The Knowledge Trap: Human Capital and Development Reconsidered*

Benjamin F. Jones†

June 2008

Abstract

This paper presents a model where human capital differences - rather than technology differences - can explain several central phenomena in the world economy. The results follow from the educational choices of workers, who decide not just how long to train, but also how broadly. A "knowledge trap" occurs in economies where skilled workers favor broad but shallow knowledge. This simple idea can inform cross-country income differences, international trade patterns, poverty traps, and price and wage differences across countries in a manner broadly consistent with existing empirical evidence. The model also provides insights about the brain drain, migration, and the role for multinationals in development. More generally, this paper shows that standard human capital accounting methods can severely underestimate the role of education in development. It shows how endogenous educational decisions can replace exogenous technology differences in a range of economic reasoning.

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

*I wish to thank Daron Acemoglu, Pol Antras, Andrew Hertzberg, Jin Li, Pablo Montagnes, Ben Olken, Scott Stern, Kane Sweeney and participants in presentations at Northwestern, Washington University, Penn State, the LSE, 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, economists often find it necessary to invoke "technology differences". This paper suggests an alternative explanation, based in human capital. I present a model where endogenously acquired skills explain many stylized facts about the world economy – including those facts often interpreted as evidence against human capital.

The model emphasizes cross-country differences in the *quality of skilled workers*. The results follow from education decisions, which have two dimensions. One is duration - how much time to put in education - which defines whether you become a skilled worker or not. The other is content - what specific knowledge to acquire. In particular, given an investment of time, one might become a "generalist" (e.g. a generalist doctor), with modest knowledge about multiple tasks, or a "specialist" (e.g. an anesthesiologist) with deep knowledge at a particular task. Quality advantages emerge in the *collective* productivity of skilled workers, where specialists working in teams bring greater collective knowledge to bear in production.

The theory thus builds on Adam Smith's foundational observation that specialization can bring high productivity. The twist is to understand why these gains may go unrealized in the educational phase. In the model, any gains from narrowly focused training are traded off against the cost of both finding specialists with complementary skills and coordinating with them in production. This tradeoff may favor breadth over depth for three reasons. First, deep, specialized knowledge may be hard to acquire locally; for example, heart surgery may be hard to learn without guidance from existing heart surgeons. Absent expert instructors, focused training provides less advantage. Second, specialization may be worthwhile only when a sufficient mass of complementary specialists already exists. For example, learning heart surgery is less useful in the absence of anesthesiologists. These two issues suggest a variety of poverty trap where local deep knowledge is a prerequisite for individuals to willingly and successfully seek deep knowledge. Finally, coordination costs in production may be especially high.[1] For any (or all) of these reasons, a low-productivity "generalist"

[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).

equilibrium may persist. I call such outcomes a "knowledge trap" because the generalist equilibrium features shallower collective knowledge.

Such quality differences are limited to skilled workers. Importantly, however, the real wages of unskilled workers also rise in rich countries. This is a general equilibrium effect that follows when the output of skilled workers is relatively abundant, making the output of unskilled workers relatively scarce. This scarcity drives up unskilled wages. More precisely, when decisions to become skilled or remain unskilled are endogenous - the duration dimension of education is a choice variable - the wage structure is pinned down in equilibrium so that although quality differences are limited to skilled workers, real income gains are shared equally by skilled and unskilled workers alike.

This equilibrium effect is crucial, because it poses significant challenges to standard human capital accounting methods. The standard approach infers cross-country skill differences from within-country returns to schooling, but in this model the entire wage distribution shifts, so that within-country wage equilibria on their own say nothing about cross-country skill differences. Estimation approaches based on immigrant behavior face similar problems. The wage gains experienced by unskilled workers who immigrate from poor to rich countries need not be explained by technology, as many authors infer; in this model, the wage gains follow simply because unskilled workers are relatively scarce in the rich country. For example, one may ask why taxi drivers earn so much more in rich countries. A natural explanation is that the taxi driver's clients - skilled workers – have a much higher opportunity cost of their time and hence will pay more for the ride.

In sum, rich countries are rich because they attain deeper collective knowledge among skilled workers. The relative scarcity of low skill means that the real wages of unskilled workers also rise in rich countries, even though such workers have no more skill in rich than poor countries. One thus finds a skill-based explanation for cross-country income differences that can also get wages right.

Furthermore, the depth versus breadth tradeoff means that (a) schooling duration is insufficient to assess skill and (b) the productivity of skilled workers is interdependent. Workers are puzzle pieces, who fit together differently in different economies. This richer perspective on skill may inform phenomena such as international trade patterns (with knowledge traps, Heckscher-Ohlin can make a comeback), migrant behavior (why do skilled immigrants often take menial jobs?), and the brain drain (why don't skilled workers move to poor countries, where they are scarce?). The model further suggests pathways for escaping poverty traps, including intriguing roles for multinationals in triggering development, which may inform growth miracles in places like Hyderabad and Bangalore.

