

GERMS, SOCIAL NETWORKS AND GROWTH

Alessandra Fogli and Laura Veldkamp*

November 2, 2011

Abstract

Social institutions undoubtedly affect a country's economy. But how does this effect operate and how much does it matter for economic development? Using network analysis tools, we explore how different social structures might affect a country's rate of technological progress. The network model also explains why societies with a high prevalence of contagious disease might evolve toward growth-inhibiting social institutions and how small initial differences can produce large divergence in incomes. Empirical work uses differences in the prevalence of diseases spread by human contact and the prevalence of other diseases as an instrument to identify an effect of social structure on technology diffusion.

How people organize themselves as a society undoubtedly affects economic activity and a country's income. But how does this effect operate and how much does it matter for development? Macroeconomists typically overlook findings of sociologists and anthropologists because social characteristics are difficult to observe, to describe formally and to quantify.¹ This paper uses tools from network analysis to explore how different social structures might affect a country's rate of technological progress. The network model also explains why societies might adopt growth-inhibiting institutions and how small differences in disease prevalence can produce large differences in incomes. Motivated by these theoretical findings, we use differences in the prevalence of diseases spread by human contact and the prevalence of other diseases as an instrument to identify an effect of social structure on technology diffusion.

There is a long history of measuring the speed of information or technology diffusion within various kinds of networks (Jackson (2008), Granovetter (2005)). Given these findings, a simple way to explain the effect of social structure on GDP is to show that some types of social networks

*Corresponding author: afogli@umn.edu, Department of Economics, University of Minnesota, 90 Hennepin Ave. Minneapolis, MN 55405. lveldkam@stern.nyu.edu, 44 West Fourth St., suite 7-77, New York, NY 10012. We thank participants at the 2010 SED meetings and seminar participants at USC and Stanford for their comments and suggestions. We thank Corey Fincher and Damian Murray for help with the pathogen data. Laura Veldkamp thanks the Hoover Institution for their hospitality and financial support through the national fellows program. Keywords: growth, development, technology diffusion, economic networks, social networks, pathogens, disease. JEL codes: E02, O1, O33, I1.

¹There is a small economics literature and a much more extensive sociology literature on the effects of social institutions on income. See e.g. Greif (1994) for economics and Granovetter (2005) for a review of the sociology literature.

disseminate new technologies more efficiently than others and append a production economy where the average technology level is related to output and income. There are two problems with this explanation. First, social contacts are presumably endogenous. If so, why would a social structure that inhibits growth persist? Second, this explanation is difficult to quantify or test. How might we determine if its effects are trivial or not? While researchers have mapped social networks in schools or on-line communities (Jackson, 2008), mapping the exact social network structure for an entire economy is not feasible.

Our theory for why some societies have growth-inhibiting social structures revolves around the idea that communicable diseases and technologies spread in similar ways - through human contact. We explore an evolutionary model, where some people favor collectivist interactions and others do not. People who form collectivist social networks, with the property that most of one's friends are friends with each other, benefit from longer expected lifetimes. With many mutual friendships, each group of friends has fewer links with the rest of the community. This limited connectivity reduces the risk of an infection entering the group of friends, allowing the participants to live longer. But it also restricts the group's exposure to new technologies, making them, on average, poorer. In contrast, having an individualist social structure with fewer mutual friendships brings the benefit of faster technology diffusion. Faster diffusion increases one's expected productivity, which encourages new entrants to adopt this trait. In countries where communicable diseases are inherently more prevalent, the longer lifetimes of collectivist types will allow them to flourish and drive out the individualist types. A social structure that inhibits the spread of disease and technology will emerge. In countries where communicable diseases are less prevalent, the individualist types will be more economically successful, new entrants will adopt their more successful trait and individualist types will drive out collectivist types.

The idea that disease prevalence and social structure are related can help to isolate and quantify the effect of social structure on technology diffusion. This is a challenging task because technology diffusion and social structure both affect each other: Technology diffusion is a key determinant of income, which may well affect a country's social structure. To circumvent this problem, we instrument for social structure using disease prevalence data. By itself, disease prevalence would be a poor instrument because it is not likely to be exogenous: higher income levels would likely translate into better health and lower disease levels. Therefore, our instrument uses differences in the prevalence of two types of disease. The first type is diseases that are spread directly from person-to-person. These diseases might plausibly affect social structure because changing one's relationships with others can prevent transmission. The second type of disease are those transmitted only by animals. Since direct human contact does not affect one's probability of infection, the prevalence of such diseases should not affect social structure. Thus, the main contribution of the

paper is to use the difference in prevalence of communicable disease and animal-transmitted disease as an instrument to measure the effect of social structure on income.

Our model helps to understand why communicable disease might be correlated with social structure and how social structure influences a country's average productivity. We isolate one particular aspect of social structure, its degree of individualism versus collectivism, while holding all other aspects of the network topology fixed. The collectivist society is populated by communities of people who mostly all know one another, and know each other's friends. In the individualistic society, agents interact, socially or economically, with others who do not know each other. This would be the case if most transactions took place in large, anonymous markets. Section 1 analyzes the epidemiology of disease and technology in each society and show that when the initial prevalence of infectious disease is higher, people form collectivist social networks, in order to reduce their chance of infection. This also reduces the growth rate of productivity.

Using historical pathogen prevalence data from Atlases of infectious disease and measures of a society's individualism from Hofstede (2001), section 3 tests the model's predictions for the relationship between disease prevalence and social structure. This establishes that disease prevalence is a powerful instrument for social structure. The section then goes on to estimate the effect of social structure on technology diffusion, using the difference in communicable and non-communicable diseases as an instrument. Finally, section 1.5 quantifies how much of the cross-country difference in technology diffusion this mechanism can explain. The paper's main finding is that a 1-standard-deviation increase in individualism increases output per worker by 23% of US output per worker.

Related literature The paper contributes to four growing literatures. A closely related literature is one that considers the effects of social structure on economic outcomes. Most of this literature considers particular firms, industries or innovations and how they were affected by the social structure in place (e.g., see Granovetter (2005) or Rauch and Casella (2001)). In contrast, this paper takes a more macro approach and studies the types of social networks that are adopted throughout a country's economy and how those affect technology diffusion economy-wide.

Thus in its scope, the paper is much more related to a second literature, that on technology diffusion. Recent work by Lucas and Moll (2011) and Perla and Tonetti (2011) use search model frameworks where every agent who searches is equally likely to encounter any other agent. What sets this paper apart is its assumption that agents only encounter those in their network. Our insights about why societies adopt networks that do not facilitate the exchange of ideas and our links to empirical measures of social structure arise because this focus on the network topology.

The third literature, on culture and its effects on national income is similarly macro in scope. Durlauf and Brock (2006) review work on social influence in macroeconomics, but bemoan that fact

that no attempt is made to incorporate social interactions in models of growth. Contemporaneous work by Tabellini (2005) and Algan and Cahuc (2007) does examine the relationship between cultural characteristics and economic outcomes. Bisin and Verdier (2001), Bisin and Verdier (2000) and Fernández and Fogli (2005) examine the transmission of culture. Cole, Mailath, and Postlewaite (1992) investigate how social norms affect savings choices, and in turn growth. But this literature typically regards culture as an aspect of preferences. Similarly, Gorodnichenko and Roland (2011) use the same measure of collectivism as we do to proxy for individuals' innovation preferences. In contrast, we examine social structure, which characterizes the set of relationships people have. Greif (1994) argues that preferences and social structure are intertwined because culture is an important determinant of a society's social structure. While this may be true, we examine a different determinant of social structure that is easily measurable for an entire country, pathogen prevalence. Our approach lends itself better to quantifying the aggregate effects of social structure on economic outcomes.

Finally, the work on the importance of political institutions by Acemoglu, Johnson, and Robinson (2002) and Acemoglu and Johnson (2005) is similar in its objectives and the approach of using pathogen prevalence to identify variation in endogenous institutions. But instead of examining political institutions, we study an equally important but distinct type of institution, social structure.

1 A Network Diffusion Model

Our model serves three purposes. First, it is meant to fix ideas. The concept of social structure is a fungible one. We want to pick a particular aspect of social structure, the degree of collectivism in a social network, to anchor our analysis on. In doing this, we do not exclude the possibility that other aspects of social or cultural institutions are important for technology diffusion and income. But we do want to be explicit about what we intend to measure.

Second, the model motivates our choice of disease as an instrument for social structure. Specifically, it explains why disease that is spread from human-to-human might influence a society's social network in a persistent way. The disease-based instrumental variable we use is a valid instrument, regardless of the veracity of this theory. The model simply offers one possible explanation for why disease and social structure might have the robust relationship we see in the data.

The third role of the model is that it helps us answer the following question: The richest countries have income and productivity levels that are 100 times higher than the poorest countries. Can differences in social structure plausibly explain such large income disparities? To answer this kind of question requires a model and some reasonably calibrated parameter values. Section 1.5 takes up this quantitative exercise.

A key feature of our model linking social structure to technological progress is that technologies spread by human contact. This is not obvious since one might think new ideas could be just as easily spread by print or electronic media. However, economists and sociologists have long noted the importance of human contact. In his 1969 presidential address, Kenneth Arrow remarks, “While mass media play a major role in alerting individuals to the possibility of an innovation, it seems to be personal contact that is most relevant in leading to its adoption. Thus, the diffusion of an innovation becomes a process formally akin to the spread of an infectious disease.” With this description of the process of technological diffusion in mind, we propose the following model.

1.1 Economic Environment

Time, denoted by $t = \{1, \dots, T\}$, is discrete and finite. At any given time t , there are n agents, indexed by their location $j \in \{1, 2, \dots, n\}$ on a circle. Each agent produces consumption goods with a technology $A_j(t)$ and labor input $l_j(t)$:

$$y_j(t) = A_j(t)l_j(t)^\alpha$$

Each healthy agent is endowed with 1 unit of labor, which they supply inelastically ($l_j(t) = 1$). Furthermore, there is no savings technology. Thus, consumption for healthy agents is $c_j(t) = y_j(t) = A_j(t)$.

An agent who catches a disease at time t loses their endowment of labor for that period ($n_j(t) = 0$ and thus $c_j(t) = 0$). At the end of period t , they die and are replaced by a new person in location j at the start of period $t + 1$. Let $\psi_j(t) = 1$ if the person in location j is sick in period t and $= 0$ otherwise and $\Psi_j(t) = \min\{s : s \geq t, \psi_j(s) = 1\}$ be the period in which the person living in location j at time t gets sick and dies.

The objective function of person j is the discounted stream of utilities over their lifetime:

$$U_j = \sum_{\tau=t}^{\Psi_j-1} \beta^{\tau-t} u(c_j(\tau)) \quad (1)$$

where β is the time discount factor and the flow utility function u is strictly increasing ($u' > 0$). Note that we need $U(0) \geq 0$ for death to be worse than low consumption. Otherwise, there will be instances where death is desirable.

Social networks Each person i is socially connected to ϕ other people. If two people have a social network connection, we call them “friends.” Let $\eta_{jk} = 1$ if person j and person k are friends and $= 0$ otherwise. To capture the idea that a person cannot infect themselves in the following

period, we set all diagonal elements (η_{jj}) to zero. Let the network of all connections be denoted N .

