Education and Knowledge Spillovers*

Paul Niehaus[†]
UC San Diego

This draft: July 12, 2012

Abstract

Educating a worker may generate knowledge spillovers by increasing what others can learn from him or by increasing what he can learn from others. This paper provides a theoretical framework for examining which mechanism is most consistent with observed education policies, common modelling assumptions, and the empirical evidence. Both views play a role in understanding public education; the second may help explain sorting, agglomeration, and evidence of skill-skill complementarity.

*I thank Julie Cullen, Vince Crawford, Roger Gordon, Karthik Muralidharan, Jim Rauch, Joel Sobel, participants at SWET 2011, and particularly Nageeb Ali for helpful suggestions. Ellen Bruno and Myungkyu Shim provided excellent research assistance. All errors are my own.

[†]Department of Economics, University of California at San Diego, 9500 Gillman Drive #0508, San Diego, CA 92093-0508. pniehaus@ucsd.edu.

1 Introduction

Educating a worker, it is often argued, generates knowledge spillovers that benefit others. This idea has affected thinking about growth, trade, and agglomeration, and appears regularly in discussions of education policy.[1] Most models of spillovers are reduced-form, however, and seem open to (at least) two interpretations.

The simplest, "viral" view is that the skills a worker acquires during their education spill over to their peers. According to this view education has positive externalities because it directly increases *what others can learn from you*.[2]

Education research suggests an alternative view in which "learning to learn" plays a central role. In Dewey's influential vision of education, for example, learned concepts were "tools by which the individual pushes most surely and widely into unexplored areas." (1902, p. 31) More recently Heckman has argued forcefully that "human capital has fundamental dynamic complementarity features. Learning begets learning. Skills acquired early on make later learning easier" and that "our economic models have to be modified to account for this" (2000, p. 8, 6). The direct benefits of learning such "foundational" skills are of course private: an increase in what one subsequently learns. Indirectly, however, this effect again increases what others can learn from you. Thus education may have positive externalities for the somewhat less obvious reason that it increases *what you can learn from others*.[3]

In practice many skills have some of both of these attributes, and it is unclear from earlier work whether or how the distinction matters. Do the aggregate implications of knowledge spillovers depend on which mechanism is operative? If so, which view is more consistent with observed education policies (e.g. universal primary education), with commonly-used reduced forms, or with the available empirical evidence? This paper addresses these questions, using simple, tractable models of the viral and foundational views to examine their theoretical implications for optimal policy, wage and production functions, and spatial equilibrium.

I study an overlapping-generations labor market in which agents can acquire discrete skills either in a competitive education sector or through spillovers. Spillovers take place in random, unpriced interactions between workers: these capture forces like observational learning or knowledge-sharing in team production (Appendix B

[1]See, for example, Heckman and Klenow (1998); McMahon (2004); Lange and Topel (2006).

[2]This view is arguably implicit in influential earlier writings such as Jacobs (1969) and Lucas (1988). More recently, Lucas (2009), Lucas and Moll (2011), and Perla and Tonetti (2011) have studied models of growth based on viral diffusion, though without modeling formal education.

[3]Along these lines, Nelson and Phelps (1966) famously argued that developing countries needed education in order to adopt advanced technologies. See also Bils and Klenow (2000) and Benhabib and Spiegel (2005).

microfounds the latter interpretation). In order to clearly expose intrinsic differences between the viral and foundational views I abstract from any other education externalities; the model should thus be seen as complementary to analyses of externalities in the education production function itself (e.g. Bils and Klenow (2000)) or of other external effects of education (see Moretti (2004b)). The economy loses skills as old workers retire but also gains skills as young workers get educated or learn through diffusion; in steady-state these forces balance.

Section 2 examines optimal policy. Under the viral skills view selective rather than universal subsidies are optimal, for the simple reason that the external returns to teaching viral skills vanish when everyone knows them. The optimal policy with respect to foundational skills, on the other hand, will often involve universal education. This is because foundational skills generate increasing returns: as more people learn them, viral skills circulate more widely, increasing the benefit of having the foundational skill. In some cases the state need only ensure coordination on the universal-education equilibrium, while in others universal education is an optimum but not an equilibrium, so that sustained intervention is necessary. Together the two views may thus help explain why states subsizide both selective training in some skills (e.g. agricultural extension via "contact farmers") and universal training in others (e.g. universal primary education).

The relative importance of these two functions changes with the rate of diffusion. The external returns to viral skills initially rise with the diffusion rate but must ultimately decline and vanish, as these skills become readily available to everyone. Coordinating investment in foundational skills, on the other hand, becomes increasingly important. The two views of spillovers thus yield qualitatively different perspectives on how public education will evolve as communication technologies improve. The tradeoff between providing skills that can diffuse and promoting their diffusion also bears on the debate over general v.s. vocational curricula, identified with the "American" and "European" systems. The spillovers perspective complements previous work, which has emphasized differences in mean earnings and in flexibility.[4]

The differential policy implications of the viral and foundational views suggest a real need for empirical work separating them. One approach to identifying foundational skills is simply to test for cross-skill spillovers directly. For example, in their study of teacher incentives in Indian primary schools Muralidharan and Sundararaman (2011) find that "positive spillovers from improvements in math and especially language led to improved scores in nonincentive subjects as well." These results bear

[4]See, among others, Psacharopoulos (1987), Glazer (1993), Bennell (1996), Goldin and Katz (2003), and Malamud and Pop-Eleches (2010) on the former point and Goldin (2001) on the latter.

out the intuitive view that language skills are foundational.

A second and more difficult question is whether the *bundle* of skills that make up public education displays either viral or foundational characteristics. Section 3 characterizes methods for addressing this question via conditional wage functions. It uses the theory to microfound empirical specifications, introduced by Rauch (1993), which augment standard Mincer models with peer characteristics such as average education. Both the viral and foundational views imply that these characteristics predict own earnings after conditioning on own education, and importantly the effects are characteristically non-linear, with the viral (foundational) view yielding negative (positive) interactions between own education and own experience, peer education, and peer experience. Which of these is the case remains an open question as the literature has not yet separated interactions due to spillovers from those due to technological complementarity and substitutability (Moretti, 2004a; Ciccone and Peri, 2006). Evidence of positive interactions can be taken as support for the foundational view, however, since neoclassical diminishing returns would tend to bias these negatively. For example, Acemoglu and Angrist (2000) find some weak evidence of skill-skill complementary in U.S. labor markets.

Finally, Section 4 examines sorting and agglomeration. While both the viral and the foundational views predict peer effects on earnings, only the foundational view can explain sorting and agglomeration per se. The reason is that while both views predict that workers prefer to locate near educated peers, only the the foundational view predicts that *educated* workers value this opportunity most – and may value it enough to be willing to pay congestion costs for it. Several aspects of this result are of interest. First, it holds without any assumption that spillover rates rise with density. The point is not that density effects are implausible – indeed, they have empirical support (Ciccone and Hall, 1996) – but rather that foundational skills generate agglomeration through a mechanism that is qualitatively different from those previously studied (Glaeser, 1999; Peri, 2002; Berliant et al., 2006). In this mechanism agglomeration is intrinsically linked with sorting.[5] Second, because it explains why local learning processes are complementary the model provides a justification for reduced-form complementarity assumptions such as those in Benabou's (1993) account of the formation of ghettos or Black and Henderson's (1999) model of urban production. Third, it implies that evidence of higher returns to education in cities (Glaeser and Mare, 2001) and increased sorting of highly-educated workers into highly-educated cities (Berry and Glaeser, 2005) indirectly support the foundational view.

[5]See Duranton and Puga (2004) for an overview of agglomeration mechanisms.

Section 5 summarizes and discusses extensions. In particular, I discuss how the theoretical framework – which is intentionally kept stark here in order to build intuition – could be enriched for quantitative analysis.

2 Knowledge Spillovers and Education Policy

Consider a continuous-time economy, $t \geq 0$ populated by a measure 1 of workers. Old workers retire at random at rate λ and are replaced by new workers. These new workers are unskilled at birth but can acquire skills in two ways. First, there is a freely available, constant-returns formal education technology; one can interpret this as skilled agents teaching unskilled ones or as unskilled ones purchasing and learning from books, training manuals, etc. As a consequence of free entry the equilibrium price of formal education is equal to the marginal cost of providing it (specified below).

Second, agents can acquire skills through informal bilateral interactions with their peers, which take place at rate $\phi > 0$. These interactions represent the pairing of co-workers through the labor market as well as social interactions outside of work; ϕ thus quantifies factors such as the rate of churn in the labor market and the breadth of agents' social networks. In spatial applications agglomeration density might be an important determinant of ϕ.

When agents with different skills interact there is some chance that skills spill over. I take it as definitional that, unlike knowledge transfers in the formal education sector, spillovers are uncompensated. As Lucas (1988) famously argued,

> "Most of what we know we learn from other people. We pay tuition to a few of these teachers, either directly or indirectly by accepting lower pay so we can hang around them, but most of it we get for free, and often in ways that are mutual without a distinction between student and teacher" (p. 38).

Lucas emphasizes that while mechanisms for internalizing spillovers do play a role – for example, accepting lower pay in order to hang around with a skilled teacher or co-worker – they are incomplete. The focus here will be on uninternalized spillovers that drive a wedge between the private and social returns to education.

If spillovers are not transactional then it is an open question how to model their likelihood. The standard approach in the literature has been to treat spillovers as exogenous events (Jovanovic and Rob, 1989; Glaeser, 1999; Peri, 2002; Berliant et al., 2006). One justification is that many spillovers are truly accidental, as for example when one worker happens to observe another worker performing some unfamiliar task. A drawback of this approach, however, is that it imposes no structure on the probability of a skill spilling over, a key primitive. As a compromise between brevity and rigor I specify exogenous spillover probabilities in the main text, but provide micro-foundations in Appendix B. The gist of these foundations is that agents interact for the purpose of team production; because each agent's payoff in these interactions depends in part on his partner's productivity he has an incentive to teach them skills. The strength of this incentive is governed by the technological and contractual features of team production; consequently the model yields natural comparative static predictions such as a positive relationship between the value of a skill and the likelihood it spills over.

The analysis proceeds in three steps: Section 2.1 examines the viral skills case, Section 2.2 adds foundational skills, and Section 2.3 summarizes implications for education policy.

2.1 Viral Knowledge

First consider the diffusion of a single viral skill. The implications generalize immediately to multiple independent skills, while interactions between skills will play an important role when foundational skills are introduced below. The cost of acquiring the skill through formal education is $c > 0$ for any agent. Allowing for heterogenous costs does not affect the substantive conclusions in this section. (It does have meaningful implications for the foundational view, which are discussed below.) The probability that the skill spills over when a skilled agent interacts with an unskilled agent is $\rho \in (0, 1]$.

Skilled agents earn a wage premium $\pi > 0$ per unit of time relative to unskilled ones, whose wage is normalized to 0. More generally the return to skill might be a function $\pi(s)$ of the fraction of workers who are skilled, with $\pi' < 0$ due to technological diminishing returns. I abstract from technology entirely here in order to highlight the scale properties intrinsic to diffusion processes, but will re-introduce it when discussing empirical applications in Section 3.