This paper is organized as follows. Section 2 introduces the core ideas. Section 3 presents a formal model, clarifying conditions for the existence of "knowledge traps" and their general equilibrium effects. Section 4 discusses several applications and relates them to existing empirical evidence in addition to new evidence about the quality of skilled workers. 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. The model provides additional perspectives on (vi) the role of multinationals in development and (vii) international trade patterns. 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 education as a source of multiple equilibria. 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, especially 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.

The primary contributions of this paper are two-fold. First, I show broadly how standard accounting methods may underestimate the role of education in skill formation. Second, I present a simple, specific mechanism, based on the division of skills, to show why large skill differences may exist and persist across countries, providing a parsimonious interpretation of many stylized facts in the world economy.

2 The Core Ideas

This section provides an introductory discussion of the core ideas in this paper. First, I introduce a "knowledge trap" to show how endogenous educational decisions can produce large cross-country differences in skill. Second, I show how standard macroeconomic accounting methods will misaccount for these skill differences. Section 3 integrates these ideas into a formal model.

2.1 A Knowledge Trap

Imagine there are two tasks, A and B, which are complementary in the production of a good. For example, the good could be heart surgery, where one task is anesthesiology and the other is the surgery itself.

Now imagine individuals must train to acquire skill, and one must decide how to use an endowment of training time. One might train as a "generalist", developing skill at both tasks. Alternatively, one might focus all their 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 a simple 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 an economy of generalists is a stable equilibrium. In a poor country, this may occur most simply because m is small. For example, becoming a skilled heart surgeon may be difficult without access to an existing skilled heart surgeon. Alternatively, coordination penalties in production may be more severe in poor countries. Hence, poor countries may feature $m'c' < 1$ while a rich country has $mc > 1$.

More subtly, an economy of generalists may persist due to thin supply of complementary specialist types. To see this, imagine you are born into a world of generalists and consider whether you would want 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 in which 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 specialist in such a team. The most income the specialist could earn is therefore $\sqrt{m}c - \frac{1}{2}$, which itself must exceed $\frac{1}{2}$ for a player to prefer training as a specialist. Hence the generalist equilibrium is stable to individual deviations if $\sqrt{m}c < 1$. 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.[2]

I call these specialization failures a "knowledge trap" because skilled workers in the generalist equilibrium have *shallower* knowledge. This doesn't mean that they have little education. For example, the generalist doctor knows something about both anesthesiology and surgery – not to mention oncology, infectious disease, psychiatry, ophthalmology, etc.

[2] 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 is that 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.

Learning something about all these different subjects may require a lot of education. But this generalist doctor will likely be far less productive than a set of specialists who work together. The specialists may have no more schooling per person, but they have much deeper knowledge about individual tasks, so that the collective body of knowledge across the specialists may be far greater. Quality differences thus follow here from the content dimension of education. To see why potentially large quality differences will not be detected by standard human capital accounting methods, we must further consider the duration dimension of education, which we turn to next.

2.2 Human Capital and Wages

A large literature has concluded that schooling variation across countries is too small to explain cross-country income differences (see Caselli 2005 for a survey). This inference is primarily drawn using the "macro-Mincer" approach, which attempts to compute human capital stocks from data on the wage-schooling relationship (e.g. Hall and Jones 1999, Bils and Klenow 2000). If workers are paid their marginal products, then the wage gain from schooling can inform how schooling influences productivity. Wage-schooling relationships are usually taken to follow the log-linear, i.e. "Mincerian", form (Mincer 1974),

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

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.[3]

To see how such *within-country* wage relationships can be misused in inferring *cross-country* skill differences, consider first that these wage structures emerge as a local equilibrium when labor supply 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

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

off, then for any two schooling levels

$$y(s) = y(s')$$

and therefore (1) follows immediately with $r_m = r$.[4] The log-linear wage structure follows through arbitrage. Individuals become skilled by investing time in education, which means giving up wages today in exchange for higher wages later. In this simple setting, the rate of return on a foregone dollar of wage income is pinned down by the expected return on investment - i.e. the discount rate.[5] Quality differences in education won't appear in the wage data, because educational duration decisions reallocate workers endogenously to ensure this equilibrium rate of return.

Now consider how one can interpret skill from wages. Imagine that there are two goods, good 1 (e.g. haircuts) produced by unskilled workers with no education and good 2 (e.g. surgery) that requires S years of training to perform. Let preferences be the same in all countries and demand for each good be downward sloping. Lastly, 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 good is then $w_1 = p_1 h_1$ and $w_2 = p_2 h_2$, and we have

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

where $w_2/w_1 = e^{rS}$ follows from income arbitrage as above.

To compare skill across countries, standard accounting methods assume that unskilled workers have the same innate skill, h_1, in all economies and estimate the skill of the educated as

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

But this method for estimating h_2 is clearly problematic. As just shown in (3), one must also confront relative prices (p_1/p_2), which are well known to differ substantially across

[4]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

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

which is just the log-linear wage structure expressed as a marginal condition.