Spread of disease Each infected person transmits the disease to each of their friends with probability π . The transmission to each friend is an independent event. Thus, if m friends are diseased at time $t - 1$, the probability of being healthy at time t is $(1 - \pi)^m$. If no friends have a disease at time $t - 1$, then the probability of contracting the disease at time t is zero.

Spread of technology Technological progress occurs when someone improves on an existing technology. To make this improvement, they need to know about the existing technology. Thus, if a person is producing with technology $A_j(t)$, they will invent the next technology with a Poisson probability λ each period. If they invent the new technology, $\ln(A_j(t+1)) = \ln(A_j(t)) + \delta$. In other words, a new invention results in a $(\delta/100)\%$ increase in productivity.

People can also learn from others in their network. If person j is friends with person k and $A_k(t) > A_j(t)$, then the next period, j can produce with k 's technology, with probability γ . If there are multiple levels of technology used by j 's social contacts, j can produce with the best of these technologies: $A_j(t+1) = \max_k \eta_{jk} A_k(t)$.

As with disease, agents' expectations about others' technology are rational. In other words, they correctly anticipate the fraction of people producing with each technology level. But they do not know who has which technologies.

1.2 Two Illustrative Networks

The previous subsection described the economic environment for a given network. Before we add a process of network evolution, it is useful to understand the properties of two networks. The evolutionary process will guide the economy to one of these two networks. The are steady states of the stochastic network process. So, understanding how disease and technologies propagate in these two networks is very informative about the long-run behavior of our economy.

The two steady-state networks are extremes along a particular dimension, their degree of collectivism. This is an aspect of a social structure that has been extensively studied by sociologists. The collectivist network is one with many mutual friendships or many instances of interdependence that are the hallmark of collectivist societies. To measure this interdependence, we can ask: If I is friends with J and with K, how often are J and K also friends? In the networks literature, a structure where I, J and K are all connected to each other is called a *triple*. Therefore, a measure of the extent of shared friendships, and thus the degree of collectivism, is the number of network triples.

To count the number of triples, we look at all the instances in a given network where one node i is connected to two other nodes j, k . Count that as a triple if j and k are connected. This triples measure is related to a common measure of network clustering: Divide the number of triples by the number of possible triples in the network to get the *overall clustering* measure (Jackson 2008).

To make our examples concrete, we will fix the number of connections ϕ to be 4. While it would also be interesting to analyze the variation in the number of connections each individual has, we restrict attention to the degree of network clustering because it corresponds most closely to our empirical measures of collectivism.

Network 1 *In the collectivist social network, each individual j is friends with the 4 people located closest to them. In other words, $\eta_{jk} = 1$ for $k = \{j - 2, j - 1, j + 1, j + 2\}$ and $\eta_{jk} = 0$ for all other k .*

The reason we propose to examine the properties of network 1 is that it is extreme in its degree of collectivism. The next result shows that there are as many triples as there are members of the network (n).

Result 1 *In the collectivist network there are n unique triples.*

The proof of this and all subsequent results are in appendix A.

At the other end of the spectrum, we examine a second network that is identical in every respect, except that it has the lowest possible degree of collectivism. We call that the individualistic network.

Network 2 *In the individualistic social network, each person is friends with the person next to them and the person 4 positions away from them, on either side. In other words, $\eta_{jk} = 1$ for $k = \{j - 4, j - 1, j + 1, j + 4\}$ and $\eta_{jk} = 0$ for all other k .*

Result 2 *In the individualistic network, there are zero triples.*

In fact, the proof shows that any network where where each person i is friends with $i - \psi$, $i - 1$, $i + 1$, and $i + \psi$, where $\psi \in \{3, \dots, n - 3\}$, has zero triples. Network 2 is simply an example of such a network.

These two network structures are particularly informative because of their starkly different numbers of triples. This stark difference facilitates matching social institution data with one or the other type of network. As we will see, networks with numbers of triples between 0 and n , are also possible along the transition path. But knowledge of the properties of these two extreme cases provides intuition about the properties of such intermediate cases as well.

1.3 Theoretical Results: Speed of Diffusion in Each Network

Disease spreads slowly in the collectivist network. The reason is that each contiguous group of friends is connected to only 4 non-group members. Those are the two people adjacent to the group, on either side. Since there are few links with outsiders, the probability that a disease within the group is passed to someone outside the group is small.

Likewise, ideas disseminate slowly. Something invented in one location takes a long time to travel to a far-away location. In the meantime, someone else may have re-invented the same technology level, rather than building on existing knowledge and advancing technology to the next level. Such redundant innovations slow the rate of technological progress and lower average consumption.

The speed at which germs and ideas disseminate can be measured by the number of social connections in the shortest path between any two people. Consider an agent in position 1 and the agent farthest away from him on the circle, agent $n/2$. If each person has 2 friends on either side of them, then agent 1 will be friends with agent 3, who will be friends with agent 5, and this person will be friends with agent 7, ect., until we reach $n/2$. Thus, if the network size n is 6, $n/2$ is 3, which could be reached in 1 step: Agent 1 and agent 3 are directly connected. If $n/2$ is 5 ($n = 10$), it could be reached in 2 steps: from 1 to 3 and 3 to 5. In general, the number of steps in this chain will be $(n - 1)/4$, if that is an integer, or otherwise the next highest integer. The distance to this farthest person in the network is called the *network diameter*.

Diameter is one measure of diffusion speed because it tells us how many periods a new idea takes to travel to every last person in the network. If each person communicates the idea to each of their friends each period, then in n/ϕ periods, the farthest person in the network will have learned the idea, along with every other agent. Since disease is spread only probabilistically, from friend to friend, the diameter gives us the smallest number of periods in which every person is infected, with positive probability.

Appendix A computes the diameter of our two networks. The diameter of network 1 (collectivist), with n nodes is $(n - 1)/4$, if that is an integer, or otherwise the next highest integer. The diameter of network 2 (individualist), with $n > 4$ nodes where each node i is connected to $i - 4$, $i - 1$, $i + 1$, and $i + 4$, is $\text{round}(n/8) + 1$. Note that, for large n , the diameter for the collectivist network is close to $n/4$, while the diameter of the individualist network is $n/8$. Therefore, as long as the network is sufficiently large, which for a country, it undoubtedly is, the individualist network will have a smaller diameter.

Another related measure of the speed of diffusion is the average path length. Instead of measuring the number of nodes in the most direct path to the farthest person, this measure computes the number of nodes in the shortest path to every person and averages those lengths. If ϕ is even and $n/(2\phi)$ is an integer then $n/(2\phi)$ is the average path length in a collectivist network. To compare

the speed of technology or disease diffusion, we compute the diameter and average path length in each network.

Appendix A also computes the average path length of our two networks. The average path length in network 1 is $1/2 + n/(2\phi)$ if ϕ is even and $n/(2\phi)$ is an integer. To describe path lengths in the individualist network, we define the operator $\text{round}(x)$ to be the integer y closest to x . In other words, if y is an integer, then $\text{round}(x) = y$ iff $x \in [y - 1/2, y + 1/2)$. In network 2, when $n/8$ is an integer, the average path length is $7/8 + n/16$. As before, as long as the network is large, the average path length in the collectivist network is close to $n/8$, while for the individualist network, it is $n/16$.

These results tell us that in the individualistic network, dissemination of ideas or disease is fast. Each group of friends is connected to many outsiders, making the probability that a disease within the group is passed to someone outside the group is high. Likewise, ideas disseminate quickly because they travel many positions around the circle each period.

To compare the benefits and drawbacks of the two networks, recall that $\Psi_j(0)$ is the number of periods that the person living in location j at time 0 will live. Similarly, let $\alpha_j(0)$ be the number of periods it takes for a new idea, introduced in period 0 to reach person j .

Result 3 *Consider two networks, an individualistic network (N2) and a collectivist network (N1). They have equal size $n > 8$, where $n/8$ is an integer, and equal degree $\phi = 4$. If $\pi = 1$ and $\sum_j \psi_j(0) = 1$, then the average lifetime $E_j[\Psi_j(0)]$ and the maximum lifetime $\max_j[\Psi_j(0)]$ are longer in network 1.*

If $\gamma = 1$, then the average discovery time $E_j[\alpha_j(0)]$ and the maximum discovery time $\max_j[\alpha_j(0)]$ are slower in network 1.

Whereas, with a collectivist network, technology invented in one location was transmitted only $\phi/2$ people further each period, in the individualist network, ideas advance 4 places at a time. Because redundant innovations are less frequent, the rate of technological progress is faster. The following result clarifies the utility trade-offs from a society that chooses one network over the other.

Result 4 *As the probability of transmitting a disease falls ($\pi \rightarrow 0$), an individual in network 2 (individualistic) achieves higher expected utility than someone network 1.*

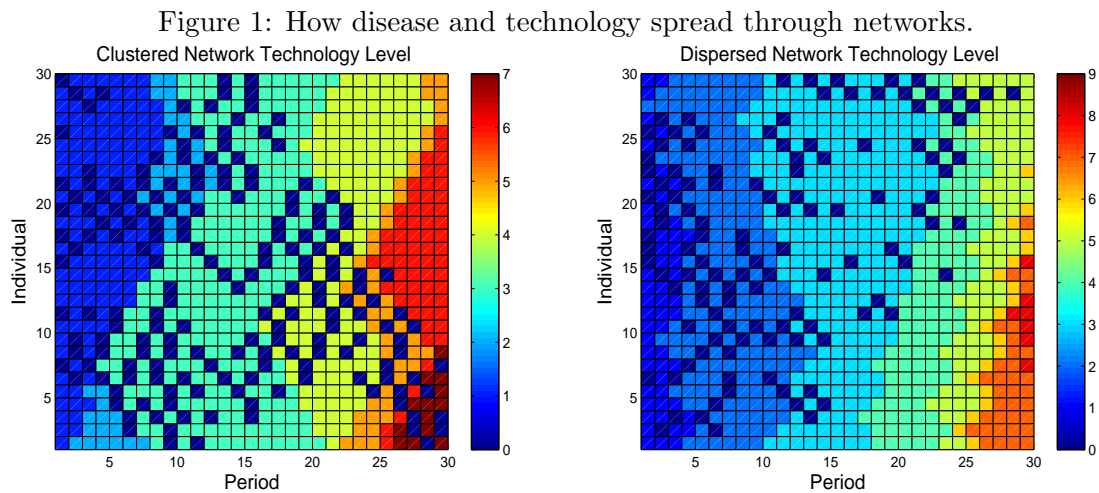
As the probability of transmitting an idea falls ($\gamma \rightarrow 0$), an individual in network 1 (collectivist) achieves higher expected utility than someone in network 2.

These results explain why ideas and germs spread more quickly in the individualistic network than in the collectivist network and gives us a glimpse into what evolutionary advantages each type of network might offer its adopters.

Could Collectivism Speed Technology Diffusion? Perhaps Arrow was not correct and technology diffusion is not a process "formally akin to the spread of infectious disease." Instead, a technology is adopted only when a person comes in contact with multiple other people who have also adopted it. It is theoretically quite possible that having many mutual friendships makes it more likely that groups of people adopt a technology together. But such a theory does not help to explain the empirical findings, which will show that collectivism is associated with slower technology diffusion. Ultimately, this model is simply a framework for helping us think about what we find in the data. While other formulations that lead to opposite conclusions are possible, they don't help us to understand the facts at hand.

