

# Technological Progress under Learning by Imitation.

Morgan Kelly\*

University College Dublin and CEPR

## Abstract

We analyse technological progress when knowledge has a large tacit component so that transmission of knowledge takes place through direct personal imitation. It is shown that the rate of technological progress depends on the number of innovators in the same knowledge network. Assuming the diffusion of knowledge to mirror the geographical pattern of trade—the greater the trade between two sites, the greater the probability that technical knowledge flows between them—we show that a gradual expansion of trade causes a sudden rise in the rate of technological progress.

JEL: O40

## 1 Introduction.

Underlying current models of technological progress is the assumption that researchers stand on the shoulders of giants by having costless access to the entire stock of human knowledge. Much of technology however, and technical skill in particular, has a large tacit, “do it like this” element. To master a skill, it is not in general sufficient to have access to blueprints, or textbooks and articles in a library: you must also have contact with people who already possess the skill; and the extent of their mastery will determine, in part, the extent of yours. This paper looks at how technology progresses when individuals learn by direct personal contact.

---

\*This paper is part of the International Trade and Investment Programme of the Geary Institute at UCD. I should like to thank the referees for their detailed and constructive criticisms of the submitted draft. All errors are mine.

Learning through imitation takes place as a follow the leader process. The skill of each individual equals the skill of the most able practitioner he knows in the previous generation, minus an imitation error that reflects individual ability. Innovators do not get to stand on the shoulders of giants, but only on the shoulders of the tallest person they know.

In Section 2 we show that the rate of technological progress under learning by imitation depends on the difficulty of imitation, and on the size of the population of innovators  $N$ . The larger is a connected group of innovators, the greater the chance that a highly able person in one generation will be matched to a highly able person in the previous generation whose skills will serve as a foundation for further progress in technology.

The influential analysis of scale effects in growth models of Jones (1995) assumes that the growth of technology is an increasing function of the labour force involved in R&D, and a diminishing function of the level of technology. The model here gives microfoundations for this specification based on a concrete form of the spillover from aggregate to individual human capital considered by Lucas (1988), namely direct personal imitation.

What determines the size of population of imitators  $N$ ? Rather than arbitrarily equating  $N$  with the population of some political or geographical unit, we suppose that the diffusion of technology mirrors the geographical pattern of trade: the larger the volume of trade between two sites, the greater probability that innovators at one site have access to technological knowledge at the other.

What we are calling imitation here is closely related to the idea of familiarization with the frontier technology in Aghion, Comin and Howitt (2006) and familiarity in Goodfriend and McDermott (1998). Technical knowledge can be transmitted through trade in a variety of ways. Flows of goods imply flows of people, allowing potential innovators to meet directly. More important, seeing a new good allows domestic producers to incorporate some of its physical features into their product, or to infer the process used to produce the good and modify their own processes accordingly. A familiar example of this process is the way that American car firms in the 1980s, faced with cheaper, more reliable Japanese imports, began to imitate the quality control and inventory practices of Japanese firms.

Trade here is assumed to follow a gravity model. In Sections 3 and 4 each site has the same population, allowing a closed form solution for the volume of trade in terms of distance, population, and transport cost. The flow of knowledge between cities creates a network where knowledge can diffuse along a chain from one connected city to another, where two cities that are not directly linked can still have knowledge flow between them through a third city to which each is directly connected.

We show that the size of the knowledge network depends critically on the volume of trade. Below a critical level of trade, the economy is split into small, fragmented knowledge networks. As the critical trade volume is reached, these local clusters coalesce into a large network that spans most sites in the economy. Consequently, a gradual rise in trade, as a result of increasing population or reduced transport cost, gives rise to a sudden increase in the rate of technical progress, as the critical number of knowledge links is reached, and the connected population of innovators  $N$  suddenly rises.<sup>1</sup> While our concern is with learning by imitation, exactly the same takeoff can occur in standard growth models where trade gives access to more intermediate goods (Rivera-Batiz and Romer, 1991). The emphasis here is on sudden growth accelerations: the model has nothing to say about reasons why different economies should have different equilibrium levels of income or rates of growth.

In Section 5 we relax the assumption that there is a fixed number of cities all with equal population, by allowing the number of cities, and the population of each, to grow through time. We show that this simple mechanism gives rise to an empirically realistic truncated Pareto distribution or power law for urban population, and that knowledge networks continue to show threshold behaviour.

While our central concern is to understand the pattern of technological progress when knowledge is tacit, the logic of learning by imitation implies that technological regress can occur if population falls. Section 6 considers technological retrogression in the context of the collapse of historical societies, showing how a self-reinforcing cycle of urban flight can cause knowledge networks to collapse.

---

<sup>1</sup>This prediction that a fall in transportation costs can lead to a sudden spurt of innovation is consistent with the rise in innovation

While the purpose of this paper is to develop a theoretical model of growth when innovators learn by imitating each other, its central empirical prediction, that a gradual expansion of trade should be associated with a sudden spurt of innovation, appears to accord with reality. In Sung Dynasty China the emergence of a national market linked by waterways was associated with a surge in innovation (Kelly, 1997), just as in eighteenth century England with the extension of turn-pikes and canals. Similarly in nineteenth century America, the completion of the Erie Canal led to a rise in patenting in adjoining areas (Sokoloff, 1988); while the first wave of globalization at the end of the nineteenth century was associated with a surge in the development and commercialization of fundamental technologies in the 1880s (Smil, 2005).

This paper draws from several different literatures: tacit knowledge, endogenous growth, interacting systems, trade and knowledge diffusion, and geography and trade. The tacitness of technical skill is stressed by Nelson and Winter (1982). The link to the endogenous growth literature, particularly Jones (1995) and Lucas (1988), was mentioned above. The effect of market expansion on growth through increased division of labour is examined by Becker and Murphy (1992), Goodfriend and McDermott (1995), and Murphy, Shleifer and Vishny (1989), while McDermott (2002) emphasises the role of trade in development, albeit through increasing returns technologies; and the takeoff caused by threshold effect in knowledge networks is analogous to the phase transitions surveyed by Brock and Durlauf (2001). The connection between trade and technology diffusion is reviewed by Keller (2004), while the modelling of geography and trade here follows Anderson and van Wincoop (2003) and Fujita, Krugman and Venables (1999).

