clock time, or $\frac{1}{2}mc$ per team member. The specialist organizational form is therefore more productive as long as $mc > 1$; that is, as long as coordination penalties do not outweigh the benefits of deeper expertise.

A "knowledge trap" occurs when the unspecialized state is a stable equilibrium, thus failing to access frontier knowledge. In a poor country, this may occur most simply because, locally, m is small. To motivate this idea, Table 1 compares the available instruction in the mechanical engineering departments of a top-ranked engineering school in East Africa (the

[4] Heart surgery requires surgical expertise (surgery), pain control (anesthesiology), as well as various complementary skills around diagnosis, infection control, and post-operative care. Microprocessor production combines microprocessor photolithography (the etching of the processor onto silicon, which draws on material science and optics), microprocessor design (including the instruction set architecture, memory, control and data path design, thermal analysis, etc), and microprocessor software (the assembler, compiler, debugger, etc) all of which draw on very different kinds of knowledge. Turbine production involves the integrated design and manufacture of turbine blades, turbofans, compressors, combustors, control systems, fuel systems, nozzles, et cetera, which draw on disparate and highly specific engineering expertise, including thermodynamics, material science, fluid mechanics, rotational and vibrational dynamics and high-heat electronics. One broadly-trained engineer working alone may be able to produce a simple integrated circuit or even a very simple turbine, but the advanced, highly productive versions (e.g. a low-power Intel Atom processor or a GE90 gas turbine) are not produced by one person. Joseph Palladino of General Electric Aircraft Engines (personal correspondence) estimates that 30-35 different disciplines are required to implement a modern jet engine. In this paper, I will consider a model with two complementary tasks to focus on the core ideas; generalizations to more tasks would make complementarities more acute.

University of Khartoum) and a top-ranked engineering school in the United States (MIT).[5] By this comparison, the training available in the U.S. appears deeper, narrower, and collectively spans much greater knowledge than the offerings in East Africa. While Khartoum offers 1 specialty area, MIT has 7 areas of study and 17 different specialized course groups within mechanical engineering. Overall, MIT offers 3.4 times as many subjects, which appear more specialized, as evident from the course titles in Table 1. This comparison is conservative, in the sense that MIT has two additional departments (Aeronautics and Astronautics, Nuclear Science and Engineering) that provide 141 additional courses in subject areas where Khartoum offers a total of 2 courses. This type of evidence suggests that the advanced knowledge underlying many high-value added industries may be difficult to access through higher education in poorer countries.[6]

While the ability to learn narrow, deep knowledge may be limited in poor countries (m is low), specialization may also be inhibited by coordination penalties ex-post in production (c is low). Such coordination penalties – downstream of skill acquisition – reduce the gains from narrow expertise and may thus dissuade workers from acquiring deep knowledge at complementary tasks.[7] Hence, most simply, poor countries may feature $m'c' < 1$ while a rich country has $mc > 1$, leading to potentially large differences in the quality of skilled labor.

More subtly, the unspecialized state may persist due to thin supply of complementary specialist types. To see this, imagine being born into an economy of generalists and consider the decision to become a specialist instead. The best you could do as a lone specialist would be to pair with an existing generalist. In such a pairing, the specialist focuses on the task where they have expertise, the generalist on the other, and the optimal output is $Y = \sqrt{m}c$. The generalist would have to be paid at least their outside option, $\frac{1}{2}$, to willingly join the

[5] Khartoum most directly draws students from Sudan, Ethiopia, and Eritrea, countries with a combined population of 142 million. Within this part of East Africa, the University of Khartoum appears to define the upper limit for technical education; it currently enrolls 16,800 undergraduate and 6,000 graduate students, while MIT enrolls 4,300 undergraduate and 6,300 graduate students.

[6] Based on Table 1, and looking just at mechanical engineering, these industries would appear to include the production of modern airplanes, helicopters, satellites, and ships, as well as industries that rely on automated manufacturing (e.g. modern automobile production, modern chemical manufacturing), MEMS technologies (e.g. optical switches in telecommunications, gyroscopes in smart phones, accelerometers in air bags, piezoelectronics in inkjet printers), and many others.

[7] Becker and Murphy (1992) discuss numerous types of coordination costs that can inhibit the division of labor.

specialist in such a team. The most income the specialist could earn is therefore $\sqrt{mc} - \frac{1}{2}$, which itself must exceed $\frac{1}{2}$ to prefer training as a specialist. Hence the unspecialized equilibrium is stable to individual deviations if $\sqrt{mc} < 1$. We thus have a potential trap: for any coordination penalty in the range $\frac{1}{m} < c < \frac{1}{\sqrt{m}}$ mutual specialization is more productive and yet the generalist equilibrium is stable.[8]

I call this set of specialization failures a "knowledge trap" because skilled workers in the generalist equilibrium have *shallower* knowledge. While generalists may still invest substantial time in training,[9] specialists acquire deeper knowledge about individual tasks, with the potential to acquire, collectively, far greater knowledge and productivity. To see the implications of these quality differences, we must further consider the quantity of skilled labor, which we turn to next.

2.2 The Quantity of Skilled Labor

We now consider the division of labor between skilled and unskilled labor, where workers choose whether or not to become skilled. This choice connects variation in the quality of skilled workers to their resulting quantity and determines the equilibrium implications. To motivate the quantity dimension, once can start with an important application: human capital accounting.

A large literature has concluded that human capital variation across countries is too small to explain cross-country income differences (see Caselli 2005 for a survey). This inference is based on the wage-schooling relationship (e.g. Klenow and Rodriguez-Clare 1997, Hall and Jones 1999). If workers are paid their marginal products, the argument goes, then wage gains from schooling should inform how schooling influences productivity.

[8] This type of knowledge trap would be resolved by mutual specialization in complementary tasks, and one may ask why this coordination problem isn't resolved naturally in the market, especially by firms. The implicit assumption for this mechanism is that important educational decisions are primarily made prior to the interactions of individuals and firms, so that firms cannot coordinate major educational investments but rather make production decisions given the skill set of the labor force. This seems a reasonable characterization empirically, since skilled workers (engineers, lawyers, doctors, etc.) typically train for many years in educational institutions that are distinct from firms, before entering the workforce. In this sense, it then falls to other institutions to solve this type of coordination problem. These issues will be discussed further in Section 4.

[9] For example, a generalist medical doctor would know something about anesthesiology, surgery, infectious disease, oncology, psychiatry, ophthalmology, etc. Learning something about all of these areas may require a lot of education.

Based on the micro-literature, wage-schooling relationships are usually taken to follow the log-linear form (Mincer 1974, Card 2003),

$$w(s) = w(s')e^{r_m(s-s')} \tag{1}$$

where s is schooling duration, $w(s)$ is the wage, and r_m is the percentage increase in the wage for an additional year of schooling.[10] In measuring human capital, standard practice then calculates $h(s)/h(s') = w(s)/w(s')$, where $h(s)$ is the average skill of workers with schooling duration s and thus provides a measure of worker quality.

To see how quantity considerations can disrupt such quality inference, consider what happens when the quantity of skilled workers is endogenous. In particular, define a worker's lifetime income as

$$y(s) = \int_s^\infty w(s)e^{-rt}dt \tag{2}$$

where individuals earn no wage income during their s years of training and face a discount rate r. If in equilibrium workers cannot deviate to other schooling decisions and be better off, such that for any two schooling levels $y(s) = y(s')$, then (1) follows directly with $r_m = r$.[11] In this setting, the log-linear wage structure in (1) follows from equilibrium labor allocation decisions. Intuitively, when individuals invest time in training, they give up wages today in exchange for higher wages later; the wage returns become pinned down by the expected return on investment - i.e. the discount rate.[12] Under this reasoning, the wage returns are decoupled from skill returns. Put another way, quantity decisions hide quality differences.

This result pins down the wage returns in a strong way. How can relative wages follow this equilibrium return, regardless of the skill returns to schooling? The answer is again, a division of labor, which allows flexibility in the prices of human capital services. In

[10] Such log-linear wage-schooling relationships have been estimated in many countries around the world (see Psacharopolous 1994).

[11] This arbitrage argument follows in the spirit of Mincer (1958). Integrating (2) gives $y(s) = \frac{1}{r}w(s)e^{-rs}$ so that $y(s) = y(s')$ implies $w(s) = w(s')e^{r(s-s')}$. Equivalently, (1) follows if workers choose schooling duration to maximize lifetime income. That is, with $s^* = \arg\max y(s)$ we have

$$w'(s^*) = rw(s^*)$$

expressing log-linearity as a marginal condition.

[12] Here the interest rate and the return to schooling are equivalent. A richer model would introduce other aspects, such as ability differences, progressive marginal income tax rates, out-of-pocket costs for education, and finite time horizons which could, for example, drive the return to schooling above the real interest rate. See Heckman et al. (2005) for a broader characterization of lifetime income.

particular, when workers produce differentiated services that face downward sloping demand, the prices of services adjust flexibly in general equilibrium. For example, imagine that there are two services, service 1 (e.g. haircuts) produced by unskilled workers with no education and service 2 (e.g. engineering) that requires S years of training to perform. Imagine as above that skill, h, and time, L, are the only inputs to production, so that $x_1 = h_1 L_1$ and $x_2 = h_2 L_2$. The marginal product for each service is then $w_1 = p_1 h_1$ and $w_2 = p_2 h_2$, and we have

$$h_2/h_1 = \frac{p_1}{p_2} e^{rS} \tag{3}$$

where $w_2/w_1 = e^{rS}$ follows from the quantity decision as above.

To compare skill across countries, traditional accounting methods typically estimate skilled returns, h_2/h_1, as

$$h_2/h_1 = e^{rS}$$

an approach that appears incomplete. As shown in (3), one must also confront the relative prices of differentiated labor services – where general equilibrium adjustments will be felt.[13] Moreover, under the innocuous assumptions that poor countries are relatively abundant in low skill and that demand is downward sloping, p_1/p_2 will be relatively small in poor countries. Hence the skill gains from education (h_2/h_1) must be adjusted upwards in rich countries relative to poor countries.[14] These observations suggest not only that wage returns do not imply skill returns, but also that the traditional method may systematically understate skill differences across countries. Skill differences may therefore play a more important role in the world economy than a large literature has suggested.

The following section presents a general equilibrium model of the division of labor, integrating analysis of the quality and quantity of skilled labor supply and exploring mechanisms

[13] Relative price differences across countries are large and motivate purchasing power parity (PPP) price corrections when comparing real incomes. Note that the relative prices of interest here are not easy to observe directly, since generally they are prices of intermediate service outputs, rather than finished goods. In Section 4, we will consider calibrations where production function assumptions allow estimation without observing these intermediate prices.

[14] The standard accounting method assumes (implicitly in most treatments) that the output of different workers are perfect substitutes. In this case $p_1 = p_2$ (effectively, there is one good only). Under this assumption, one could estimate h_2/h_1 based purely on w_2/w_1. However, this assumption is unrealistic if we believe that worker types are less than perfect substitutes, as suggested by an extensive labor literature (e.g. Katz and Autor 1999) as well as prima facie evidence in the economy that different workers produce very different types of services. These literatures will be discussed further in Section 4.

for endogenous differences across countries. Section 4 details several applications and considers established empirical evidence from the model's perspective.

3 The Model

Imagine a world where workers are born, invest in skills, and then work, possibly in teams. They can work in one of two sectors. One sector requires only unskilled labor, and output is insensitive to the education level of the worker. Output in the other sector depends on formal education.

The key decision problem for the individual is what skills to learn. Skill type is chosen to maximize expected lifetime income. Once educated, the worker enters the labor force and produces output, which occurs efficiently conditional on the education decisions made and the ability to form appropriate teams. The educational decision is thus the key to the model.

3.1 Environment

There is a continuum of individuals of measure L. Individuals are born at rate $r > 0$ and die with hazard rate r, so that L is constant. Individuals are identical at birth and may either start work immediately in the unskilled sector or invest S years of time to undertake education. If they choose to educate themselves, they may develop skill at two tasks, A and B. We denote an individual's skill level $h = \{h_A, h_B\}$. An individual may choose to become a "generalist" and learn both skills, developing skill level $h = \{h, h\}$. Alternatively, one may focus on a single skill and develop deeper but narrower expertise, attaining skill level $h = \{mh, 0\}$ or $h = \{0, mh\}$ where $m > 1$.

3.1.1 Timing

For the individual, the sequence of events is:

1. The individual is born.

2. The individual makes an educational decision, becoming one of four types of workers[15]

[15] For simplicity, the model is developed where skilled workers – generalists or specialists - choose the

(a) Type U workers ("unskilled") undertake no education, $s^U = 0$, and have skill level $h^U = \{0, 0\}$.

(b) Type G workers ("generalists") undertake $s^G = S$ years of education and learn both tasks, developing skill level $h^G = \{h, h\}$.

(c) Type A workers ("A-specialists") focus $s^A = S$ years on task A, developing skill level $h^A = \{mh, 0\}$.