1.4 An Illustrative Numerical Example

To illustrate the mechanics of technology and disease diffusion, we turn to a small-scale illustrative example. The next subsection will describe parameter values that match features of the data. Here, we hand-pick some of the parameter values (in particular the rate of technology arrival) to make it easier to visualize diffusion taking place. Figure 1.4 illustrates the diffusion of technology and disease. Each box represents a person/date combination. Time is on the horizontal axis. People are lined up on the vertical axis according to their location. In the first period (first column of boxes on the left), everyone starts with the same technology level. But there are a few agents who have a disease (the darkest boxes).



The darkest boxes indicate individuals who acquired the disease in period t and therefore have zero time- t productivity. Warmer colors indicate higher levels of technology.

By the second period, new ideas start to arrive. In the second column of boxes, there are a couple of lighter-colored boxes that indicate that these agents have reached the next technology

level. In the collectivist network (left figure), some agents who are adjacent to or 2 places away from agents that were sick in period 1 are now sick. In the individualistic network (right figure), some agents who are adjacent to or 4 places away from agents that were sick in period 1 are now sick. In period 3, the new ideas that arrived in period 2 start to diffuse to nearby locations. In the collectivist network, some individuals are still using the initial technology level in period 8. In the individualistic network, all the healthy agents have adopted the second technology level after period 5.

After 30 periods, the most technologically advanced agents in the collectivist network only realize 7 steps in the quality ladder. In the individualistic network, some agents operate at 9 steps. If each innovation represents a 5% productivity increase, being two steps further represents a 10% higher degree of productivity.

This example is meant to illustrate the mechanics of the model. It makes clear that an individualistic network spreads ideas more efficiently, but that it also spreads germs more efficiently. Of course, this is just an illustrative example. It is a comparison of the maximum level of technology from a small number of agents. To get a sense of the aggregate effect, the following calibration exercise averages the technology level over many agents and many simulations.

1.5 Quantitative Results: How Much Effect Might Networks Have on Output?

A potential concern about using this model to explain income differences across countries is the worry that its predicted effect is trivial, compared to the vast differences in incomes across countries. This concern is not misplaced since because cross-country differences in incomes can be 100-fold.

What our calibration exercise shows is that changing a society's social network structure has a small effect on the annual diffusion rate. Over time, small effects cumulate. The result is large differences in productivity levels in the long run. Thus, changes in network choice produce differences in technology diffusion rates which could explain a modest part of the disparity in countries' incomes.

1.5.1 Parameter Choice

To know whether changes in the network can produce big differences in technology levels, we need to choose some realistic parameter values for our model and analyze the simulated model outcomes. The key parameters in the model are the probabilities of disease and technology transmission, the initial pathogen prevalence rate and the rate of arrival of new technologies. These parameters are summarized in table 1.

For the initial pathogen prevalence rate, we use the annual tuberculosis death rate in China, a country where the disease was endemic. Tuberculosis is the most common cause of death in our

Table 1: Parameters and their calibration targets

	Parameter	Value	Target
Initial disease prevalence	$\text{Prob}(n_j(0) = 0)$	0.5%	TB death rate in China
Disease transmission probability	p_ψ	31%	disease steady state in individualistic netwk
Technology arrival rate	λ	5%	2% growth rate in low-disease country
Technology transfer probability	p	50%	Half-diffusion in 19 years (Comin et. al. '06)

sample. Note that this is a mortality rates, not an infection rate. Since individuals who get sick in the model die, this is the relevant comparison. Also, it is a conservative calibration because it would be easier to get large effects out of the higher disease prevalence rates. The probability of disease transmission is chosen to make the initial prevalence rate equal to the steady state rate of infection. Thus, the economy starts with a given fraction of the population being sick and each sick person represents an independent 31% risk (π) of passing the disease on to everyone that person is friends with.

Everyone starts with a technology level of 1. But each period, there is a chance that any given person may discover a new technology that raises their productivity by one percent. The rate of arrival of new technologies is calibrated so that the individualistic network economy (more likely to be the developed economy in the data) grows at a rate of 2% per year. The probability of transmitting a new technology to each friend (λ) is chosen to explain the fact that for the average technology, the time between invention and when half the population has adopted the technology is approximately 19 years (Comin, Hobijn, and Rovito, 2006).

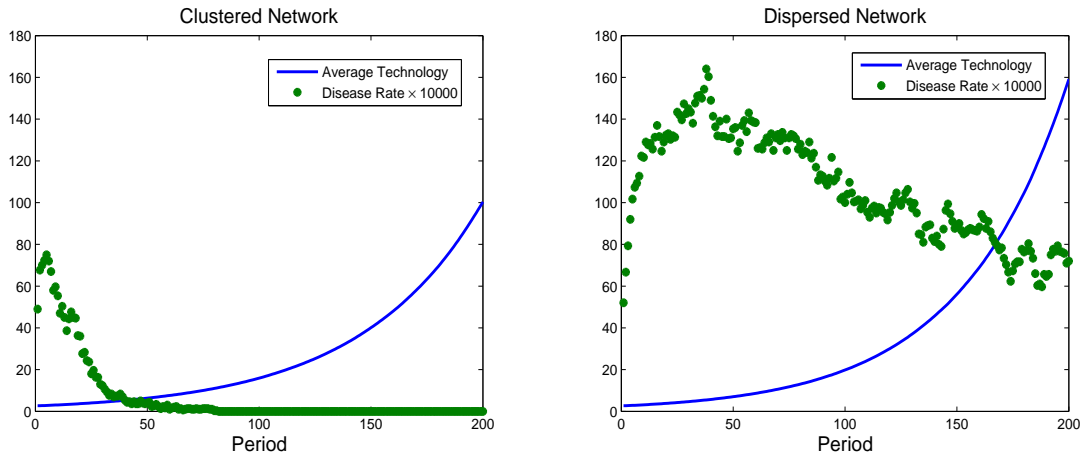
We simulate the high and low disease prevalence economies each with collectivist and individualistic networks. In this example, the economy consists of 1000 people, each with 4 friends. We average the results from 30 independent runs.

1.5.2 Simulation Results

While idea transmission facilitates reaching higher levels of productivity, disease prevalence diminishes productivity. To see the net effect of these two forces, we simulate the model many times and examine the average outcomes.

Figure 1.5.2 plots the average disease prevalence (times 10,000) and the average technology level for the whole population over 200 years. These are results for the high-disease calibration (0.5% initial infection rate). The fraction of the population infected with disease is significantly higher in

Figure 2: Average disease prevalence and productivity



Prevalence of disease ($\times 10,000$) and average technology level in a collectivist network (left) and a individualistic network (right).

the individualistic network society. In fact, the collectivist networks inhibit the spread of disease so much that it becomes extinct in this calibration.

However, having a individualistic network results in technology that grows at 2.0% per year. This is true by construction because it was one of the calibration targets. But the economy with the collectivist network grows at only 1.8% per year. While the difference in growth rates is small, in time, it produces large level differences. After 200 years, the average level of technology is about 60% higher in the individualistic network than in the collectivist network. This simple example makes the point that a difference in network structure can create a small friction in technology diffusion. When cumulated over a long time horizon, this small friction has the potential to explain large differences in countries' incomes.

1.6 Network Evolution

So far, we have simply described diffusion properties of two networks. This leaves open the question of why some societies might adopt one type of network or the other. To provide one possible explanation, we next consider an evolutionary model where the network structure evolves as agents die and new ones are born in their place.

1.6.1 Model

Preferences, production, endowments and the diffusion processes for technology and disease are the same as in the fixed-network model.

In addition, at each date t , each person j can be one of two types: They are either a collectivist $\tau_j(t) = co$ or an individualist $\tau_j(t) = in$. All agents are linked to the two people adjacent to them. In addition, they are linked to at least one other person. Which other people depends on their type and the type of their neighbors. Collectivists form links with those adjacent to them and someone four spaces to their right. For example, if the person is in location j , they are linked to $j - 1$, $j + 1$ and $j + 4$. Individualists form links with those adjacent to them and someone two spaces to their right. For example, if the person is in location j , they are linked to $j - 1$, $j + 1$ and $j + 2$. In addition, a person of either type might be linked to nodes $j - 2$ and/or $j - 4$, depending on whether the agents in those locations are individualist or collectivist. In other words, a person's own type governs their links to the right (with indices higher than yours, except near n), others' types govern links to the left.

A person's type is fixed throughout their lifetime. The network structure only changes when someone dies. There are two reasons an individual can die. First, they can acquire the disease. Someone who acquires the disease at time t has zero output in period t . At the end of period t , they die. Second, they can die stochastically, for non-disease related reasons (accident, old age, etc.). With probability δ , each person has an accident and dies at the end of each period. This probability is independent across time and individuals. When someone at node j dies in period t , then at the start of period $t + 1$, a new person inhabits that node.

A newborn person inherits the best technology from the set of people that the parent was socially connected to. He also inherits the type of the person with that best technology. In other words, if the person at node j is socially connected to nodes $\{k : \eta_{jk}(t) = 1\}$ and dies at time t , the new person at node j at time $t + 1$ will start with technology $\max_{\{k:\eta_{jk}(t)=1\}} A_{kt}$. Let k^* be the argument that maximizes this expression (i.e. the friend with the highest time- t technology), then the time- $t + 1$ type of the person is the same as the time- t type of person k^* : $\tau_j(t + 1) = \tau_{k^*st}(t)$.

The idea behind this process is that evolutionary models often have the feature that more "successful" types are passed on more frequently. At the same time, we want to retain the network-based idea that one's preferences are shaped by one's community. Therefore, in the model, the process by which one inherits the collectivist or individualist trait is shaped by one's community, the social network, and the by relative success (relative income) of the people in that network.

1.6.2 Theoretical Results

The question we want this model to answer is: Why do some societies end up with a collectivist network even though it inhibits growth? What features might influence the long-run network equilibrium? These results are designed to show that a higher initial prevalence of disease makes it more likely that a society will end up with a collectivist network, like that in N1.

The first set of results show that eventually, the economy always converges to either the fully collectivist network (N1) or the fully individualist one (N2).

Result 5 *With probability 1, there exists a finite time T such that $\tau_j(t) = \tau_k(t) \forall k$ and $\forall t > T$.*

In other words, after some date T , everyone will have the same type forever after. They might all be individualist or all be collectivist. But everyone will be the same. The reason for this is that since traits are inherited from neighbors, when a trait dies out, it never returns. The state where all individuals have the same trait is an absorbing state. Since there are a finite number of states, and whenever there exists a j, k such that $\tau_j(t) \neq \tau_k(t)$, every state can be reached with positive probability in a finite number of steps, then with probability one, at some finite time, an absorbing state is reached and the economy stays there forever after.

Similarly, having zero infected people is an absorbing state. Since that state is always reachable from any other state, with positive probability, it is the unique steady state.

Result 6 *With probability 1, there exists a finite time T such that $\psi_j(t) = 0 \forall j$ and $\forall t > T$.*

What these results tell us is that which network type will prevail is largely dependent on which dies out first, the individualist trait, or the disease. When there is a positive probability of infection, people with individualist networks have shorter lifetimes, on average. If disease is very prevalent, it kills all the individualists off and the society is left with a collectivist network forever after. If disease is not very prevalent, the transmission rate is low, or by good luck, it just dies out quickly, individualists will survive. Since they are more economically successful, they are more likely to pass on their individualist trait. So, the economy is more likely to converge to an individualist network. This is not a certain outcome because passing on a trait is a stochastic process. It is always possible that there is an unlikely sequence of events that causes individualists to become extinct, even if the disease itself is no longer present.