In a steady state a constant fraction $e \in [0, 1]$ of new agents get educated and a constant fraction $s \in [0, 1]$ of agents are skilled at any time t. When s is constant the continuation payoffs for skilled and unskilled workers are also time-invariant and so each agent purchases the skill in the formal education sector either at the beginning of her life or not at all.[6] For (s, e) to be a steady state the rate at which skilled agents

[6]This is strictly optimal for agents who derive positive expected surplus from purchasing the skill and weakly optimal for marginal, indifferent agents. If agents were finite-lived then of course all agents'

leave the economy must equal the rate at which new skilled workers enter plus the rate at which previously unskilled workers become skilled, or

$$\lambda s = \lambda e + \phi\rho s(1-s) \tag{1}$$

This equation has a unique stable positive root

$$s(e) \equiv \frac{-(\lambda-\phi\rho)+\sqrt{(\lambda-\phi\rho)^2+4\lambda\phi\rho e}}{2\phi\rho} \tag{2}$$

satisfying $s'(e) > 0$ and $s''(e) < 0$.[7] Note that $s(0) \geq 0$ as $\phi\rho \gtreqless \lambda$: if the diffusion rate is higher than the turnover rate then knowledge of the skill is self-sustaining in the population even without any formal education. The shorthand $q \equiv \frac{\lambda}{\phi\rho}$ will be useful for keeping track this property.

The continuation payoffs for skilled and unskilled workers as functions of the steady-state education level e are characterized by Hamilton-Jacobi-Bellman equations

$$\begin{aligned}
\lambda V(1,e)dt &= \pi dt \\
\lambda V(0,e)dt &= \phi\rho s(e)(V(1,e)-V(0,e))dt
\end{aligned}$$

with solutions

$$V(1,e) = \frac{\pi}{\lambda} \tag{3}$$

$$V(0,e) = \frac{\pi}{\lambda}\left(\frac{\phi\rho s(e)}{\lambda+\phi\rho s(e)}\right) \tag{4}$$

Note that $\frac{\phi\rho s(e)}{\lambda+\phi\rho s(e)}$ is the probability that an uneducated worker is now skilled, and so $\pi\frac{\phi\rho s(e)}{\lambda+\phi\rho s(e)}$ is the average wage of an uneducated worker. The private rate of return on education is $\Delta(e) \equiv V(1,e) - V(0,e) = \pi/(\lambda+\phi\rho s(e))$, which is decreasing in the spillover rate since spillovers provide a substitute learning mechanism. It is incentive-compatible for agents to pay for formal education if and only if the private return exceeds the cost c.

Definition 1. *A decentralized equilibrium is an* $\hat{e}$ *such* $\hat{e} = 1$ *and* $\Delta(\hat{e}) > c$, $\hat{e} = 0$ *and* $\Delta(\hat{e}) < c$, *or* $\hat{e} \in (0,1)$ *and* $\Delta(\hat{e}) = c$.

A social planner, in contrast, maximizes the welfare of a representative generation.

preferences over the timing of education would be strict.

[7] If $e = 0$ and $\phi\rho > \lambda$ then $s = 0$ is an additional root but is unstable.

Definition 2. *A centralized optimum is a solution to* $\max_e eV(1,e)+(1-e)V(0,e)-ec$

Since steady-state output is

$$\begin{aligned}\pi s(e) &= e\pi + (1-e)\pi\left(\frac{s(e)-e}{1-e}\right)\\ &= e\pi + (1-e)\pi\left(\frac{\phi\rho s(e)}{\lambda+\phi\rho s(e)}\right)\\ &= \lambda\left[eV(1,e)+(1-e)V(0,e)\right]\end{aligned}$$

this is equivalent to maximizing steady-state output net of education investment costs in the special case where $\lambda = 1$.[8] The marginal social return to education is

$$\underbrace{\Delta(e)}_{\text{Private Return}} + \underbrace{\Gamma(e)}_{\text{External Return}} - \underbrace{c}_{\text{Marginal Cost}}$$

where

$$\begin{aligned}\Gamma(e) &\equiv e\frac{\partial V(1,e)}{\partial e} + (1-e)\frac{\partial V(0,e)}{\partial e}\\ &= (1-e)\pi\left[\frac{\phi\rho}{(\lambda+\phi\rho s(e))^2}\right]s'(e)\end{aligned} \tag{5}$$

is the positive external return to education: unskilled agents stand a better chance of becoming skilled in a more skilled population.[9]

Examining their respective expressions (and noting that $s''(e) < 0$) it is apparant that the private and social returns to education are both decreasing. Decreasing returns are due to saturation effects. As more agents become skilled, any given agent will have more opportunities to learn the skill from a peer, which lowers his private rate of return to getting educated. At the same time he will have fewer opportunities to teach the skill to others, which lowers the external social value of educating him. These forces guarantee a unique decentralized equilibrium.[10]

Proposition 1. *There is a unique equilibrium level of education* $\hat{e}$ *satisfying*

1. $\hat{e} = 0$ *if* $c \geq \frac{\pi}{\max\{\lambda,\phi\rho\}}$

[8]As with all OLG models there is ambiguity as to what welfare weights the planner should assign to different generations. Since in a steady-state all generations attain the same utility level, however, Definition 2 is without further loss of generality.

[9]Readers familiar with Ciccone and Peri's (2006) reduced-form accounting framework for external returns to human capital will recognize this as an instance of their Equation 1.

[10]This contrasts the model's transmission-driven externality from models with a conformity motive such as Glaeser et al. (1996) in which strong enough social interactions typically generate multiple equilibria.

2. $\hat{e} = 1$ *if* $c \leq \frac{\pi}{\lambda + \phi\rho}$

3. $\hat{e} \in (0, 1)$ *and* $\Delta(\hat{e}) = c$ *otherwise*

Proof. See Appendix A for this and subsequent proofs. □

and a unique optimal level of education

Proposition 2. *There is a unique optimal level of education* e^* *satisfying*

1. $e^* = 0$ *if* $c \geq \frac{\pi}{|\lambda - \phi\rho|}$

2. $e^* = 1$ *if* $c \leq \frac{\pi}{\lambda + \phi\rho}$

3. $e^* \in (0, 1)$ *and* $\Delta(e^*) + \Gamma(e^*) = c$ *otherwise*

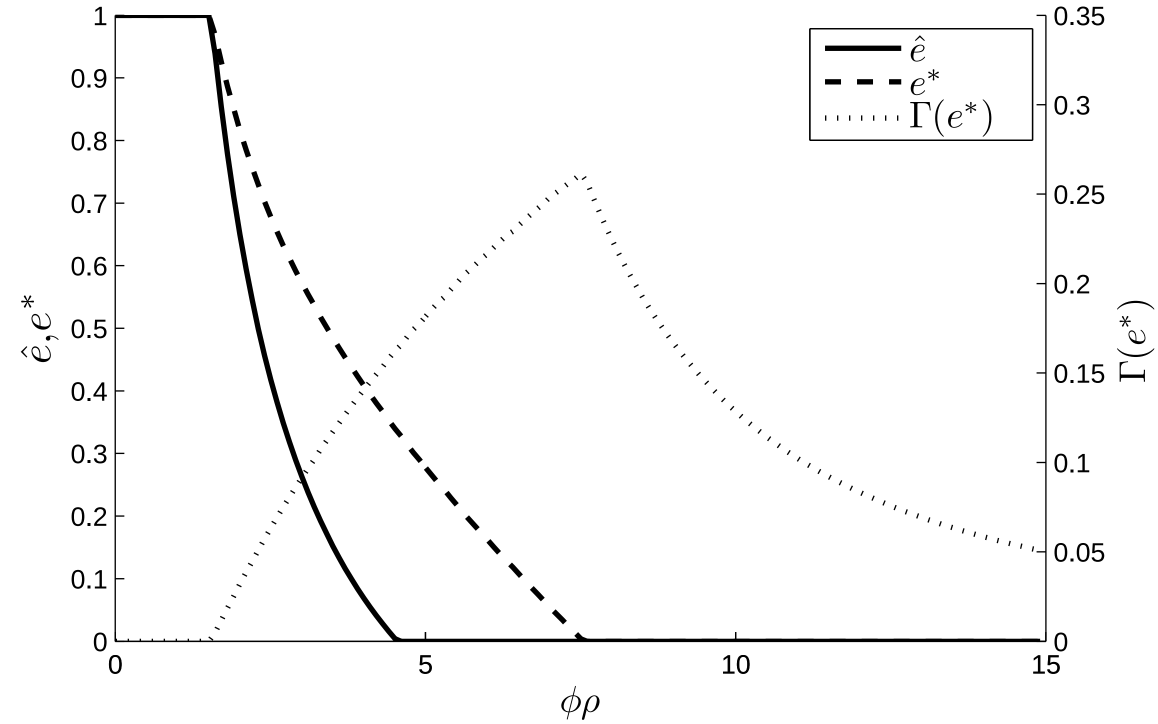
The optimal level of education is typically greater than the equilibrium level because of the uninternalized spillover benefits of education. Notice, however, that the conditions under which universal education is optimal coincide with those under which it is the equilibrium outcome. In other words, universal education is optimal if and only if it is the laissez-faire equilibrium. This is a consequence of saturation: when everyone is educated there is no scope for spillovers and so the private and public returns to education coincide. Notice also that universal education is less likely to be optimal the higher is the diffusion rate $\phi\rho$ – in other words, spillovers *weaken* the case for universal education. This captures the idea that in the presence of spillovers one can educate a subset of the population and let their skills diffuse to the rest.

One further insight about the wedge between private and social returns can be obtained from examining Equation 5 above. Notice that the external returns to education vanish both as $\phi\rho \to 0$ and also as $\phi\rho \to \infty$. The first point is obvious: externalities are due to spillovers, so without spillovers there are no externalities. The second is more subtle. To see why externalities also vanish as the rate of diffusion becomes very large, imagine a limiting economy in which all knowledge is pooled, so that knowledge becomes available to everyone as soon as a single agent acquires it. This "folk knowledge" will circulate widely and in perpetuity even if no one learns it in school ($e = 0$). Given this, a planner would never pay anyone to acquire it. The following proposition summarizes these points:

Proposition 3. *The level of under-education* $e^* - \hat{e}$ *and the optimal Pigouvian subsidy* $\Gamma(e^*)$ *both approach 0 as either* $\phi\rho \to 0$ *or* $\phi\rho \to +\infty$ *(provided in the former case that there is a unique equilibrium at* $\phi\rho = 0$*).*

Figure 1 illustrates an example. The equilibrium and optimum levels of education coincide at 1 when the spillover rate is sufficiently low. For intermediate spillover

Figure 1: Equilibrium, Optimum, and Externalities by Spillover Rate



Plots the equilibrium level $\hat{e}$ and optimum level e^* of education (left-hand axis) and the external returns to education at the optimum $\Gamma(e^*)$ (right-hand axis) as functions of the spillover rate $\phi\rho$. Other parameters are fixed at $\pi = \lambda = 3$, $c = 2/3$.

rates the optimum and equilibrium are interior and the optimal level of education strictly exceeds the equilibrium level. Finally, for sufficiently high spillover rates the optimum and equilibrium again coincide at 0. Note that $e^* = 0$ for finite values of $\phi\rho$; this is because the social returns to educating the first worker approach 0, and thus eventually fall below c, as saturation increases.

2.2 Foundational Knowledge

In the viral skills view interactions between skills are not important, in the sense that the costs and benefits of acquiring a new skill do not depend on what one already knows. It is often argued, however, that skills build on each other and that a good deal of formal education is really "learning to learn". For example, arithmetic is a foundational skill that children must master before they can progress to more advanced mathematical skills as well as to various applied ones such as calculating compound interest or estimating fertilizer requirements for a farm plot. Heckman and Carneiro (2003) argue that this is a general feature of human capital accumulation: "learning begets learning, skills (both cognitive and non-cognitive) acquired early on facilitate later learning." Their view is consistent with stylized facts from the technology adoption literature: better-educated individuals are almost always earlier adopters and countries with more tertiary education appear to benefit more from foreign R&D.[11] It is also consistent with daily experience: people with similar education and training usually find communication easier and more productive.