(d) Type B workers ("B-specialists") focus $s^B = S$ years on task B, developing skill level $h^B = \{0, mh\}$.

3. The individual enters the workforce.

(a) Unskilled workers (type U) go to work immediately in the unskilled sector.

(b) Skilled workers (types G, A, B) enter the skilled sector after S years and may choose to work alone or pair with other skilled workers.

i. Unpaired skilled workers randomly meet other unpaired skilled workers with hazard rate λ.

ii. If paired and your partner dies (at rate r), then you become unpaired again.

3.1.2 Preferences

Expected utility is given by

$$U^k = \int_0^\infty u(C^k(t))e^{-rt}dt$$

where $u(C)$ is increasing and concave. The effective rate of time preference is given by r, the hazard rate of death, which is equivalent to the discount rate.[16] This equivalence implies that an individual's consumption does not change across periods, by the standard Euler equation.[17]

same duration of education. The model could be alternatively developed where specialists undertake longer education than generalists (e.g. a Ph.D. on top of an undergraduate degree). That potentially increased level of realism increases the complexity of the exposition but does not add substantial theoretical insights and is therefore left aside.

[16] There is no physical capital in this model, so there is no rental rate of capital. However, there are loans, since players are born with no wealth and therefore those in school must borrow to consume. We imagine a zero-profit competitive annuity market where individuals hand over rights to their future lifetime income, W, upon birth in exchange for a payment, a, every period. This payment must be $a = rW$ by the zero profit condition. Therefore, the rate of interest on loans is the same as the hazard rate of death.

[17] The Euler equation is $\frac{du'(C)/dt}{u'(C)} = r - r = 0$, so that $u(C)$ and hence C are constant with time.

Let preferences across goods be

$$C^k(x_1, x_2) = \left(\gamma x_1^\rho + (1-\gamma)\, x_2^\rho\right)^{1/\rho} \tag{4}$$

where x_1 is the good produced by the unskilled sector, x_2 is the good produced by the skilled sector, and $\varepsilon = \frac{1}{1-\rho}$ is the elasticity of substitution between goods, which we assume is finite.

3.1.3 Income

The expected present value of lifetime income for a worker of type k is

$$W^k = \int_{s^k}^{\infty} rV^k e^{-r\tau} d\tau \tag{5}$$

where $s^k \in \{0, S\}$ is the duration of education. Time subscripts are suppressed because we will focus on steady-state equilibria. V^k is the value of being a type k worker at the moment your education is finished, which is the expected value of being an unpaired worker of type k. For unskilled workers, $rV^U = w_1$, where w_1 is the wage earned from producing the unskilled good. For skilled workers we have

$$rV^k = w_2^k + \lambda \sum_{j \in \Omega^k} \Pr(j) \left(V^{kj} - V^k\right) \tag{6}$$

The flow value of being unpaired, rV^k, equals the wage from working alone, w_2^k, in the skilled sector plus the expected marginal gain from a possible pairing. You meet other unpaired skilled workers at rate λ, and the unpaired skilled worker is type j with probability $\Pr(j)$. We assume a uniform chance of meeting any particular unpaired skilled worker, so that

$$\Pr(j) = L_p^j / L_p \tag{7}$$

where L_p^j is the measure of workers of type j who are unpaired and $L_p = \sum_j L_p^j$.[18] You accept the match if $V^{kj} \geq V^k$ and reject otherwise, which defines the "acceptance set",

[18] Note that this specification guarantees that the aggregate rate at which type k people bump into type j people ($\lambda \Pr(j) L_p^k$) is the same as the rate at which type j people bump into type k people ($\lambda \Pr(k) L_p^j$). Specifically,

$$\lambda \Pr(j) L_p^k = \lambda \left(L_p^j / L_p\right) L_p^k = \lambda \left(L_p^k / L_p\right) L_p^j = \lambda \Pr(k) L_p^j$$

$\Omega^k \subset \{G, A, B\}$, the set of types that a player of type k is willing to match with. If you reject, you remain in the matching pool. If you accept, you leave the matching pool and earn V^{kj}, which is defined

$$rV^{kj} = w_2^{kj} - r\left(V^{kj} - V^k\right) \tag{8}$$

The flow value of being paired, rV^{kj}, is equal to the wage you receive in this pairing, w_2^{kj}, less the expected loss from becoming a solo worker again, which occurs when your partner dies (with probability r).

Paired workers split the value of their joint output by Nash Bargaining, dividing the joint output such that

$$w_2^{kj} = \arg\max_{\hat{w}^{kj}} \left(V^{kj} - V^k\right)^{1/2} \left(V^{jk} - V^j\right)^{1/2} \tag{9}$$

Meanwhile, a solo worker earns the total value of his output when working alone.

3.1.4 Output

Sector 1 produces a simple good, x_1, with unskilled labor and with no advantage to skill in tasks A or B. Each worker in sector 1 produces with the technology

$$x_1 = z$$

per unit of clock time.

Sector 2 produces a good where skill at tasks A and B matters. Workers in sector 2 may work alone or with a partner, with the production function

$$x_2 = zc(n)\left(H_A^\alpha + H_B^\alpha\right)^{1/\alpha}, \; H_k = \sum_i t_i^k h_i^k \tag{10}$$

where $\sigma = \frac{1}{1-\alpha}$ is the elasticity of substitution between the two skills and we assume $\sigma \leq 1$, so that both inputs are necessary for positive production.[19] The term $c(n) \in [0, 1]$ captures the coordination penalty from working in a team of size $n \in \{1, 2\}$. Without loss of generality set $c(1) = 1$ and $c(2) = c$. The time devoted by individual i to task k is t_i^k, and members of a team split their time across tasks to produce maximum output.

[19] The CES production function in (10) is used for simplicity. The theory can be developed from a more general production function, $x_2 = c(n)f(H_A, H_B)$, where $f(H_A, H_B)$ is a symmetric, constant returns to scale function. Gross complements ($\sigma \leq 1$) provides substantial tractability but is not a necessary condition for the main results.

3.2 Equilibrium

An equilibrium is a decision by each worker that maximizes her utility given the decisions of other workers. The choice involves (a) maximizing lifetime income, and (b) maximizing utility of consumption given this lifetime income. We look at equilibria where all players of skilled type k have the same matching policy Ω^k that is constant with time.

It is convenient to define the equilibrium in terms of aggregate variables. Let L^k be the measure of living individuals who have chosen to be type k, and let L_q be the measure of workers actively producing the good of type q. Let X_q^S, X_q^D, and p_q respectively be the total supply, total demand, and price of good q.

Definition 1 *A steady-state equilibrium consists of W^k, V^k, C^k, L^k for all worker types $k \in \{U, G, A, B\}$; V^{kj}, Ω^k, L_p^k for all skilled worker types $k, j \in \{G, A, B\}$; and L_q, X_q^S, X_q^D, p_q for each good $q \in \{1, 2\}$ such that*

1. *(Income maximization: Choice of worker type)* $W^k \geq W^j \ \forall k \in \{U, G, A, B\}$ *such that* $L^k > 0$, $\forall j \in \{U, G, A, B\}$

2. *(Income maximization: Matching policy)* $j \in \Omega^k$ *for any* $j \in \{G, A, B\}$ *such that* $V^{kj} \geq V^k$, $\forall k \in \{G, A, B\}$

3. *(Consumer optimization)* $C^k(x_1, x_2) \geq C^k(x_1', x_2') \ \forall x_1, x_2, x_1', x_2'$ *such that* $p_1 x_1 + p_2 x_2 \leq rW^k$ *and* $p_1 x_1' + p_2 x_2' \leq rW^k$, $\forall k \in \{U, G, A, B\}$

4. *(Market clearing)* $X_q^D = X_q^S \ \forall q \in \{1, 2\}$

5. *(Steady–state)* L^k *is constant* $\forall k \in \{U, G, A, B\}$ *and* L_p^k *is constant* $\forall k \in \{G, A, B\}$

We will further focus on equilibria in the "full employment" setting, where $\lambda \to \infty$.

3.3 Analysis

We analyze the equilibria in this model in two stages. First, we focus on the skilled sector. We investigate two different equilibria that can emerge in the organization of skilled labor, a "generalist" equilibrium and a "specialist" equilibrium. Second, we introduce the unskilled sector and demand to close the economy.

3.3.1 Organizational Equilibria in the Skilled Sector

The value of being a skilled worker of type k at the moment one's education is complete is, from (6) and (8),

$$V^k = \frac{1}{r}\frac{w_2^k + \frac{\lambda}{2r}\sum_{j\in\Omega^k}\Pr(j)w^{kj}}{1+\frac{\lambda}{2r}\sum_{j\in\Omega^k}\Pr(j)} \tag{11}$$

so that the value of being a type k worker depends on (a) the wage you earn if you work alone, w_2^k, (b) the wage you can earn in pairings you are willing to accept, w_2^{kj}, and (c) the rate such pairings occur, $\lambda\Pr(j)$. To solve this model, we consider the wages and pairings that can be supported in equilibrium.

The equilibrium definition requires that no individual be able to deviate and earn higher income. Hence we must have $W^k = W$ for all active worker types in any equilibrium and therefore, by (5),

$$V^k = V \text{ for all } k \in \{G, A, B\}$$

That is, each type of skilled worker must have the same expected income upon finishing school. If one type did better than the others, an individual would switch to become this type.

This common value, V, means that in any equilibrium individuals have the same outside option when wage bargaining. Defining x_2^{kj} as the maximum output individuals of type k and j can produce when working together, it then follows from Nash Bargaining, (9), that in any accepted pairing $V^{kj} = V^{jk}$ and

$$w_2^{kj} = \frac{1}{2}p_2 x_2^{kj} \tag{12}$$

so that in equilibrium a worker team splits its joint output equally. Meanwhile, if skilled workers work alone, then they earn the total product, so that

$$w_2^k = p_2 x_2^k \tag{13}$$

where x_2^k is the maximum output an individual of type k can produce when working alone.

These results lead to a limited set of matching behaviors that can exist in equilibrium.

Lemma 1 *(Matching Rules) In equilibrium, matching behavior is either* $\left\{\Omega^A, \Omega^B, \Omega^G\right\} = \{\{B\}, \{A\}, \{\varnothing\}\}$ *or* $\left\{\Omega^A, \Omega^B, \Omega^G\right\} = \{\{B, G\}, \{A, G\}, \{A, B\}\}$

Proof. See appendix. ■

This result states in part that types never match with themselves. This is intuitive because matching with one own's type provides no productivity advantage but incurs co-ordination costs. The lemma also states that a specialist is always willing to match with the other specialist type in equilibrium. This is intuitive because an AB pairing produces the highest wages. A second, intuitive equilibrium property follows from the symmetry between specialists and their desire not to be unemployed.

Lemma 2 *(Balanced Specialists) In equilibrium,* $L^A = L^B$*.*

Proof. See appendix. ■

This lemma limits the class of possible equilibria. If L^s is the total mass of skilled workers, then we can distinguish three potential equilibria: (1) a "generalist" equilibrium where $\{L^A, L^B, L^G\} = \{0, 0, L^s\}$; (2) a "specialist" equilibrium where $\{L^A, L^B, L^G\} = \{\frac{1}{2}L^s, \frac{1}{2}L^s, 0\}$; and (3) a "mixed" equilibrium where $\{L^A, L^B, L^G\} = \{L', L', L^s - 2L'\}$ for some L' such that $0 < L' < \frac{1}{2}L^s$.

Proposition 1 *(Knowledge Trap) With full employment, where* $\lambda \to \infty$*, a "generalist" equilibrium exists iff* $x_2^{AG} \leq 2x_2^G$ *and a "specialist" equilibrium exists iff* $x_2^{AB} \geq 2x_2^G$*. With full employment, any "mixed" equilibrium limits to the "generalist" equilibrium. For some parameter values, both a generalist and specialist equilibrium can exist. These equilibria are summarized in Figure 1.*

Proof. See appendix. ■

The intuition for these results is straightforward. As $\lambda \to \infty$, workers meet at such a high rate that they match instantaneously in equilibrium and are never unemployed. Hence skilled workers choose matches based simply on wages. In the "generalist" case, skilled workers earn $w_2^G = p_2 x_2^G$. If a player deviates to be a specialist, say type A, then the best he can do is pair with an existing generalist and earn $p_2 x_2^{AG} - w_2^G$.[20] Hence, a world of generalists is an equilibrium iff $p_2 x_2^{AG} - w_2^G \leq w_2^G$, or

$$x_2^{AG} \leq 2x_2^G$$

[20] With full employment, the deviating player captures the joint output net of the other player's outside wage. With finite λ, the possibility of unemployment further affects the wage bargain - see Appendix.