## 2 Learning by Imitation.

There is a single general purpose technology that determines the quality of all goods produced: multiple technologies do not change things materially but need more notation.  $N$  producers use this technology. In Sections 3 and 4 we determine  $N$ , but for now we treat it as given.

Individuals devote a fixed amount of effort to acquiring human capital. Ability depends on the individual's personal quality, and on the ability of the person he im-

itates. The ability  $a_{it}$  of individual  $i$  in generation  $t$  reflects his success in imitating the most able individual he knows in the previous generation whose ability is  $a_{t-1}^{\max}$

$$a_{it} = e_{it} a_{t-1}^{\max} \quad (1)$$

where  $e_{it}$  represents an imitation error: most agents will not be as able as their exemplar (the average Delta guitarist was inferior to Charlie Patton, the average MIT PhD was not as smart as Paul Samuelson) so in most cases  $e_{it} < 1$ .<sup>2</sup>

Taking logs, where  $\alpha_{it} = \log a_{it}$  and  $\varepsilon_{it} = -\log e_{it}$ , the minus highlighting the fact that it is an imitation error and people are usually worse than their exemplar,

$$\alpha_{it} = \alpha_{t-1}^{\max} - \varepsilon_{it}. \quad (2)$$

This log error is normally distributed across imitators with mean  $\mu_t$  and variance  $\sigma_t^2$ ,  $\varepsilon_{it} \sim N(\mu_t, \sigma_t^2)$ .

The distribution of log ability  $\alpha$  in each generation is therefore normally distributed with mean

$$A_t = E(\alpha_{t-1}^{\max}) - \mu_t \quad (3)$$

and variance  $\sigma_t^2$ . Define the growth rate of average human capital or technology as  $g_t = E(a_t)/E(a_{t-1}) - 1$ . For a given difficulty of imitation reflected in the parameters  $\mu$  and  $\sigma$ , the rate of technical progress depends on the size of the pool of innovators:

**Proposition 1.** *If the variance of learning is constant,  $\sigma_{t-1}^2 = \sigma_t^2$ , the rate of technical progress under learning by imitation is*

$$g_t = \left( -c_1 + c_2 \sqrt{\log N_{t-1}} \right) \sigma_{t-1} - \mu_t. \quad (4)$$

---

<sup>2</sup>The role of the most talented person in the previous generation recalls [Murphy, Shleifer and Vishny \(1991\)](#), although the analysis here hinges on the fact that the linkage is stochastic rather than deterministic. The multiplicative, rather than additive error term, allows the relative variation of talent to remain constant (people are only, say, half as good on average as their exemplar) as ability rises.

*Proof.* If there were  $N_{t-1}$  producers last period, the expected value of the first term in (3), from Cramér (1946, 374–375), is

$$E(\alpha_{t-1}^{\max}) = A_{t-1} + \varsigma_{t-1} \left( \sqrt{2 \log N_{t-1}} - \frac{\log(\log N_{t-1}) + \log 4\pi + 2E(v_1)}{2 \sqrt{2 \log N_{t-1}}} \right) \quad (5)$$

where  $-v_1$  has log-gamma distribution with expected value  $E(v_1) = 0.577$  (Johnson, Kotz and Balakrishnan, 1995, 89–90).

The expression in brackets in (5) can be approximated to an accuracy of more than 0.01 percent for  $N > 20$  by  $(-c_1 + c_2 \sqrt{\log N_{t-1}})$ , where  $c_1 = 0.624$  and  $c_2 = 1.482$ . It follows from (3) and (5) that the change in average log ability from one generation to the next is  $A_t - A_{t-1} = (-c_1 + c_2 \sqrt{\log N_{t-1}}) \sigma_{t-1} - \mu_t$ .

Because  $\log E(a_t) = A_t + \sigma_t^2$ , if  $\sigma_{t-1}^2 = \sigma_t^2$  the change in expected log ability is approximately equal to the growth in the level of average skill  $A_t - A_{t-1} = \log E(a_t) - \log E(a_{t-1}) \approx g_t$ .  $\square$

Note that the growth rate is increasing in  $\sigma$  the variability of the imitation error. Just like a call option, all that matters is upside potential: the possibility for very smart imitators to appear.

Imitation difficulty, reflected in the parameters  $\mu_t$  and  $\sigma_{t-1}$ , depends on the characteristics of the technology, and on the characteristics of the population of imitators. Imitation difficulty rises with the level of technology when technological advance consists of refinements of existing techniques, with each advance demanding greater exactitude: steel making requires more precise control of temperature and selection of raw materials than iron smelting; and screws require more precise machining than nails. Alternatively,  $\mu$  may rise and  $\sigma$  fall through what Jones (1995) calls “fishing out”: the easiest innovations are made first, so that a constant rate of innovation requires greater effort. By contrast, serendipitous technological advances that result from a single clever insight, such as sewing needles, stirrups or double-entry book keeping, are trivially imitated once invented.

Imitation difficulty also reflects the characteristics of producers: their skills, needs, and cognitive patterns; and the social institutions they inhabit. The greater is the skill of individual producers, through formal education, division of labour,

and the efficiency that individuals skills are matched with occupations; the lower will be the difficulty of imitation for a given technology.<sup>3</sup>

If  $\mu$  is increasing and  $\sigma$  diminishing in the level of technology  $A$ , the basic growth equation (4) implies that the rate of technological progress is increasing in the number of innovators  $N$  and diminishing in the level of technology  $A$ ,  $g = f(N, A)$ . This is the form of the technological progress function assumed by [Kremer \(1993\)](#) and [Jones \(1995\)](#). By giving a concrete form to the spillover from aggregate to individual human capital considered by [Lucas \(1988\)](#), namely personal imitation, the approach here delivers microfoundations for that analysis. The rate of technological progress is slightly different: the change in average ability rather than in the total stock of knowledge, reflecting personal transmission of skill instead of access to the entire stock of human knowledge.