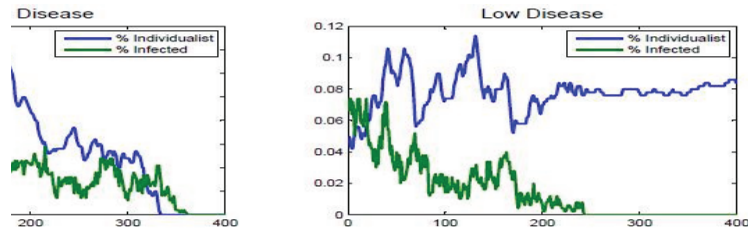
1.6.3 Numerical Results

What we ultimately want to know from this evolutionary model is how the initial prevalence of disease can alter the long-run state of the network. To do that, we want to calculate the probability of arriving at each state state (where all individuals have the same type) and see how that probability changes in response to changes in disease prevalence. However, the probability of the network arriving at each absorbing state is not something we can characterize analytically. To characterize the probability of a single stochastic process crossing one boundary before another is a difficult problem. Here, there are two interacting stochastic processes, one for disease and one

for network types. Both have absorbing states. That added complexity makes characterizing the crossing probabilities an intractable problem.

What we can do is examine the probability of each network steady-state in the context of our numerical example. We use that same parameters as before. (See Table 1.) We set the initial fraction of travelers to 5% and simulate the economy for 500 periods 250 times. Figure 1.6.3 reports the average (averaged over the runs) fraction of agents that have the individualist type in each period, and the average fraction of agents infected with the disease each period for two initial conditions. The low disease economy starts with 0.05%

Figure 3: Node types and infection rates over time.



Prevalence of disease and the fraction of individualist nodes at each date. The economy on the left differs only because it began with a higher rate of disease prevalence. Because of this difference, the individualist agents typically die out and the network becomes collectivist.

1.6.4 Linking Model and Data

Of course, the model could be made richer in many dimensions, it could include network choice, random connections between people meant to represent anonymous market transactions, strong and weak social ties between people, or more nuanced health states. But none of these extensions is likely to change the basic prediction that the network with the most triples diffuses germs and ideas more rapidly. It is this basic idea that forms the basis for the empirical work.

The next step uses the following prediction prediction of the model as the basis for selecting instruments that can help us quantify the effect of social structure on productivity: High disease-prevalence societies are more likely to become collectivist. Since the diameter can be interpreted as the minimum time it takes a disease to spread from one carrier to every last person in the network, proposition 3 tells us that collectivist networks offer longer disease transmission times. Thus, for

sufficiently high disease rates, collectivism offers residents of these societies higher expected utility because it protects them against the spread of disease. Of course, collectivism itself slows infections. Thus, one might argue that collectivism should be associated with lower disease prevalence. But as long as social structure is not exogenous, and there is some selection (choice or natural selection) in networks, this secondary effect is not the dominant one.² Thus, the theory predicts that high-disease societies should be collectivist. These collectivist societies should have slower technology diffusion and therefore lower average productivity.

2 Data

Our theory is about the relationship between three main variables: pathogen prevalence, social structure, and the technological frontier. The section describes how these three variables are measured.

2.1 Measuring Pathogen Prevalence

To measure the prevalence of disease, we use the historical prevalence of 9 pathogens: leishmanias, leprosy, trypanosomes, malaria, schistosomes, filariae, dengue, typhus and tuberculosis. We choose these diseases because we have good worldwide data on their incidence, and they are serious, potentially life-threatening diseases that people would go to great length to avoid. By necessity, a contemporary source was used to estimate the prevalence of tuberculosis, but the prevalence of the remaining eight pathogens was estimated on the basis of old atlases of infectious diseases and other historical epidemiological information provided in Rodenwaldt and Bader's (1952-1961) World-Atlas of Epidemic Diseases and in Simmons, Whayne, Anderson, and Horack's (1944) Global Epidemiology.³ They used a 4-point coding scheme: 0 = completely absent or never reported, 1 = rarely reported, 2 = sporadically or moderately reported, 3 = present at severe levels or epidemic levels at least once.⁴ The prevalence of tuberculosis was based on a map contained in the National Geographic Society's (2005) Atlas of the World, which provided incidence information in each region for every 100,000 people. These scores, calculated for a total of 230 geo-political regions of the world, were summed up to in an overall index that summarizes a country's overall historical

²It is like a price effect on quantity: If there is high demand, suppliers set a high price, which should lower demand. The secondary effect of price on demand is smaller. High demand is, on net, associated with a high price.

³Detailed information about the occurrence of various diseases across the world was collected by the Medical Intelligence Division of the United States Army.

⁴In the rare cases in which these two epidemiological sources provided contradictory information, priority was placed on data provided by the older source (Simmons et al., 1944). In cases in which the relevant maps were unavailable (this was especially true for leprosy) or insufficiently detailed (this was especially true for many of the Pacific island nations), prevalence ratings were informed also by verbal summaries found in Simmons et al.

pathogen prevalence⁵. Figure 7 shows the historical world-wide distribution of pathogens according to the overall index. This is the data we use for most of our analysis, including our IV estimations.

For comparison, we used the same method to create an alternative measure of pathogen prevalence based explicitly on contemporary information. Data were obtained from the Global Infectious Diseases and Epidemiology Online Network (<http://www.gideononline.com/>), which reports current distributions of infectious diseases in each country of the world. (The database is updated weekly; the data used were obtained during the period April–June 2007.) The data cover seven classes of pathogens (leishmanias, trypanosomes, malaria, schistosomes, filariae, spirochetes and leprosy) and coded the relative prevalence of each specific pathogenic disease within each class. A total of 22 specific pathogenic diseases were coded, each on the same three-point prevalence scale. These values were summed within each region to create a composite index estimating the contemporary prevalence of pathogens.⁶ The countries with the highest pathogen prevalence are Brazil, India, China, Nigeria and Ghana. Countries with the lowest prevalence include Canada, Switzerland, Luxembourg, Hungary and Sweden.

Disease reservoirs To identify the effect of disease on social structure, we will also use the difference in the prevalence of various types of diseases. The “reservoir” of a diseases is a term from epidemiology that refers to the long-term host of a pathogen. When an animal hosts the disease, it may or may not become ill from it. We first follow Smith, Sax, Gaines, Guernier, and Guban (2007) and Thornhill, Fincher, Murray, and Schaller (2010) by distinguishing diseases according to their reservoirs:

Human-specific Many infectious agents known to afflict mankind are currently entirely restricted to human reservoir hosts (i.e., contagious only between persons), even though they historically may have arisen in other species, such as measles which originated in cattle. In our data set, the only pathogen that resides exclusively in humans is filariae.

Zoonotic Infectious agents that develop, mature, and reproduce entirely in non-human hosts, but nonetheless have the potential to spill over and infect human populations, are referred to herein as zoonotic infectious agents. Humans are a dead-end host for infectious agents in this group. Our zoonotic pathogens are schistosomes and typhus.

⁵The majority of these regions are nations (e.g., Albania, Zimbabwe); others are territories or protectorates (e.g., Falkland Islands, New Caledonia) or culturally distinct regions within a nation (e.g., Hawaii, Hong Kong).

⁶The reliability of the two pathogen prevalence indices are indicated by high correlations with a similar index created by Gangestad & Buss (1993) to assess pathogen prevalence within a smaller sample of 29 regions ($r = 0.89$ and 0.83 , for the historical and contemporary pathogen prevalence indices, respectively). Correlation between our two indices is 0.77 .

Multi-host Some infectious agents can use both human and non-human hosts to complete their lifecycle. We call these “multi-host” pathogens. The multi-host pathogens in our data are leishmaniasis, leprosy, trypanosomes, malaria, dengue and tuberculosis.

Since multi-host and human-specific pathogens can reside in humans, they have the potential to affect the relative benefits of a social network. Zoonotic pathogens are not carried by people, only by other animals. Therefore, their prevalence is less likely to affect the benefits of any particular social structure. Therefore, for the purposes of our analysis, we will group human-specific and multi-host diseases together and compare their effects to those of zoonotic diseases. This difference is measured using the historical prevalence data.

Disease vectors Another way of comparing diseases is according to their “vector.” The vector is the animal or substance that carries the disease from one person to another. In many cases, the vector is also the reservoir. But sometimes not. For example, some diseases are hosted by mammals, but are spread when an insect bites the animal and then bites another shortly after. The insect is not the long-term host of the disease, but it was the vector that spread it.

Separating diseases by vector gets at the heart of our theory because whether or not a disease affects the benefits of a social network depends on how it is spread. At the same time, just because a disease is spread when a head lice bug jumps from one person’s head to another, rather than directly from person-to-person, doesn’t mean that social contact was irrelevant to the transmission. The question of whether social contact and social structure affect disease transmission depends on how mobile the vector is. This is further complicated by the fact that a disease often has more than one possible vector.

Therefore, in this comparison, we use only a subset of the diseases for which we have data so that we can make a clean comparison of two diseases that are clearly spread by social contact and one that although contagious, does not require any social contact. We measure the prevalence of each type using the historical data.

Human vector The two human diseases that have human vectors in our data set are tuberculosis and leprosy. Tuberculosis is spread when a person coughs or sneezes and another person inhales airborne droplets containing the disease. Leprosy is transmitted by prolonged close contact or by nasal droplet.

Aquatic vector One disease that is clearly not spread by social contact is schistosomiasis. This is a parasitic infection by a type of flatworm called a schistosome. It is spread when a person urinates or defecates and the larva are consumed by a snail, which is an intermediate host to the disease. Only after the parasite emerges from the snail can it infect a new human host.

The parasite lives in water and directly penetrates the skin of a person that comes in contact with it. Since the substance that conveys the disease to humans is water, water is the vector and we call such a disease “aquatic.”

Societies have long known that tuberculosis and leprosy are spread by human contact. Both have been widely prevalent throughout history. Active tuberculosis, if left untreated, kills about one-half of its victims. Nearly one-third of the world population is estimated to be a latent carrier of the disease.

We choose schistosomiasis because it never spreads from person-to-person, can travel long distances in moving water, and should not affect one’s choice of social contacts in any way. Although the mortality rate is not very high, schistosomiasis damages internal organs and affects cognitive development in children. According to the World Health Organization, it is the second most socioeconomically devastating parasitic disease after malaria.