In the "specialist" case, skilled workers produce in teams and earn a wage $w_2^{AB} = \frac{1}{2}p_2 x_2^{AB}$. If a player deviates to be a generalist, then he could either (a) work alone and earn w_2^G or (b) pair with an existing specialist and earn $p_2 x_2^{AG} - w_2^{AB}$. The latter option cannot be worthwhile. In particular, since $x_2^{AG} < x_2^{AB}$, deviating to be a generalist only to pair with a specialist is not better than remaining as a specialist in the first place. We therefore only need consider the first case, where the deviating generalist works alone. Hence, this world of specialists is an equilibrium iff $w_2^G \leq w_2^{AB}$, or

$$x_2^{AB} \geq 2x_2^G$$

These existence conditions can be rewritten in terms of the model's exogenous parameters, using the production functions, where the condition for specialist stability, $x_2^{AB} \geq 2x_2^G$, is simply $mc \geq 1$, and the condition for generalist stability, $x_2^{AG} \leq 2x_2^G$, is $mc \leq \left(\frac{2}{1+m^{\frac{1-\sigma}{\sigma}}}\right)^{\frac{\sigma}{\sigma-1}}$. The equilibria are plotted in Figure 1.

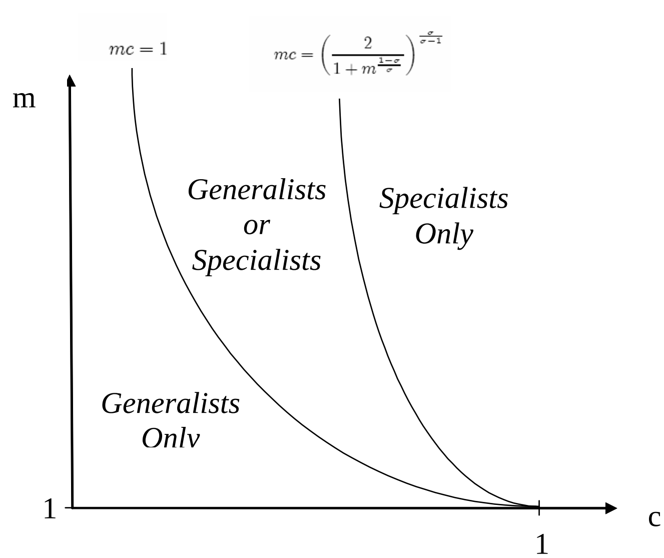

Figure 1: The Knowledge Trap

A country where coordination costs are low (i.e. high c), or the skill gains from narrow training are large (i.e. high m) will tend towards the specialist equilibrium. A country where coordination costs are high or gains from focused training are modest will tend towards the generalist equilibrium. The failure to develop deep specialists could therefore be viewed as institutional problems, where the important policy parameters are m and c, as will be discussed below. There are also, however, regions of the parameter space where different equilibria may emerge even if m and c are the same, providing the possibility of multiple, pareto-ranked equilibria. In general, a country with specialized skilled workers is mc times more productive than an economy with generalist skilled workers. Moreover, the ratio of income between generalist and specialist equilibria is potentially unbounded even where both are stable.

Corollary 1 *(Gains from Specialization) Output in the skilled sector is mc times larger in a "specialist" equilibrium than in a "generalist" equilibrium. Moreover, the range of potential combinations mc where both a generalist and specialist equilibria exist is unbounded from above.*

Proof. See appendix. ■

Note the important roles of (1) coordination costs and (2) task complementarity in supporting a sub-optimal generalist equilibrium. Deviating to become a specialist only to pair with an existing generalist is less appealing when coordination costs are high (i.e. smaller c) or complementarities of tasks are high (i.e. smaller σ). With sufficient coordination costs or complementarity, m (and hence mc) can become unboundedly large, so that the generalist case is stable even though the specialist organization produces unboundedly higher income. For example, with Leontief task aggregation ($\sigma = 0$), mc can be unboundedly large for arbitrarily small coordination costs.

Lastly, note the role of a "thick market" problem for supporting a robust generalist equilibrium despite large mc. The generalist equilibrium is stable to the extent that finding a complementary specialist type is challenging were you to deviate yourself. With finite λ, the generalist equilibrium is stable to trembles where positive masses of specialists appear, because the search friction impedes easy matching. The convenient case of "full employment", where $\lambda \to \infty$, is the limit of trembling hand perfect equilibria.[21]

[21] In the limit, the model still features a "needle in a haystack" friction where, although search is extremely

3.3.2 The Equilibrium Economy

Given the possible organizational equilibria in the skilled sector, we now consider the influence of this organizational equilibrium on the economy at large. Denote with the superscript n the organizational equilibrium in the skilled sector, where $n = G$ defines the "generalist" outcome and $n = AB$ defines the "specialist" outcome. The equilibrium in the skilled sector will influence the endogenous outcomes in both the skilled and unskilled sectors, including labor allocations, prices, and wages.

The first result concerns wages.

Lemma 3 *(Log-linear Wages). In any full employment equilibrium*

$$w_2^n = w_1^n e^{rS} \tag{14}$$

Proof. See appendix. ■

This functional form follows from (a) exponential discounting and (b) the opportunity cost of time. Through endogenous decisions to become skilled or unskilled, an identical Mincerian wage structure emerges regardless of the organizational equilibrium in the skilled sector.

Given this wage relationship, we can now pin down prices. In equilibrium, workers in each sector are paid

$$\begin{aligned} w_1^n &= p_1^n z \\ w_2^n &= p_2^n z A^n \end{aligned}$$

where skilled workers' productivity depends on their organizational equilibrium,

$$A^n = 2^{\frac{1}{\sigma-1}} h \times \begin{cases} 1, & n = G \\ mc, & n = AB \end{cases}$$

Therefore, using the wage ratio, the price ratio on the supply side is determined as a function of exogenous parameters[22]

$$\frac{p_1^n}{p_2^n} = A^n e^{-rS} \tag{15}$$

rapid ($\lambda \to \infty$) there are so many workers (a continuum) that one cannot expect to find a particular worker in finite time.

[22] The price ratio is determined entirely by the supply side because both the skilled and unskilled sectors exhibit constant returns to scale.

Now consider the demand side to close the model. With CES preferences, aggregate demands are such that

$$\frac{X_1^n}{X_2^n} = \left(\frac{\gamma}{1-\gamma}\right)^{\varepsilon} \left(\frac{p_1^n}{p_2^n}\right)^{-\varepsilon}$$

Market clearing implies $p_1^n X_1^n = w_1^n L_1^n$ and $p_2^n X_2^n = w_2^n L_2^n$ so that labor allocations are also pinned down given relative prices

$$\frac{L_1^n}{L_2^n} = \left(\frac{\gamma}{1-\gamma}\right)^{\varepsilon} \left(A^n e^{-rS}\right)^{1-\varepsilon} e^{rS} \tag{16}$$

where L_q^n is the measure of people actively working in sector q.[23]

Real income per-capita, $y^n = Y^n/L$, is also pinned down given relative prices[24]

$$y^n = z\left(\gamma^{\varepsilon} + (1-\gamma)^{\varepsilon} \left(A^n e^{-rS}\right)^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}} \tag{17}$$

and we can define human capital's contribution to output as $H^n = y^n L/z$.[25]

4 Applications and Discussion

This section examines several applications of the model. One application uses general equilibrium reasoning to show why human capital can play a much larger role in the world economy than traditional accounting estimates suggest. A series of further applications show that "knowledge traps" may provide a parsimonious interpretation of several stylized facts in the world economy while also suggesting novel mechanisms that can obstruct economic development.

[23] There are also a number of students who are training in sector 2 and not yet active workers. Given the hazard rate of death r, we have $e^{rS}L_2^n$ people currently training and working in sector 2, so that total labor supply is $L = L_1^n + e^{rS}L_2^n$.

[24] Real national income (Y^n) is given by $p^n Y^n = w_1^n L_1 + w_2^n L_2$, where the aggregate price level is $p^n = \left(\gamma^{\varepsilon}\left(p_1^n\right)^{1-\varepsilon} + (1-\gamma)^{\varepsilon}\left(p_2^n\right)^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}$. Real per-capita income ($y^n = Y^n/L$) is

$$y^n = \frac{w_1^n}{p^n}\left(\frac{L_1^n}{L} + \frac{w_2^n}{w_1^n}\frac{L_2^n}{L}\right) = \frac{w_1^n}{p^n}$$

Thus average per-capita income is equivalent to the real wage in the low-skilled sector. This follows in equilibrium because workers' net present value of lifetime wage income is equivalent at birth. We can alternatively write this in terms of sector 2 wages, since $w_1^n = e^{-rS}w_2^n$.

[25] Note that the model, which considers two final goods in consumption, is equivalent to a model that considers a single final good and treats x_1 and x_2 as two intermediates. In that interpretation, where (4) is now a production function instead of a preference aggregator, we would write the production function as $y^n = z\left(\gamma\left(L_1^n\right)^{\rho} + (1-\gamma)\left(A^n L_2^n\right)^{\rho}\right)^{1/\rho}/L$, which can be shown to be equivalent to (17), where total factor productivity is interpreted as z, and the contribution of human capital is $H^n = \left(\gamma\left(L_1^n\right)^{\rho} + (1-\gamma)\left(A^n L_2^n\right)^{\rho}\right)^{1/\rho}$.

4.1 Wages, Prices, and Labor Allocations

When people choose to be highly educated, any excessive wage gains to the highly-educated can be arbitraged away by an increase in the supply of such workers. In the model, this choice problem generates the log-linear "Mincerian" wage structure and pins the skilled wage premium to the interest rate, as in (14).[26]

One key implication is that two countries can have vastly different mappings between schooling duration and skill and yet have identical wage returns to schooling in equilibrium. In fact, skill differences are hidden by the wage structure. It is prices and labor supply that shift to ensure the equilibrium wage-schooling relationship.

Corollary 2 *(Balassa-Samuelson) Prices adjust in the model such that* $\frac{p_1^{AB}/p_2^{AB}}{p_1^G/p_2^G} = mc$, *and labor supply adjusts such that* $\frac{L_1^{AB}/L_2^{AB}}{L_1^G/L_2^G} = (mc)^{1-\varepsilon}$.

Proof. See appendix. ■

This result says that low-skilled services will be cheaper in the poor country. This feature of the equilibrium may be appealing, as it provides a Balassa-Samuelson effect in relative prices (e.g. Harrod 1933, Balassa 1964, Samuelson 1964). The model may thus inform the standard observation that certain goods are relatively cheap in poor countries, an effect that motivates the need for PPP price corrections when comparing real income across countries.[27] The knowledge trap model provides a basis for this phenomenon, where low-skilled goods (e.g. haircuts) are relatively cheap in a poor country because low skill is (endogenously) relatively abundant there.[28]

Together, the equilibrium price and labor supply adjustments decouple the wage returns to schooling from the skill-gains from schooling. Among other applications, these results

[26] Note that this simple perspective suggests a positive correlation between interest rates and returns to schooling across countries. In fact, the literature has suggested both (a) higher interest rates in poor countries (e.g. Banerjee and Duflo 2005) and (b) higher rates of return to schooling in poor countries (Psacharapolous 1994).

[27] Classic explanations for this price phenomenon imagine exogenous cross-country differences in technology (Balassa 1964, Samuelson 1964) or factor endowments (Bhagwati 1984).

[28] Note that the model considers price differences between a final good completely produced through skilled labor and a final good completely produced through unskilled labor. In looking at microeconomic price data, one would consider input mixes of skilled and unskilled labor away from these extremes, which would attenuate the observed price differences in final goods. In a generalization of the model, the observable price effect would appear such that goods that use unskilled labor relatively intensively would be relatively expensive in rich countries.

challenge the traditional macro-Mincer calibration method, as discussed in the following sub-section.

4.2 Human Capital Stocks

Many analyses have concluded that human capital plays a relatively modest role in explaining the wealth and poverty of nations, leaving residual variation in total factor productivity as a major explanation (see Caselli 2005 for a review). This conclusion is reached using the "macro-Mincer" method to account for human capital (Klenow and Rodriguez-Clare 1997, Hall and Jones 1999). In this method, each economy's human capital is calculated as the labor supply at each level of education, weighted by the average wage at that education level; i.e., in this paper's notation, $H^n_{Mincer} = L^n_1 + e^{rS} L^n_2$. The returns to education are taken as e^{rS}, and countries differ in their human capital to the extent that they have more or less educated workers.[29] To see how this method can mis-account for human capital, first consider an example.

Example 1 *With Cobb-Douglas aggregation ($\varepsilon = 1$), it follows from Corollary 2 that $L^{AB}_1/L^{AB}_2 = L^G_1/L^G_2$, so that the labor allocation does not vary with the skill gains from education, mc. Therefore, the macro-Mincer human capital stock calculation, $H^n_{Mincer} = L^n_1 + e^{rS} L^n_2$, would not vary with the skill gains from education. Mincerian accounting would therefore suggest no role for human capital, even should human capital explain unboundedly large income differences across countries.*[30]

[29] Some calibrations also allow e^{rS} to vary across countries, based on observed educational returns. To focus on the core methodological issue, the following theoretical results will abstract from variation in r. The calibration evidence cited below incorporates such variation as well.