[5]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 drive the return to schooling above the real interest rate. See Heckman et al. (2005) for a broader characterization of lifetime income.

countries.[6] And it is easy to see how ignoring relative prices might substantially understate human capital differences. 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. Failing to account for these price differences will dampen cross-country skill differences compared to the case where we assume prices are the same everywhere.[7]

These observations suggest that skill differences might explain rather more about the world economy than a large literature has suggested. The following section presents a general equilibrium model, integrating the quality differences of knowledge traps with endogenous schooling duration decisions. Section 4 then details several applications and reconsiders established empirical evidence from the model's perspective.

3 The Knowledge Trap Economy

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.

[6] Such relative price differences are large and motivate the need for purchasing power parity (PPP) price corrections when comparing real incomes across countries.

[7] The standard accounting method assumes that the output of different skilled 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. More realistically, any number of high school students are unlikely to successfully perform angioplasty, assemble a jet engine, or write a contract consistent with the UCC. Different types of workers produce different types of goods that face downward sloping demand. Hence, skill endowments will matter in making inferences about human capital. This will be discussed further in Section 4.

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,
   (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 Income

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

$$W^k = \int_{s^k}^{\infty} rV^k e^{-r\tau} d\tau \tag{4}$$

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. This is defined by the Bellman equation,

$$rV^k = w^k + \lambda \sum_{j \in \Omega^k} \Pr(j) \left(V^{kj} - V^k\right) \tag{5}$$

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

$$\Pr(j) = L_p^j / L_p \tag{6}$$

where L_p^j is the measure of workers of type j who are unpaired and $L_p = \sum_j L_p^j$.[8] You accept the match if $V^{kj} \geq V^k$ and reject otherwise, which defines the "acceptance set", $\Omega^k \subset \{G, A, B\}$, 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

$$rV^{kj} = w^{kj} - r\left(V^{kj} - V^k\right) \tag{7}$$

[8] 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,

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

The flow value of being paired, rV^{kj}, is equal to the wage you receive in this pairing, w^{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

$$w^{kj} = \arg\max_{\hat{w}^{kj}} \left(V^{kj} - V^{k}\right)^{1/2} \left(V^{jk} - V^{j}\right)^{1/2} \tag{8}$$

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

3.1.3 Output

There are two output sectors. 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

$$x_1 = z$$

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

$$x_2 = zc(n)\left(H_A^{\alpha} + H_B^{\alpha}\right)^{1/\alpha}, \; H_k = \sum_i t_i^k h_i^k \tag{9}$$

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.[9] The term $c(n) \in [0,1]$ captures the coordination penalty from working in a team of size n. 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.

3.1.4 Preferences

Utility is given by

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

[9] The CES production function in (9) 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 (see also Footnote 36).

where $u(C)$ is increasing and concave and the rate of time preference, r, is given by the hazard rate of death. Individuals have access to a competitive annuity market which pays an interest rate on loans of r.[10] The equivalence of the interest rate and the rate of time preference implies that an individual's consumption does not change across periods, by the standard Euler equation.[11] Let preferences across goods be

$$C^k(x_1, x_2) = \left(\gamma x_1^\rho + (1-\gamma)\, x_2^\rho\right)^{1/\rho} \tag{10}$$

where $\varepsilon = \frac{1}{1-\rho}$ is the elasticity of substitution between goods, which we assume is finite.

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 stationary 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 $V^{kj} \geq V^k$, $\forall k \in \{G, A, B\}$*

[10]There is no 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.

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

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 (5) and (7),

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

so that the value of being a type k worker depends on (a) the wage you earn if you work alone, w^k, (b) the wage you can earn in pairings you are willing to accept, w^{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 (4),

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

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, (8), that in any accepted pairing $V^{kj} = V^{jk}$ and

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

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

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

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 coordination 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 $\left\{L^A, L^B, L^G\right\} = \{0, 0, L^s\}$; (2) a "specialist" equilibrium where $\left\{L^A, L^B, L^G\right\} = \left\{\frac{1}{2}L^s, \frac{1}{2}L^s, 0\right\}$; and (3) a "mixed" equilibrium where $\left\{L^A, L^B, L^G\right\} = \{L', L', L^s - 2L'\}$ for some L' such that $0 < L' < \frac{1}{2}L^s$.

Proposition 3 *(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 $w^G = p_2 x_2^G$. 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 $p_2 x_2^{AG} - w^G$.[12] Hence, a world of generalists is an equilibrium iff $p_2 x_2^{AG} - w^G \leq w^G$, or

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

In the "specialist" case, skilled workers produce in teams and earn a wage $w^{AB} = \frac{1}{2} p_2 x_2^{AB}$. If a player deviates to be a generalist, then he could either (a) work alone and earn w^G or (b) pair with an existing specialist and earn $p_2 x_2^{AG} - w^{AB}$. 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 $w^G \leq w^{AB}$, 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.