2.2 A Sociological Measure of Clustering: Collectivism

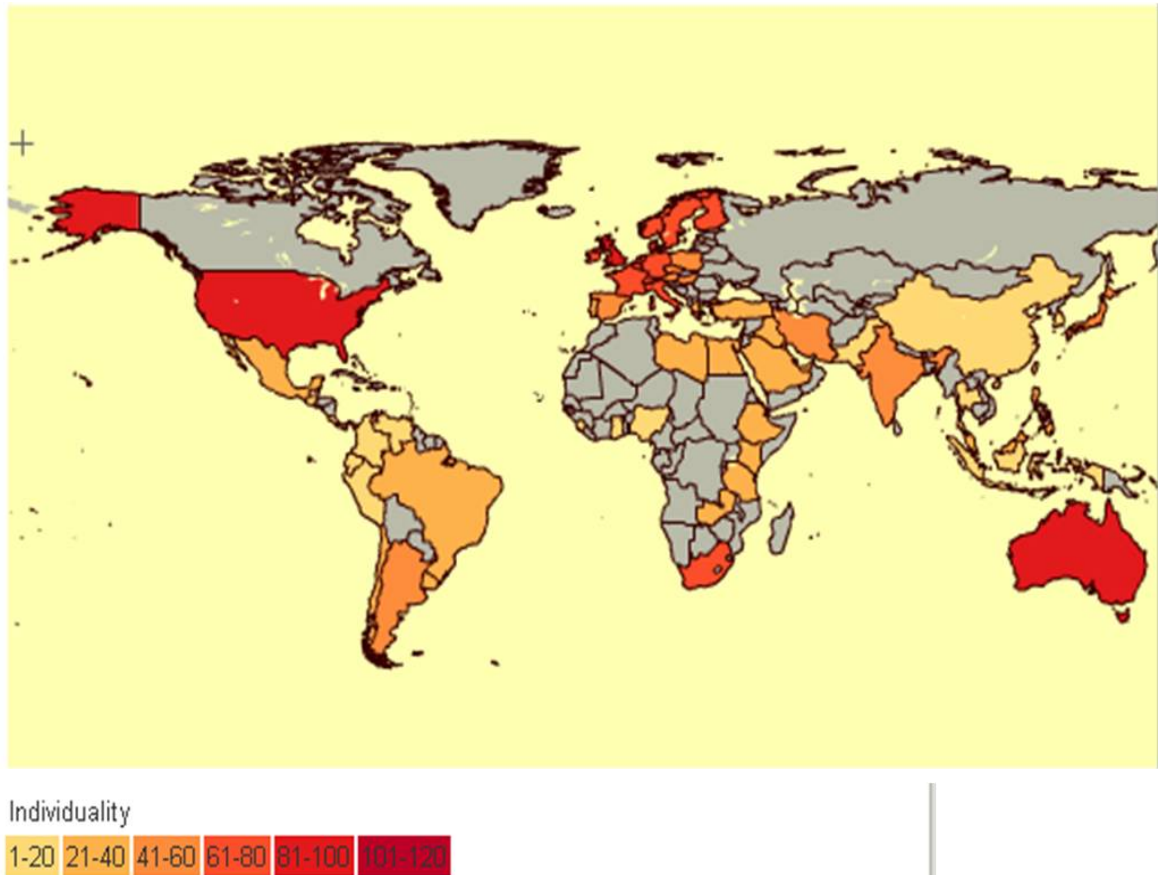
Collectivism is defined as a social pattern of closely linked individuals; interdependent members of a collective. Collectivistic societies are ones in which individuals are integrated into communities. What distinguishes communities from sets of people with random ties to each other is that in communities, people have mutual friendships. In other words, it is common that two friends have a third friend in common. This is the sense in which they are interdependent.

Individualism is the opposite of collectivism. Individualistic societies are ones where the ties between individuals are loose. Everyone is expected to look after him/herself and immediate family members. In individualistic societies, people interact through market mechanisms. Through markets, they interact with a variety of people who are unlikely to know each other. Thus, individualistic societies are ones where social networks have fewer mutual friendships.

To measure where various societies fall on the individualism/collectivism spectrum, Hofstede (2001) performed a survey of IBM employees worldwide. He used the 33 survey questions to construct an index of individualism that ranges from between 0 (strongly collectivist) to 100 (strongly individualist). Figure 2.2 summarizes the findings of his survey in a color-coded map. The most individualist countries are Canada, Netherlands, United Kingdom, Australia and United States. The most collectivist countries are Guatemala, Ecuador, Panama, Venezuela, Columbia and Pakistan.

Collectivism as strong social norms Another way to interpret collectivism and the notion of interdependence that it entails is to relate it to the strength of social norms. Perhaps being members of an integrated collective means adopting similar behaviors and norms. In fact, Hofstede’s individualism index is highly correlated with measures of social conformity in the GSS survey.

Figure 4: Map of Hofstede’s individualism index.



Social conformity is easier to sustain in collectivist networks. Coleman (1988) shows that the presence of effective norms and thus the accumulation of social capital depend on network “closure.” Closure is present when your friends are also your friends’ friends. In other words, it depends on the presence of triples. Coleman explains that people enforce strong group norms through collective punishments of deviators. If j observes i deviating from a social norm, then j can directly contact other friends of i to enact some joint retribution for the misdeed. When collective punishments are implementable, conforming behavior is easier to sustain than if punishments must be implemented in an uncoordinated way. Thus, if we interpret collectivism as strong social conformity, such collectivism is more likely to emerge in networks with many triples.

2.3 Measuring the Technological Frontier

We use the cross-country historical adoption of technology (hereafter CHAT) data set developed by Comin, Hobijn, and Rovito (2006). CHAT covers the diffusion of about 115 technologies in over

150 countries during the last 200 years. We use the number of adopted technologies per country to measure how far up the technological ladder the country’s most advanced agents are. This measure seems to reliably capture countries’ technological ranking because there are universal leaders and universal followers. In other words, countries’ ranking in terms of their speed of adoption is stable across technologies and over time.

3 Empirical Results: How Much Do Networks Affect Technology?

Our objective is to better understand how social structure affects technology diffusion and how large that effect is on economic development. The difficulty is that economic development also can potentially change the social structure. The challenge is to isolate each of these two effects. To do this, we consider the following structural model:

$$A = \alpha_1 + \alpha_2 S + \epsilon \tag{2}$$

where A is productivity, or more specifically, the degree of technology adoption, S is social structure (individualism), as measured by the Hofstede index, the α ’s are unknown coefficients, ϵ is a mean-zero residual orthogonal to S , and

$$S = \gamma_1 + \gamma_2 A + \gamma_3 x + \eta \tag{3}$$

where the γ ’s are unknown coefficients, x is a collection of other variables that determine social structure, and η is a mean-zero residual orthogonal to A and x . The coefficient of interest is α_2 , which measures the effect of social structure S on productivity A .

This model recognizes the endogeneity problem inherent in estimating the relationship between A and S . It incorporates our main hypothesis, that social structure S matters for productivity A , but it also reflects the idea that perhaps productivity (an income) can cause social structure to change as well. Because A depends on S and S depends on A , an OLS estimate would be biased.

Our key identifying assumption that will allow us to isolate the effect of social structure on productivity is that the x variables affect productivity only through social structure, and not directly. In other words, we assume that $E[x\epsilon] = 0$. We will carefully choose the x variables we use to make this case plausible and we will then test the assumption using a test of over-identifying restrictions.

Note that we do not need to know all the determinants of social structure x . We will not estimate equation (3). Rather, any subset of the x variables can serve as valid instruments for S . Similarly, we do not need to observe S . A proxy variable with random measurement noise is

sufficient for an unbiased instrumental variables estimate of the coefficient α_2 (see Hall and Jones (1999)).

Before we look at the effect of social structure on technology diffusion, we first establish that our instruments, based on historical disease prevalence, have an empirical relationship with social structure that justifies their use as instrumental variables.

3.1 First-Stage Regressions: Disease and Social Institutions

We will explore two different instrumental variables specifications. Each one will have two instruments. We want to have more than one instrument in each estimation because it will allow us to test the orthogonality of each instrument with the residual in equation (2). Each specification has latitude (distance from the equator) as one of the instruments. Since it is an immutable feature of a country, it is clearly not endogenous. Although some have argued that there is a direct effect of latitude on productivity through climate and disease, we will test the assumption that latitude is orthogonal to ϵ and be unable to reject that hypothesis. Latitude does affect social structure because geography and climate are key determinants of where people locate vis-a-vis each other.

The second instrument in each specification will be a difference in the prevalence of two types of diseases. To understand why these differences might be economically relevant for social structure, we first examine the relationship between disease and social structure.

Figure 5 illustrates the positive statistical correlation between individualism and the prevalence of pathogenic disease.

Figure 5: Hofstede's individualism index plotted against pathogen prevalence.

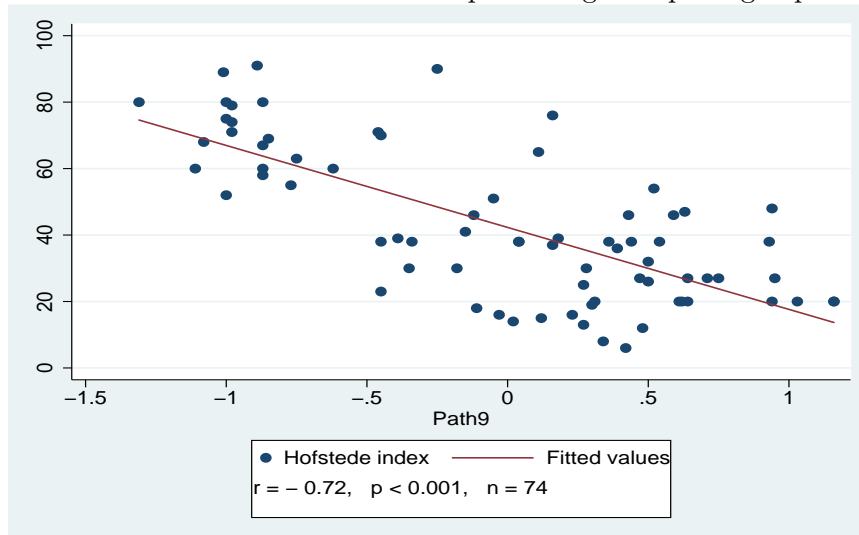


Table 2 quantifies this relationship. Columns 1 and 2 show that pathogen prevalence and

individualism are related in a statistically significant way. The first column uses our contemporary pathogen data while the second column uses the historical data from the 1930's.

The negative sign on the pathogen coefficient means that the increased presence of pathogens is associated with a less individualistic (more collectivist) society. That is consistent with our theory because the more collectivist society, with its greater propensity for network triples, would be a more effective structure for inhibiting the spread of disease. The explanatory power of pathogens is large; the R^2 of the regression is over 50%. The economic magnitudes are also large. A one-unit increase in our pathogen measure corresponds to one disease being endemic (always widespread), instead of epidemic (occasionally widespread). Having one more disease consistently prevalent corresponds to an individualism index that is between 8 and 43 points lower, depending on the specification.

Even though the contemporaneous data probably has less measurement error, the historical pathogen data has a stronger statistical relationship with social structure. The contemporaneous data could be reflecting more reverse causality since a less individualistic social structure in the 1970's could be responsible for lower disease prevalence today. That would explain why the coefficient is more positive (less negative).

Table 2: **First-stage regressions of pathogen prevalence variables on individualism index**

Dependent variable	Individualism (S)			
	OLS	OLS	OLS	OLS
Pathogens (contemporary)	-2.16 (0.31)			
Pathogens (historical)		-2.65 (0.30)		
Human - zoonotic Pathogens			-1.15 (0.60)	
Human - aquatic Pathogens				-3.75 (1.35)
Latitude			0.73 (0.16)	0.89 (0.10)
Constant	111.05	76.81	28.16	24.98
R^2	0.39	0.51	0.57	0.59
Observations	76	78	78	78

The table reports OLS estimates of the γ coefficients in $S = \gamma_1 + \gamma_2 x$, where the x variables are listed in the first column of the table.

Of course, it is possible that both disease and social structure are governed by GDP, or that higher population density lends itself to a different social structure and more disease prevalence. To explore whether pathogens might have an effect, beyond that governed by GDP and density, we

also estimated auxiliary regressions where we control for GDP and population density. We use the figures from 1970, the same time as the Hofstede survey was being collected. Controlling for GDP and population density only slightly lessens the significance of the relationship between disease and social structure.