[30] Moreover, a regression of per-capita income on average schooling duration would also show no relationship. With Cobb-Douglas preferences ($\varepsilon = 1$) the average schooling in a population is

$$s^n = S\frac{L^n_2}{L} = (1-\gamma)Se^{-rS}$$

a constant independent of which equilibrium is attained. For average schooling to be positively associated with income (which it is), we require the elasticity of substitution between skilled and unskilled labor to be greater than 1, as supported by the literature discussed further below. Then countries with high quality skilled-labor (i.e. specialization) will see an endogenous increase in the supply of such skilled workers.

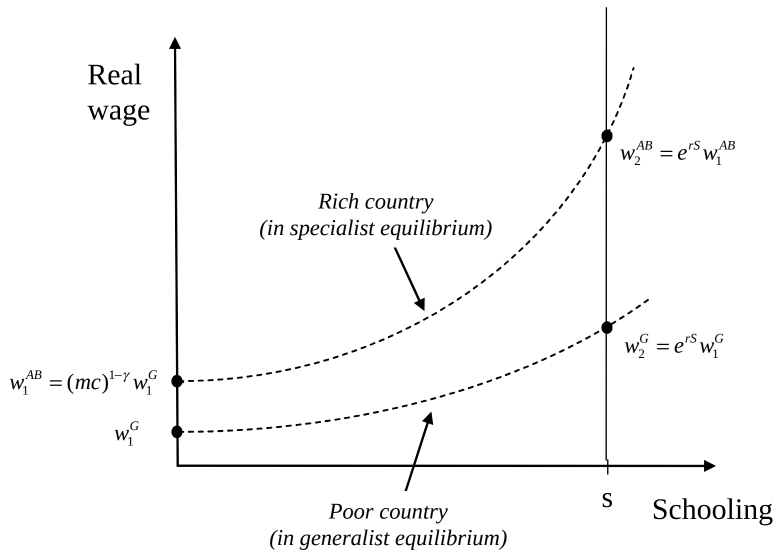

Figure 2: Equilibrium Wage-Schooling Relationships

The general intuition can be stated as follows. With downward sloping demand for different labor classes, countries that are very good at producing high skill will find that goods and services produced by low-skill workers are scarce, which drives up low-skilled wages. In particular, with relative wages pinned down by the discount rate, as in (14), workers allocate themselves so that the percentage wage gains for skilled and unskilled workers rise or fall in equal proportion. Wages are Mincerian in each country, but this within-country equilibrium does not inform human capital differences across countries. Rather, the wage-schooling relationship shifts vertically depending on the skilled equilibrium. This is shown in Figure 2 for the Cobb-Douglas case, in which price adjustments fully offset productivity differences, requiring no labor adjustment.

Because Mincerian accounting rules out the scarcity effect on unskilled wages, it will in fact systematically understate human capital differences across countries given the observed allocations of labor. Define the ratio of actual human capital differences across countries to the Mincerian calculation of these differences as

$$R_H = \frac{H^{AB}/H^G}{H^{AB}_{Mincer}/H^G_{Mincer}}$$

Lemma 4 *(Mincer as Lower Bound)* $R_H \geq 1$ *for all* $\varepsilon \in [0, \infty]$*. Moreover,* $\lim_{\varepsilon \to 1} R_H = \infty$

for a given labor allocation $L_1^G/L_1^{AB} \neq 1$.

Proof. See appendix. ■

This lemma states that Mincerian human capital accounting is only a lower-bound on the actual human capital differences across countries. The lemma further says that the magnitude of the underestimate may be arbitrarily large, depending on the elasticity of substitution between skilled and unskilled labor. The reasoning follows from Corollary 2. For example, fixing the observed labor allocation, $\left(L_1^{AB}/L_2^{AB}\right) / \left(L_1^G/L_2^G\right) < 1$, reducing ε towards 1 calls for greater mc, which makes for a larger human capital difference between these countries.[31] Put another way, once educational attainment is seen as a choice problem, it is natural to ask why so many more workers seek higher education in rich countries. The larger supply of such workers is reconciled in equilibrium by larger skill gains from schooling. As ε falls, the human capital differences must increase to compensate if we are to explain the observed supply of skilled workers.

It is clear that the elasticity of substitution between skilled and unskilled labor becomes a key parameter in assessing the role of human capital. The literature suggests values of $\varepsilon \in [1, 2]$.[32] Calibrations using such parameterizations are extensively explored in a separate paper (Jones, forthcoming), which shows that residual TFP differences are no longer necessary to explain cross-country income differences when $\varepsilon = 1.5$. Meanwhile, Caselli and Coleman (2006) use realistic values of ε and calibrate separate productivity terms for skilled and unskilled workers across countries.[33] They find an enormous productivity advantages of skilled workers in rich countries while the productivity of unskilled workers is no higher there.[34] This calibration is consistent with the knowledge trap model: very large productivity advantages in rich countries that are limited to skilled workers.[35]

[31] In practice, we see $\left(L_1^{AB}/L_2^{AB}\right) / \left(L_1^G/L_2^G\right) < 1$, which is consistent with $\varepsilon > 1$.

[32] See, e.g., the review by Katz and Autor (1999). The relatively well identified estimates of Ciccone and Peri (2005) suggest $\varepsilon \approx 1.5$.

[33] These authors use the production function $y = k^\alpha \left[(A_u L_u)^\rho + (A_s L_s)^\rho\right]^{\frac{1-\alpha}{\rho}}$, which is the analogue of (4) in this paper with the addition of physical capital, k.

[34] The calibrations imply that skilled productivity differences between the richest and poorest countries are on the order of 100, depending on choices of ε. One reason to emphasize the division of labor - where skilled laborers are vessels of advanced ideas - is because this explanation seems capable of producing such large productivity differences, whereas simpler conceptions of quality in the educational production function may face greater difficulty.

[35] Among other recent calibrations, notable contributions include Erosa et al. (2010) and Manuelli and Seshadri (forthcoming), which infer human capital stocks by calibrating sophisticated models of endogenous

4.3 Immigrant Wages and Occupations

An alternative approach to assessing human capital's role in cross-country income differences considers immigration. If human capital differences were critical, it has been argued that immigrants should experience significant wage penalties in the rich country's economy, since immigrants bring their human capital with them. Noting that immigrants from poor to rich countries earn wages broadly similar to workers in the rich country, authors have thus concluded that human capital plays at most a modest role in explaining productivity differences across countries (Hendricks 2002). However, this estimation approach, as implemented, sidesteps important considerations that follow with labor differentiation.[36]

The intuition can be developed as follows. Take a skilled worker ("i") from a poor country and place them in the rich country.

Corollary 3 *(Immigrant Wage Bounds) While the skill ratio, h_2^i/h_2^{AB}, at the skilled task may be arbitrarily close to zero, the wage ratio is bounded away from zero such that $w^i/w_2^{AB} \geq e^{-rS}$.*

Proof. See appendix. ■

The basic intuition, once again, is that relative wages no longer reflect relative skills. Here, however the intuition is also slightly subtler. In migrating to a new economy, the migrant can also reallocate their labor effort across tasks. With the new prices in the rich country, the individual's occupational choice may naturally change. To the extent that a skilled migrant can always work in a lower-skilled task (e.g. as a taxi driver, security guard, etc.), the wage a worker experiences is bounded from below by the local wage schedule. Since Mincerian wage returns within countries are modest, this immigration-based wage accounting will provide only limited scope for variation (i.e. based on e^{rS}) that is not

human capital acquisition over the life-cycle and/or across heterogenous individuals. These papers infer large quality differences in human capital across countries, thereby reducing the role of TFP differences. They do not consider a skilled versus unskilled distinction in the production function, which is central to this paper's analysis and connects directly to the calibrations of Caselli and Coleman (2006) and Jones (forthcoming).

[36] The main estimates in Hendricks (2002) assume workers output at different skill classes are perfect substitutes, thus eliminating any effect of scarcity on the wages of the unskilled. To the extent calibrations with less then perfect substitutes are considered, the elasticity of substitution between skilled and unskilled labor is assumed to be at least 5, well above the consensus estimates in the literature that range between 1 and 2.

informative of actual skill variation at skilled tasks.

The model provides more specific results as follows.

Corollary 4 *(Immigrant Wages and Occupations) An unskilled worker who migrates from a poor to a rich country will earn a higher real wage. The skilled generalist who migrates from a poor to a rich country will work in the unskilled sector and earn the unskilled wage, which may provide more or less real income than staying at home.*

Proof. See appendix. ■

The knowledge trap model predicts that low-skilled immigrants, who are the majority of immigrants, will enjoy (a) much higher real wages than they left behind and (b) face no wage penalty in the rich economy vis-a-vis other unskilled workers. Their wage gains follow naturally when the low-skilled immigrant moves to a place where his labor type is relatively scarce.

More subtly, the corollary says that skilled immigrants from poor countries are unable to find local specialists willing to team with them. Moreover, these immigrants won't work alone; the specialized equilibrium of the rich country raises the low-skilled wage enough to make unskilled work a more enticing alternative to the unspecialized immigrant than using his education. Hence, for example, we can see highly educated immigrants who drive taxis.

While the literature does not appear to have examined migration outcomes through this lens, descriptive facts can be assembled using U.S. census data. The appendix consider these data and shows two important facts (Figure A1). First, the location of higher education appears to matter. Highly educated U.S. immigrants (who likely received their education outside the U.S.) experience wage penalties of 50% or more compared to similarly-educated U.S.-born workers. Moreover, such skilled immigrants tend to shift down the occupational ladder into jobs that are typically filled by those with substantially less education. Second, by contrast, the location of secondary education does not appear to matter. Wages or occupational types do not differ by birthplace or immigration age for workers with an approximately high-school level education.

These stylized facts are consistent with the above corollary and, more broadly, with the idea that human capital differences across countries exist primarily among the highly educated, as implied by the model and suggested by independent literature discussed above. Together, the two primary methods used to argue against substantial roles for human capital – the macro-Mincer approach and the immigration approach - do not reject this paper's human capital theory but rather appear to provide evidence in concert with it.

4.4 Mechanisms of Poverty

The theory in this paper emphasizes that the stock of productive knowledge in an advanced economy is too great for one person to know. If workers organize themselves to learn specialized pieces of this knowledge, then they can collectively access the knowledge frontier. But workers may fail to organize into deep, narrow skills, avoiding reliance on teamwork and failing to embody advanced knowledge. This section further considers challenges to collective skill improvement from the perspective of the model.[37]

4.4.1 The Quality of Higher Education

Income differences across countries may persist if countries are in different regions of Figure 1. Countries with $mc < 1$ have shallow knowledge and remain in poverty. This may occur if acquiring deep skills is hard in poor countries (m is small). One can think of m as a policy parameter, where, for example, m increases through public investment in higher education. Small m also follows naturally if knowledge acquisition is limited by local access to others with deep skill - i.e. expert teachers. For example, becoming skilled at protein synthesis will be difficult without access to existing skilled protein synthesists: their lectures, advice, the ability to train in their laboratories, etc. In this setting, we can imagine a simple, further type of knowledge trap. If we write m^n, where $m^G < m^{AB}$, then countries that start in the generalist equilibrium will remain there if $m^G c < 1$.

Escaping such a trap involves importing skill from abroad to train local students or sending students abroad and hoping they will return. Both approaches face an incentive problem however, since those with deep skills will earn higher real wages by remaining in

[37] The following discussion focuses on poverty-inducing mechanisms that follow from the supply of human capital. Coordination costs (the parameter c) may also be important, especially should coordination costs be more severe in poorer countries. Lastly, market size may be an important limiting factor in specialization. This last possiblity may also bear consideration in confronting data but is not incorporated in the model for focus and brevity.

the rich country. The model thus suggests a "brain drain" phenomenon.

Corollary 5 *(Brain Drain) Once trained as a specialist in the rich country, one will prefer to stay.*

Proof. See appendix. ■

Specialists in rich countries prefer to stay because they can work with complementary specialists there and thus earn higher wages. Hence students who migrate to the U.S. for their Ph.D.'s face real wage declines if they go home - even though they are scarce at home. Related, it is clear that students from rich countries do not migrate to developing countries for their education, even though university and living expenses are considerably lower. This may further suggest that the quality of education is low.[38]

This result suggests that wage subsidies or other incentives may be required to attract skilled experts to the poor country and improve local training. China, for example, has been actively engaging in such policies (see, e.g., Zweig 2006).

4.4.2 The Coordination of Higher Education

Should poor countries produce high-quality higher education, there is still an organizational challenge. Countries may be in the middle region of Figure 1, facing the same parameters m and c but sitting in different equilibria. Here a country cannot escape poverty without creating thick measures of specialists with complementary skills.[39] This may be hard. Any intervention must convince initial cohorts of students to spend years in irreversible investments as specialists, which would be irrational if complementary specialists were not expected. Hence we need a "local push".[40,41] Yet it is not obvious what institutions have

[38] I thank Kevin Murphy for pointing this out.

[39] One could alternatively construe the "trap" as being a deterministic function of the initial conditions, where a sufficent mass of specialists of each type creates a stable, high income state, while insufficient supply of specialists creates a stable, low income state.