If this is correct, formal education may generate spillovers less because the skills taught in school diffuse than because they *catalyze* the diffusion of other, viral skills. To formalize this idea consider two skills, one foundational and one viral. The results will again generalize immediately to the case of many viral skills. The foundational skill does not itself spill over but is prerequisite for learning or using the viral one. The viral skill spills over with probability ρ as before, except that only agents who already know the foundational one can acquire it. This sharp dichotomy is clearly exaggerated, as in reality many skills have both "viral" and "foundational" features. It serves a purely analytic function here, juxtaposing the viral and foundational views as sharply as possible. For quantitative applications one could of course allow for a more continuous distinction.[12]

Agents now choose at birth whether to invest in both skills, only the foundational skill, or neither. Let e_{11} be the proportion that invest in both, e_{10} the proportion investing in the foundational skill, and so on. The vector $e = (e_{11}, e_{10}, e_{00})$ represents the distribution of initial education levels while $s = (s_{11}, s_{10}, s_{00})$ represents the

[11] See Schultz (1975), Rogers (1983), Skinner and Staiger (2005) and Coe et al. (2008), among others. Winarto (2004) provides a colorful illustration of the complementarity of foundational and applied skills in her case study of Integrated Pest Management adoption in Indonesia. She found that innumerate farmers had difficulty implementing IPM techniques because the requisite calculations made them "dizzy" (pp. 118,155,173).

[12] Interestingly, there is also evidence that "the capacity for change in the foundations of human skill development and neural circuitry is highest earlier in life and decreases over time" (Knudsen et al., 2006). Cognitive researchers believe that this is particularly true of language skills, which children learn rapidly without any formal instruction but which adults must invest great effort to acquire (Newport, 1990; Pinker, 2000), and of cognitive and emotional development during the early years of childhood (Shonkoff and Phillips, eds, 2000). These constraints are reflected in corporate training practices: for example, a Motorola vice president once stated that "Motorola will train people to grow proficient in scientific and technical disciplines... what we cannot do is train them to be trainable." (Avishai, 1996) Formally, the microfoundations discussed in Appendix B predict that foundational skills do not spill over if they are sufficiently costly to explain.

steady-state skill distribution in the population.[13] A steady state is defined by

$$\begin{aligned} s_{11}\lambda &= e_{11}\lambda + \phi\rho s_{10}s_{11} \\ s_{10}\lambda + \phi\rho s_{10}s_{11} &= e_{10}\lambda \\ s_{00}\lambda &= e_{00}\lambda \end{aligned} \tag{6}$$

Workers who start life without the foundational skill never learn the viral one and so their population proportion remains constant. Workers who obtain the foundational skill through education, however, may subsequently learn the viral one as well.[14] This motivates the following value functions:

$$\begin{aligned} &\lambda V(11,e)dt = \pi dt \\ &\lambda V(10,e)dt = \phi\rho s_{11}(e)[V(11,e) - V(10,e)]dt \\ &\lambda V(00,e)dt = 0 \end{aligned}$$

Note that nothing here depends qualitatively on the assumption that the foundational skill does not generate any direct returns; if the foundational skill yields a flow payoff of $\pi_1 < \lambda c_1$ then the statements below go through replacing c_1 with $c_1 - \pi_1/\lambda$. There are now two private returns, the return to learning the foundational and viral skills:

$$\Lambda(e) \equiv V(10,e) - V(00,e) = \frac{\pi}{\lambda}\left[\frac{\phi\rho s_{11}(e)}{\lambda + \phi\rho s_{11}(e)}\right] \tag{7}$$

$$\Delta(e) \equiv V(11,e) - V(10,e) = \frac{\pi}{\lambda + \phi\rho s_{11}(e)} \tag{8}$$

Note that the bracketed term in (7) is the probability that a randomly selected worker with only a foundational education has also acquired the viral skill, or equivalently the expected fraction of his lifetime for which such a worker expects to possess the viral skill.

The costs of teaching the foundational and viral skills using the formal education technology are c_1 and c_2, respectively. The definition of a decentralized equilibrium

[13] If it were feasible to learn but not to use the viral skill on its own then there would also be types s_{01} (e_{01}), but nothing that follows changes as there are none of these types either in equilibrium or at an optimum.

[14] As above, a worker could pay to acquire skills through formal education later in life, but will always at least weakly prefer to do so at birth.

is analogous to that above.[15] The social planner's problem is

$$\max_{e} \sum_{h \in \{00,10,11\}} e_h V(h,e) - c_1(e_{11}+e_{10}) - c_2(e_{11})$$

and the social return to foundational education is

$$\underbrace{[V(10,e)-V(00,e)]}_{\Lambda(e)} + \underbrace{\sum_{h \in \{00,10,11\}} e_h \left(\frac{\partial V(h,e)}{\partial e_{10}} - \frac{\partial V(h,e)}{\partial e_{00}} \right)}_{\Gamma(e)}$$

Here the first term $\Lambda(e)$ is the private return and the remaining terms, collectively labeled $\Gamma(e)$, are learning externalities.

To understand the properties of equilibria and optima it is useful to first characterize the mechanics of the relationship between s and e. Lemma 1 establishes the most important property: the fraction of agents who know the productive combination of skills — and hence total "output" in the economy — is increasing in the number who are taught the foundational skill at an increasing rate:

Lemma 1. *There exists a unique stable steady state of the system (6) except in the knife-edge case $e_{10} = \frac{\lambda}{\phi\rho}$ and $e_{11} = 0$ when there are only unstable steady states. Moreover if education levels are perturbed from a steady-state at $e = (e_{11}, e_{10}, e_{00})$ to $(e_{11}, e_{10}+\epsilon, e_{00}-\epsilon)$ then $\frac{\partial^2 s_{11}}{\partial \epsilon^2} \geq 0$ at $\epsilon = 0$.*

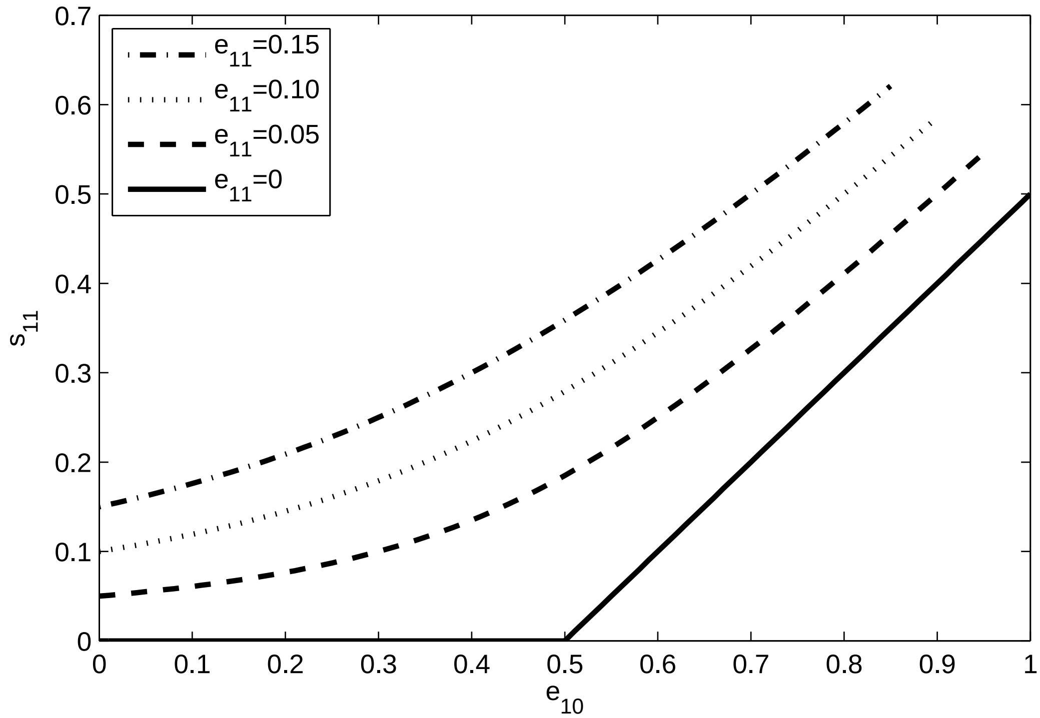
Lemma 1 captures a second generic property of communication, the *weakest link* effect. Consider an agent 1 who knows the viral skill, and suppose 1 interacts with 2, then 2 with 3, and so forth. The knowledge originally held by agent 1 can reach the nth agent if all agents $2, \ldots, n-1$ have the requisite foundational knowledge, but not if there is a single "weak link" in this chain. If each of these agents has the foundational knowledge with independent probability x then agent n's chance of learning is proportional to x^{n-2}, which increases more than proportionately with x.[16]

Recall that in the viral model it was possible for the viral skill to perpetuate itself purely through spillovers if the spillover rate exceeded the turnover rate λ. With

[15] There are nine potential cases: (1) $\Delta < c_2$, $\Lambda < c_1$, and $e_{00} = 1$; (2) $\Delta < c_2$, $\Lambda = c_1$, and $e_{10} + e_{00} = 1$; (3) $\Delta < c_2$, $\Lambda > c_1$, and $e_{10} = 1$; (4) $\Delta = c_2$, $\Lambda < c_1$, and $e_{00} = 1$; (5) $\Delta = c_2$, $\Lambda = c_1$, and $e_{11} + e_{10} + e_{00} = 1$; (6) $\Delta = c_2$, $\Lambda > c_1$, and $e_{11} + e_{10} = 1$; (7) $\Delta > c_2$, $\Lambda < c_1$, and either $e_{00} = 1$ or $e_{11} = 1$ depending on $\Delta + \Lambda \lesseqgtr c_1 + c_2$; (8) $\Delta > c_2$, $\Lambda = c_1$, and $e_{11} = 1$; (9) $\Delta > c_2$, $\Lambda > c_1$, and $e_{11} = 1$.

[16] Niehaus (2011) illustrates this effect in a technology adoption example but does not examine education choices. The mechanism is also related to the role of "fixed agents" in Glaeser et al.'s (1996) model of social interaction, as these agents follow their private signals and thus break up behavioral cascades.

Figure 2: Smooth and Discrete Increasing Returns



Plots total steady-state output $s_{11}(e)$ as a function of the proportion of agents with only a foundational education (e_{10}) and for different proportions with a complete education (e_{11}). Other parameters are fixed at $\lambda = 1$, $\phi\rho = 2$.

complementary foundational skills the threshold spillover rate becomes a function of the level of foundational education. As a result, increases in the level of foundational education can qualitatively change the equilibrium from one with no viral skills to one with positive levels. Specifically, if $e_{11} = 0$ then one can show that when $\phi\rho < \lambda/e_{10}$ then the unique stable steady state is $s_{11} = 0, s_{10} = e_{10}$, while if $\phi\rho > \lambda/e_{10}$ then the unique stable steady state is $s_{11} = e_{10} - \lambda/(\phi\rho), s_{10} = \lambda/(\phi\rho)$. The intuition here is as follows: if agents are not learning the viral skill in school ($e_{11} = 0$) the economy will have a steady state in which no one knows it. For this to be the stable state it must hold that if a small fraction of the population acquired the viral skill (e.g. through learning-by-doing) they would die out faster than their knowledge diffused to the rest of the population. Whether this holds depends on, among other things, the percentage e_{10} of agents who can use the viral skill. Figure 2 depicts this: when $e_{11} = 0$ there is a threshold level of foundational education above which viral skills begin to circulate.

One expects the weakest link effect to push towards increasing returns, but it is a priori unclear whether it is strong enough to offset the saturation effects that are still present.