Differences in disease reservoirs Another identifying assumption is that while productivity and GDP may affect disease prevalence, even 40 years prior, it affects many diseases similarly. Thus, the difference in the prevalence of one type of disease or another is exogenous with respect to GDP. The first difference we consider is based on the disease reservoir. We consider diseases that reside in humans (human-specific plus multi-host) versus diseases that reside exclusively in non-human animals (zoonotic diseases). The difference between the prevalence of these two types of diseases is our first instrument.

To explain economically why the reservoir-based difference in disease prevalence is likely to be a good predictor of social structure, we use the model. It tells us that when disease can be spread through a social network, there is a higher expected utility from adopting a collectivist social network. Diseases that reside in humans could be spread through social networks. Diseases that reside only in other animals are not spread through social networks and have no effect on the expected utility from adopting one form of network or another. Thus, the difference in the prevalence of these two diseases should, according to the model, predict the benefits of collectivism versus individualism. To the extent that societies can choose their networks or that high-utility networks are more likely to survive over time, this difference in disease should predict social structure.

Statistically, this difference is likely to be a powerful instrument because the average correlation of individualism with each disease carried by humans is much larger in magnitude than the average correlation with each of the zoonotic diseases (-0.53 vs. -0.29). The fact that the correlations are negative tells us that higher disease prevalence is associated with more collectivist societies. This is consistent with our theory, which tells us that societies should choose these networks when disease is more prevalent because they are better suited to protect against contagion.

In the first-stage regression, both latitude and differences in disease prevalence are highly-significant predictors of social structure (table 2, column 3).

Differences in disease vectors The second type of difference in disease prevalence we consider is the prevalence of diseases spread by humans (humans are the vector), minus the prevalence of disease spread by water (aquatic vector).

Just as with the previous difference, the model gives the economic reason why the vector-based difference in disease prevalence is likely to be a good predictor of social structure. Diseases that

are spread from human-to-human are spread through social networks. Diseases that are passed on only after they are hosted by another animal and then transmitted by contaminated water are not spread through social networks and have no effect on the expected utility from adopting one form of network or another. Thus, to the extent that societies can choose their networks or that high-utility networks are more likely to survive over time, the vector-based difference in disease should also predict social structure.

Statistically, this difference is likely to be a powerful instrument because the average correlation of individualism with each disease conveyed by humans is larger in magnitude than the correlation with the aquatic disease (-0.55 vs. -0.41).

In the first-stage regression, both latitude and vector-based differences in disease prevalence are highly-significant predictors of social structure (table 2, column 4).

Even with the difference in diseases, one might be concerned about endogeneity of the instrument. While greater levels of development spur public health initiatives, these measures prevent the human transmission and water-borne transmission of diseases. Likewise, better health care lowers mortality rates from both types of diseases. If anything, there is a tendency for clean water initiatives to be one of the first public health measures a country adopts when its income rises. If this were the case, then there would be a negative correlation between aquatic illness and income and therefore a positive correlation between the difference in human and water diseases and income. This would make the regression coefficient of the difference in diseases on individualism, which is positively correlated with income, more positive. To the extent that we find a significant negative coefficient, it is despite any reverse causality coming from clean-water initiatives spurred by economic development.

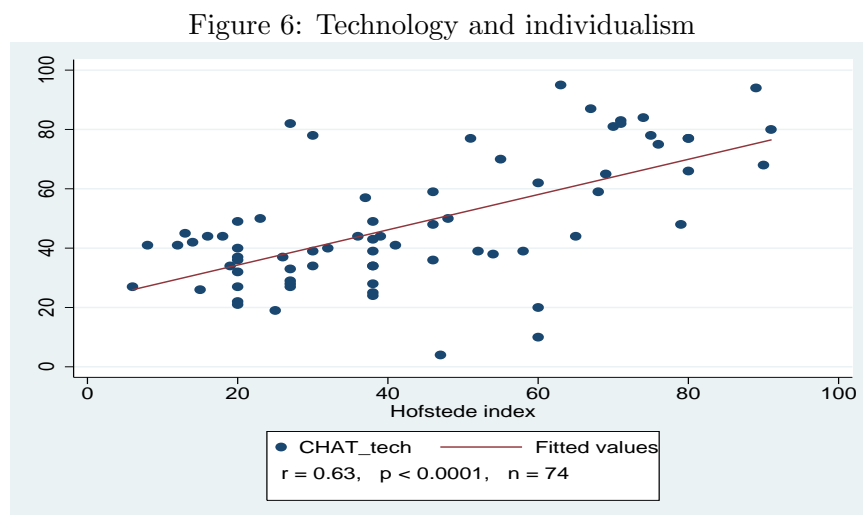
The other hypothetical cause for concern might be that higher productivity and the accompanying higher income cause the social structure to change. In particular, a richer, more modern society is more likely to be market-based and individualist. The change in social structure could affect the difference in disease prevalence by facilitating the transmission of diseases spread from human-to-human. Notice that this logic does not imply that differences in disease x are correlated with the estimation error ϵ in (2). This story suggests that social structure S depends on A , something already represented in our specification (equation 3), and it suggests that there should be an additional equation representing the idea that the instrument x depends on social structure: $x = \psi_1 + \psi_2 S + \nu$. In this structure, as long as $e[\epsilon\nu] = 0$, x is still a valid instrument for S . In other words, as long as productivity affects the difference in disease through social structure, rather than directly, *it does not invalidate the use of disease differences as instruments*. It only suggests that γ_2 is perhaps not an unbiased estimator of the effect of disease on social institutions. Our estimates suggest that more disease is associated with less individualism. If individualism spreads

disease, then this estimate is downwards-biased. In other words, the true effect of disease on social institutions would be larger than the one we estimate.

These results are important for the next stage, identifying an effect of institutions on technology diffusion. But they are also important on their own because they point to reason why countries may have chosen different social institutions. These results are consistent with some effect of communicable pathogens on social structure. Perhaps social structures have evolved, in part, as a defense against the spread of disease. But further statistical work would need to be done to say conclusively that disease prevalence is part of the reason why some societies have adopted social structures that inhibit technological diffusion and growth.

3.2 Main Results: Social Institutions and Technology Diffusion

Our main result is to quantify the effect of social structure on technology diffusion. Figure 6 illustrates the statistical relationship between social structure and the speed of technology diffusion. It reveals that more individualistic societies (those with little interdependence of relationships) tend to also be societies where technologies diffuse quickly. In interpreting this correlation, reverse causality is obviously a concern. Faster technology diffusion raises incomes, which might well change the social structure. Likewise, the economic development that results from technology diffusion could produce a wave of urbanization, which influences social structure. Therefore, we use the differences in pathogen prevalence as an instrument for social structure.



Comin, Hobijn, and Rovito (2006)'s technology adoption (CHAT) measure (vertical axis) plotted against Hofstede's individualism index (horizontal axis).

The first two columns of table 3 show that the degree of individualism in a country's social

structure has a large effect on a country’s level of technology. A 1-standard deviation in the Hofstede index is 28.5. Thus, a 1-standard deviation increase in individualism results in $28.5 \cdot 0.717 = 16.3$ additional technologies that are adopted by a country by the end of the 20th century (or 16.6 using `diff_vec` as an instrument). The mean of the CHAT measure of technology is 48. Thus, a degree of individualism that is 1 standard deviation above the average is associated with technology adoption that is 34% (35%) above the average level of technology.

Table 3: **Social Structure and Technology (main result)**

Dependent variable:	Technology			Solow Residual		Output per worker	
Instruments:	<code>diff_res</code> latitude	<code>diff_vec</code> latitude	none (OLS)	<code>diff_res</code> latitude	<code>diff_vec</code> latitude	<code>diff_res</code> latitude	<code>diff_vec</code> latitude
Individualism	0.717* (0.112)	0.728* (0.110)	0.586* (0.084)	0.0067* (0.0018)	0.0074* (0.0017)	0.0103* (0.0013)	
Over-ID test	2.66 Accept	0.711 Accept		0.042 Accept	1.90 Accept	1.00 Accept	0.30 Accept
R^2	0.57	0.38	0.40	0.21	0.20	0.54	0.54
N	75	75	75	67	67	67	67

The first row reports α_2 coefficient from an IV estimation of $A = \alpha_1 + \alpha_2 S + \epsilon$. Technology A is Comin, Hobijn, and Rovito (2006)’s measure of the number of technologies adopted in a country. Individualism S is the Hofstede index. `diff_res` is the difference between pathogen prevalence, based on the reservoir. It is the prevalence of diseases carried by humans, minus the prevalence of zoonotic disease. `diff_vec` is the difference between pathogen prevalence, based on the vector. It is the prevalence of diseases transmitted by humans, minus the prevalence of aquatic disease. The over-ID test is a Sargan test statistic. The null hypothesis is that the instruments are uncorrelated with ϵ . Accept means that null hypothesis cannot be rejected at the 5% or even the 10% confidence level. * indicates significance at the 5% level.

The Sargan test statistics (in the row labeled over-ID test) are chi-square statistics on the test of the null hypothesis that the instruments are uncorrelated with the regression residual ϵ . For every IV specification, we cannot reject this null hypothesis at the 5% or even the 10% level. Although a couple of the specifications have p-values between 10 and 20% (columns 1 and 5), the rest have p-values ranging from 30-85%, suggesting that the instruments are likely to be valid. We also computed Basmann statistics. They were quite close in value to the Sargan statistics in every instance.

Our baseline results use technology adoption as a dependent variable because it is most closely related to our theory. To interpret these results economically, it is helpful to re-run the estimation with productivity measures that are more familiar to macroeconomists: the Solow residual and output per worker. We again estimate the effect of individualism on out productivity, instrumenting individualism with the two differences in diseases. The coefficients in the Solow residual estimation tell us that a 1-standard-deviation increase in the Hofstede index corresponds to a 0.15 (for `diff_res`, 0.17 for `diff_vec`) increase in productivity. Since the Solow residual is measured as a fraction of

its value for the US, this tells us that a 1-standard-deviation increase in individualism increases productivity by 15% (17%) of the US value of productivity. For output per worker, the effects are even larger. A 1-standard deviation increase in individualism increases output per worker by 0.23, which represents an increase of 23% of US output per worker (same value for both instruments).

Measurement error. One potential concern with the quantitative interpretation of our results is that we believe that individualism is measured with error. This leads us to ask how much variation there is in true individualism, versus measured individualism. One way to infer the size of measurement error is to compare an OLS estimate, which suffers from attenuation bias, with an IV estimate, which is unbiased even in the presence of measurement error. Hall and Jones (1999) show that the square root of the ratio of the OLS estimate and IV estimates of the coefficient on individualism provide an unbiased estimate of the ratio of standard deviation of true individualism S with to the standard deviation of measured individualism \hat{S} :

$$plim \left(\frac{\hat{\beta}_{OLS}}{\hat{\beta}_{IV}} \right)^{1/2} = \frac{\sigma_S}{\sigma_{\hat{S}}}. \quad (4)$$

Column 3 of table 3 estimates the OLS regression of individualism on technological adoption. We find that the ratio of our OLS to our IV estimate is 0.82. This means that about 1/5th of the variation in measured individualism is measurement noise. Thus, the magnitude of our results may well be overstated, by up to 20%. But this still leaves a sizeable economic effect.