Following [Romer \(1990\)](#) we have treated knowledge as a public good by allowing everyone to imitate the most able practitioner in the previous generation, but nothing vital hinges on this. If imitators follow social prestige rather than technical ability and imitate the  $k$ -th most able individual in the previous generation who happens to have the highest social status, equation (5) still holds, except that now  $-E(v_1)$  is replaced by  $-E(v_k) = 0.577 - \sum_{j=1}^{k-1} 1/j$  so that equation (4) for the change in average ability holds but with slightly different values of  $c_1$  and  $c_2$ .

If individuals are assigned to mentors at random—for instance if ability has a small heritable component and if skills are transmitted within families—(5) still holds, averaging over  $E(v_k)$  by the fraction assigned to mentors at each rank in the distribution. Finally, if there is sorting of learners by ability, in the manner of university admissions, with the top few percent assigned to the most able mentor, the next few percent to the second most able, and so on; the analysis continues to hold, focusing on each group of learners separately. There will be one version of equation (4) for the top group; another, typically with different values of  $\mu$  and  $\sigma^2$ , for the second group; and so on.

---

<sup>3</sup>[Mokyr \(2002\)](#) emphasises the importance of the empirical method taken from science, which [Margolis \(1987\)](#) traces back to Copernicus, for the development of European technology; while [Huff \(2003\)](#) argues that the development of autonomous intellectual institutions is what gave European science a resilience absent in Islam and China.

### 3 A gravity model of knowledge diffusion.

From (4) the growth of technical skill under imitation depends critically on the size of the pool of innovators  $N$ . The larger is the connected network of innovators, the greater the chance that a high ability individual in one generation will have the opportunity to acquire the skills of a high ability person in the previous generation. What determines  $N$ ?

We assume that the diffusion of technical knowledge reflects the pattern of trade, and model trade patterns in a standard gravity model with CES preferences (Anderson and van Wincoop, 2003).<sup>4</sup> To derive closed form solutions, we make things symmetric across sites. There is a fixed number  $C$  of cities spread at random on a plain of area  $R$ , each with the same population  $n$ . We show that the results continue to hold with a more realistic distribution of population in Section 5.

Each city specializes in the production of a subset of goods that, by symmetry, we can think of as one good. Each good comes in a range of qualities  $\alpha_{ik}$  reflecting the skill of worker  $k$  at site  $i$  that produced it. If  $x_{ijk}$  is the amount of region  $i$  good of quality  $k$  consumed in region  $j$ , region  $j$  consumers choose  $x_{ijk}$  to maximize

$$\left( \sum_i \alpha_{ik} x_{ijk}^{(\epsilon-1)/\epsilon} \right)^{\epsilon/(\epsilon-1)} \quad (6)$$

subject to the budget constraint  $\sum_i p_{ijk} x_{ijk} = y_j$ . We suppose that every location  $j$  receives equal quantities of the output of each worker so that the price of each unit of good  $i$  is proportional to its log quality  $p_{ijk}/p_{ijl} = \alpha_k/\alpha_l$ . Therefore we can suppose that each region produces a homogeneous good of quality  $\bar{\alpha}_i = E(\alpha_i)$ , the average quality of producers at the site. Goods incur iceberg costs in transit: of one unit of a good shipped from  $i$  to  $j$ , a fraction  $1 - 1/t_{ij}$  is lost in transit so  $p_{ij} = t_{ij} p_i$ .

Expenditure on good  $i$  at site  $j$  is

$$e_{ij} = \left( \frac{p_i t_{ij}}{\bar{\alpha}_i P_j} \right)^{1-\epsilon} y_j \quad (7)$$

---

<sup>4</sup>Eaton and Kortum (2002) develop a model of gravity and technology, but in their framework the technology of each site is drawn from a Frechet distribution independently of its neighbors.



where the price index

$$P_j = \left[ \sum_i (p_i t_{ij} / \bar{\alpha}_i)^{1-\epsilon} \right]^{1/(1-\epsilon)}. \quad (8)$$

Market clearing requires that  $y_j = \sum_i e_{ij}$ . Assuming that transportation costs between sites are symmetric, [Anderson and van Wincoop \(2003\)](#) show that

$$P_j^{1-\epsilon} = \sum_i P_i^{\epsilon-1} \eta_i t_{ij} \quad (9)$$

where  $\eta_i = y_i / \sum_i y_i$ , and derive the gravity equation for the value of goods shipped from  $i$  to  $j$

$$e_{ij} = \frac{y_i y_j}{\sum_i y_i} \left( \frac{t_{ij}}{P_i P_j} \right)^{1-\epsilon}. \quad (10)$$

Following [Fujita, Krugman and Venables \(1999, Chapter 4\)](#) we suppose that producing quantity  $q$  of a good requires  $l = f + cq$  workers, and that units are chosen so that marginal cost  $c = (\epsilon - 1)/\epsilon$ , making  $p_i = w_i$  and  $q = l$ ; while the zero profit condition implies that each site with population  $n$  produces  $n/f\epsilon$  varieties of good ([Fujita, Krugman and Venables, 1999](#), p. 54).

The  $C$  sites have identical populations  $n$  and endowments and differ only in their technology level  $\bar{\alpha}_i$  and transportation costs  $t_{ij}$ . Because our concern here is with the development of technology through time rather than the distribution of economic activity through space, we will assume that these spatial differences are negligible: each site has the same technology level  $\bar{\alpha}$ ; and the same average transport cost

$$T_j = T = \left( \sum_i t_{ij}^{1-\epsilon} \right)^{1/(1-\epsilon)}. \quad (11)$$

Equal transport costs require that cities on the edge of the surface face the same costs as those in the centre. This can be achieved by allowing the the surface to become unbounded so that every site is equally a central point or, by placing the points to be on a sphere rather than a plain so that, again, no point is a central or edge point.