[12] 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.

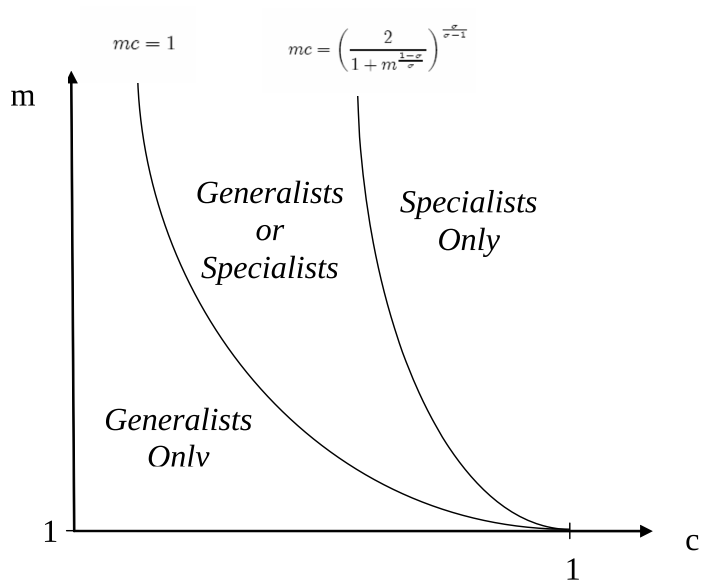

Figure 1: The Knowledge Trap

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. This model thus can produce multiple, pareto-ranked equilibria. Moreover, the ratio of income between generalist and specialist equilibria is potentially unbounded even where both are stable.

Corollary 4 *(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.

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.[13]

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 5 *(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

[13] In the limit, the model still features a "needle in a haystack" friction where, although search is extremely rapid ($\lambda \to \infty$) there are so many workers (a continuum) that one cannot expect to find a particular worker in finite time.

each sector are paid

$$
\begin{aligned}
w_1^n &= p_1^n z \\
w_2^n &= p_2^n 2^{\frac{1}{\sigma-1}} zh \times \begin{cases} 1, & n = G \\ mc, & n = AB \end{cases}
\end{aligned}
$$

Therefore, using the wage ratio, the price ratio on the supply side is determined as a function of exogenous parameters[14]

$$
\frac{p_1^n}{p_2^n} = 2^{\frac{1}{\sigma-1}} he^{-rS} \times \begin{cases} 1, & n = G \\ mc, & n = AB \end{cases} \tag{15}
$$

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

$$
\frac{L_1^n}{L_2^n} = \left(\frac{\gamma}{1-\gamma}\right)^{\varepsilon} \left(\frac{p_1^n}{p_2^n}\right)^{1-\varepsilon} e^{rS} \tag{16}
$$

where L_q^n is the measure of people actively working in sector q.[15]

Lastly, real income is also pinned down given relative prices

$$
y^n = w_1^n / p^n = z \left(\gamma^{\varepsilon} + (1-\gamma)^{\varepsilon} \left(\frac{p_1^n}{p_2^n}\right)^{\varepsilon - 1} \right)^{\frac{1}{\varepsilon-1}} \tag{17}
$$

where the aggregate price level, p^n, is $p^n = \left(\gamma^{\varepsilon} \left(p_1^n\right)^{1-\varepsilon} + (1-\gamma)^{\varepsilon} \left(p_2^n\right)^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}$.[16]

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

[15] 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 $L = L_1^n + e^{rs} L_2^n$.

[16] Real national income (Y^n) is given by $p^n Y^n = w_1^n L_1 + w_2^n L_2$, so that real per-capita income (y^n) is

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

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 $w_1^n = e^{-rs} w_2^n$.

4 Applications and Discussion

This section clarifies the model's implications and discusses several phenomena and stylized facts from the model's perspective. We begin with prices, wages, and income differences across countries, and then consider immigration, poverty traps, the role of multinationals in development, and extensions to international trade and growth.

4.1 Prices, Wages, and Income Differences Across Countries

To understand cross-country income differences, we begin with prices and wages. We then build from this foundation to discuss real output differences.

4.1.1 Relative Prices

A central observation in development is that certain goods are relatively cheap in poor countries (e.g. Harrod 1933, Balassa 1964, Samuelson 1964).[17] This observation motivates the need for PPP price corrections when comparing real income across countries. The knowledge trap model provides an endogenous mechanism to understand this price phenomenon, where low-skilled goods (e.g. haircuts) are relatively cheap in a poor country because low skill is relatively abundant there. In particular, the price ratio is given by (15), so that

$$\frac{p_1^{AB}/p_2^{AB}}{p_1^G/p_2^G} = mc$$

and the low-skilled good is mc times cheaper in the poor country.[18]