Lemma 2. *Consider a stable steady state $e = (e_{11}, e_{10}, e_{00})$ and perturb education levels to $(e_{11}, e_{10} + \epsilon, e_{00} - \epsilon)$. The private rate of return to education is globally increasing ($\frac{\partial}{\partial \epsilon}\Lambda(e) \geq 0$), and the social rate of return is globally increasing ($\frac{\partial}{\partial \epsilon}[\Lambda(e) + \Gamma(e)] \geq 0$) if the spillover rate $\phi\rho$ is sufficiently low.*

To understand this, note that the private return to a foundational education depends only on the fraction of agents who know the viral skill, which increases with the educated share due to the weakest link effect (Lemma 1). The social return is more complicated as it also depends on the number of people who could potentially learn the viral skill via a marginal educated agent. This may either increase with the share of educated agents (since they are potential recipients of the viral skill) or decrease (since they will meet more educated peers who can relay viral skills to them). The latter effect weakens as the spillover rate falls; the proof shows that it weakens uniformly enough that returns can be globally increasing.

The fact that private returns are increasing is of interest since it implies the possibility of multiple equilibria and low-education traps: when few people are educated there is little viral knowledge in the population for an educated person to learn and so the return to education is low.

Proposition 4. *The set of decentralized equilibria depends on parameters as follows:*

1. *If $\frac{\pi}{\lambda} \geq c_1 + c_2$ then there is a unique equilibrium in which all agents obtain at least a foundational education ($\hat{e}_{11} + \hat{e}_{10} = 1$).*
2. *If $c_1 < \frac{\pi}{\lambda} < c_1 + c_2$ but $\phi\rho > \frac{\lambda\pi}{\pi - c_1\lambda}$ then there are two stable equilibria, one in which no agent gets any education ($\hat{e}_{00} = 1$) and one in which all agents get only a foundational education ($\hat{e}_{10} = 1$).*
3. *In all other cases there is a unique equilibrium in which no agent gets any education ($\hat{e}_{00} = 1$).*

Notice that multiple equilibria are possible only when the spillover rate $\phi\rho$ is high enough. The intuition for this is as follows. For no education to be an equilibrium, it must be that acquiring both skills costs more than they are worth. Given this, the only way that full foundational education can also be an equilibrium is if workers with the foundational skill are very likely to acquire the viral skill via diffusion – so the spillover rate must be high.

Proposition 5. *The optimal level of foundational education depends on parameters as follows:*

1. *If* $\frac{\pi}{\lambda} \geq c_1 + c_2$ *then universal foundational education is optimal* ($e^*_{11} + e^*_{10} = 1$).
2. *If* $c_1 < \frac{\pi}{\lambda} < c_1 + c_2$ *then there exists* $(\phi\rho)^*$ *such that if* $\phi\rho \geq (\phi\rho)^*$ *then universal foundational education is optimal and otherwise no education is optimal. Moreover for* c_2 *low enough* $(\phi\rho)^* < \frac{\lambda\pi}{\pi - c_1\lambda}$, *so that universal education may be optimal although it is not an equilibrium.*
3. *If* $\frac{\pi}{\lambda} \leq c_1$ *then no education is optimal.*

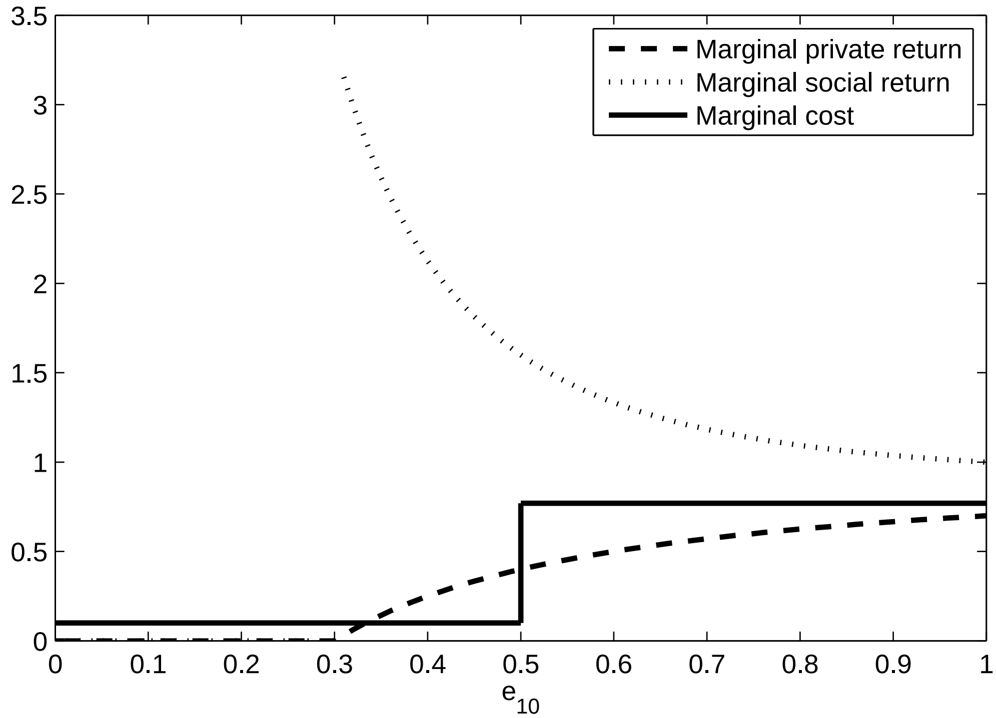
Comparing and summarizing, universal education may be both optimal and an equilibrium, neither optimal nor an equilibrium, or optimal but not an equilibrium. Intervention may be necessary in an economy with foundational skills either to ensure that the high-education equilibrium is selected or to keep the economy in a high-education state that is *not* an equilibrium, and the former case can persist even as $\phi\rho$ grows large. The intuition for the latter case is that coordination on universal foundational education is less attractive under laissez-faire than would be efficient because of inadequate private incentives to learn the complementary viral skill. By subsidizing the acquisition of viral skills the planner can raise the social return to foundational education.[17]

What if the costs of getting educated are different for different workers? Cost heterogeneity naturally tends to push equilibria (and optima) away from the corners: for example, if the marginal costs of educating an additional worker rise faster than the marginal private (social) benefits over some range then there can be an equilibrium (optimum) with positive but less-than-universal education in a foundational skill. What is more interesting, however, is that heterogenous costs can *strengthen* the argument for subsidized universal education, in the sense that the only equilibria may involve less-than-universal education while the only optimum involves universal education.

Figure 3 illustrates such a case. In this example there are two types of worker, half with a low cost $\underline{c}_1$ and the other half with a high cost $\overline{c}_1$ of acquiring the foundational skill; the marginal cost of foundational education is therefore a step function. The marginal private and social returns to education are 0 up to the point where the viral skill becomes self-sustaining and then strictly positive thereafter. The higher cost

[17] The multiple equilibrium problem here is related to those in Kremer (1993) and Acemoglu (1996) which are also driven by noisy matching combined with a form of skill-skill complementarity – in those cases, in the production function. See Appendix B for a microfoundation that nests both kinds of complementarity.

Figure 3: High-Cost Types May Need Subsidies



Plots marginal costs and benefits of foundational education as a function of the proportion of agents with only a foundational education (e_{10}). Other parameters are $\pi = \lambda = 1$, $\phi\rho = 10/3$, $\underline{c}_1 = 0.1$, and $\overline{c}_1 = 0.77$.

$\overline{c}_1$ is sufficiently high that none of the high-cost workers would ever choose to get a foundational education, so the only (stable) equilibria are at $e_{10} = 0$ and $e_{10} = 1/2$. It can be socially optimal, however, for *all* agents to obtain a foundational education. The intuition is that while educating a high-cost type may cost more than it increases his expected productivity, doing so also removes a "missing link" from the diffusion process and thus speeds up the rate at which his (low-cost) peers acquire viral skills.[18]

2.3 Implications for Education Policy

Optimal levels of education are salient policy issues: many countries set explicit targets for education participation at different levels (Wolf, 2004). Interestingly, some targets are universal and others are not. For example, universal primary education has

[18]Sufficient conditions are $\phi\rho > \lambda$, $c_2 > \pi/(\phi\rho - \lambda)$ for all workers, $0 < \underline{c}_1 < \frac{\pi}{\lambda}\left(1 - \frac{\lambda}{\phi\rho}\right) < \overline{c}_1$, $\underline{c}_1$ sufficiently small, $\overline{c}_1$ sufficiently close to $\frac{\pi}{\lambda}\left(1 - \frac{\lambda}{\phi\rho}\right)$, and α sufficiently large. Proof available on request.

been endorsed in the Jomtien Declaration[19] and is the second Millenium Development Goal.[20] On the other hand, agricultural extension services are typically provided to a small fraction of interested farmers. What is striking is that both of these approaches have been rationalized as optimal responses to knowledge spillovers. Proponents of subsidized universal education appeal to externalities (e.g. Paquette (1995)), yet so do proponents of selective extension, whose idea is precisely to exploit spillovers by providing knowledge to a few agents and then letting it diffuse more broadly.[21] Can such arguments be mutually compatible?

Interpreted through the lens of the analysis above, these arguments rest on different views of the education-spillover nexus. Under the simpler "viral" view spillovers not only do not rationalize subsidized universal education, they *weaken* the case for it. If intervention is optimal at all, the best intervention is to subsidize some agents to learn and let the rest learn from them. This corresponds exactly to the arguments advanced for agricultural extension (for example).[22] On the other hand, universal education in a foundational skill may be optimal, including in situations where it is not an equilibrium, so that the foundational view can potentially rationalize interventions like subsidized primary education. Thus while neither view of spillovers explains the mix of observed education policies on its own, together they perform reasonably well.

The model also provides some perspective on how optimal policy may change as communication speeds up with the increasing density of human settlement and the spread of electronic means of communication. Comparing Proposition 3 with Propositions 4 and 5 shows that the two views of spillovers imply different answers. Under the viral view the need for intervention in education markets must ultimately vanish, as knowledge circulates so rapidly through informal channels that further formal training is not cost-effective.[23] Under the foundational view, on the other hand, the state has an enduring role to play in ensuring that agents coordinate on acquisition of the foundational skills that enable diffusion and social learning. The two views thus imply qualitatively different future roles for education policy.

As this discussion highlights, empirical work on spillovers will be most valuable if

[19] `http://www.unesco.org/education/wef/en-conf/Jomtien\%20Declaration\%20eng.shtm`

[20] `http://www.un.org/millenniumgoals/education.shtml`

[21] For example, Feder et al. (2004) discuss how "farmer-to-farmer diffusion effects are expected to bring about cost-effective knowledge dissemination." (46)

[22] In application the choice of which agents to "seed" with a skill or idea is important. See Domingos and Richardson (2001), Kempe et al. (2003) and Campbell (2010) for network analyses.

[23] "Communication technologies" here should be thought of as encompassing both the capacity to transmit information and the capacity to locate information worth transmitting. For example, the world wide web would have less impact on the rate at which useful knowledge diffuses without search and indexing technologies to help users navigate its contents.

it not only tests whether they occur (i.e. whether $\phi\rho > 0$) but also quantifies their rate. While this has not been an explicit goal of existing studies, many do implicitly provide tests for complete knowledge pooling. For example, Conley and Udry (2010) test whether farmers who are closer in a social network to an innovating farmer are more likely to mimic profitable innovations. This test would fail to reject if there were no social learning, but also if information were fully pooled within the village, since in that case everyone would mimic profitable experiments regardless of social distance. Conley and Udry's results thus implicitly define some upper bound on the rate of knowledge diffusion. With further structural modeling one could explicitly define and estimate a contextually appropriate concept of the spillover rate for use in policy design.