In this symmetric case, from (9) every site has price level

$$P = C^{1/2(\epsilon-1)} T^{1/2} \quad (12)$$

and producer price and wage from (8) of

$$p = \bar{\alpha} C^{1/2(\epsilon-1)} T^{-1/2}. \quad (13)$$

Nominal income at each site is  $y = np$  and each site has a share  $\eta_j = y_j / \sum_i y_i = 1/C$  of world income. The quantity of goods shipped from  $i$  to  $j$ ,  $x_{ij} = e_{ij} / p_i t_{ij}$  is therefore

$$x_{ij} = \frac{n}{T^{1-\epsilon}} t_{ij}^{-\epsilon} \quad (14)$$

If sites  $i$  and  $j$  are a distance  $d_{ij}$  apart, we suppose that transport costs between them are  $t_{ij} = \theta d_{ij}$  for  $j \neq i$  and  $t_{ii} = \theta d_0$  where  $d_0 > 0$ , so that

$$x_{ij} = \frac{n}{\theta D^{1-\epsilon}} d_{ij}^{-\epsilon} \quad (15)$$

where  $D_j = \left( \sum_i d_{ij}^{1-\epsilon} \right)^{1/(1-\epsilon)}$ .

## 4 Knowledge networks.

$C$  cities are spread at random on a surface of area  $R$  giving a settlement density of  $\delta = C/R$ . The probability that technological knowledge flows directly between two cities  $i$  and  $j$  is an increasing function of the volume of trade between them  $\pi_{ij} = h(x_{ij})$ , so from (15)  $\pi_{ij} = g(n/\theta, d_{ij})$  which is increasing in the ratio of population to transport cost  $n/\theta$  and diminishing in distance  $d_{ij}$ .

Take an arbitrary site and label it as the origin. For any other site at location  $y \in \mathbb{R}^2$ , the probability that they are connected is  $g(n/\theta, d_{0y})$ . The number of sites connected directly to the origin is a Poisson process with parameter

$$\nu = \delta \int g(n/\theta, d_{0y}) dy. \quad (16)$$

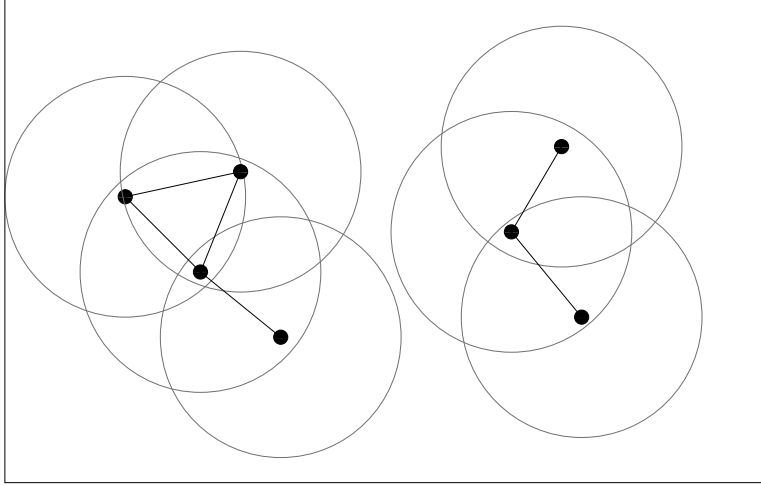


Figure 1: A technological network in the case where each city, represented by a dot, shares knowledge with all neighbours in a fixed radius.

The probability that each site connects directly to  $k$  other sites is

$$p_k = \frac{e^{-\nu} \nu^k}{k!}. \quad (17)$$

To rule out trivial behaviour, we assume  $0 < \nu < \infty$ .

For example, suppose each site shares knowledge directly with all other sites in a circle of radius  $\rho$ : this network is illustrated in Figure 1 on page 11. Then  $h(x_{ij}) = 1$  for  $x_{ij} \geq \bar{x}$  and 0 otherwise; and  $g(n/\theta, r) = 1$  for  $r \leq \rho$  where  $\rho = \left(n/(\theta \bar{x} D^{1-\epsilon})\right)^{1/\epsilon}$ , and 0 otherwise. The average number of knowledge links per site is  $\nu = \delta \pi \rho^2$ . Alternatively, if  $h(x_{ij}) = \left(1 + \exp(x_{ij}^{-1/\epsilon})\right)^{-1}$  the probability of linkage depends logarithmically on distance

$$g(n/\theta, d_{ij}) = \frac{1}{1 + \exp(c^{-1/\epsilon} d_{ij})} \quad (18)$$

where  $c = n/\theta D^{1-\epsilon}$  is the volume of trade between sites that are a unit distance apart. The average number of neighbours linked to each site is then  $\nu = \delta c^{2/\epsilon} \pi^3/6$ .

Each city is linked directly to other cities, that are linked in turn to other cities, giving rise to a connected network of cities through which technical knowledge can diffuse. It is obvious that the larger is the connectivity parameter  $\nu$ , the larger

will be the resulting knowledge networks. What is less immediately obvious is that the size of connected clusters rises discontinuously with  $\nu$  (Meester and Roy, 1996, Theorem 6.1):

**Proposition 2.** *For the number of sites  $C$  large there is a critical value  $\nu^*$  for  $\nu$ . For  $\nu < \nu^*$ , an infinitely large connected cluster of sites exists with probability zero; for  $\nu > \nu^*$ , an infinite cluster exists with probability one.*

The intuition for the result is that as  $\nu$  rises at first, small connected islands first appear and grow slowly. As these islands continue to grow a critical stage is reached when, instead of swallowing up isolated points, they all start to bump into each other and coalesce into a large continent. This sudden threshold behaviour is generic for random graphs (Bollobás and Thomason, 1986).

In consequence the size of networks of potential imitators  $N$  changes suddenly as the volume of trade rises. In economies with low volumes of trade due to low population  $n$  or high transport cost  $\theta$ , the average number of knowledge links per city  $\nu$  is small, and the economy is split into small isolated clusters of communicating cities. This limits the possibility that highly talented individuals in one generation will be matched with highly talented individuals in the previous generation. As the critical number of connections  $\nu^*$  is reached, these isolated knowledge networks rapidly coalesce into a single network that spans most sites in the economy, increasing the pool of innovators who can learn from the most talented producer in the previous generation.

It is not necessary to assume that knowledge diffuses without friction across connected networks. The important point is that there is a sudden increase in network size so that, even if information flows imperfectly, there are still many more innovators in each generation being matched together.