4.1.2 Wages

The flexibility of prices and labor supply meanwhile ensures that equilibrium wage gains from education are limited. Because people *choose* to be highly educated, excessive wage gains to the highly-educated can be arbitraged away by an increase in the supply of such workers. In the model's simple formulation, the cost of education is driven by foregone

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

[18] Relative price differences are often noted as a phenomenon of relatively cheap services (e.g. haircuts) in poor countries. One may also observe that investment goods appear relatively expensive compared to consumption goods in poor countries (Hsieh and Klenow, 2005). Testing whether skill endowments underly these cross-country price differences is an interesting empirical question left to future work.

wages during the training phase. This generates the log-linear "Mincerian" wage structure and pins the skilled wage premium to the interest rate, as in (14).[19]

More generally, with output prices and labor supply both adjusting, the wage returns to schooling no longer capture the underlying human capital in an economy, as discussed next.

4.1.3 Real Income Differences

Many authors have concluded that physical and human capital cannot explain the wealth and poverty of nations, leaving residual variation in "total factor productivity" as a major explanation (see Caselli 2005 for a review). The above model challenges this conclusion by reconsidering the role of human capital.

Standard human capital accounting uses Mincerian wage structures to count up human capital stocks by assuming that different skill classes produce perfect substitutes (Hall and Jones 1999; Bils and Klenow 2000). Under this strong assumption, the prices of any worker's output is the same and one can ignore output prices in inferring human capital. In this paper, we build instead from the viewpoint where skilled and unskilled workers are not perfectly substitutable (retail workers cannot perform heart surgery). Demands for the outputs of different skill classes are downward sloping so that output prices adjust when the quantity or quality of a skill class changes. Countries that are very good at producing high skill will find that goods produced by low-skill workers are scarce, which drives up low-skilled wages. In fact, with relative wages pinned down by the discount rate, as in (14), workers will allocate themselves between skilled and unskilled careers so that the percentage wage gains for skilled and unskilled workers rise or fall in equal proportion. Wages are Mincerian in each country, but the wage-schooling relationship shifts vertically depending on the skilled equilibrium. This is shown in Figure 2 for the case of Cobb-Douglas preferences.[20]

[19] Note that this simple perspective suggests a positive correlation between interest rates and returns to schooling across countries. In fact, the literature finds 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). Also see footnote 5.

[20] An interesting feature of the model is that a country's average educational attainment need not even be positively associated with average income. For example, with Cobb-Douglas preferences ($\varepsilon = 1$) the average schooling in a population is

$$s^n = S\frac{L_2^n}{L} = (1-\gamma)Se^{-rS}$$

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

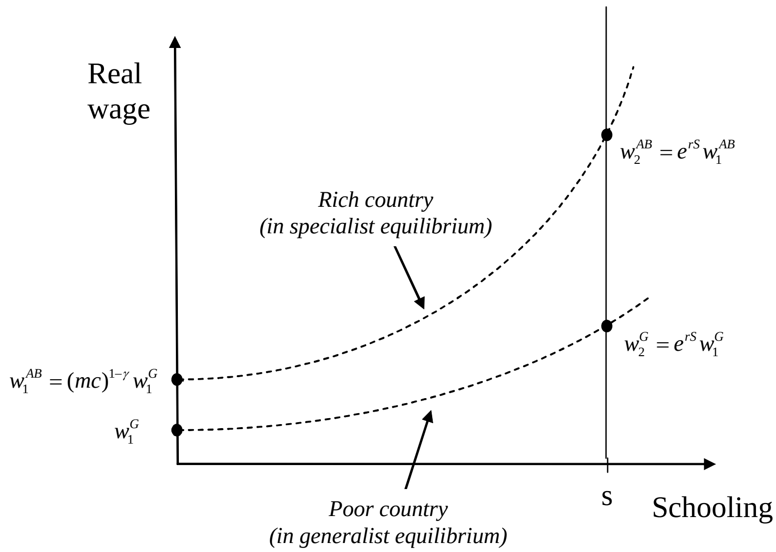

Figure 2: Equilibrium Wage-Schooling Relationships

To properly incorporate the effects of skill in a cross-country analysis, one must consider the elasticity of substitution between skilled and unskilled labor. Consensus estimates suggest an elasticity between 1 and 2 (not infinity, as the perfect substitutes approach assumes). Observing this fact, Caselli and Coleman (2006) reconsider development accounting in a setting where labor scarcity affects wages. They estimate separate productivity terms for skilled (s) and unskilled (u) workers using 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 (10) in this paper with the addition of physical capital, k.[21] With this estimation strategy, they find that the productivity advantage of rich countries is limited to skilled workers. In particular, they find an enormous productivity advantage of skilled workers (A_s) in rich countries while the productivity of unskilled workers (A_u) is no higher there.[22] The "knowledge trap" model suggests exactly this effect. The theory shows how

greater than 1. Then countries with high quality skilled-labor (i.e. specialization) will see an endogenous increase in the supply of such skilled workers.