3.3 Could Social Structure Really Change in Response to Disease?

The idea that people might choose their social circles based on disease avoidance might sound far-fetched. But researchers in animal behavior have long known that other species choose their mates with health considerations in mind (Hamilton and Zuk, 1982). Furthermore, primate research has shown that the animals most similar to human beings behave similarly to the agents in our model. Their mating strategies, group sizes, social avoidance and barriers between groups are all influenced by the presence of socially transmissible pathogens (Loehle, 1995).

One might also question whether historical societies knew enough about contagion to make informed choices about social networks. Yet, historical documents reveal a reasonable understanding of epidemiology. For example, in the sixteenth century, when smallpox reached the Americas and became a global phenomenon, people understood that the skin lesions and scabs that accompany smallpox could transmit the disease. They knew that survivors of smallpox and other infections were immune to re-infection. The practice of inoculation, whereby people were intentionally exposed to disease was practiced in China, Africa and India. Similarly, the plague was recognized to

be contagious. Therefore, control measures focused primarily on quarantine and disposal of dead bodies. Even two thousand years ago, in biblical times, leprosy was understood to be contagious. Lepers, or suspected lepers, were forced to carry a bell to warn others that they were coming. Thus, the idea that one should avoid contact with others who carry particular contagious diseases is not just a modern idea.

4 Conclusions

Measuring the effect of social structure on the economic development of countries is a challenging task. Social structure is difficult to measure and susceptible to problems with reverse causality. We use a theory of social network evolution to identify properties of social networks that can be matched with data and to select promising instrumental variables that can predict network structure. The theory predicts that societies with higher disease prevalence are more likely to become collectivist: Their social networks will have dense connections within a group, but few connections to non-group members. Such networks inhibit disease transmission, but they also inhibit idea transmission. This model guides us to choose sociological measures of individualism and collectivism to measure the properties of interest of social networks. It also suggests that disease prevalence might be a useful instrument for social network because it is one important concern that societies incorporate when they choose their network.

The problem with using disease prevalence as an instrument in a regression of social structure on technological diffusion is that technology affects income and societies with higher incomes have better public health programs that prevent the spread of disease. Thus, pathogen prevalence is not exogenous. To avoid endogeneity problems, we use historical prevalence of disease as an instrument for recent social structures. We also use difference in disease prevalence as an instrument. Since diseases that are spread from person-to-person can be avoided by choosing one's social contacts carefully, these types of diseases should affect social network formation. But diseases that are spread by contaminated water or infected animals should not affect social networks because the social network structure has no effect on the probability of contracting the disease. Thus, the difference in the prevalence of these two diseases should predict social network formation. But this difference should be exogenous with respect to technology and income. As incomes increase, public health programs prevent both the spread of human-to-human disease and develop clean water and hygiene programs that inhibit the spread of other diseases. Thus, technology diffusion and income should not affect the difference in rates of disease prevalence. Using historical disease and difference in disease rates as instruments, we find that social structure has a significant effect on technology diffusion and an economically meaningful effect on incomes.

References

- ACEMOGLU, D., AND S. JOHNSON (2005): “Unbundling Institutions,” *Journal of Political Economy*.
- ACEMOGLU, D., S. JOHNSON, AND J. ROBINSON (2002): “Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distributions,” *Quarterly Journal of Economics*, CXVII(4), 1231–1294.
- ALGAN, Y., AND P. CAHUC (2007): “Social attitudes and Macroeconomic performance: An epidemiological approach,” Paris East and PSE Working Paper.
- BISIN, A., AND T. VERDIER (2000): “Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits,” *Quarterly Journal of Economics*, 115(3), 955–988.
- BISIN, A., AND T. VERDIER (2001): “The Economics of Cultural Transmission and the Evolution of Preferences,” *Journal of Economic Theory*, 97(2), 298–319.
- COLE, H., G. MAILATH, AND A. POSTLEWAITE (1992): “Social Norms, Savings Behavior, and Growth,” *Journal of Political Economy*, 100(6), 1092–1125.
- COMIN, D., B. HOBIJN, AND E. ROVITO (2006): “Five Facts You Need to Know About Technology Diffusion,” NBER Working Paper 11928.
- DURLAUF, S., AND W. BROCK (2006): “Social Interactions and Macroeconomics,” in *Post-Walrasian Macroeconomics: Beyond the Dynamic Stochastic General Equilibrium Model*, ed. by D. Colander. New York: Cambridge University Press.
- FERNÁNDEZ, R., AND A. FOGLI (2005): “An Empirical Investigation of Beliefs, Work and Fertility,” NBER Working Paper 11268.
- GORODNICHENKO, Y., AND G. ROLAND (2011): “Culture, institutions and the wealth of nations,” University of California at Berkeley Working Paper.
- GRANOVETTER, M. (2005): “The Impact of Social Structure on Economic Outcomes,” *The Journal of Economic Perspectives*, 19(1), 33–50.
- GREIF, A. (1994): “Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies,” *Journal of Political Economy*, 102, 912–950.
- HALL, R., AND C. JONES (1999): “Why Do Some Countries Produce So Much More Output per Worker than Others?,” *Quarterly Journal of Economics*, 114, 83–116.
- HAMILTON, W., AND M. ZUK (1982): “Heritable True Fitness and Bright Birds: A Role for Parasites?,” *Science*, 218, 384–387.
- HOFSTEDE, G. (2001): *Culture’s consequences : comparing values, behaviors, institutions, and organizations across nations*. Sage Publications.

- JACKSON, M. (2008): *Social and Economic Networks*. Princeton University Press.
- LOEHLE, C. (1995): “Social barriers to pathogen transmission in wild animal populations,” *Ecology*, 76(2).
- LUCAS, R., AND B. MOLL (2011): “Knowledge Growth and the Allocation of Time,” NBER Working Paper 17495.
- PERLA, J., AND C. TONETTI (2011): “Endogenous Risk and Growth,” NYU working paper.
- RAUCH, J., AND A. CASELLA (2001): *Networks and Markets*. Russell Sage, first edn.
- SMITH, K., D. SAX, S. GAINES, V. GUERNIER, AND J.-F. GUGAN (2007): “Globalization of Human Infectious Disease,” *Ecology*, 88(8), 1903–1910.
- TABELLINI, G. (2005): “Culture and Institutions: Economic Development in the Regions of Europe,” CESifo Working Paper No.1492.
- THORNHILL, R., C. FINCHER, D. MURRAY, AND M. SCHALLER (2010): “Zoonotic and Non-Zoonotic Diseases in Relation to Human Personality and Societal Values: Support for the Parasite-Stress Model,” *Evolutionary Psychology*, 8(2), 151–169.

A Proofs of Propositions

Proof of result 1 *In a collectivist network, where $\phi = 4$, there are n unique triples.*

Claim 1: Any three adjacent nodes are a triple.

Proof: Consider nodes j , $j + 1$ and $j + 2$. Since every node is connected to its adjacent nodes, $j + 1$ is connected to j and $j + 2$. And since every node is also connected to nodes 2 places away, j is connected to $j + 2$. Since all 3 nodes are connected to each other, this is a triple.

Claim 2: Any sets of 3 nodes that are not 3 adjacent nodes are not a triple.

Proof: Consider a set of 3 nodes. If the nodes are not adjacent, then two of the nodes must be more than 2 places away from each other. Since in a collectivist network with $\phi = 4$, nodes are only connected with other nodes that are 2 or fewer places away, these nodes must not be connected. Therefore, this is not a triple.

Thus, there are n unique sets of 3 adjacent nodes (for each j there is one set of 3 nodes centered around j : $\{j - 1, j, j + 1\}$). Since every set of 3 adjacent nodes is a triple and there are no other triples, there are n triples in the network. \square

Proof of result 2 *In an individualistic network, where where each person i is connected to $i - \psi$, $i - 1$, $i + 1$, and $i + \psi$, where $\psi > 2$, there are zero triples.*

Proof: Consider each node connected to an arbitrary i , and whether it is connected to another node, which is itself connected to i . In addition to being connected to i , node $i - \psi$ is connected to $i - 2\psi$, $i - \psi - 1$, and $i - \psi + 1$. None of these is connected to i . Node $i - 1$ is also connected to $i - 2$, $i - \psi - 1$ and $i + \psi - 1$. But none of these is connected to i . Node $i + 1$ is also connected to $i + 2$, $i - \psi + 1$ and $i + \psi + 1$. But none of these is connected to i . Finally, node $i - \psi$ is also connected to $i + \psi - 1$, $i + \psi + 1$ and $i + 2\psi$. But none of these is connected to i . Therefore, there are no triples among any connections of any arbitrary node i . \square

Diameter of network 1. Proof: Without loss of generality, consider the agent in the last position, the agent with location n on the circle. *Case 1: n even.* If n is even, then the farthest node from n is $n/2$. If each person is connected to the ϕ closest people, where ϕ is even, then they are connected to $\phi/2$ people on either side. Therefore, the shortest path will be the one that advances $\phi/2$ places around the circle, at each step in the path, until it is within $\phi/2$ nodes of its end point. For example, agent n reach $\phi/2$ in one step, ϕ in two steps and $n/2$ in $(n/2)/(\phi/2) = n/\phi$ steps, if n/ϕ is an integer. If dividing n by ϕ leaves a remainder m , then one step in the path to reach $n/2$ must

be only $m < n/2$ nodes away. Thus, when n is even, the shortest path to the furthest node $n/2$ is $\text{ceil}(n/\phi)$, where $\text{ceil}(x) = x$ if x is an integer, and is otherwise, the next largest integer.

Case 2: n odd. If n is odd, then $(n-1)/2$ and $(n+1)/2$ are equally far from node n . Each is $(n-1)/2$ nodes away. Following the same logic as before, the shortest path will be the one that advances $\phi/2$ places around the circle, and reaches the furthest node in $\text{ceil}((n-1)/2)/(\phi/2) = \text{ceil}((n-1)/\phi)$ steps.

Lastly, note that when n is even, $\text{ceil}(n/\phi) = \text{ceil}((n-1)/\phi)$. Note that, since $\phi > 1$ and both ϕ and n are integers, $\text{ceil}(n/\phi)$ and $\text{ceil}((n-1)/\phi)$ will only differ if $(n-1)/\phi$ is an integer, so that adding $1/\phi$ to it will make $\text{ceil}(n/\phi)$ the next largest integer. But if ϕ is even and $(n-1)/\phi$ is an integer, then $n-1$ must be even, which makes n odd. Thus, $\text{ceil}(n/\phi) = \text{ceil}((n-1)/\phi)$. \square

Average path length in network 1. Proof: Without loss of generality, consider the distance from the last node, n . n can be connected to nodes 1 through $\phi/2$ and $n-1$ through $n-\phi/2$ in 1 step. More generally, it can be connected to nodes $(s-1)\phi/2 + 1$ through $s\phi/2$ and $n-(s-1)\phi/2 - 1$ through $n-s\phi/2$, in s steps. For each s , there are ϕ nodes for which the shortest path length to n is s steps. We know from result 1 that when ϕ is even and n/ϕ is an integer, the longest path length (the diameter) is n/ϕ . Thus, the average length of the path from n to any other node is $1/n \sum_{s=1}^{n/\phi} \phi s$. Using the summation formula, this is $(\phi/n)(n/\phi)(n/\phi + 1)/2 = 1/2 + n/(2\phi)$. \square

Diameter of network 2. *The diameter of an individualistic network, with $n > 4$ nodes where each node i is connected to $i-4$, $i-1$, $i+1$, and $i+4$, is $\text{round}(n/8) + 1$.*

Proof: Without loss of generality, consider distances from the agent located at node n . n can reach nodes 1, 4, $n-1$ and $n-4$ in one step. It can reach nodes 2, 3, 5, 8 and $n-2$, $n-3$, $n-5$ and $n-8$ in two steps. In any number of steps $s > 1$, agent n can reach nodes $4(s-2) + 2$, $4(s-1) - 1$, $4(s-1) + 1$, $4s$ (moving clockwise around the circle) as well as $n-4(s-2) - 2$, $n-4(s-1) + 1$, $n-4(s-1) - 1$, $n-4s$ (moving counter-clockwise).