Figure 2 shows a simulation with  $C = 1024$  sites spread at random on a  $32 \times 32$  square so settlement density  $\delta = R/C = 1$ . The connection probability declines logistically with distance (18). Figure 2 plots the fraction of sites that are in the largest cluster against the volume of trade between two sites a unit distance apart, assuming an elasticity of substitution  $\epsilon = 1$ . As the volume of trade rises from 0.5 to 0.7, the fraction of sites in the largest cluster rises from 0.1 to 0.8. Increasing the elasticity  $\epsilon$  reduces the average number of connections  $\nu$  for a given trade volume

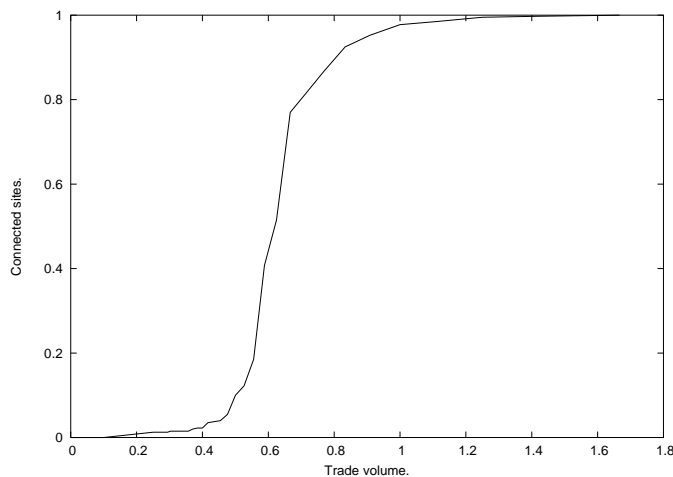


Figure 2: Proportion of sites in largest connected cluster versus trade volume. Trade volume is measured between two sites at distance one apart, assuming an elasticity of substitution  $\epsilon = 1$ .

$c$ , shifting the curve to the right. Figure 3 gives two snapshots of the system when trade between sites one unit apart is 0.55 and 0.6. It can be seen that adding extra linkages causes clusters of sites to link into a single cluster that goes right across the economy. As the number of sites in the network rises, the takeoff becomes more sudden.

Market expansion does not limit economies to just one takeoff in human capital accumulation. Repeated takeoffs can be modelled in two ways. First, a new round of development can occur where points that have connected into a large network can be thought of as being fused together into a single, compound point, and get to join with other compound points in a second round. The development of internal markets through canals and railways in the first half of the nineteenth century linked cities in individual countries into national networks, that can be treated as individual economies. The development of steamships and electric telegraphs in the second half of the nineteenth century joined these internally articulated national economies into an international network.

Alternatively, different levels of technology may have different connectivity functions  $h$ . Simple technologies can have high probabilities of connection  $h$  at low volumes of trade  $x$  allowing a global network to appear early; whereas more

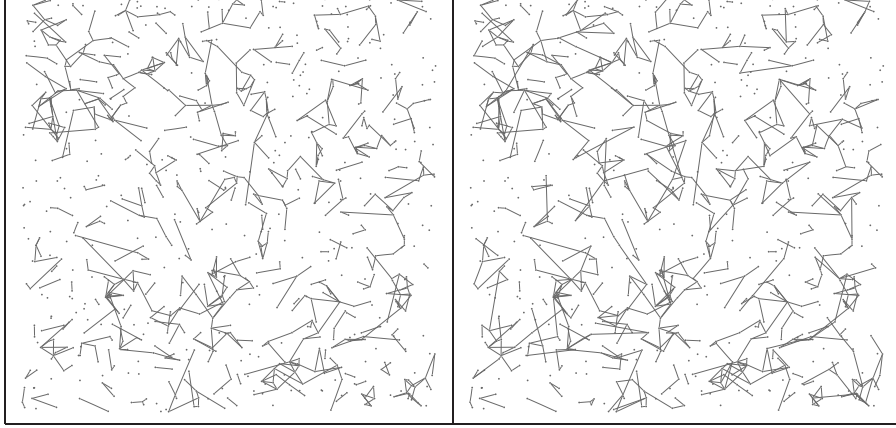


Figure 3: Connections between 1024 cities, when trade between sites one unit apart equals 0.55 and 0.6.

advanced technologies have lower  $hs$ , and require higher populations and lower transportation costs for a connected network to appear.

## 5 Distribution of settlements.

There are two, complementary approaches to analysing the role of cities in economies: as nodes in commercial networks, and as central places supplying services to surrounding areas ([Hohenberg and Lees, 1995](#), 47–73). So far, to make the equations of the gravity model tractable through symmetry, we have focused on cities as trade nodes, assuming that there was a fixed number  $C$  of cities with an equal population  $n$ . We now allow each city to function also as a central place in an urban hierarchy, with its own satellite villages and towns. We suppose that existing settlements give rise to new settlements at a constant rate, and that all settlements grow at a constant rate. This simple process causes population to be distributed across settlements according to a truncated power law.

In an interval of length  $dt$  each existing city gives rise to a new city with probability  $\lambda dt$ . Starting with one settlement at time 0, there will be an expected number  $\exp \lambda T$  after time  $T$  has elapsed. The initial size of each city is  $n_0$  which we normalize to unity.

Following Gibrat's law ([Gabaix, 1999](#); [Mitzenmacher, 2002](#)), settlements grow at a constant multiplicative rate  $\gamma$ : settlements develop independently of each other, and Malthusian pressures and overcrowding do not impede their growth. The Appendix generalizes the growth process to geometric Brownian motion. Integrating the distribution function we have immediately

**Proposition 3.** *At time  $T$  after the first settlement is established, the probability distribution of population  $n$  across settlements is*

$$F(n) = K(1 - n^{-\lambda/\gamma}) \quad 1 \leq n \leq e^{\lambda T} \quad (19)$$

where  $K = (1 - \exp(-\lambda T))^{-1}$ .

In other words, population follows a truncated Pareto distribution or power law. When new cities emerge at the same rate that population grows we have an exponent of minus one: Zipf's law.