[21] They calculate L_u and L_s by aggregating workers within lower schooling ranges (L_u) and upper schooling ranges (L_s) with perfect substitutes assumed within each range.

[22] In their prefered specification, Caselli and Coleman further argue that unskilled workers are actually *less* productive in rich than poor countries. An explanation for this phenomenon may be selection on ability within labor markets (see Footnote 26 below). More generally, their result for unskilled workers is sensitive to the calibration parameters and how one defines "unskilled worker". If one classifies such workers as having less than high school or less than college-level education, then unskilled workers in their calibration become mildly more productive in rich countries. What appears highly robust about their specification is that skilled workers have enormous productivity advantages in rich countries.

human capital differences can produce large skilled labor productivity differences and hence provides a human-capital based interpretation of cross-country income differences.[23]

4.2 Immigrant Wages and Occupations

An alternative approach to assessing human capital's role in cross-country income differences is to examine what happens when workers trained in poor countries are placed in rich countries. If human capital differences were critical, authors argue, then such workers should experience significant wage penalties in the rich country's economy. Noting that immigrants from poor to rich countries earn wages broadly similar to workers in the rich country, authors have concluded that human capital plays at most a modest role in explaining productivity differences across countries (Hendricks 2002). However, this estimation approach as implemented falls into the same pitfall of standard accounting approaches, by limiting the effect of scarce labor supply.[24]

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. Indeed, why would education matter for the uneducated, working as taxi drivers, retail workers, and farm hands? Wage gains follow naturally when the low-skilled immigrant moves to a place where his labor type is relatively scarce. The over-riding role of scarcity, rather than productivity, for unskilled workers is corroborated by the calibration discussed above. The potentially more informative implications of the knowledge trap model lie among skilled immigrants.

[23] Another important calibration is Manuelli and Sheshadri (2005), who estimate human capital by considering it as an endogenous choice variable. They find large quality differences in human capital across countries that, once accounted for, require little or no TFP differences. Their estimation suggests large advantages in the quality of education in rich countries even at entrance to primary school. This skill advantage at very low-education levels differs from the "knowledge trap" approach, which emphasizes difference that are limited to the highly skilled and differs from the Caselli and Coleman calibration, where skill differences exist only among those with more education. Manuelli and Sheshadri's imputed quality differences at all skill levels appear to follow because their model does not allow for the relative price effects that occur when skilled and unskilled workers produce different intermediate or final goods.

[24] 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, which is far above the consensus estimates in the literature that range between 1 and 2. The calibration of Caselli and Coleman (2006) (see discussion above) uses the consensus range of elasticities, thus capturing the effects of scarce labor supply and finding no general TFP advantages in rich countries.

Corollary 6 *(Immigrant Workers) 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. ■

Skilled immigrants, as generalists, are unable to find local specialists willing to team with them. Moreover, they 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 immigrant generalist than using his education. Hence, for example, we can see immigrant Ph.D.'s who drive taxis.

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.[25] 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.

Figure 3a shows two important facts. First, 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 more. Meanwhile there is no wage penalty if the immigrant arrived early enough to receive higher education in the United States. Second – and conversely – the location of high-school education does not matter. 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 the "knowledge trap" suggests.[26]

[25] The data and methods are detailed further in the Appendix.

[26] 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

Figure 3b considers related evidence based on occupation type. To construct this graph, each occupation in the census is first categorized by the modal level of educational attainment for workers in that occupation. For example, taxi drivers typically have high school degrees, physicians typically have professional degrees, and physicists typically have Ph.D.'s. The figure shows the propensity for workers with professional or doctoral degrees to work in different occupations. 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 3a, which shows that late immigrants with professional or Ph.D. degrees earn average wages no better than a locally educated college graduate.

This evidence is consistent with the "knowledge trap" model but inconsistent with a pure technology story, in which the location of education would not matter. More broadly, the evidence is consistent with the idea that human capital differences across countries exist primarily among the highly educated, as suggested independently by the calibration of Caselli & Coleman (2006) discussed above.

4.3 Poverty Traps

Unlike poverty trap models that envision aggregate demand externalities, such as big push models (e.g. Murphy et al., 1989), knowledge traps can be overcome locally, when workers achieve greater collective skill.[27] 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, and I consider here challenges to collective skill improvement from the perspective of the model.

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.

[27] 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.

4.3.1 The Quality of Higher Education

Income differences across countries may persist if countries are in different regions of Figure 1. Countries with $mc < 1$ will 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 the rich country. The model thus suggests a "brain drain" phenomenon.

Corollary 7 *(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.[28] 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.[29]

[28] More generally, the fact that skilled workers do not substantially flow from rich to poor countries seems self-evident, and potentially mysterious since skilled workers are scarce in poor countries. A natural explanation is that some complementary inputs are missing. The "knowledge trap" viewpoint suggests that other skilled workers are an important complementary input, although of course one can imagine other "missing" inputs that are complementary with skilled labor.