3 Peer Effects on Earnings

The differential policy implications of the viral and foundational views suggest a real need for empirical work separating them. The evidence on technology adoption discussed above is consistent with the view that some components of measured "education" serve a foundational function, but does not make clear which ones. The most direct way forward is to generate exogenous variation in one skill and test for impacts on the acquisition of others. Muralidharan and Sundararaman (2011) provide a clean illustration of this idea: they find that paying teachers based on students' math and language test scores not only increased those scores but also increased student performance on other non-incentivized subjects. These findings support the intuitive view that language skills are foundational.

A second and more difficult question is whether the bundle of skills currently delivered by public education systems display either viral or foundational characteristics. Intuitively, evidence on peer effects on earnings should be diagnostic in this regard. Following Rauch (1993) a number of studies have estimated augmented Mincer models that include the characteristics of workers' (geographic) peers as predictors. (Acemoglu and Angrist, 2000; Moretti, 2004a) Specifically, Rauch estimated

$$w_{ij} = \alpha + x_{ij}\beta + z_j\gamma + \epsilon_{ij} \tag{9}$$

where w_{ij} is the wage of individual i in city j, the x_{ij} are individual characteristics including education and experience, and the z_j are common characteristics of city j, including average education and experience levels. The rationale for including z_j is human capital externalities, with knowledge spillovers a leading example. (9) is linear because the reduced-form Roback model underlying it does not predict any particular non-linear structure; subsequent work has, on the other hand, estimated interaction terms. The key question for our purposes is whether the viral and foundational views leave distinctive "footprints" in these estimates.

To address this question we must first specify which quantities in the model correspond to the concept of "education" measured in the data. If education were a perfect measure of a worker's human capital then peer characteristics could have no further predictive power. Thus, for a model like (9) to make sense education must refer to an *initial* investment in human capital that may subsequently have been augmented by spillovers. In this case peer characteristics are valid regressors because they are determinants of those unobserved spillovers.

Suppose first that education consists primarily of viral skills. In this case the natural motivation for including peer characteristics is that we observe whether or not a worker was educated but not whether they subsequently became skilled through spillovers. Formally, if $e(i) \in \{0,1\}$ is the *initial* skill level of an agent i aged τ and $s(i) \in \{0,1\}$ his current skill level then we are interested in the conditional expectation of his earnings, $\mathbb{E}[\pi s(i)|e(i), e, \tau]$. To reintroduce the possibility of general-equilibrium decreasing returns to skilled labor, let the skill premium be a non-increasing function $\pi(s)$ of overall skill levels. Then the wage regression function is

$$\mathbb{E}[\pi(s)s(i)|e(i), e, \tau] = \pi(s(e))\left[e(i) + (1 - e(i))\left(1 - \exp(-\phi\rho s(e)\tau)\right)\right] \qquad (10)$$

where $(1 - \exp(-\phi\rho s(e)\tau))$ is the probability that agent i learns the skill from a peer during an interval of length τ.

Now suppose instead that education consists primarily of foundational skills. Then the natural rationale for including peer characteristics in a Mincer regression is to pick up the effects of viral skills acquired from peers. Letting $e_h(i) \in \{0,1\}$ indicates whether or not individual i aged τ learned skill vector $h \in \{00, 10, 11\}$ in school, and $s_h^t(i) \in \{0,1\}$ indicates his current skill set, the wage regression function

$$\begin{aligned} &\mathbb{E}[\pi(s_{11})s_{11}(i)|e_{11}(i) + e_{10}(i), e, \tau] \\ &= \pi(s_{11}(e))\left[(e_{11}(i) + e_{10}(i))\left(\frac{e_{11}}{e_{11} + e_{10}} + \frac{e_{10}}{e_{11} + e_{10}}\left(1 - \exp(-\phi\rho s_{11}(e)\tau)\right)\right)\right] \end{aligned} \qquad (11)$$

expresses expected earnings as a function of observed foundational skills.[24] Here

[24]Wages conditional on $e(i)$ are

$$\mathbb{E}[\pi(s_{11})s_{11}(i)|e_{11}(i), e_{10}(i), e, \tau] = \pi(s_{11}(e))\left[e_{11}(i) + e_{10}(i)\left(1 - \exp(-\phi\rho s_{11}(e)\tau)\right)\right] \qquad (12)$$

$(1 - \exp(-\phi\rho s_{11}(e)\tau))$ is the probability that a worker with the foundational skill learns the viral skill from a peer during an interval of length τ.

To build intuition, compare these expressions under the assumption of constant returns to skilled labor, i.e. $\pi(s) = \pi$. Both models predict a positive relationship between earnings and own education and between earnings and average education e. Beyond this point their implications diverge. The viral view predicts substitutability between individual and peer education: peer skill levels matter most for those who received the least education because they stand to benefit most from spillovers. Longer experience τ gives the unskilled worker more opportunities to realize this benefit and hence experience and average education are complements. By the same logic own experience substitutes for own education. Lastly, local average experience – inversely measured by the turnover rate λ – tends to raise local knowledge rates and therefore substitutes for own formal education while complementing own experience. In contrast, the foundational skills model predicts complementarity between own and average education: the average level of education is most important for educated workers as they are the ones capable of learning and using the skills that other educated people have. Similarly, experience complements own and peer education, since opportunities to learn are more valuable for a worker with stronger foundational skills.

In principle these characteristic differences, summarized in Table 1, could be used to test between them. The chief challenge is that earnings may also depend on the interaction between own and peer education for purely neoclassical reasons (Moretti, 2004a; Ciccone and Peri, 2006). To see this, consider the viral skills model and suppose that that the skill premium $\pi(s)$ is strictly decreasing in the population skill level s. Then peer education e has both positive and negative effects on individual earnings in (10): it raises the probability of becoming skilled but lowers the skill premium:

$$\frac{\partial}{\partial e}\left[\mathbb{E}[\pi(s)s(i)|e(i)=1,e,\tau] - \mathbb{E}[\pi(s)s(i)|e(i)=0,e,\tau]\right]$$
$$= \pi'(s)s'(e)\exp(-\phi\rho s(e)\tau) - \pi(s)\phi\rho\tau s'(e)\exp(-\phi\rho s(e)\tau)$$

This expression remains unambiguously negative, so that a negative estimated interaction is still consistent with the model. It would be unclear, however, whether to attribute such an estimate to technological decreasing returns or the saturation effect. Now consider the foundational skills model. Peer education e_{10} has no effect on the

Estimating this equation directly would of course require decomposing education into its viral and foundational components.

Table 1: Predicted Interaction Effects in Wage & Productivity Models

	Own Education	Own Experience
Own Experience	$(-), [+]$	
Average Education	$(-), [+]$	$(+), [+]$
Average Experience	$(-), [+]$	$(+), [+]$

Summarizes the predictions of the viral skills model (in parenthesis) and the foundational skills model [in brackets] for interaction terms in wage or productivity models. A "+" indicates complementarity, a "-" substitutability. All main effects are positive.

earnings of the uneducated, while its effect on the earnings of the educated is

$$\pi'(s_{11}(e))\frac{\partial s_{11}}{\partial e_{10}}\left(\frac{e_{11}}{e_{11}+e_{10}}+\frac{e_{10}}{e_{11}+e_{10}}\left(1-\exp(-\phi\rho s_{11}(e)\tau)\right)\right) + \pi(s_{11}(e))\frac{e_{10}}{e_{11}+e_{10}}\phi\rho\tau\frac{\partial s_{11}}{\partial e_{10}}\exp(-\phi\rho s_{11}(e)\tau)$$

The first term here is negative due to decreasing returns, but the second is positive due to the weakest link effect. Thus while a negative interaction between own and peer education is consistent both with viral spillovers and with technological decreasing returns, a positive interaction is consistent only with the foundational skills view.

The evidence available in the literature following Rauch (1993) is tantalizingly mixed. Moretti (2004a) estimates that college graduates have a stronger positive effect on the wages of less educated workers, which is consistent with the viral view (but could also capture neoclassical diminishing returns). Acemoglu and Angrist (2000), on the other hand, find some weak evidence of a positive interaction between own- and peer high-school education. Since diminishing returns would yield the opposite result, this provides some tentative support for the foundational view. More recently Ciccone and Peri (2006) and Iranzo and Peri (2009) have proposed a novel "constant-composition" approach for estimating external returns to education in the presence of factor supply effects. Their approach does not permit separate identification of external effects on different categories of workers, however, precisely because it involves aggregating wage elasticities across those categories. The question thus remains open.

4 Sorting and Agglomeration

Because both the viral and foundational views predict peer effects on earning, one might imagine that they also have implications for worker's *choice* of peers, and thus for spatial sorting and agglomeration. Indeed, the notion that workers choose to bear the congestion costs and higher rents associated with living in dense areas in order to benefit from spillovers has a long pedigree (Marshall, 1890; Jacobs, 1969; Glaeser et al., 1992).

An illustrative model may help shed light on this issue. Fix workers' initial education levels at e but suppose that they can choose at birth between one of two locations, A and B. A steady-state consists of a division of each generation between locations such that if m^l is the mass of agents locating in l then $m^A e^A + m^B e^B = e$ (which implies $m^A + m^B = 1$). This yields a steady-state skill distribution $s(e^l)$. Workers must bear lifetime congestion costs $C(m^l)$ when they live in location $l \in \{A, B\}$, with $C' > 0$ and $C'' > 0$. The steady state is an equilibrium if all agents' location choices are privately optimal given their own education levels and the state (e^A, e^B, m^A, m^B).

First consider the viral skills case. Workers with an initial education who locate in location l earn a payoff $\pi/\lambda - C(m^l)$ which depends on l only via C, so they always choose to live in the smaller location. Given this, workers without an initial education also strictly prefer the smaller location, since it is less congested and must have at least as high a proportion of skilled workers. Thus in equilibrium all workers prefer the smaller location, and so the only equilibrium is one in which the population is distributed evenly.

Now consider the foundational skills model. Workers without foundational education cannot benefit from spillovers and so will always choose the least congested location. Workers who do have the foundational skill, however, potentially face a tradeoff between locations with less congestion and locations with a higher proportion of educated workers, where their chances of acquiring the viral skill are higher. Consequently there can be equilibria in which one location attracts a disproportionate share of the population.

Consider, purely for illustration, a simple case where $\pi = \lambda = 1$, $q = \frac{\lambda}{\phi\rho} < 1$, $e_{11} = 0$, $e_{10} > 1 - q/2$, and $C(m^l) = (m^l)^2$. These conditions imply that the viral skill will only circulate in a location if and only if it has a high-enough concentration of educated workers. Hypothesize an equilibrium in which location A attracts $1/2 < m^A < e_{10}$ educated workers and no uneducated workers, so that $e^A = (0, 1, 0)$ while $e^B = \left(0, \frac{e_{10} - m^A}{1 - m^A}, \frac{e_{00}}{1 - m^A}\right)$. This ensures that the viral skill circulates in location A, while if $m^A > \frac{e_{10} - q}{1 - q}$ there are too few educated workers in B for the viral skill to circulate there. Uneducated workers are clearly best-responding in this scenario by choosing the less congested location. Educated workers, on the other hand, must be indifferent between the two locations, which holds if

$$V(10,(0,1,0)) - C(m^A) = V\left(10, \left(0, \frac{e_{10} - m^A}{1 - m^A}, \frac{e_{00}}{1 - m^A}\right)\right) - C(1 - m^A) \quad (13)$$

Substituting in for V and C and simplifying, this is

$$m^A = 1 - \frac{q}{2} > \frac{1}{2} \quad (14)$$

This condition is consistent with other maintained assumptions provided $e_{10} > 1 - \frac{q}{2} > \frac{e_{10}-q}{1-q}$, which is satisfied for a full-dimensional subset of parameters (e.g. in a neighborhood of $e_{10} = q = 3/4$). Thus there are equilibria with agglomeration, in the sense that congestion costs are not minimized, because educated workers value the company of their peers enough to endure congestion costs.