Proposition 3 generalizes the standard result that an exponentially growing process observed at exponentially distributed times has a Pareto distribution ([Johnson, Kotz and Balakrishnan 1994](#), 608; [Reed 2001](#)) by allowing time to have a finite beginning. This rules out the usual tail of unboundedly large cities and gives expected city size a finite value

$$E(n) = \frac{K\lambda}{\gamma - \lambda} [-1 + \exp(\lambda - \gamma)T]. \quad (20)$$

Trade is described by the same model of Section 3. Now that sites have different populations  $n_i$  the gravity equation for expenditure (10) continues to hold but the price  $p_i$  of each region's output is weighted by the number of varieties  $n_i/F\epsilon$  it produces; and the price index in larger sites will be lower, reflecting the smaller share of goods that incur transport costs. While closed form solutions are no longer possible, the volume of trade between two sites will be increasing in their populations, and diminishing in the cost of transport. It follows that the probability that a city at the origin with population  $i$  will communicate with a city of population  $j$  at

location  $y$  is  $\pi_{ij} = g(\theta, y, i, j)$  so the probability that a city at the origin communicates with a city at location  $y$  is

$$g(\theta, d_{0y}) = \int_1^{\exp \gamma T} \int_1^{\exp \gamma T} g(\theta, d_{0y}, i, j) f(i) f(j) dj di \quad (21)$$

where  $f$  is the density of city sizes corresponding to the distribution function (19). The analysis of network size then goes through as in Section 4

## 6 Technological retrogression.

The concern so far has been with explaining technological progress under learning by imitation. However, our basic equation for the change in skill (4) also implies that technology can regress if population falls.<sup>5</sup>

From (4) there is a critical population needed to maintain the existing level of technology

$$N^* = \exp \left[ \frac{1}{c_2^2} \left( \frac{\mu_t}{\sigma_{t-1}} + c_1 \right)^2 \right] \quad (22)$$

This critical population increases rapidly with the difficulty of imitating the technology given by the inverse of the coefficient of variation  $\mu/\sigma$ . For  $\mu/\sigma = 3$ ,  $N^* = 395$ ; for  $\mu/\sigma = 5$ ,  $N^* = 1.8$  million. Consequently, if imitation difficulty rises with the level of technology, a fall in population will cause technological retrogression.<sup>6</sup>

### 6.1 Locational choice.

To understand societal collapse, we now allow each household the choice between engaging in market activity, which allows the consumption of tradeable goods but requires tax to be paid, and a rural, subsistence existence that gives reservation

---

<sup>5</sup>The retrogression of technology to simpler forms, together with a fall in the quality and homogeneity of artifacts, are features common to the collapse of urban societies such as Harappa, Mesopotamia, Mycenae, and the Western Roman empire in the Old World; and the Maya, Olmec, Chacoans, and Hohokam in the New (Tainter, 1988, 20).

<sup>6</sup>Henrich (2002) argues that this process can explain the loss of basic technologies among the aboriginal Tasmanians



utility  $\bar{U}_j$ . We suppose for concreteness that there is the same potential population  $n$  at each of the  $C$  sites in the economy.

A household in a city divides expenditure to maximize utility

$$U_i = A_i M^\beta F^{1-\beta} \quad (23)$$

where  $M$  is a basket of manufactured goods with Dixit-Stiglitz utility given by (6) and price index  $P$  (8); and  $F$  represents food and fuel coming from the agricultural sector. The parameter  $A_i$  represents inherent utility of city life and varies randomly across households: this ensures that not everyone deserts the city simultaneously.

Agricultural goods cost  $p_F$  and the household pays a fraction  $\tau$  of its income in tax to the government. Each unit of agricultural goods requires  $l_F$  units of labour to produce so  $p_F = l_F w$  and, from the zero profit condition, household income  $Y = w$ . As before, quantities are normalized so that  $w = p$ . The household receives indirect utility from (12) and (13)

$$\begin{aligned} U_i &= A_i \beta^\beta (1-\beta)^{1-\beta} \frac{(1-\tau)Y}{P^\beta p_F^{1-\beta}} \\ &= A_i \beta^\beta (1-\beta)^{1-\beta} \frac{1-\tau}{l_F^{1-\beta}} \left( \frac{\bar{\alpha}}{\theta D} \right)^\beta. \end{aligned} \quad (24)$$

Welfare is increasing in technology  $\bar{\alpha}$ , and diminishing in tax rates  $\tau$ , the labour requirement of agriculture  $l_F$ , and the cost of transportation  $\theta$ .

Households will desert the city if utility falls below the autarky threshold  $\bar{U}$ . This threshold utility has a distribution function across households of  $H(\bar{U})$  with associated density  $h(\bar{U})$ . If a city has potential population of  $n$ , its actual population reflects the fraction of households whose utility lies above the threshold for urban living  $L = n \int_0^{\bar{U}} h(\bar{U}) d\bar{U}$ .  $L$  now replaces  $n$  as the relevant population term for the volume of trade and network size in Sections 3 and 4.

Any factor that reduces the return to market activity (24) can induce a civilizational collapse if it drives urban population  $L$  below the value needed to maintain the critical number of knowledge linkages  $\nu^*$ . Once the threshold is passed the economy splits into local knowledge networks with small populations below  $N^*$ , causing technological knowledge to regress, and further reducing the utility of ur-

ban living  $U$ . We focus on two causes of urban collapse that have received particular attention: ecological decline, and increasing taxation in response to military pressure.

Deteriorations in climate and ecology increase the labour requirement of agriculture  $l_F$ , reducing the payoff to market activity (24).<sup>7</sup> The impact will be more immediate if utility (23) is generalized to a Geary-Stone form  $U = M^\beta(F - \bar{F})^{1-\beta}$  where  $\bar{F}$  is the subsistence quantity of agricultural goods, so that a fall in agricultural productivity that drives  $l_F \bar{F}$  above households' labour endowments causes cities to be abandoned immediately.

## 6.2 Political factors.

When not a vector of epidemic disease, the destructive power of pre-industrial armies was limited, and the effects of military conflict are principally through taxation and geographical disunity. Military conquest can split knowledge networks, causing regression in isolated clusters such as the cities of Western Europe after the disintegration of the western Roman empire. Even without military collapse, the taxes needed to maintain an effective army can lead to a flight of population from cities, causing urban networks to collapse.