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

This result suggests that wage subsidies or other incentives may be required to attract skilled experts to the poor country and improve local training.

4.3.2 The Coordination of Higher Education

Even if poor countries can 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. This may be hard. Any intervention must convince initial cohorts of students to spend long years in irreversible investments as specialists, which would be irrational if complementary specialists were not expected. Hence we need a "local push".[30] Yet it is not obvious what institutions have 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.[31] 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.

4.3.3 Multinationals and Poverty Traps

Intra-firm trade can allow for production teams that span national borders, and I discuss here how a multinational can play a unique role in helping countries escape poverty.

[30] 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 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.

[31] 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 Ph.D.'s are produced in educational institutions, not in firms.

Corollary 8 *(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.

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 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.[32] Here, the "visionary" leader need not recreate Microsoft but simply produce a sufficient quantity of one specialist type that Microsoft will hire. To the extent that a thick supply of this specialist type triggers complementary specialization locally, the local economy may escape from the trap broadly.[33]

4.4 Generalizations: International Trade and Growth

Knowledge traps may also 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 so-called "Leontieff Paradox" (Leontieff 1953, Maskus 1985, Bowen et al. 1987, etc.).

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

[33] With only two types of specialists, the emergence of one type in the poor country triggers the emergence of the other, and the poor country will become rich. With more than two specialist types, 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 offhsoring and persistently "cheap engineers".

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.[34] Yet these comparative advantages - based in specialization - won't appear in standard labor classifications.[35] With knowledge traps, rich countries are net exporters of skilled goods not simply because they have more skilled workers, but because their skilled workers have so much more skill.[36]

Finally, knowledge traps may help explain divergence over time in cross-country per-capita income levels, thus confronting the typical empirical regularity in growth (Jones 1997, Pritchett 1997). To introduce growth in the model, one can let m increase with time as advanced economies accumulate ideas.[37] In fact, advanced economies appear to accumulate an enormous amount of codified knowledge, which is associated with increasing specialization and team-orientation among knowledge workers (Jones 2005, Wuchty et al. 2007). As one measure, the ISI Web of Science indexes 1.3 million science and engineering journal articles published by US knowledge workers from 2001 to 2005. Divergence follows naturally to the extent that such new knowledge is economically important and yet workers in poor countries, for any of the reasons detailed above, do not learn these specialized ideas.

5 Conclusion

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

[34] 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 equilibrium exports the high skilled good (2).

[35] 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).

[36] Several studies find that augmenting factor-endowment differences with technology differences can produce more successful models empirically (Trefler 1993 and 1995, Harrigan 1997). These results, like the cross-country income literature, leave unexplained technology differences as key explanatory forces. In the knowledge trap model, productivity differences are seen as the result of different, endogenous endowments of skill.

[37] In an endogenous growth framework, some fraction of skilled workers would produce productivity enhancing ideas that lead to growth in m. An accumulation of knowledge in an economy 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 (2005) for such a growth model.

shows how large differences in the quality of skilled labor may (a) persist across economies yet (b) not appear in the wage structure. Together, these ideas show how standard human capital accounting may severely underestimate cross-country skill differences. Building from a simple conception of human capital, the model provides an integrated perspective on cross-country income differences, poverty traps, relative wages, price differences, trade patterns, migrant behavior, and other phenomena in a way that appears broadly consistent with important facts. Future work will test the model's implications against other hypotheses on each of these dimensions.

This paper speaks directly to a long-running debate over the roles of "human capital" and "technology" in explaining income differences across countries. The model is centered on human capital, but because it directly embraces "knowledge", it also comes close to some conceptions of "technology". I close by further considering these distinctions.

In this paper, education is conceived of as the acquisition of knowledge: workers are born with no knowledge, and education is the process of embodying existing knowledge (techniques, methods, facts, theories, blueprints) into empty minds. Rich countries are more productive because they load deeper knowledge into these minds than poor countries do. If we equate knowledge with "technology", then human capital and technology appear tightly related. Human capital is the embodiment of technology into the labor force, much as a microprocessor is the embodiment of technology into silicon. It is thus the emphasis on embodiment, rather than the role of "ideas", that distinguishes this paper from other approaches.

The focus on human capital does not suggest that technology is not an important, distinct concept. Technology can be conceived as the set of discovered techniques, methods, facts, models, et cetera that limits what can be embodied in minds or machines. At the frontier of the world economy, technological progress, the expansion of this set, may still drive economic development, but even here the effects of knowledge will likely be felt – and understood – not through a disembodied process but rather through the embodiment of these ideas into the people and machines that actually produce things.