[40] Note that the type of trap allowed in the model differs from poverty traps that envision aggregate demand externalities (e.g. Murphy et al. 1989). Rather, knowledge traps can be overcome locally, when workers achieve greater collective skill. A challenge for aggregate demand models is that many poor economies are quite open to trade or have large GDP on their own despite low per-capita GDP, so it is unclear that aggregate demand is a credible obstacle. Meanwhile, booms are often local, whether it is city-states like Hong Kong or Singapore, or cities within countries, like Bangalore, Hyderabad, and Shenzhen, which have led growth in India and China. Yet such booms are also rare, which suggests that local coordination problems are themselves not trivial to overcome.

[41] Some authors see such coordination failures as easily solved due to trembling hand type arguments (e.g. Acemoglu 1997). However, there are several reasons to think that small "trembles" are unlikely

the incentives or knowledge to coordinate such a push. A firm may have little incentive to make these investments when students can decamp to other firms.[42] Public institutions may not produce the right incentives either. Developing deep expertise requires time, so that the fruits of educational investments may not be felt for many years, depressing the interest of public leaders (or firms), who may have short time horizons. Even if local leaders wish to intervene, it may be challenging to envision the set of skills to develop, especially if there are many required skills and deep knowledge does not exist locally. These difficulties suggest a need for "visionary" public leaders. They also suggest an intriguing role for multinationals in triggering escapes from poverty.

Multinationals and Poverty Traps Intra-firm trade can allow for production teams that span national borders, suggesting that a multinational may play a unique role in helping countries develop deep skills.

Corollary 6 *(Desirable Cheap Specialists) A firm of specialists in a rich country would hire specialists in poor countries, if they could be found.*

Proof. See appendix. ■

This result follows because the skilled wage in the poor country is held down by the Mincerian wage equilibrium, making a specialist there attractive. Hence, production would shift to incorporate a skilled specialist in the poor country if such a type existed. But now we have a cross-border coordination problem. A multinational will only be able to find these specialists if they exist in sufficient measure, and no one in the poor country will want to become such a specialist unless the multinational will be able to find them.

---

to undo a generalist equilibrium. First, we are considering many years of education for an individual, so that a "tremble" must be rather large. Second, while we consider two tasks for simplicity, there may be $N > 2$ tasks needed for positive output, which would then require simultaneous trembles over many specialties. Third, with greater search frictions in the market (smaller λ), trembles must occur over a large mass of workers. Fourth, in tradeable sectors, one must leap to the skill equilibrium of the rich countries to compete internationally - small skill trembles won't suffice.

[42] Contracts may help here, but labor contracts that prevent workers from departing a firm (i.e., in an extreme form, slavery) are typically illegal. Labor market frictions may allow firms to do some training if frictions give the firm some monopsony power (e.g. Acemoglu and Pischke 1998). Still, it is clear that foundational training and specialization,such as the Ph.D., occur in educational institutions, prior to engagement with firms.

The interesting aspect is that a multinational allows the local educational institutions to avoid producing all required specialities locally. The multinational provides the complementary worker types from abroad. For example, in working to initiate the economic boom in Hyderabad, governor Naidu both subsidized a vast expansion in engineering education and personally convinced Bill Gates to employ these workers in Microsoft's global production chain, so that computer programmers in Hyderabad now team with other skilled specialists in advanced economies.[43] Here, the "visionary" leader need not recreate the multinational but simply produce a sufficient quantity of one specialist type multinationals will hire.[44]

4.5 Generalizations

The emphasis on the division of labor among skilled workers also suggests natural generalizations to trade, labor, and growth contexts. In particular, by emphasizing a link between skilled workers, the division of labor, and the acquisition of advanced ideas, the model provides a general foundation for thinking about skill bias, with several applications.

First, knowledge traps may provide a useful perspective on comparative advantage. The factor endowment model of trade, Heckscher-Ohlin, explains why Saudi Arabia exports oil but is famously poor at predicting trade flows based on capital and labor endowments – the "Leontief Paradox" (Leontief 1953, Maskus 1985, Bowen et al. 1987, etc.). International trade analysis, much like cross-country income analysis, has therefore relied on substantial residual productivity terms to explain the empirical patterns (e.g. Trefler 1993, 1995, Harrigan 1997). With knowledge traps, the rich country has a comparative advantage in the skilled good while the poor country has a comparative advantage in the low-skilled good.[45] Yet these comparative advantages - based in the division of labor - won't appear in

[43] See, e.g., Bradsher, Keith. "A High-Tech Fix for One Corner of India", *The New York Times*, December 27, 2002, p. B1.

[44] With only two types of specialists, the emergence of one type in the poor country can trigger the emergence of the other, and the poor country will become rich. With more than two specialist types, or with an inability to train locally in the other skill, the emergence of one type may not inspire the local creation of the other types. Here, a multinational can continue to employ a narrow type of skilled specialists in one country without triggering a general escape from poverty. Here we will see both off-shoring and persistently "cheap engineers".

[45] In terms of the model, we can consider two small open economies who can trade both goods 1 and 2. With world prices, p_1/p_2, such that

$$\frac{p_1^G}{p_2^G} < \frac{p_1}{p_2} < \frac{p_1^{AB}}{p_2^{AB}}$$

the country in the generalist equilibrium exports the low-skilled good (1) while the country in the specialist standard calculations of labor endowments.[46] The division of labor model allows a human-capital interpretation of residual productivity terms, where rich countries are net exporters of skilled goods not simply because they have more skilled workers (quantity), but because their skilled workers have much more collective skill (quality).

The emphasis on skilled workers as the vessels of advanced ideas also suggests a natural generalization to skill-biased technical change. Along the growth path in advanced economies, the empirical tendency for skilled wage premiums to hold steady, or even rise, despite large increases over time in skilled labor supply is consistent with the rising quality of skilled labor compared to unskilled labor (see, e.g. Katz and Autor 1999). This tendency would occur naturally in a generalization of the model. In particular, if growth is associated with the creation of new ideas and consequent expansion of frontier knowledge, which can be modeled as an increase in m, then growth is intrinsically skill-biased.

Similar reasoning would also predict cross-country income divergence over time. To the extent that workers in poor countries, organized for general knowledge, do not access this deepening set of ideas, cross-country divergence in per-capita income becomes the natural outcome empirically, as is the usual case (Jones 1997, Pritchett 1997). As with skill-biased technical change, this dynamic result would follow from the same extension of the model, where advanced ideas drive growth and these advanced ideas are accessed in the workforce through increasingly differentiated skills, as shown empirically in some contexts.[47]

5 Conclusion

This paper offers a human-capital based interpretation of several phenomena in the world economy and therefore a possible guide to core obstacles in development. The model

equilibrium exports the high skilled good (2).

[46] For example, the degree of specialization won't appear in designations like "professional" or "highly educated" worker, which can explain why attempts to save Heckscher-Ohlin through finer-grained classifications of labor endowments have failed (e.g. Bowen et al. 1987).

[47] In an endogenous growth framework, some fraction of skilled workers would produce productivity enhancing ideas that lead to growth in m. Such accumulation of knowledge may require innovators to become more specialized along the growth path, so that the number of tasks at the frontier (2 in this model) becomes endogenous and increases with time. See Jones (2009) for such a growth model and for empirical evidence that knowledge workers in the U.S. become increasing specialized with time and work in larger teams. See also Agrawal et al. (2013) for evidence that specializatison and collaboration both increase when there is a positive shock to extant knowledge.

shows how endogenous differences in the division of labor may persist across economies, emphasizing the importance of skilled workers as vessels of ideas and high productivity. As one application, the theory shows how standard human capital accounting may severely underestimate cross-country skill differences. More broadly, the model may provide an integrated perspective on cross-country income differences, labor allocations, wages, price differences, migrant behavior, poverty traps, and other phenomena in a way that appears broadly consistent with important facts.

Integrating across many such facts, one sees a consistency in favor of substantial "skill bias" associated with rich economies - whether in calibration exercises such as Caselli and Coleman (2006), the immigration evidence discussed above, the evidence for skill-biased technical change over time in advanced economies, or the comparative advantage of poor countries in low-skill, labor-intensive tasks. In this sense, theories for skill bias may be especially useful avenues for research on economic development. The "knowledge trap" is one such theory. Its basic premise is that advanced productive knowledge is too great for one person to know, so that the implementation of advanced ideas into production relies on a division of labor. Skill-bias follows naturally on these heels. By suggesting specific mechanisms, including institutional mechanisms, that disrupt the collective acquisition and employment of advanced knowledge, the theory further suggests tangible, micro-empirical avenues for future work. Institutional parameters in the theory further suggest how one can connect capital accumulation and institutional reasoning in explaining development patterns.

This paper also speaks directly to a long-running debate over the roles of "human capital" and "technology" in explaining income differences across countries. I close by further considering this distinction. Much existing literature imagines human capital and ideas as distinct inputs into a production function and, using macro-Mincer accounting, suggests a modest role for human capital, pushing education toward the periphery in understanding key issues in development. What is called technology, the residual, has consequently occupied a central position and is often imagined as a set of techniques, methods, facts, models, et cetera that impact production. At root, this paper attempts to reconfigure this debate and, in some sense, sidestep it. This paper shows how human capital may play a central role while also embracing the critical importance of ideas. People are born with empty minds, and human capital is seen as the process of acquiring knowledge. Rather than conceiving technology as a distinct, disembodied input to production, this paper imagines that ideas are embodied in people. It is thus the emphasis on embodiment, rather than the role of "ideas", that distinguishes this paper from other approaches. Overall, the theory provides a framework where human capital, ideas, and institutions are all essential features.

6 Appendix

Proof of Lemma (Matching Rules)

Proof. The lemma follows from five intermediate results.

(1) Workers are never willing to match with their own type ($k \notin \Omega^k \forall k$)

In equilibrium, all skilled types have some $V > 0$. A type k never matches with type k if $V^{kk} < V$. For As or Bs, the joint output when teaming with one's own type is zero. Hence (8) implies $V^{AA} = V^{BB} = \frac{1}{2}V < V$. Therefore, neither As or Bs will match with their own type. For Gs, (8) implies $V^{GG} = \frac{1}{2}w_2^{GG}/r + \frac{1}{2}V$. Noting that $V \geq w_2^G/r$ (G's income if he never matches, from (6)) and that $w_2^{GG} < w_2^G$ (GG matches provide no skill advantage but incur a coordination penalty), it follows that $V^{GG} < V$. Hence no type will match with her own type.

(2) Type k is willing to match with type j iff type j is willing to match with type k ($k \in \Omega^j \iff j \in \Omega^k$)

A type k is willing to match with type j if $V^{kj} \geq V$. With the Nash Bargaining Solution and common V in equilibrium, it follows that $V^{kj} = V^{jk}$. Hence $k \in \Omega^j \iff j \in \Omega^k$.

(3) As are willing to match with Gs iff Bs are willing to match with Gs ($G \in \Omega^A \iff G \in \Omega^B$)

As are willing to match with Gs if $V^{AG} \geq V$. In equilibrium, $V^{AG} = V^{BG}$. This follows from (8) because with (a) common V and (b) $x_2^{AG} = x_2^{BG}$, Nash Bargaining implies $w_2^{AG} = w_2^{BG}$. Hence, $V^{AG} \geq V \iff V^{BG} \geq V$, so that As are willing to match with Gs iff Bs are willing to match with Gs.

(4) If an A or B is willing to match with Gs, then the A or B is also willing to match with the complementary specialist type ($G \in \Omega^A \Rightarrow B \in \Omega^A$ and $G \in \Omega^B \Rightarrow A \in \Omega^B$)

If As are willing to match with Gs, then $V^{AG} \geq V$ and $w_2^{AG} = \frac{1}{2}p_2 x_2^{AG}$. But $w_2^{AB} = \frac{1}{2}p_2 x_2^{AB} \geq \frac{1}{2}p_2 x_2^{AG} = w_2^{AG}$ and hence, from (8), $V^{AB} \geq V^{AG}$. Hence A will also be willing to match with Bs: $G \in \Omega^A \Rightarrow B \in \Omega^A$. A symmetric argument demonstrates that $G \in \Omega^B \Rightarrow A \in \Omega^B$.

(v) As and Bs must match ($\Omega^k \neq \{\varnothing\}$ for $k = A, B$)

This result follows because tasks A and B are gross complements in production. Hence, As or Bs who work in isolation do not produce positive output and earn no income.[48]

With these five properties, the only remaining, possible equilibrium matching policies are $\{\Omega^A, \Omega^B, \Omega^G\} = \{\{B\}, \{A\}, \{\varnothing\}\}$ or $\{\Omega^A, \Omega^B, \Omega^G\} = \{\{B, G\}, \{A, G\}, \{A, B\}\}$. ■

[48] Gross complements, $\sigma \leq 1$, is a (strong) sufficient condition for this result but is not necessary. If $\sigma > 1$, then positive production becomes possible when a specialist works alone. Nevertheless, it can be shown that, with $\sigma > 1$, As and Bs still prefer to match in equilibrium so long as $c > (1/2)^{\frac{1}{\sigma-1}}$; i.e. matching occurs as long as coordination costs are not too severe (c is not too small) or the elasticity of substitution between tasks is not too great (σ is not too large). The paper focuses on the case of $\sigma \leq 1$ to enhance tractability, brevity and intuition.