We suppose that the government uses tax revenue to recruit an army of size  $S$ . Suppose that a force of  $S_i$  directly engages an enemy force of size  $E_i$ . It inflicts casualties on the enemy at a rate  $k_S$  while the enemy inflicts casualties at rate  $k_E$ :  $\dot{S}_i = -k_E E_i$ ,  $\dot{E}_i = -k_S S_i$ , so that the loss rate relative to the enemy is  $k_E E_i^2 / k_S S_i^2$ . This is the Lanchester square law: the effective forces, defined in terms of relative casualty rates, on each side are  $k_S S_i^2$  and  $k_E E_i^2$ .

While the outcome of an engagement on part of a battlefield from these equations is deterministic, the outcome of a campaign reflects additional factors such as skill and luck in concentrating forces, disease, hunger, weather, and other fortunes of war. In a campaign where an army of total size  $S$  faces a total enemy force of

---

<sup>7</sup>While there are few civilizations whose decline has not been attributed to climatic change (Tainter, 1988, 44–51), strong evidence implicates prolonged drought in the collapse of the Akkadian and classical Maya, states and the Chacoan pueblo culture (deMenocal, 2001); while salinization due to irrigation has been blamed for the abandonment of Mesopotamian cities (Postgate, 1995, 181); and deforestation appears central to the collapse of Easter Island (Brander and Taylor, 1998).

size  $E$ , we suppose that the probability of victory is proportional to effective forces  $Pi_V = k_S S^2 / (k_S S^2 + k_E E^2)$ . The value of winning is  $V$  which will generally be greater for a defensive war than an offensive one.

There are  $N$  taxpayers with income  $p$ . Each soldier costs  $c_S > 1$  so the direct cost of the army is  $c_S S$ . Given national income of  $Np$  we suppose that the perceived cost to the government of raising each denarius, in taxpayer discontent and defections, is proportional to the share of military spending in national income  $c_\tau c_S S / pN$ , so the cost of spending  $c_S S$  on an army is  $C(S) = c_\tau c_S^2 S^2 / pN$ .

The government's problem is to choose  $S$  to maximize  $Pi_V V - C(S)$  which implies that

$$S = \left( \max \left( \frac{E}{c_S} \sqrt{\frac{k_E k_S p N V}{c_\tau}} - k_E E^2, 0 \right) \right)^{1/2}. \quad (25)$$

This is positive and increasing in the number of enemies  $E$ , their effectiveness  $k_E$ , population  $N$ , and the payoff to victory  $V$ , as long as the first term is positive which will be the case so long as the number of taxpayers is very much larger than the effective enemy force.

The tax rate is  $\tau = c_S S / pN$ . From (25) it follows that the tax rate is of the order  $N^{-3/4}$ . Tax rates rise rapidly as population falls. As a consequence, an epidemic induced decline in population accompanied by increased military pressure, such as occurred in third century Rome and seventh century Byzantium, can set off a cycle where rising taxes induce urban flight, increasing the tax burden on the remaining population.

## 7 Conclusions.

Since Adam Smith's observation that the division of labour is limited by the extent of the market, economics has been aware of the close links between technological skill and trade. The goal of this paper was to use this linkage to provide a set of explicit microfoundations for the production function for technology. Whereas existing models assume that all technical knowledge is available to all researchers (requiring that technical knowledge can be stored, transmitted, and retrieved losslessly and costlessly), this paper began with the premise that technical knowledge has a large tacit component that must be transmitted by direct personal contact.

It showed that under such learning by imitation the rate of technological progress, or regress, depended on the size of the population of innovators sharing the same knowledge network. If knowledge networks reflect trade—the greater the volume of trade between sites, the greater the probability that producers at one site have knowledge of the technology at the other site—we demonstrated a threshold in the size of knowledge networks. As the volume of trade rises to a critical volume because of rising population or lower transport costs, the size of knowledge networks suddenly rise, leading to a jump in the rate of technical progress.

## Appendix: Population distribution under geometric Brownian motion.

Suppose that settlement size evolves as geometric Brownian motion

$$dn(t) = \gamma n(t) + \varsigma n(t)B(t)$$

so that the population of settlements of age  $t$  is lognormally distributed

$$\log n(t) \sim N\left((\gamma - \varsigma^2)t, \varsigma^2 t\right).$$

The age of settlements is exponentially distributed with parameter  $1/\lambda$  and maximum  $T$  so the density of settlement sizes is

$$f(n) = \int_0^T \lambda e^{-\lambda t} \frac{1}{\sqrt{2\pi\varsigma^2 t}} \frac{1}{n} \exp\left(-\frac{(\log n - (\gamma - \varsigma^2)t)^2}{2\varsigma^2 t}\right) dt.$$

Substituting  $u^2 = t$

$$f(n) = \frac{\lambda}{\varsigma} \sqrt{\frac{2}{\pi}} n^{a-\lambda-1} \int_0^{\sqrt{T}} \exp(-au^2 - bu^{-2}) du$$

where  $a \equiv \lambda + (\gamma - \zeta^2)/2\zeta^2$  and  $b \equiv (\log n)^2/2\zeta^2$ . Solving the integral

$$f(n) = \frac{\lambda}{\zeta \sqrt{2a}} n^{a-\lambda-1} \left\{ e^{-2\sqrt{ab}} \Phi \left( \sqrt{2aT} - \sqrt{\frac{2b}{T}} \right) - e^{2\sqrt{ab}} \Phi^c \left( \sqrt{2aT} + \sqrt{\frac{2b}{T}} \right) \right\}$$

where  $\Phi$  is the standard normal distribution and  $\Phi^c = 1 - \Phi$ . Expanding the  $e^{2\sqrt{ab}}$  terms

$$f(n) = \begin{cases} \frac{\lambda}{\zeta \sqrt{2a}} \left\{ n^{a-\lambda-1-d} \Phi \left( \sqrt{2aT} - \sqrt{\frac{2b}{T}} \right) - n^{a-\lambda-1+d} \Phi^c \left( \sqrt{2aT} + \sqrt{\frac{2b}{T}} \right) \right\} & n \geq 1 \\ \frac{\lambda}{\zeta \sqrt{2a}} \left\{ n^{a-\lambda-1+d} \Phi \left( \sqrt{2aT} - \sqrt{\frac{2b}{T}} \right) - n^{a-\lambda-1-d} \Phi^c \left( \sqrt{2aT} + \sqrt{\frac{2b}{T}} \right) \right\} & n < 1 \end{cases}$$

where  $d \equiv \sqrt{2a}/\zeta$ . For  $T$  large and  $\log n$  small relative to  $T$ , the  $\Phi$  term is close to 1 and the  $\Phi^c$  term is close to zero. This gives a density  $f(n)$  that is again close to a power law, with different distributions on either side of the initial city size of 1.