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 (7) implies $V^{AA} = V^{BB} = \frac{1}{2}V < V$. Therefore, neither As or Bs will match with their own type. For Gs, (7) implies $V^{GG} = \frac{1}{2}w^{GG}/r + \frac{1}{2}V$. Noting that $V \geq w^G/r$ (G's income if he never matches, from (5)) and that $w^{GG} < w^G$ (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 (7) because with (a) common V and (b) $x_2^{AG} = x_2^{BG}$, Nash Bargaining implies $w^{AG} = w^{BG}$. 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 $w^{AG} = \frac{1}{2}p_2 x_2^{AG}$. But $w^{AB} = \frac{1}{2}p_2 x_2^{AB} \geq \frac{1}{2}p_2 x_2^{AG} = w^{AG}$ and hence, from (7), $V^{AB} \geq V^{AG}$. 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.[38]

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

[38] 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 $0 = [\Pr(A) - \Pr(B)]\left[w^{AB} + \frac{\lambda}{2r}\Pr(G)\left(w^{AB} - w^{AG}\right)\right]$. 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.[39] 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, $w^{BA} \geq \frac{1}{2}p_2x_2^{AB}$. 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 $\{L^A, L^B, L^G\} = \{0, 0, L^s\}$.

In this case,

$$rV = w^G$$

where $w^G = p_2x_2^G$.

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

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

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 $\{L^A, L^B, L^G\} = \left\{\frac{1}{2}L^s, \frac{1}{2}L^s, 0\right\}$.

In this case,

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

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

[39] 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 $rV^G = w^G = p_2 x_2^G$. If you do match with specialists, then $rV^G = (w^G + \frac{\lambda}{2r} w^{GA}) / \left(1 + \frac{\lambda}{2r}\right)$, where $w^{GA} = \frac{1}{2} p_2 x_2^{AG} - \frac{1}{2} r(V - V^G)$ 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 $\left\{L^A, L^B, L^G\right\} = \{L', L', L^s - 2L'\}$. There are two sub-cases: (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

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

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

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

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

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

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 (9), 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 (4) 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 (Immigrant Workers)

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 $mc + c$.[40] 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 $\frac{w^{AB}/p_1^{AB}}{w^G/p_1^G}$ (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 $w' = p_2^G(x_2^{AG} - x_2^G)$. Since the poor country is in a generalist equilibrium, we must have $x_2^{AG} \leq 2x_2^G$ which implies that $w' \leq p_2^G x_2^G = w^G$. 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

[40] 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 $\sigma \to \infty$.

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. ■

Proof of Corollary (Desirable Cheap Specialists)

Proof. Think of the firm as a specialist in the rich country. He earns $w^{AB} = \frac{1}{2}p_2^{AB}x_2^{AB}$. 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 $w = p_2^{AB}x_2^{AB} - p_2^{AB}x_2^G$, 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 3

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.[41] There are 2.8 million individuals in this sample, including 320 thousand individuals who immigrated to the United States.

The wage-schooling relationships in Figure 3a 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 3a 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.

To construct Figure 3b, the modal educational attainment is first determined for each of the 511 occupational classes in the data (using the IPUMS variable "occ"). Occupations are

[41] 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.)

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 3b shows the propensity of individuals with professional or doctoral degrees to work in occupations with the given modal educational attainment.

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.

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

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

[7] 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.

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

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

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

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

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

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

[14] 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.

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

[16] 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.

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

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

[19] 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.

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

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

[22] Jones, Benjamin F. "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" NBER Working Paper #11360, 2005.

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

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

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

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

[27] Manuelli, Rodolfo and Ananth Seshadri, "Human Capital and the Wealth of Nations", mimeo, University of Wisconsin, June 2005.

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

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

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

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

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

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

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

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

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

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

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

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

Figure 3a: Do Skilled Immigrants Experience Wage Penalties?
The Wage-Schooling Relationship

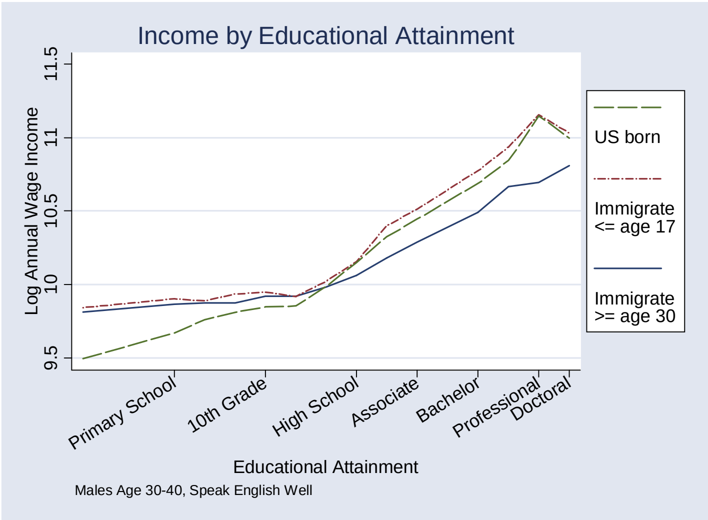

Figure 3b: Do Skilled Immigrants use their Education?
Typical Educational Level of Occupation for Workers with Professional or Doctoral Degrees