One interesting aspect of this example is that it works without assuming that spillover rates rise with density. Glaeser (1999), Peri (2002), and Berliant et al. (2006) provide models in which agglomeration emerges because it speeds up the diffusion of knowledge. Density effects are surely an important part of the story. The example shows that they may not be the only part; the foundational view provides a complementary interpretation in which agglomeration is intrinsically linked with sorting.

The example also underscores the local complementarity implied by the foundational view. In this sense it microfounds reduced-form complementarity assumptions found elsewhere in the literature on cities. In particular, Benabou's (1993) account of the formation of ghettos is driven by the assumption that investments in education are local complements. He assumes that externalities within the education process itself make education cheaper for workers whose peers are also purchasing it. Interpreting the "education process" broadly to include the diffusion of viral skills, however, this is exactly what the foundational view predicts. Complementarity also plays a central role in Black and Henderson's (1999) model of urban production and growth.

Finally, the example opens up additional avenues for empirically evaluating the two views of spillovers. Evidence that the returns to education are higher in cities than in rural areas (Glaeser and Mare, 2001) and that college-educated workers are increasingly locating in cities with high baseline education levels (Berry and Glaeser, 2005) is inconsistent with the viral view of spillovers but consistent with the foundational view.

5 Conclusion

Previous research on education and knowledge spillovers has not taken a stance on the mechanism by which they are related. Drawing on education research, this paper formalizes and characterizes two alternative interpretions. One possibility is that education gives workers viral skills that subsequently diffuse to their less-skilled peers. Another possibility is that education gives workers foundational skills that enhance their ability to learn from more-skilled peers. While these views seem qualitatively different, it is unclear from earlier work which is more consistent with observed education policies, with standard reduced-form modelling assumptions, or with the available evidence.

The analysis shows that both views have a role to play in explaining education policy: they rationalize selective training in viral skills along with universal training in foundational ones. The foundational view produces complementarities that are consistent with a number of reduced-form models of sorting, agglomeration, and growth. The relevant evidence is mixed and incomplete, but a number of pieces support the relevance of the foundational view.

While the theoretical framework in this paper is kept stark in order to facilitate intuition-building, it could readily be enriched and extended to facilitate quantitative analysis. For example, one could model agent i's skills as an arbitrary vector $s(i)$ yielding payoffs $\pi(s(i))$ and generating a vector of spillover probabilities $\rho(s(i), s(j))$ when agents i and j interact. A flexible setup such as this would not yield analytical insights as readily but would capture the fact that many skills have both "viral" and "foundational" attributes. It would also allow one to consider deeper hierarchies of foundational skills, each facilitating the mastery of the next.

References

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

_ **and Joshua Angrist**, "How Large are the Social Returns to Education? Evidence from Compulsory Attendance Laws.," in Daron Acemoglu, ed., *NBER Macroeconomics Annual*, number 15 2000.

Avishai, Bernard, "Companies can't make up for failing schools," *Wall Street Journal*, July 1996, p. A12.

Benabou, Roland, "Workings of a City: Location, Education, and Production," *The Quarterly Journal of Economics*, August 1993, *108* (3), 619–52.

Benhabib, Jess and Mark M. Spiegel, "Human Capital and Technology Diffusion," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 1 of *Handbook of Economic Growth*, Elsevier, 2005, chapter 13, pp. 935–966.

Bennell, Paul, "General verses Vocational Secondary Education in Developing Countries: A Review of the Rates of Return Evidence," *Journal of Development Studies*, December 1996, *33*, 230–247.

Berliant, Marcus, Robert R. Reed III, and Ping Wang, "Knowledge exchange, matching, and agglomeration," *Journal of Urban Economics*, July 2006, *60* (1), 69–95.

Berry, Christopher R. and Edward L. Glaeser, "The Divergence of Human Capital Levels Across Cities," NBER Working Papers 11617, National Bureau of Economic Research, Inc September 2005.

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

Black, Duncan and Vernon Henderson, "A Theory of Urban Growth," *Journal of Political Economy*, April 1999, *107* (2), 252–284.

Campbell, Arthur, "Tell Your Friends! Word of Mouth and Percolation in Social Networks," Technical Report, Yale University 2010.

Ciccone, Antonio and Giovanni Peri, "Identifying Human-Capital Externalities: Theory with Applications," *Review of Economic Studies*, 04 2006, *73* (2), 381–412.

_ **and Robert E Hall**, "Productivity and the Density of Economic Activity," *American Economic Review*, March 1996, *86* (1), 54–70.

Coe, David, Elhanan Helpman, and Alexander Hoffmeister, "International R&D Spillovers and Institutions," *mimeo, Harvard University*, December 2008.

Conley, Timothy G. and Christopher R. Udry, "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, March 2010, *100* (1), 35–69.

Dewey, John, *The Child and the Curriculum*, University of Chicago Press, 1902.

Domingos, Pedro and Matt Richardson, "Mining the network value of customers," in "Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining" KDD '01 ACM New York, NY, USA 2001, pp. 57–66.

Duranton, Gilles and Diego Puga, "Micro-foundations of urban agglomeration economies," in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4 of *Handbook of Regional and Urban Economics*, Elsevier, April 2004, chapter 48, pp. 2063–2117.

Feder, Gershon, Rinku Murgai, and Jaime B. Quizon, "Sending Farmers Back to School: The Impact of Farmer Field Schools in Indonesia," *Review of Agricultural Economics*, 02 2004, *26* (1), 45–62.

Glaeser, Edward L., "Learning in Cities," *Journal of Urban Economics*, September 1999, *46* (2), 254–277.

Glaeser, Edward L and David C Mare, "Cities and Skills," *Journal of Labor Economics*, April 2001, *19* (2), 316–42.

_ **, Bruce Sacerdote, and Jose A Scheinkman**, "Crime and Social Interactions," *The Quarterly Journal of Economics*, May 1996, *111* (2), 507–48.

_ **, Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer**, "Growth in Cities," *Journal of Political Economy*, December 1992, *100* (6), 1126–52.

Glazer, Nathan, "A Human Capital Policy for the Cities," *Public Interest*, Summer 1993, (112), 27–49.

Goldin, Claudia, "The Human-Capital Century And American Leadership: Virtues Of The Past," *The Journal of Economic History*, June 2001, *61* (02), 263–292.

_ **and Lawrence F. Katz**, "The "Virtues"; of the Past: Education in the First Hundred Years of the New Republic," NBER Working Papers 9958, National Bureau of Economic Research, Inc September 2003.

Heckman, James and Pedro Carneiro, "Human Capital Policy," NBER Working Papers, National Bureau of Economic Research, Inc February 2003.

_ **and Peter Klenow**, "Human Capital Policy," in M. Boskin, ed., *Policies to Promote Capital Formation*, Stanford, CA: Hoover Institution, 1998.

Heckman, James J., "Policies to foster human capital," *Research in Economics*, March 2000, *54* (1), 3–56.

Iranzo, Susana and Giovanni Peri, "Schooling Externalities, Technology, and Productivity: Theory and Evidence from U.S. States," *The Review of Economics and Statistics*, November 2009, *91* (2), 420–431.

Jacobs, Jane, *The Economy of Cities*, New York: Vintage, 1969.

Jovanovic, Boyan and Rafael Rob, "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, October 1989, *56* (4), 569–82.

Kempe, David, Jon Kleinberg, and Éva Tardos, "Maximizing the spread of influence through a social network," in "Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining" KDD '03 ACM New York, NY, USA 2003, pp. 137–146.

Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff, "Economic, Neurobiological and Behavioral Perspectives on Building America's Future Workforce," Working Paper 12298, National Bureau of Economic Research June 2006.

Kremer, Michael, "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, August 1993, *108* (3), 551–75.

Lange, Fabian and Robert Topel, *The Social Value of Education and Human Capital*, Vol. 1 of *Handbook of the Economics of Education*, Elsevier, November

Lucas, Robert E., "Ideas and Growth," *Economica*, 02 2009, *76* (301), 1–19.

— **and Benjamin Moll**, "Knowledge Growth and the Allocation of Time," NBER Working Papers 17495, National Bureau of Economic Research, Inc October 2011.

Lucas, Robert Jr., "On the mechanics of economic development," *Journal of Monetary Economics*, July 1988, *22* (1), 3–42.

Malamud, Ofer and Cristian Pop-Eleches, "General Education versus Vocational Training: Evidence from an Economy in Transition," *The Review of Economics and Statistics*, 06 2010, *92* (1), 43–60.

Marshall, Alfred, *Principles of Economics*, Macmillan, 1890.

McMahon, Walter W., "The Social and External Benefits of Education," in Geraint Johnes and Jill Johnes, eds., *International Handbook on the Economics of Education*, Edward Elgar, 2004.

Moretti, Enrico, "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data," *Journal of Econometrics*, 2004, *121* (1-2), 175–212.

— , "Human capital externalities in cities," in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4 of *Handbook of Regional and Urban Economics*, Elsevier, April 2004, chapter 51, pp. 2243–2291.

Muralidharan, Karthik and Venkatesh Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39 – 77.

Nelson, Richard R. and Edmund S. Phelps, "Investment in Humans, Technological Diffusion, and Economic Growth," *The American Economic Review*, 1966, *56* (1/2), pp. 69–75.

Newport, Elissa L., "Maturational constraints on language learning," *Cognitive Science*, 1990, *14* (1), 11 – 28.

Niehaus, Paul, "Filtered Social Learning," *Journal of Political Economy*, 2011, *119* (4), 686 – 720.

Paquette, Jerry, "Universal Education: Meanings, Challenges, and Options into the Third Millennium," *Curriculum Inquiry*, 1995, *25* (1), pp. 23–56.

Peri, Giovanni, "Young workers, learning, and agglomerations," *Journal of Urban Economics*, November 2002, *52* (3), 582–607.

Perla, Jesse and Christopher Tonetti, "Endogenous Risk and Growth," Technical Report, New York University May 2011.

Pinker, Steven, *The language instinct: how the mind creates language*, New York: HarperCollins, 2000.

Psacharopoulos, George, "To vocationalize or not to vocationalize? That is the curriculum question," *International Review of Education*, 1987, *33*, 187–211.

Rauch, James, "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," *Journal of Urban Economics*, November 1993, *34* (3), 380–400.

Rogers, Everett, *Diffusion of Innovations*, 3rd. ed., New York: Free Press, 1983.

Schultz, Theodore W, "The Value of the Ability to Deal with Disequilibria," *Journal of Economic Literature*, September 1975, *13* (3), 827–46.

Shonkoff, Jack and Deborah Phillips, eds, *From Neurons to Neighborhoods: the Science of Early Childhood Development*, Washington, D.C.: National Academy Press, 2000.

Skinner, Jonathan and Douglas Staiger, "Technology Adoption From Hybrid Corn to Beta Blockers," NBER Working Papers 11251, National Bureau of Economic Research April 2005.

Winarto, Yunita, *Seeds of Knowledge: The Beginning of Integrated Pest Management in Java*, New Haven: Yale University Southeast Asian Studies, 2004.

Wolf, Alison, "Education and Economic Performance: Simplistic Theories and their Policy Consequences," *Oxford Review of Economic Policy*, Summer 2004, *20* (2), 315–333.