Proof of Lemma (Balanced Specialists)

Proof. (I) First consider the case where $L^A > 0$ and $L^B > 0$.

1. In equilibrium $V^A = V^B$. Let $\{\Omega^A, \Omega^B, \Omega^G\} = \{\{B, G\}, \{A, G\}, \{A, B\}\}$. Equating $V^A = V^B$ using (11) implies $0 = [\Pr(A) - \Pr(B)] \left[w_2^{AB} + \frac{\lambda}{2r} \Pr(G) \left(w_2^{AB} - w_2^{AG}\right)\right]$. Hence $\Pr(A) = \Pr(B)$ in equilibrium. If, alternatively, $\{\Omega^A, \Omega^B, \Omega^G\} = \{\{B\}, \{A\}, \{\varnothing\}\}$, it follows directly from $V^A = V^B$ using (11) that $\Pr(A) = \Pr(B)$.

2. Next we show that $\Pr(A) = \Pr(B)$ implies $L^A = L^B$. The probability of meeting a worker of type j is $\Pr(k) = L_p^k / L_p$. To analyze L_p^k, the mass of type k workers who are unmatched, note that workers enter and leave the matching pool by four routes. Workers enter the matching pool either because (a) they finish their studies or (b) their partner dies. Workers exit the pool either by (c) dying themselves or (d) pairing with other workers. These flows are defined as follows.

(a) There are L^k people in the population of type k. In steady state, they are born at rate rL^k and survive to their graduation with probability e^{-rs}. The rate at which new graduates enter the matching pool is therefore $rL^k e^{-rs}$.

(b) There are $L^k e^{-rs} - L_p^k$ type k workers currently matched in teams. Since workers die at rate r, the rate of reentry into the matching pool is $r\left(L^k e^{-rs} - L_p^k\right)$.

(c) Type k workers in the matching pool die at rate rL_p^k.

(d) Type k workers in the matching pool match other unpaired workers at rate $\lambda L_p^k \sum_{j \in \Omega^k} \Pr(j)$.

Summing up these routes in and out of the matching pool, we have

$$\dot{L}_p^k = 2rL^k e^{-rs} - 2rL_p^k - \lambda L_p^k \sum_{j \in \Omega^k} \Pr(j) \tag{18}$$

In steady-state, $\dot{L}_p^k = 0$, which implies that $L_p^k = \left[1 + \frac{\lambda}{2r} \sum_{j \in \Omega^k} \Pr(j)\right]^{-1} e^{-rs} L^k$. The ratio of probabilities for an A and B meeting is therefore

$$\frac{\Pr(A)}{\Pr(B)} = \frac{1 + \frac{\lambda}{2r} \sum_{i \in \Omega^B} \Pr(i)}{1 + \frac{\lambda}{2r} \sum_{l \in \Omega^A} \Pr(l)} \frac{L^A}{L^B} \tag{19}$$

It then follows directly, given the allowable matching rules defined by Lemma 1, that $\Pr(A) = \Pr(B)$ implies $L^A = L^B$.

(II) Second, consider the case where $L^A > 0$ and $L^B = 0$.

We rule this case out by contradiction. Since As earn zero if they work alone, As must match in equilibrium. Hence an equilibrium with $L^A > 0$ and $L^B = 0$ would require $L^G > 0$ with As and Gs matching. In equilibrium, common V then implies from (11) that

$$rV = \frac{\frac{\lambda}{2r} \Pr(G) \frac{1}{2} p_2 x_2^{AG}}{1 + \frac{\lambda}{2r} \Pr(G)} \tag{20}$$

Now consider a player who deviates to type B. This player could choose to match only with Gs and earn the same V.[49] Hence, when meeting an A, the B deviator would have no worse outside option than V. Hence, if B chose to match with an A, $w_2^{BA} \geq \frac{1}{2}p_2x_2^{AB}$. Hence if the B deviator chose to match with As or Gs then

$$rV^B \geq \frac{\frac{\lambda}{2r}\Pr(A)\frac{1}{2}p_2x_2^{AB} + \frac{\lambda}{2r}\Pr(G)\frac{1}{2}p_2x_2^{AG}}{1 + \frac{\lambda}{2r}\Pr(A) + \frac{\lambda}{2r}\Pr(G)} > rV$$

where the strict inequality follows because $x_2^{AB} > x_2^{AG}$. Therefore, by contradiction, there is no equilibrium with $L^A > 0$, $L^B = 0$. By a symmetric argument there is no equilibrium where $L^A = 0$, $L^B > 0$.

Hence in equilibrium the model must feature $L^A = L^B$. ■

Proof of Proposition (Knowledge Traps)

Proof. Consider the "generalist", "specialist", and "mixed" cases in turn.

(I) The "generalist" case, where $\left\{L^A, L^B, L^G\right\} = \{0, 0, L^s\}$.

In this case,

$$rV = w_2^G$$

where $w_2^G = p_2x_2^G$.

Now consider whether an (infinitesimal) individual would deviate to a specialist type, say type A. The type A worker earns $w_2^A = 0$ when working alone. Hence from (11) $rV^A = \left[\frac{\lambda}{2r}/\left(1 + \frac{\lambda}{2r}\right)\right] w_2^{AG}$, where $w_2^{AG} = \frac{1}{2}p_2x_2^{AG} - \frac{1}{2}r(V - V^A)$ from the Nash Bargaining Solution. Solving these to eliminate w_2^{AG} gives

$$rV^A = \frac{\frac{\lambda}{2r}}{2 + \frac{\lambda}{2r}}\left(p_2x_2^{AG} - w_2^G\right)$$

Workers won't deviate if $rV \geq rV^A$, or (after some algebra)

$$x_2^{AG} \leq 2x_2^G\left(1 + \frac{2r}{\lambda}\right)$$

If this condition holds, the "generalist" case is an equilibrium. With full employment, $\lambda \to \infty$, the "generalist" case is an equilibrium iff $x_2^{AG} \leq 2x_2^G$.

(II) The "specialist" case, where $\left\{L^A, L^B, L^G\right\} = \left\{\frac{1}{2}L^s, \frac{1}{2}L^s, 0\right\}$.

In this case,

$$rV = \frac{\frac{\lambda}{2r}}{2 + \frac{\lambda}{2r}}w_2^{AB}$$

where $w_2^{AB} = \frac{1}{2}p_2x_2^{AB}$.

[49] If a player deviates to type B and chooses $\Omega^B = \{G\}$, then $rV^B = \frac{\frac{\lambda}{2r}\Pr(G)w^{BG}}{1+\frac{\lambda}{2r}\Pr(G)}$. Nash Bagaining implies $w^{BG} = \frac{1}{2}p_2x_2^{BG} - \frac{1}{2}r(V - V^B)$. With V given in (20), and noting $x_2^{BG} = x_2^{AG}$, it then follows that $rV^B - rV = 0$. In this setting, deviating to be a player of type B and using the same matching policy as the existing As provides the same income as the existing players receive.

The "specialist" case is an equilibrium iff $rV \geq rV^G$. If you deviate to be a generalist and don't match with specialists, then $rV^G = w_2^G = p_2 x_2^G$. If you do match with specialists, then $rV^G = (w_2^G + \frac{\lambda}{2r} w_2^{GA}) / \left(1 + \frac{\lambda}{2r}\right)$, where $w_2^{GA} = \frac{1}{2} p_2 x_2^{AG} - \frac{1}{2} r(V - V^G)$ from the Nash Bargaining Solution.

Assuming Gs match with As and Bs the condition that $rV \geq rV^G$ is therefore (after some algebra)

$$x_2^{AB} \geq \left(\frac{2 + \frac{\lambda}{2r}}{1 + \frac{\lambda}{2r}} \right) \left(\frac{4r}{\lambda} x_2^G + x_2^{AG} \right)$$

Assuming alternatively that Gs do not match, the condition that $rV \geq rV^G$ is

$$x_2^{AB} \geq \left(1 + \frac{4r}{\lambda} \right) 2x_2^G$$

So the condition for the specialist case to be an equilibrium is

$$x_2^{AB} \geq 2x_2^G \max \left[1 + \frac{4r}{\lambda}, \left(\frac{2 + \frac{\lambda}{2r}}{1 + \frac{\lambda}{2r}} \right) \left(\frac{2r}{\lambda} + \frac{x_2^{AG}}{2x_2^G} \right) \right]$$

As $\lambda \to \infty$, the specialist case is an equilibrium iff $x_2^{AB} \geq \max \left[2x_2^G, x_2^{AG} \right]$. Noting that $x_2^{AB} > x_2^{AG}$, the binding condition can therefore only be $x_2^{AB} \geq 2x_2^G$with full employment.

(III) The "mixed" case, where $\left\{ L^A, L^B, L^G \right\} = \left\{ L', L', L^s - 2L' \right\}$. There are two subcases: (i) Gs do not match with As and Bs and (ii) Gs do match with As and Bs (see Lemma 1).

(i) If Gs do not match, then the equivalence of rV across worker types in equilibrium requires, using (11), that

$$\frac{\frac{\lambda}{2r} P w_2^{AB}}{1 + \frac{\lambda}{2r} P} = w_2^G \tag{21}$$

where $P = \Pr(A) = \Pr(B)$, $w_2^G = p_2 x_2^G$, and with the Nash Bargaining Solution $w_2^{AB} = \frac{1}{2} p_2 x_2^{AB}$.

(ii) If Gs do match, then the equivalence of rV across worker types in equilibrium requires that

$$\frac{\frac{\lambda}{2r} \left[P w_2^{AB} + (1 - 2P) w_2^{AG} \right]}{1 + \frac{\lambda}{2r} \left[1 - P \right]} = \frac{w_2^G + \frac{\lambda}{2r} 2P w_2^{GA}}{1 + \frac{\lambda}{2r} 2P} \tag{22}$$

where w_2^{AB} and w_2^G are as in (i) and, with the Nash Bargaining Solution, $w_2^{AG} = \frac{1}{2} p_2 x_2^{AG}$.

Deviating to another worker type has no effect on payoffs, since players are infinitesimal. These cases thus exist as equilibria if (a) a player would not change her matching policy and (b) there exists a $P \in [0, 1/2]$ that satisfies equality of income between specialists and generalists.

Comparing a Gs payoff when he doesn't match with the payoff when he does (the RHS of equations (21) and (22)), it is clear that $x_2^{AG} \geq 2x_2^G$ is necessary for G to match in equilibrium, and $x_2^{AG} \leq 2x_2^G$ is necessary for G not to match in equilibrium. Rearranging

(21), we can define an equilibrium value P^* as

$$P^* = \frac{2r}{\lambda(\frac{x_2^{AB}}{2x_2^G} - 1)}$$

where $P \in [0, 1/2]$ is necessary for an equilibrium to exist. Thus the "mixed" case where Gs do not match is an equilibrium iff $x_2^{AG} \leq 2x_2^G$ (Gs do not want to match), $x_2^{AB} \geq 2x_2^G$ ($P^* \geq 0$), and $\lambda \geq 4r\left[\frac{1}{2}x_2^{AB}/x_2^G - 1\right]^{-1}$ ($P^* \leq 1/2$).

As $\lambda \to \infty$ (full employment), $P^* \to 0$, so that this "mixed" equilibrium converges towards the "generalist" equilibrium.

If G does match in equilibrium, then rearranging (22) produces a quadratic in P, with either 0, 1, or 2 roots such that $P \in [0, 1/2]$. With some algebra, we can define an equilibrium value $\hat{P}$ as

$$\hat{P} = \frac{-\frac{2r}{\lambda}(\frac{x_2^{AB}}{2x_2^G} - 4\frac{x_2^{AG}}{2x_2^G} + 1) \pm \sqrt{\left(\frac{2r}{\lambda}\right)^2\left(\frac{x_2^{AB}}{2x_2^G} - 4\frac{x_2^{AG}}{2x_2^G} + 1\right)^2 + 8\frac{2r}{\lambda}\left(\frac{x_2^{AB}}{2x_2^G} - \frac{x_2^{AG}}{2x_2^G}\right)(\frac{2r}{\lambda} + 1 - \frac{x_2^{AG}}{2x_2^G})}}{4\left(\frac{x_2^{AB}}{2x_2^G} - \frac{x_2^{AG}}{2x_2^G}\right)} \tag{23}$$

The "mixed" case where Gs do match with As and Bs is an equilibrium iff $x_2^{AG} \geq 2x_2^G$ (Gs match with As and Bs) and $\hat{P} \in [0, 1/2]$. It can be shown that as many as 2 such equilibria are possible for some parameter values.