## References

- Aghion, Philippe, Diego Comin and Peter Howitt. 2006. “When Does Domestic Saving Matter for Economic Growth?” Working Paper 12275 NBER. [2](#)
- Anderson, James E. and Eric van Wincoop. 2003. “Gravity with Gravitas: A Solution to the Border Puzzle.” *American Economic Review* 93:170–192. [4](#), [8](#), [9](#)
- Becker, Gary S. and Kevin M. Murphy. 1992. “The Division of Labor, Coordination Costs, and Knowledge.” *Quarterly Journal of Economics* 107:1137–1160. [4](#)
- Bollobás, Bela and A. Thomason. 1986. “Threshold Functions.” *Combinatorica* 7:35–38. [12](#)
- Brander, James A. and M. Scott Taylor. 1998. “The Simple Economics of Easter Island: A Ricardo-Malthus Model of Renewable Resource Use.” *American Economic Review* 88:119–138. [18](#)

- Brock, William and Steven N. Durlauf. 2001. Interactions-Based Models. In *Handbook of Econometrics, Volume 5*, ed. James J. Heckman and Edward Leamer. Amsterdam: North Holland. [4](#)
- Cramér, Harald. 1946. *Mathematical Methods of Statistics*. Princeton: Princeton University Press. [6](#)
- deMenocal, Peter B. 2001. “Cultural Responses to Climate Change during the Late Holocene.” *Science* 292:667–673. [18](#)
- Eaton, Jonathan and Samuel Kortum. 2002. “Technology, Geography, and Trade.” *Econometrica* 70:1741–1779. [8](#)
- Fujita, Masahisa, Paul Krugman and Anthony J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press. [4](#), [9](#)
- Gabaix, Xavier. 1999. “Zipf’s Law for Cities: An Explanation.” *Quarterly Journal of Economics* 114:739–767. [15](#)
- Goodfriend, Marvin and John McDermott. 1995. “Early Development.” *American Economic Review* 85:116–133. [4](#)
- Goodfriend, Marvin and John McDermott. 1998. “Industrial Development and the Convergence Question.” *American Economic Review* 88:1277–1289. [2](#)
- Henrich, Joseph. 2002. Demography and Cultural Evolution: Why Adaptive Cultural Processes Produced Maladaptive Losses in Tasmania. Working paper Department of Anthropology, Emory University. [16](#)
- Hohenberg, Paul M. and Lynn Hollen Lees. 1995. *The Making of Urban Europe, 1000–1994*. Revised ed. Cambridge, MA: Harvard University Press. [14](#)
- Huff, Toby E. 2003. *The Rise of Early Modern Science: Islam, China and the West*. Cambridge: Cambridge University Press. [7](#)
- Johnson, Normal L., Samuel Kotz and N. Balakrishnan. 1994. *Continuous Univariate Distributions, Volume 1*. Second ed. New York: Wiley. [15](#)

- Johnson, Normal L., Samuel Kotz and N. Balakrishnan. 1995. *Continuous Univariate Distributions, Volume 2*. Second ed. New York: Wiley. 6
- Jones, Charles I. 1995. "R&D Based Models of Economic Growth." *Journal of Political Economy* 103:759–784. 2, 4, 6, 7
- Keller, Wolfgang. 2004. "International Technology Diffusion." *Journal of Economic Literature* . Forthcoming. 4
- Kelly, Morgan. 1997. "The Dynamics of Smithian Growth." *Quarterly Journal of Economics* 112:939–964. 4
- Kremer, Michael. 1993. "Population Growth and Technological Change. One Million B.C. to 1990." *Quarterly Journal of Economics* 108:681–716. 7
- Lucas, Jr., Robert E. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22:3–42. 2, 4, 7
- Margolis, Howard. 1987. *Patterns, Thinking, and Cognition*. Chicago: University of Chicago Press. 7
- McDermott, John. 2002. "Development Dynamics: Economic Integration and the Demographic Transition." *Journal of Economic Growth* 7:371–409. 4
- Meester, Ronald and Rahul Roy. 1996. *Continuum Percolation*. Cambridge: Cambridge University Press. 12
- Mitzenmacher, Michael. 2002. A Brief History of Generative Models for Power Law and Lognormal Distributions. Working paper Division of Engineering and Applied Sciences, Harvard University. 15
- Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton: Princeton University Press. 7
- Murphy, Kevin, Andrei Shleifer and Robert Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97:1003–1026. 4

- Murphy, Kevin M., Andrei Shleifer and Robert W. Vishny. 1991. "The Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics* 106:503–530. [5](#)
- Nelson, Richard R. and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge MA: Harvard University Press. [4](#)
- Postgate, J. N. 1995. *Early Mesopotamia*. London: Routledge. [18](#)
- Reed, William J. 2001. "The Pareto, Zipf and other Power Laws." *Economic Letters* 74:15–19. [15](#)
- Rivera-Batiz, Luis and Paul Romer. 1991. "Economic Integration and Endogenous Growth." *Quarterly Journal of Economics* 106:531–555. [3](#)
- Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98:S71–S102. [7](#)
- Smil, Vaclav. 2005. *Creating the Twentieth Century: Technical Innovations of 1867-1914 and Their Lasting Impact*. Oxford: Oxford University Press. [4](#)
- Sokoloff, Kenneth L. 1988. "Inventive Activity in Early Industrial America: Evidence from Patent Records, 1790-1846." *Journal of Economic History* 48:813–850. [4](#)
- Tainter, Joseph A. 1988. *The Collapse of Complex Societies*. Cambridge: Cambridge University Press. [16](#), [18](#)