A Proofs

Proof of Proposition 1

Given decreasing returns, the equilibrium level of education is either 0 (if the net private return $\Delta(e) - c$ is negative at $e = 0$), 1 (if the net private return is positive at $e = 1$, or an interior $e \in (0, 1)$ if neither of these conditions hold. The first condition varies depending on the relative values of $\phi\rho$ and λ. If $\phi\rho \leq \lambda$ then $s(0) = 0$ and $\Delta(0) = \frac{\pi}{\lambda}$, while if $\phi\rho > \lambda$ then $s(0) = \frac{\phi\rho - \lambda}{\phi\rho}$ and $\Delta(0) = \frac{\pi}{\phi\rho}$. Thus $\Delta(e) < c$ for all $e > 0$, implying $\hat{e} = 0$, if and only if $c \geq \frac{\pi}{\max\{\lambda, \phi\rho\}}$. The second condition depends only on $\Delta(1) = \frac{\pi}{\lambda + \phi\rho}$ and so $\Delta(e) - c > 0$ for all $e < 1$ if and only if $c \leq \frac{\pi}{\lambda + \phi\rho}$, in which case $\hat{e} = 1$ must hold in equilibrium. In intermediate cases $\frac{\pi}{\lambda + \phi\rho} < c < \frac{\pi}{\max\{\lambda, \phi\rho\}}$ we have $\Delta(1) - c < 0$ but $\Delta(0) - c > 0$ and since Δ is smoothly decreasing there must exist a unique $\hat{e} \in (0, 1)$ satisfying $\Delta(\hat{e}) - c = 0$.

Proof of Proposition 2

Given decreasing social returns, there are again three cases depending on the sign of the net social return $\Delta(e) + \Gamma(e) - c$ evaluated at $e = 0$ and at $e = 1$. Differentiating (2) and substituting into (5) we can write the external return as

$$\Gamma(e) = (1 - e)\pi \left[\frac{\phi\rho}{(\lambda + \phi\rho s(e))^2} \right] \frac{\lambda}{\sqrt{(\lambda - \phi\rho)^2 + 4\lambda\phi\rho e}} \tag{15}$$

Evidently $\Gamma(1) = 0$ so that $e^* = 1 \Leftrightarrow \Delta(1) \geq c \Leftrightarrow \frac{\pi}{\lambda + \phi\rho} \geq c$. At $e = 0$ the gross social return is

$$\frac{\pi}{\lambda + \phi\rho s(0)} + \pi \left[\frac{\phi\rho}{(\lambda + \phi\rho s(0))^2} \right] \frac{\lambda}{|\lambda - \phi\rho|} \tag{16}$$

If $\phi\rho \leq \lambda$ then $s(0) = 0$ and this simplifies to $\frac{\pi}{\lambda - \phi\rho}$. If on the other hand $\phi\rho > \lambda$ then $s(0) = \frac{\phi\rho - \lambda}{\phi\rho}$ and this simplifies to $\frac{\pi}{\phi\rho - \lambda}$. Thus the gross return at $e = 0$ is $\frac{\pi}{|\lambda - \phi\rho|}$.

Proof of Proposition 3

Part I: $\phi\rho \to 0$. Since $s(e)$ and $s'(e)$ are bounded with respect to e the private return to education $\Delta(e)$ and the social return to education $\Delta(e) + \Gamma(e)$ converge uniformly to π/λ (see Equation 15). If $\pi/\lambda > c$ then this implies that for $\phi\rho$ small enough both the private net return $\Delta(e) - c$ and the social net return $\Delta(e) + \Gamma(e) - c$ are positive for all e and thus $e^* = \hat{e} = 1$. If on the other hand $\pi/\lambda < c$ then for $\phi\rho$ small enough $\Delta(e) - c$ and $\Delta(e) + \Gamma(e) - c$ are negative for all e and thus $e^* = \hat{e} = 0$. In the knife-edge case $\pi/\lambda = c$ there is potentially a discontinuity in the limit as at

$\phi\rho = 0$ any value of e is an equilibrium and an optimum.

Part II: $\phi\rho \to \infty$. Equation 2 implies that for $\phi\rho > \lambda$ we have $s(0) = \frac{\phi\rho-\lambda}{\phi\rho}$ and therefore $\Delta(0) = \frac{\pi}{\phi\rho} \to 0$ so that the decentralized equilibrium is $\hat{e} = 0$ for $\phi\rho$ sufficiently large. Similarly by examination of (15) the optimal subsidy $\Gamma(0)$ at $e = 0$ approaches 0 and thus since $\Delta(0) + \Gamma(0) \to 0$ the socially optimal education level is 0 for $\phi\rho$ sufficiently large.

Proof of Lemma 1

Uniqueness. Using the notation $q \equiv \frac{\lambda}{\phi\rho}$ we can write the system as

$$0 = -qs_{11} + qe_{11} + s_{10}s_{11} \tag{17}$$

$$0 = -qs_{10} + qe_{10} - s_{10}s_{11} \tag{18}$$

$$0 = -s_{00} + e_{00} \tag{19}$$

Solving the second equation for s_{10}, substituting into the first, and solving for s_{11} yields a quadratic with roots

$$s_{11} = \frac{-(q - e_{11} - e_{10}) \pm \sqrt{(q - e_{11} - e_{10})^2 + 4qe_{11}}}{2} \tag{20}$$

It remains to be seen which of these is stable and within $[0, 1]$. For stability, the Jacobian

$$J(e) = \begin{pmatrix} -q + s_{10} & s_{11} & 0 \\ -s_{10} & -q - s_{11} & 0 \\ 0 & 0 & -q \end{pmatrix} \tag{21}$$

of the system must have only strictly negative eigenvalues; its eigenvalues are $-q$ (twice) and $s_{10} - s_{11} - q$, so this requires $q + s_{11} - s_{10} > 0$. Substituting in $s_{10} = e_{11} + e_{10} - s_{11}$ this is $q + 2s_{11} - e_{11} - e_{10} > 0$; substituting for s_{11} from (20) this is $\pm\sqrt{(q - e_{11} - e_{10})^2 + 4qe_{11}} > 0$. Thus only the + root can be stable and it is stable unless $e_{10} = q$ and $e_{11} = 0$.

Increasing returns. Let $\overline{s}$ be a stable steady state for given $\overline{e}$ and perturb this to $e_{10} = \overline{e}_{10} + \epsilon$, and $e_{00} = \overline{e}_{00} - \epsilon$. Totally differentiating with respect to ϵ,

$$q\frac{\partial s_{11}}{\partial \epsilon} = \frac{\partial s_{10}}{\partial \epsilon}s_{11} + s_{10}\frac{\partial s_{11}}{\partial \epsilon} \tag{22}$$

$$q\frac{\partial s_{10}}{\partial \epsilon} = q - \frac{\partial s_{10}}{\partial \epsilon}s_{11} - s_{10}\frac{\partial s_{11}}{\partial \epsilon} \tag{23}$$

whose solution is

$$\frac{\partial s_{11}}{\partial \epsilon} = \frac{s_{11}}{q + s_{11} - s_{10}} \tag{24}$$

which is positive since $q + s_{11} - s_{10} > 0$ for stability. Likewise

$$\frac{\partial s_{10}}{\partial \epsilon} = 1 - \frac{\partial s_{11}}{\partial \epsilon} = \frac{q - s_{10}}{q + s_{11} - s_{10}} \tag{25}$$

The denominator is again positive by the stability condition. With some effort the numerator can also be shown to be positive. Specifically, from the + root of Equation 20 and $s_{10} = e_{11} + e_{10} - s_{11}$ we have, after re-arranging some of the terms under the square root,

$$s_{10} = \frac{(q + e_{11} + e_{10}) - \sqrt{(q + e_{11} + e_{10})^2 - 4qe_{10}}}{2} \tag{26}$$

When $e_{11} = 0$ this reduces to

$$s_{10} = \begin{cases} q & \text{if } q \leq e_{10} \\ e_{10} & \text{if } q > e_{10} \end{cases} \tag{27}$$

so that $q \geq s_{10}$ for all e_{10}. Combining this with the fact that

$$\frac{\partial s_{10}}{\partial e_{11}} = \frac{1}{2}\left(1 - \frac{(q + e_{11} + e_{10})}{\sqrt{(q + e_{11} + e_{10})^2 - 4qe_{10}}}\right) \leq 0 \tag{28}$$

we can conclude that $q \geq s_{10}$ always. This lets us conclude that $\frac{\partial s_{10}}{\partial \epsilon} \geq 0$.

Turning to second derivatives,

$$q\frac{\partial^2 s_{11}}{\partial \epsilon^2} = \frac{\partial^2 s_{10}}{\partial \epsilon^2} s_{11} + 2\frac{\partial s_{11}}{\partial \epsilon}\frac{\partial s_{10}}{\partial \epsilon} + s_{10}\frac{\partial^2 s_{11}}{\partial \epsilon^2} \tag{29}$$

Substituting $\frac{\partial^2 s_{11}}{\partial \epsilon^2} = -\frac{\partial^2 s_{10}}{\partial \epsilon^2}$ and solving yields

$$\frac{\partial^2 s_{11}}{\partial \epsilon^2} = \frac{2s_{11}(q - s_{10})}{(q + s_{11} - s_{10})^3} \tag{30}$$

which is also positive, again invoking the stability requirement $q + s_{11} - s_{10} > 0$ and the fact derived above that $q \geq s_{10}$.

Proof of Lemma 2

Defining the probability that a worker with a foundational education has acquired the viral skill

$$\delta(e) \equiv \frac{\phi\rho s_{11}(e)}{\lambda + \phi\rho s_{11}(e)} \tag{31}$$

The private return to education is simply $\frac{\pi}{\lambda}\delta(e)$ and this evidently increases with ϵ since $\frac{\partial s_{11}}{\partial \epsilon} > 0$. Social welfare can be written as

$$\frac{\pi}{\lambda}[e_{11} + \delta(e)e_{10}] - c_1(e_{11} + e_{10}) - c_2(e_{11}) \tag{32}$$

which has second derivative with respect to ϵ proportional to $2\frac{\partial \delta(e)}{\partial \epsilon} + e_{10}\frac{\partial^2 \delta(e)}{\partial \epsilon^2}$, or

$$\left[\frac{2\lambda\phi\rho}{(\lambda + \phi\rho s_{11})^2}\right]\frac{\partial s_{11}}{\partial \epsilon} + \left[\frac{e_{10}\lambda\phi\rho}{(\lambda + \phi\rho s_{11})^2}\right]\frac{\partial^2 s_{11}}{\partial \epsilon^2} - \left[\frac{2e_{10}\lambda(\phi\rho)^2}{(\lambda + \phi\rho s_{11})^3}\right]\left(\frac{\partial s_{11}}{\partial \epsilon}\right)^2 \tag{33}$$

Multiply this by $(q + s_{11} - s_{10})^2(\lambda + \phi\rho s_{11})^3$, divide by $\left(2\lambda\phi\rho\frac{\partial s_{11}}{\partial \epsilon}\right)$, and substitute in (24) and (30) to get

$$\begin{aligned}
&(q + s_{11} - s_{10})^2(\lambda + \phi\rho s_{11}) + e_{10}(q - s_{10})(\lambda + \phi\rho s_{11}) - e_{10}\phi\rho s_{11}(q + s_{11} - s_{10}) \\
&\quad = (q + s_{11} - s_{10})^2(\lambda + \phi\rho s_{11}) + e_{10}(q - s_{10})\lambda - e_{10}\phi\rho s_{11}^2 \\
&\quad = \phi\rho\left[(q + s_{11} - s_{10})^2(q + s_{11}) + e_{10}(q - s_{10})q - e_{10}s_{11}^2\right] \qquad (34)
\end{aligned}$$

As $s_{10} < q$ and all terms within the square brackets other than q are bounded, this is unambiguously positive for q sufficiently large.

Proof of Proposition 4

Clearly if $\frac{\pi}{\lambda} \geq c_1 + c_2$ it is not an equilibrium for any agent to get no education, so $e_{11} + e_{10} = 1$. Conditional on all agents obtaining foundational education the returns to viral education are decreasing continuously as in Section 2.1 and so there must be a unique equilibrium.