As $\lambda \to \infty$ (full employment), it follows directly from (23) that $\hat{P} \to 0$, so that any such "mixed" equilibrium also converges towards the "generalist" equilibrium. ■

Proof of Corollary (Gains from Specialization)

Proof. Output per specialist is $\frac{1}{2}p_2x_2^{AB} = p_2mc2^{\frac{1}{\sigma-1}}zh$ and output per generalist is $p_2x_2^G = p_22^{\frac{1}{\sigma-1}}zh$, so that the ratio of these outputs is $\frac{1}{2}p_2x_2^{AB}/(p_2x_2^G) = mc$. Hence the first part. For the second part, recall that the condition for the generalist equilibrium to be stable is $x_2^{AG} \leq 2x_2^G$ with full employment. Using the production function (10), this condition is equivalently written in terms of underlying parameters as $mc \leq \left(\frac{2}{1+m^{\frac{1-\sigma}{\sigma}}}\right)^{\frac{\sigma}{\sigma-1}}$. Recalling that tasks A and B are gross complements in production ($\sigma \leq 1$), it follows that $\lim_{m\to\infty}\left(\frac{2}{1+m^{\frac{1-\sigma}{\sigma}}}\right)^{\frac{\sigma}{\sigma-1}} = \infty$. Hence the maximum possible mc for which generalists exist in a stable equilibrium is unbounded from above. ■

Proof of Lemma (Log-Linear Wages)

Proof. Given that individuals have the same choice set at birth and maximize income, they must be indifferent across career choices so that $W^k = W$ for all worker types. With full employment, this income arbitrage means from (5) that

$$\int_0^\infty w_1^n e^{-rt}dt = \int_s^\infty w_2^n e^{-rt}dt \tag{24}$$

where $w_1^n = rV^U$ is the wage paid in the unskilled sector and $w_2^n = rV$ is the wage paid in the skilled sector. Integrating (24) gives $w_2^n = w_1^n e^{rs}$. ■

Proof of Corollary (Balassa-Samuelson)

Proof. The price adjustment follows directly from (15). The labor supply adjustment then follows directly from (16). ■

Proof of Lemma (Mincer Accounting as Lower Bound)

Proof. In the model, $H^{AB}/H^G = Y^{AB}/Y^G$. Skilled workers are $mc > 1$ times more skilled in the AB case than the G case. From (17) and (16), we write

$$\frac{H^{AB}}{H^G} = \frac{L_1^{AB}\left(1 + e^{rS}\frac{L_2^{AB}}{L_1^{AB}}\right)^{\frac{\varepsilon}{\varepsilon-1}}}{L_1^G\left(1 + e^{rS}\frac{L_2^G}{L_1^G}\right)^{\frac{\varepsilon}{\varepsilon-1}}} \tag{25}$$

Recalling that $H_{Mincer}^n = L_1^n + e^{rS}L_2^n$, we can manipulate (25) to write the ratio of the true human capital ratio to the Mincerian calculation, $R_H = \frac{H^{AB}/H^G}{H_{Mincer}^{AB}/H_{Mincer}^G}$, as

$$R_H = \left(\frac{1 + e^{rS}\frac{L_2^{AB}}{L_1^{AB}}}{1 + e^{rS}\frac{L_2^G}{L_1^G}}\right)^{\frac{1}{\varepsilon-1}}$$

Consider the case where $\varepsilon \in [1, \infty]$. From Corollary 2, $L_2^{AB}/L_1^{AB} \geq L_2^G/L_1^G$, with strict inequality if $\varepsilon > 1$. Given the observed labor allocations, it follows that $\lim_{\varepsilon \to 1^+} R_H = \infty$ and that R_H declines in ε. Further $\lim_{\varepsilon \to \infty} R_H = 1$. Hence, $R_H \geq 1$ given $\varepsilon > 1$.

Consider the case where $\varepsilon \in [0, 1]$. From Corollary 2, $L_2^{AB}/L_1^{AB} \leq L_2^G/L_1^G$, with strict inequality if $\varepsilon < 1$. Given the observed labor allocations, it follows that $\lim_{\varepsilon \to 1^-} R_H = \infty$ and that R_H increases in ε. Further $\lim_{\varepsilon \to 0} R_H > 1$. Hence, $R_H > 1$ given $\varepsilon \leq 1$.

In sum, over $\varepsilon \in [0, \infty]$ it follows that $R_H \geq 1$. Moreover, for a given labor allocation $L_1^G/L_1^{AB} \neq 1$, $\lim_{\varepsilon \to 1} R_H = \infty$. ■

Proof of Corollary (Immigrant Wages Bounds)

Proof. The high skill immigrant has skill h_2^G. The ratio h_2^G/h_2^{AB} may be arbitrarily close to zero by Corollary 1. However, if the skilled immigrant chooses to work at the unskilled task, the wage will be $w^i = w_1^{AB}$. In the rich country equilibrium, this wage is $w_1^{AB} = w_2^{AB}e^{-rS}$. Hence the immigrant earns at least $w^i = w_2^{AB}e^{-rS}$. ■

Proof of Corollary (Immigrant Wages and Occupations)

Proof. The low-skilled immigrant earns a higher real wage by moving to the rich country because, from (17)

$$\frac{w_1^{AB}/p^{AB}}{w_1^G/p^G} = \frac{y^{AB}}{y^G} > 1$$

Hence an unskilled worker who migrates from a poor to a rich country will earn a higher real wage.

Now consider skilled immigrants.

Note first that the skilled generalist who migrates will never team with a specialist in the rich country. Rather, he would always prefer to work alone, since he must give up too much of the joint product to convince a specialist to partner with him. In particular, he would earn $p_2^{AB} x_2^G$ alone, while in a team (with full employment) he would earn $p_2^{AB}\left(x_2^{AG} - \frac{1}{2}x_2^{AB}\right)$, and there are no parameter values where $x_2^G < x_2^{AG} - \frac{1}{2}x_2^{AB}$. To see this, write this condition as $1 < x_2^{AG}/x_2^G - \frac{1}{2}x_2^{AB}/x_2^G$. Note that $\frac{1}{2}x_2^{AB}/x_2^G = mc$ and that x_2^{AG}/x_2^G can be no greater than $mc + c$.[50] Hence the condition is equivalently $1 < c$, which contradicts the assumption of the model that there are coordination costs in production, $c < 1$.

Next, note that working alone as a generalist in the rich country is never preferred to staying in the poor country. In either country, the generalist produces x_2^G units of output per unit of time. Given that this good is relatively expensive in the poor country (i.e. recall that $p_2^G/p_1^G = mc\left(p_2^{AB}/p_1^{AB}\right)$), the real income is higher working as the generalist in the poor country.

Lastly, note that the generalist may still prefer to migrate and work in the unskilled sector. This occurs when the real wage gain across countries for unskilled work $\frac{w_1^{AB}/p_1^{AB}}{w_1^G/p_1^G}$ (see above) is larger than the real wage gain locally for skilled work, e^{rs}, which is more likely the greater the income differences between the countries; for example, the greater the gains from specialization, mc.

In sum, skilled generalists may or may not be better off migrating to rich countries, but if they do they will work in the unskilled sector. ■

Proof of Corollary (Brain Drain)

Proof. The specialist who moves to the poor country will earn a wage $w_2' = p_2^G(x_2^{AG} - x_2^G)$. Since the poor country is in a generalist equilibrium, we must have $x_2^{AG} \leq 2x_2^G$ which implies that $w_2' \leq p_2^G x_2^G = w_2^G$. Hence, the skilled worker who moves from the rich to the poor country will earn a wage no greater than the skilled worker wage in the poor country. Now note that skilled workers receive a higher real wage in the rich country than the poor country because, from (14) and (17),

$$\frac{w_2^{AB}/p^{AB}}{w_2^G/p^G} = \frac{y^{AB}}{y^G} > 1$$

Hence, specialists in the rich country will prefer to stay. ■

[50] This follows because x_2^{AG}/x_2^G is increasing in σ, attaining a maximum $x_2^{AG}/x_2^G = mc + c$ as $\sigma \to \infty$.

Proof of Corollary (Desirable Cheap Specialists)

Proof. Think of the firm as a specialist in the rich country. He earns $w_2^{AB} = \frac{1}{2} p_2^{AB} x_2^{AB}$. If he can alternatively form a cross-border team by locating an (off-equilibrium) specialist in the poor country, then he can earn at least $w_2 = p_2^{AB} x_2^{AB} - p_2^{AB} x_2^{G}$, where he need provide the specialist in the poor country no more than x_2^G, the going rate for generalists in that country. Hence, hiring a specialist in the poor country makes sense iff $x_2^{AB} - x_2^G \geq \frac{1}{2} x_2^{AB}$ or $x_2^{AB} \geq 2x_2^G$, which is just the condition for specialists to exist in the first place in the rich country. ■

Data and Analysis for Figure A1

Friedberg (2000) demonstrates that the source of education does matter to immigrant wages, but the literature does not appear to have looked explicitly at higher education. Descriptive facts can be assembled however using census data. I divide individuals in the 2000 U.S. Census into three groups: (1) US born, (2) immigrants who arrive by age 17, and (3) immigrants who arrive after age 30. The idea is that those who immigrated by age 17 likely received any higher education in the United States, while those who immigrated after age 30 likely did not.

Data on wages and occupations is taken from the 1% microsample of the 2000 United States census, which is available publicly through www.ipums.org.[51] There are 2.8 million individuals in this sample, including 320,000 individuals who immigrated to the United States.

The wage-schooling relationships in Figure A1a are the predicted values from the following regression

$$\ln w_i = \alpha + \beta MALE + Age_{fe} + English_{fe} + Group_{fe} + Education_{fe} + Group_{fe} \times Education_{fe} + \varepsilon_i$$

where w_i is the annual wage, $MALE$ is a dummy equal to 1 for men and 0 for women, Age_{fe} are fixed effects for each individual age in years, $English_{fe}$ are fixed effects for how well the individual speaks English (the IPUMS "speakeng" variable which has 6 categories), $Education_{fe}$ are fixed effects for highest educational attainment (the IPUMS "educ99" variable, which has 17 categories) and $Group_{fe}$ are fixed effects for three different groups: (1) US born, (2) immigrants who arrive by age 17, (3) immigrants who arrive age 30 or later. Figure A1a plots predicted values from this regression, plotting the log wage against educational attainment for each of the three groups. For comparison purposes, the predicted values focus on males between the ages of 30 and 40 who speak English at least well.

[51] Integrated Public Use Microdata Series (Steven Ruggles, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. *Integrated Public Use Microdata Series: Version 3.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004.)

Figure A1a shows that, controlling for age and English language ability, the location of higher education appears to matter. Among highly educated workers, those who immigrate after age 30 experience significant wage penalties, of 50% or greater. Meanwhile there is no wage penalty if the immigrant arrived early enough to receive higher education in the United States. Second – and conversely – wages do not differ by birthplace or immigration age for workers with an approximately high-school level education. Hence, the location of education matters for high skill workers but not so much for low skill workers, as this paper's model suggests.[52]

To construct Figure A1b, the modal educational attainment is first determined for each of the 511 occupational classes in the data (using the IPUMS variable "occ"). Occupations are then grouped according to modal educational attainment. For example, lawyers are grouped with doctors as typically having professional degrees, and taxi drivers are grouped with security guards as typically having high school degrees. For each of the three groups defined for the $Group_{fe}$ above, Figure A1b shows the propensity of individuals with professional or doctoral degrees to work in occupations with the given modal educational attainment.

We see that US born workers and early immigrants have extremely similar occupational patterns. However, late immigrants with professional or doctoral degrees have a much smaller propensity to work in occupations that rely on such degrees. Instead, they tend to shift down the occupational ladder into jobs that require only college degrees and even, to a smaller extent, into occupations typically filled by those with high school or less education. (This pattern is further reflected in Figure A1a, which shows that late immigrants with professional or Ph.D. degrees earn average wages no better than a locally educated college graduate.) Overall, these immigration findings are consistent with the model.

References

[1] Acemoglu, Daron. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation", *Quarterly Journal of Economics*, August 1996, 111 (3), 779-804.

[2] Acemoglu, Daron. "Training and Innovation in an Imperfect Labor Market", *Review of Economic Studies*, July 1997, 64, 445-464.

[3] Acemoglu, Daron and Jorn-Steffen Pischke. "Why do Firms Train? Theory and Evidence", *Quarterly Journal of Economics*, 1998, 113 (1), 79-120.

[4] Acemoglu, Daron, Pol Antras, and Elhanan Helpman. "Contracts and Technology Adoption", mimeo, Harvard University, 2006.

[52] Note also that immigrants with high school or less education have extremely similar wage outcomes regardless of immigration age. This further suggests that early-age immigrants are an adequate control group for late-age immigrants, highlighting that differing labor market outcomes only occur at higher education levels. Lastly, it is clear that very-low education immigrants (e.g. primary school) do significantly better then very-low educated US born workers. Such limited education is very rare among the US born and likely reflects individuals with developmental difficulties, which may explain that wage gap.