If $\frac{\pi}{\lambda} < c_1 + c_2$ it is never optimal for an agent to pay to learn both skills, so $e_{00} + e_{10} = 1$. By Lemma 2 the returns to foundational education in this context are non-decreasing and hence either $e_{00} = 1$ or $e_{10} = 1$ (there may be a third, interior equilibrium but it will be unstable). The former is always an equilibrium since when $e_{00} = 1$ it must be that $s_{11} = 0$ and so there is no return to obtaining only a foundational education. For the latter to be an equilibrium it most hold that $\Lambda(e_{10} = 1) \geq c_1$. It follows from (27) that if $q \geq 1$ then the only stable steady state when $e_{10} = 1$ has $s_{11} = 0$ and so there is no return to foundational education. If instead $q < 1$ then the return to foundational education is

$$\frac{\pi}{\lambda}\left(\frac{\phi\rho(1-q)}{\lambda+\phi\rho(1-q)}\right) = \frac{\pi}{\lambda}\left(1-\frac{\lambda}{\phi\rho}\right) \tag{35}$$

which exceeds c_1 if $\phi\rho > \lambda\pi/(\pi - c_1\lambda)$. Since this condition also implies $q < 1$ it is both necessary and sufficient.

Proof of Proposition 5

Social welfare is

$$e_{11}\left(\frac{\pi}{\lambda} - c_1 - c_2\right) + e_{10}\left(\frac{\phi\rho s_{11}(e)}{\lambda+\phi\rho s_{11}(e)}\frac{\pi}{\lambda} - c_1\right) \tag{36}$$

If $\pi/\lambda \leq c_1$ then this expression is at most 0 (when $e_{11} = e_{10} = 0$), so no foundational education is optimal. If $\pi/\lambda \geq c_1 + c_2$ then social welfare is increasing in e_{11} holding e_{10} fixed and so $e_{11} < 1 - e_{10}$ cannot be optimal. There remains the intermediate case $c_1 + c_2 > \pi/\lambda > c_1$.

Without spillovers ($\phi\rho = 0$) there is no value to providing any form of education other than comprehensive education that includes both foundational and viral skills, so $e_{10} = e_{01} = 0$. Social welfare is

$$e_{11}\left(\frac{\pi}{\lambda} - c_1 - c_2\right) \tag{37}$$

which is evidently maximized at $e_{11} = 0$. Now consider the case with positive spillovers; let $\delta^c \equiv \frac{\lambda c_1}{\pi} < 1$ be the critical value of δ at which the return to providing foundational education e_{10} turns positive. Suppose that $\delta(\tilde{e}_{11}, \tilde{e}_{10}) = \delta^c$ for some $(\tilde{e}_{11}, \tilde{e}_{10})$. Then since $\frac{\partial\delta}{\partial e_{10}} > 0$ it cannot be optimal to have $e_{10} \in [\tilde{e}_{10}, 1 - \tilde{e}_{11})$, since the net return is positive in that range. Consequently either $e_{10} < \tilde{e}_{10}$ or $e_{10} = 1 - \tilde{e}_{11}$. Moreover, $0 < e_{10} < \tilde{e}_{10}$ cannot be optimal since it yields a payoff of

$$e_{11}\left(\frac{\pi}{\lambda} - c_1 - c_2\right) + e_{10}\left(\delta\frac{\pi}{\lambda} - c_1\right) \tag{38}$$

both terms of which are negative when $\delta < \delta^c$, so that $e_{11} = e_{10} = 0$ does better. We can conclude, therefore, that the optimum is either no education ($e_{11} = e_{10} = 0$) or universal foundational education ($e_{11} + e_{10} = 1$); universal foundational education is optimal if and only if there is a universal foundational education policy that yields strictly positive social surplus. By applying the envelope theorem to the social value function (36) we see that this property is monotone in $\phi\rho$, so to establish the existing of a threshold $(\phi\rho)^*$ it is sufficient to show that it holds for $\phi\rho$ large enough. Consider therefore the case $e_{10} = 1$ and $\phi\rho > \frac{\lambda}{1-\lambda}$, so that $q < 1$. It follows from (27) that in this case the unique stable equilibrium is $s_{11} = e_{10} - q = 1 - q$. Social welfare is

$$\frac{\phi\rho(1-q)}{\lambda + \phi\rho(1-q)} \frac{\pi}{\lambda} - c_1 = (1-q)\frac{\pi}{\lambda} - c_1 \tag{39}$$

Since $q \to 0$ as $\phi\rho \to \infty$ this is positive for $\phi\rho$ sufficiently large.

To show that $(\phi\rho)^*$ may be less than $\frac{\lambda\pi}{\pi - \lambda c_1}$ we will show that the social welfare generated by universal education can be strictly positive when $\phi\rho = \frac{\lambda\pi}{\pi - \lambda c_1}$ and then appeal to continuity in $\phi\rho$. First note that when $\phi\rho = \frac{\lambda\pi}{\pi - \lambda c_1}$ then setting $e_{10} = 1$ yields social welfare of 0, since the first condition is the one that sets the expected payoff of agents with only foundational education to 0. Now consider the social return to the viral skill evaluated at this point:

$$\frac{\pi}{\lambda}(1 - \delta(e)) - c_2 + \frac{\pi}{\lambda}\frac{\partial\delta(e)}{\partial e_{11}} = \left(\frac{\pi}{\lambda} - c_1 - c_2\right) + \frac{\pi}{\lambda}\frac{\partial\delta}{\partial e_{11}}$$

The second term is strictly positive and the first can be made arbitrarily close to 0 by decreasing c_2; hence for c_2 small enough the social (but not the private) return to learning viral skills is positive. This implies that the planner can achieve strictly positive social welfare for some $e_{11} \in (0, 1)$ as required.

B Diffusion Through Team Production

This section outlines a team-theoretic microfoundation for the assumptions about knowledge spillovers made in the main text. The idea is straight-forward: sharing knowledge is costly, but agents occasionally have short-term incentives to share because they are matched with peers for *team production*. Compensating an individual based on aggregate output yields sub-optimal effort incentives but also generates positive incentives for "helping," and in particular for sharing knowledge. To take a familiar example, a professor has incentives to teach skills to her research assistant that will make him more productive and increase their joint research output.

Introducing team production has the attractive consequence that whether skills do or do not spill over is determined by the interaction of individual optimization with the team production function, adding empirical content to the model's driving assumptions. However, team production also becomes a new source of returns to knowledge that must be accounted for. As I will discuss, Kremer (1993) and Acemoglu (1996) have already studied the aggregate implications of these returns, and I will not repeat that analysis here. Instead I take the limit of the team production model as the benefits of team interaction and the costs of sharing knowledge fall proportionally; this preserves incentives for knowledge-sharing and yields, asymptotically, the simpler model studied in the main text.

Consider two agents i and j who have been matched together. In the case of the viral skills model, let $s(i)$ indicate whether agent i possesses the viral skill. In the case of the foundational skills model there are potentially two skills to track. Since only agents with the foundational skill can learn the viral skill, however, we may as well assume that both agents possess the foundational skill, and so again let $s(i)$ indicate whether agent i possesses the viral skill. Then i and j produce joint output given by a symmetric production function

$$y = 2\theta f(s(i), s(j)) \tag{40}$$

where $\theta \sim G$ is a stochastic productivity shock. The scaling factor 2 is purely to simplify subsequent expressions. The agents share output between them so that at least one agent's payoff will typically depend on the other agent's skill level. (This property would fail to hold only if the production function were additively separable and both agent's contributions to total output were perfectly distinguishable.) For simplicity I assume that output is shared equally between the workers, so that both receive $\theta f(s(i), s(j))$. Acemoglu (1996, Appendix 1) provides a bargaining/search micro-foundation for proportional profit sharing in a similar environment.

Team production generates an incentive for knowledge-sharing, which agents weigh against the costs of communicating. Let $\tilde{c}$ be the cost to a skilled agent of communicating this skill to the unskilled agent. Then when a skilled and an unskilled worker are matched the skilled worker will teach his skill to the unskilled worker if $\theta[f(1,1) - f(1,0)] \geq \tilde{c}$, so the probability of a spillover is $\rho \equiv \mathbb{P}_G\left(\{\theta : \theta[f(1,1) - f(1,0)] \geq \tilde{c}\}\right)$.

While this nearly completes the analysis, the payoffs from team interactions must also be taken into account when computing value functions. I illustrate this for the viral case; the analysis for the foundational case is exactly analogous. Let $\overline{f}_{s,s'}$ be the mean instantaneous payoff that an agent with skill level s obtains from interacting with an agent with skill s', net of any communication costs incurred.

$$\begin{aligned}
\overline{f}_{11} &= \mathbb{E}_G[\theta f(1,1)] \\
\overline{f}_{10} = \overline{f}_{01} &= \mathbb{E}_G[\max\{\theta f(1,1) - \tilde{c}, \theta f(1,0)\}] \\
\overline{f}_{00} &= \mathbb{E}_G[\theta f(0,0)]
\end{aligned}$$

Then

$$\begin{aligned}
\lambda V(1,e)dt &= \left(\pi + \phi[s(e)\overline{f}_{11} + (1 - s(e))\overline{f}_{10}]\right) dt \\
\lambda V(0,e)dt &= \left(\phi[s(e)\overline{f}_{01} + (1 - s(e))\overline{f}_{00}] + \phi s(e)\rho(V(1,e) - V(0,e))\right) dt
\end{aligned}$$

and the private return to becoming skilled is

$$\Delta(e) = \frac{\pi + \phi s(e)(\overline{f}_{11} - \overline{f}_{01}) + \phi(1 - s(e))(\overline{f}_{10} - \overline{f}_{00})}{\lambda + \phi\rho s(e)}$$

The $s(e)$ in the denominator represents the saturation effect, which remains a source of decreasing returns. There is a second force, however, determined by the sign of $\overline{f}_{11} - \overline{f}_{01} - \overline{f}_{10} + \overline{f}_{00}$; if $\overline{f}$ is supermodular then this will tend to generate increasing returns. When will $\overline{f}$ be supermodular? At a minimum the underlying production function f must itself be supermodular, since the opportunity to communicate makes one agent's knowledge *more* substitutable for the other's than it would be without communication.[25] Skills could be substitutes in production if, for example, one worker needs to read a manual to figure out what both should do. Alternatively skills could

[25] If f is submodular then

$$\begin{aligned}
\overline{f}_{11} + \overline{f}_{00} - 2\overline{f}_{10} &= \overline{\theta} f(1,1) + \overline{\theta} f(0,0) - 2\mathbb{E}_G[\max\{\theta f(1,1) - \tilde{c}, \theta f(1,0)\}] \\
&\leq \overline{\theta} f(1,1) + \overline{\theta} f(0,0) - 2\overline{\theta} f(1,0) \\
&\leq 0
\end{aligned}$$

be complements if, as in Kremer (1993), both agents need to perform a task well in order for the final output to have any value. Kremer shows how this assumption combined with imperfect matching of workers to each other generates aggregate increasing returns. Acemoglu (1996) presents a related argument in which imperfect matching of skills to machines generates increasing returns to both factors. The potential for increasing returns here is due precisely to this mechanism.

One can suppress the Kremer-Acemoglu mechanism and isolate the spillover effects of interest by making the costs and benefits of team interaction small in proportion to the other benefits of becoming skilled. Formally, let $f = \kappa\hat{f}$ and $\tilde{c} = \kappa\hat{c}$ for some fixed production function f and cost $\hat{c}$. The probability ρ of a spillover is invariant to κ, but $\overline{f}_{s,s'} \to 0$ as $\kappa \to 0$ for any (s, s'). The model studied in the text is thus a limit case of the teams model as the relative importance of team interactions falls.