[5] Agrawal, Ajay, Avi Goldfarb, and Florenta Teodoridis. "Does Knowledge Accumulation Increase the Returns to Collaboration?" NBER Workding Paper No. 19694, December 2013.

[6] Balassa, Bela. "The Purchasing-Power-Parity Doctrine: A Reappraisal", *Journal of Political Economy*, December 1964, 72, 584-96.

[7] Banerjee, Abhijit and Esther Duflo. "Growth Theory Through the Lens of Development Economics", *The Handbook of Economic Growth*, Aghion and Durlauf eds., Elsevier, 2005.

[8] Barro, Robert J. and Lee, Jong-Wha. "International Data on Educational Attainment: Updates and Implications." *Oxford Economic Papers*, 2001, 53(3), 541–63.

[9] Becker, Gary S. and Kevin M. Murphy. "The Division of Labor, Coordination Costs, and Knowledge", *Quarterly Journal of Economics*, November 1992, 107 (4), 1137-1160.

[10] Bhagwati, Jagdish. "Why Are Services Cheaper in Poor Countries?", *Economic Journal*, June 1984, 94, 279-285.

[11] Bils, Mark, and Peter J. Klenow. "Does Schooling Cause Growth?", *American Economic Review*, December 2000, 90 (5), 1160-1183.

[12] Bowen, Harry, Edward Leamer, and Leo Sveikauskas. "Multicountry, Multifactor Tests of the Factor Abundance Theory", *American Economic Review*, December 1987, 77 (5), 791-809.

[13] Caselli, Francesco. "Accounting for Cross-Country Income Differences", *The Handbook of Economic Growth*, Aghion and Durlauf eds., Elsevier, 2005.

[14] Caselli, Francesco and Wilbur J. Coleman. "The World Technology Frontier", *American Economic Review*, June 2006, 96 (3), 499-522.

[15] Card, David. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69 (5), 2003, 1127-1160.

[16] Ciccone, Anotonio and Kiminori Matsuyama. "Start-Up Costs and Pecuniary Externalities As Barriers to Economic Development", *Journal of Development Economics*, 1996, 49, 33-59.

[17] Erosa, Andres, Diego Restuccia, and Tatyana Koreshkova, "How Important is Human Capital? A Quantitative Theory Assessment of World Income Inequality," *Review of Economic Studies*, October 2010, 77 (4), 1421-49.

[18] Friedberg, Rachel. "You Can't Take It With You? Immigrant Assimilation and the Portability of Human Capital", *Journal of Labor Economics*, April 2000, 18 (2), 221-251.

[19] Garicano, Luis. "Hierarchies and the Organization of Knowledge in Production", *Journal of Political Economy*, October 2000, 108 (5), 874-904.

[20] Hall, Robert and Charles Jones. "Why Do Some Countries Produce So Much More Output Per Worker Than Others?", *Quarterly Journal of Economics*, February 1999, 114 (1), 83-116.

[21] Harrod, Roy F. *International Economics*. Cambridge Economic Handbooks, London: Cambridge University Press, 1933.

[22] Harrigan, James. "Technology, Factor Supplies, and International Specialization: Estimating the Neoclassical Model," *American Economic Review*, September, 1997, 87 (4), 475-494.

[23] Heckman, James J., Lochner, Lance J., and Petra E. Todd. "Earnings Functions: Rates of Return and Treatment Effects: The Mincer Equation and Beyond", *IZA Discussion Paper Series* No. 1700, August 2005.

[24] Hendricks, Lutz. "How Important Is Human Capital for Development? Evidence from Immigrant Earnings", *American Economic Review*, March 2002, 92 (1), 198-219.

[25] Hsieh, Chang-Tai and Peter Klenow. "Relative Prices, Relative Prosperity", NBER Working Paper #9701, 2003.

[26] Jones, Benjamin F. "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" *Review of Economic Studies* 76 (1), January 2009.

[27] Jones, Benjamin F. "The Human Capital Stock: A Generalized Approach," *American Economic Review*, forthcoming.

[28] Jones, Charles I. "On the Evolution of the World Income Distribution", *Journal of Economic Perspectives*, Summer 1997, 11 (3), 19-36.

[29] Katz, Lawrence F. and Autor, David H. “Changes in the Wage Structure and Earnings Inequality,” in Orley Ashenfelter and David Card,eds., *Handbook of labor economics* Vol. 3A. (Amsterdam: Elsevier Science, North-Holland), 1999, 1463–555.

[30] Kim, Sunwoong. "Labor Specialization and the Exent of the Market", *Journal of Political Economy*, June 1989, 97 (3), 692-705.

[31] Klenow, Peter J., and Andres Rodriguez-Clare. “The Neoclassical Revival in Growth Economics: Has It Gone Too Far?”, in: B.S. Bernanke, and J.J. Rotemberg, eds., *NBER Macroeconomics Annual* (MIT Press, Cambridge), 1997, 73-103.

[32] Kremer, Michael. "The O-Ring Theory of Economic Development". *Quarterly Journal of Economics*, August 1993, 108 (3), 551-575.

[33] Leontief, Wassily W. "Domestic Production and Foreign Trade: The American Capital Position Re-examined." *Proceedings of the American Philosophical Society*, September 1953, 97, 332-349.

[34] Manuelli, Rodolfo and Ananth Seshadri, "Human Capital and the Wealth of Nations," *American Economic Review*, forthcoming.

[35] Marshall, Alfred. *Principles of Economics*. London: MacMillan. 1920.

[36] Maskus, Keith. "A Test of the Heckscher-Ohlin-Vanek Theorem: The Leontief Commonplace", *Journal of International Economics*, November 1985, 9, 201-212.

[37] Mincer, Jacob. "Investment in Human Capital and Personal Income Distribution", *Journal of Political Economy*, August 1958, 66 (4), 281-302.

[38] Mincer, Jacob. *Schooling, Experience, and Earnings*, New York: Columbia University Press, 1974.

[39] Murphy, Kevin P., Andrei Shleifer, and Robert W. Vishny. "Industrialization and the Big Push", *Journal of Political Economy*, October 1989, 97 (5), 1003-1026.

[40] Pritchett, Lant. "Divergence, Big Time", *Journal of Economic Perspectives*, Summer 1997, 11 (3), 3-17.

[41] Psacharopoulos, George. "Retruns to Investment in Education: A Global Update", *World Development*, 1994, 22 (9), 1325-43.

[42] Rodriguez-Clare, Andres. "The Division of Labor and Economic Development", *Journal of Development Economics*, 1996, 49, 3-32.

[43] Samuelson, Paul. "Theoretical Notes on Trade Problems", *Review of Economics and Statistics*, May 1964, 46, 145-54.

[44] Saxenian, Annalee. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press. 1994.

[45] Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations*. 1776.

[46] Summers, Robert and Heston, Alan. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988." *Quarterly Journal of Economics*, 1991, 106(2), 327–68.

[47] Trefler, Daniel. "International Factor Price Differences: Leontief Was Right!", *Journal of Political Economy*, 1993, 101 (6), 961-987.

[48] Trefler, Daniel. "The Case of the Missing Trade and Other Mysteries", *American Economic Review*, December 1995, 85 (5), 1029-1046.

[49] Wuchty, Stefan, Benjamin F. Jones and Brian Uzzi. "The Increasing Dominance of Teams in the Production of Knowledge", *Science*, 2007, 318, 1036-39.

[50] Zweig, David. "Competing For Talent: China's Strategies to Reverse the Brain Drain", *International Labour Review*, 2006, 145, 65-89.

Table 1: Access to Advanced Knowledge: An Example

	University of Khartoum Department of Mechanical Engineering	Massachusetts Institute of Technology Department of Mechanical Engineering
Degrees	Undergraduate, Masters, Ph.D.	Undergraduate, Masters, Ph.D.
Faculty	20 (14 professors, 6 lecturers)	160 (93 professors, 62 lecturers, 5 technical instructors)
Subfields	1 (M.Sc. in Energy Engineering)	7 "Areas" and 17 different course groups
Courses	51	175[b]
	(Complete list)	(Examples)
	Undergraduate Courses (34 courses)	*"Core Undergraduate Subjects" (13 courses)*
	Thermodynamics I, II, III; Fluid Mechanics I, II	*"System Dynamics and Control" (17 courses)*
	Mechanics of Materials I, II; Manufacturing Process I, II	(e.g. Robotics, Biomechanics and Neural Control of Movement)
	Mechanical Engineering Laboratory I, II, III	*"Energy and Power Systems" (11 courses)*
	Mechanical Engineering Design I, II, III	(e.g. Internal Combustion Engines, Superconducting Magnets)
	Hydraulic Machines I, II, Thermal Power Engineering I, II	*"Dynamics and Acoustics" (12 courses)*
	Dynamics of Mechanical Systems I, II	(e.g. Acoustics & Sensing, Mechanical Vibration, Surface Wave Dynamics)
	Mechanics of Machines, Machine Elements	*"Solid Mechanics and Materials" (10 courses)*
	Exposing Information Technology, Computer Applications	(e.g. Mechanics of Continuous Media, Plates and Shells, Structural Impact)
	Heat Transfer, Heat and Mass Transfer	*"Computational Engineering" & "Experimental Engineering" (14 courses)*
	Engineering Economics, Engineering Management,	(e.g. Computational Geometry, Molecular Modeling & Sim of Mechanics)
	Industrial Management	*"Fluid Mechanics and Combustion" (11 courses)*
	Refrigeration and Air Conditioning, A/C Systems	(e.g. Marine Hydrodynamics, Hydrofoils and Propellers)
	Gas Dynamics	*"MEMS and Nanotechnology" (5 courses)*
	M.Sc. in Energy Engineering (17 courses)	(e.g. Design & Fabrication of MEMS, Submicrometer & Nanometer Tech)
	Energy Science, Numerical Techniques & Computations	*"Thermodynamics" & "Heat and Mass Transfer" (9 courses)*
	Energy Economics & Management,	(e.g. Nano-to-Macro Transport Processes, Radiative Transfer)
	Instrumentation & Experimental Techniques	*"Oceanographic Engineering and Acoustics" (9 courses)*
	Hydro Power Plants, Advanced I.C. Engines	(e.g. Marine Bio-Acoustics and Geo-Acoustics)
	Steam Power Plants, Gas Turbine Power Plants	*"Design" & "Naval Architecture" (15 courses)*
	Solar and Wind Power Plants, Nuclear Power Plants[b]	(e.g. Principles of Naval Ship Design, Mechatronics)
	Rocket and Aircraft Propulsion[b], Novel Power Systems	*"Bioengineering" (13 courses)*
	Energy Systems Control, Advanced Heat Transfer	(e.g. Cell-Matrix Mechanics, Biomaterials: Tissue Interactions)
	Storage and Transportation of Energy	*"Manufacturing" & "Engineering Management" (19 courses)*
	Computational Fluid Dynamics, Combustion Engineering	(e.g. Tribology, Fabrication Technology)
		"Optics" & Other (17 courses)
		(e.g. Optical Microscopy & Spectroscopy for Biology and Medicine)

Notes: (a) Count of MIT mechanical engineering faculty does not include 56 research scientists and post-docs; (b) MIT provides 97 further courses in a separate department, "Aeronautics and Astronautics", whereas Khartoum provides one course on that topic, Rocket and Aircraft Propulsion, listed within mechanical engineering. Similarly, MIT provides 44 further courses in a separate department, "Nuclear Science and Engineering", whereas Khartoum provides one relevant course, Nuclear Power Plants, listed within mechanical engineering. Sources: MIT Bulletin (2007-2008) and www.uofk.edu (accessed, 9/2011).

Figure A1a: Do Skilled Immigrants Experience Wage Penalties?
The Wage-Schooling Relationship

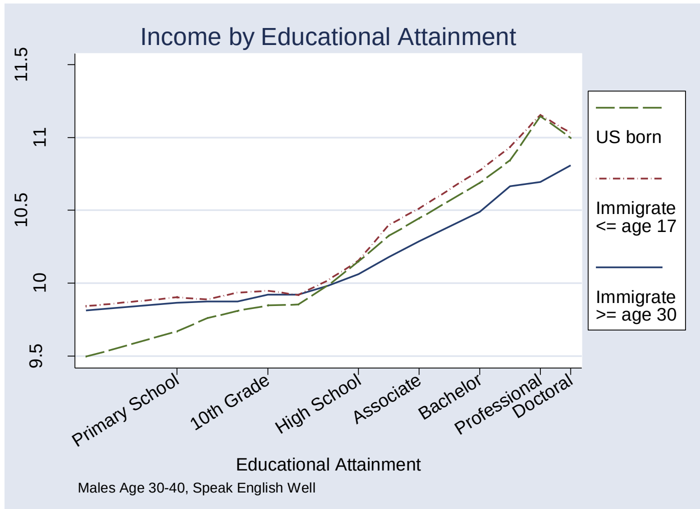

Figure A1b: Do Skilled Immigrants use their Education?
Occupations of Workers with Professional or Doctoral Degrees