Let the operator $\text{floor}(x)$ be the largest integer y such that $y \leq x$. Define $\tilde{n} \equiv 4 * \text{floor}(n/8)$. Then $\tilde{r} \equiv n - 2 * \tilde{n}$ is the remainder when n is divided by 8. There are eight cases to consider, one for each possible value of \tilde{r} .

Case 1: $\tilde{r} = 0$. If the total number of nodes in the network n is a multiple of 8, then it takes $(1/4) * n/2$ steps to connect node n with node $n/2$, the geographically farthest node in the network. But it takes one more step to reach $n/2 - 1$, $n/2 + 1$. The nodes $n/2 - 2$ and $n/2 + 2$ can be reached in 2 steps from $n/2 - 4$ and $n/2 + 4$, each of which is one step closer to n than $n/2$ is. Thus, every node can be reached in $n/8 + 1$ steps, making the diameter of the network $n/8 + 1$.

Case 2: $\tilde{r} = 1$. In this case, \tilde{n} and $\tilde{n} + 1$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps. But it takes one more step to reach $\tilde{n} - 1$, $\tilde{n} - 2$, $\tilde{n} + 2$ or $\tilde{n} + 3$. Since $\tilde{n} = 4\text{floor}(n/8)$, $\tilde{n}/4 = \text{floor}(n/8)$, and thus the diameter is one step more than that, which is $\text{floor}(n/8) + 1$.

Case 3: $\tilde{r} = 2$. In this case, \tilde{n} and $\tilde{n} + 2$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps. But it takes one more step to reach $\tilde{n} - 1$, $\tilde{n} - 2$, $\tilde{n} + 1$, $\tilde{n} + 3$ or $\tilde{n} + 4$. Thus, the diameter is again $\text{floor}(n/8) + 1$.

Case 4: $\tilde{r} = 3$. In this case, \tilde{n} and $\tilde{n} + 3$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. It is still the case that it takes one more step to reach $\tilde{n} - 1$, $\tilde{n} - 2$ and $\tilde{n} + 1$. $\tilde{n} + 2$ can be reached in one additional step from $\tilde{n} + 3$, as can $\tilde{n} + 4$. And $\tilde{n} + 5$ can be reached in 2 additional steps from $\tilde{n} + 4$, which is one step closer to n than $\tilde{n} + 3$. Thus, every node can still be reached in $\text{floor}(n/8) + 1$ steps.

Case 5: $\tilde{r} = 4$. In this case, \tilde{n} and $\tilde{n} + 4$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. But now, getting to $\tilde{n} + 2$ requires 2 additional steps. Thus, the diameter of this network is $\text{floor}(n/8) + 2$.

Case 6: $\tilde{r} = 5$. In this case, \tilde{n} and $\tilde{n} + 5$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. Getting to either $\tilde{n} + 2$ or $\tilde{n} + 3$ requires 2 additional steps. Thus, the diameter of this network is $\text{floor}(n/8) + 2$.

Case 7: $\tilde{r} = 6$. In this case, \tilde{n} and $\tilde{n} + 6$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. In one additional step, one can connect from \tilde{n} to $\tilde{n} + 1$ or $\tilde{n} + 4$ or from $\tilde{n} + 6$ to $\tilde{n} + 2$ or $\tilde{n} + 5$. It takes two additional steps from \tilde{n} to connect to $\tilde{n} + 3$. Thus, the diameter of this network is $\text{floor}(n/8) + 2$.

Case 8: $\tilde{r} = 7$. In this case, \tilde{n} and $\tilde{n} + 7$ are equally far away from n in the network. Each requires $\tilde{n}/4$ steps to reach. In one additional step, one can connect from \tilde{n} to $\tilde{n} + 1$ or $\tilde{n} + 4$ or from $\tilde{n} + 7$ to $\tilde{n} + 3$ or $\tilde{n} + 6$. It takes two additional steps from either \tilde{n} or $\tilde{n} + 7$ to connect to $\tilde{n} + 2$ or $\tilde{n} + 5$. Thus, the diameter of this network is $\text{floor}(n/8) + 2$.

The one condition that encapsulates all 8 of these cases is $\text{diameter} = \text{round}(n/8) + 1$. To see this, recall that \tilde{r} is the remainder when n is divided by 8. When this remainder is zero, then $(n/8) + 1 = \text{round}(n/8) + 1$. When this remainder is less than 4, then $\text{floor}(n/8) + 1 = \text{round}(n/8) + 1$. When this remainder is 4 or more (4-7), then $\text{round}(n/8) = \text{floor}(n/8) + 1$, and therefore $\text{floor}(n/8) + 2 = \text{round}(n/8) + 1$. Thus, in each case of the 8 cases, the diameter of the network is equal to $\text{round}(n/8) + 1$. \square

Average path length of network 2. *In the example individualistic network, when $n/8$ is an integer, the average path length is $7/8 + n/16$. This is less than the average path length in a collectivist network with $\phi = 4$, when the network is large ($n > 6$).*

Proof: Without loss of generality, consider distances of each node from node n . n can reach 4 different nodes: 1, 4, $n - 1$ and $n - 4$ in one step. It can reach 8 different nodes 2, 3, 5, 8 and $n - 2$, $n - 3$, $n - 5$ and $n - 8$ in two steps. More generally, for a number of steps $s \geq 2$, agent n can reach 8 new nodes with each step. These nodes are: $4(s - 2) + 2$, $4(s - 1) - 1$, $4(s - 1) + 1$, $4s$ (moving clockwise around the circle) as well as $n - 4(s - 2) - 2$, $n - 4(s - 1) + 1$, $n - 4(s - 1) - 1$, $n - 4s$ (moving counter-clockwise). This rule holds until the number of steps s reaches $n/8$, the number of steps to travel approximately half way around the circle. At that point, the number of additional nodes that can be reached in an additional step depends on the size of the network. There are 8 cases to consider.

Recall that $\tilde{n} \equiv 4 * \text{floor}(n/8)$ and that $\tilde{r} \equiv n - 2 * \tilde{n}$ is the remainder when n is divided by 8. There are eight cases to consider, one for each possible value of \tilde{r} .

If the total number of nodes in the network n is a multiple of 8, then it takes $n/8$ steps to connect node n with node $n/2$. Using the algorithm above, it also takes $n/8$ steps to connect with nodes $n/2 - 6$, $n/2 - 5$, $n/2 - 3$, $n/2 + 6$, $n/2 + 5$ and $n/2 + 3$. But this is 7 total nodes instead of 8 total nodes because when the total number of steps being considered is $n/8$ ($s = n/8$) nodes $4s$ and $n - 4s$ are both equal to node $n/2$.

It takes one more step to reach $n/2 - 1$, $n/2 + 1$. The nodes $n/2 - 2$ and $n/2 + 2$ can be reached in 2 steps from $n/2 - 4$ and $n/2 + 4$, each of which is one step closer to n than $n/2$ is. Thus, 4 additional nodes can be reached in $n/8 + 1$ steps.

Counting up, there is 1 node (n) reachable in zero steps, 4 nodes reachable in 1 step, 8 nodes reachable in s steps for $s \in \{2, 3, \dots, n/8 - 1\}$, 7 nodes reachable in $n/8$ steps and 4 nodes reachable in $n/8 + 1$ steps. That makes the average path length $1/n$ times the sum of all the path lengths to the n nodes: $1/n[4 + 8 \sum_{s=2}^{n/8-1} s + 7 * n/8 + 4 * (n/8 + 1)]$. Applying the summation formula, $8 \sum_{s=2}^{n/8-1} s = 8(n/8)(n/8 - 1)/2 - 8$, where the -8 corrects for the fact that the sum begins at $s = 2$, rather than at $s = 1$. Substituting in this formula and collecting terms, this is $1/n[4 + 8(n/8)(n/8 - 1)/2 - 8 + 11n/8 + 4] = 1/8n[n(n - 8)/2 + 11n] = 7/8 + n/16$. \square

Proof of result 3 *For a large network ($n > 8$) where $n/8$ is an integer, the individualistic network has a smaller diameter and a shorter average path length than a collectivist network with equal size n and equal degree $\phi = 4$.*

Proof: We begin with the diameter. If $n/8$ is an integer, then n must also be a multiple of 4. Since the diameter of the collectivist network with $\phi = 4$ is $(n - 1)/4$ or the next highest integer, that integer-valued diameter will be $n/4$. Likewise, in the individualistic network, $\text{round}(n/8) + 1 = n/8 + 1$. Thus, the diameter of the individualistic network is smaller iff $n/8 + 1 < n/4$, which is true iff $n > 8$.

Next, we turn to average path length. If $n/8$ is an integer and $\phi = 4$, then result ?? tells us that the average path length of a collectivist network is $1/2 + n/8$. Result ?? tells us that the average path length of the individualistic network is $7/8 + n/16$. Thus, the individualistic path length is smaller iff $7/8 + n/16 < 1/2 + n/8$, which is true iff $n > 6$. Thus, $n \geq 8$ is a sufficient condition for the individualistic network to have a shorter average path length.

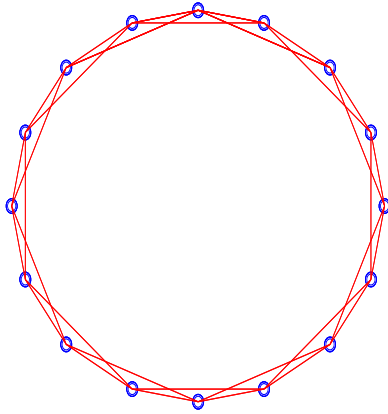
Since each disease is transmitted with probability 1, the average lifetime is the average path length from the point of origin of the disease. Since the network is symmetric, this is simply the average path length. Similarly, the maximum lifetime is the diameter.

Similarly, since the technology is also transmitted with probability 1, the average and maximum discovery times are also the average path length and diameter. \square

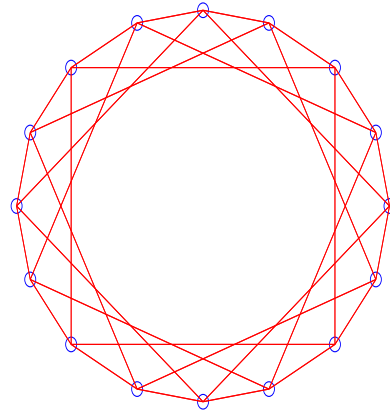
coming soon. Proofs of the remaining results are in progress.

Both results follow from the fact that zero disease, all individualistic and all collectivist are all absorbing states. From any other state, there is a positive probability of transiting to each absorbing state. And, there are a countably finite number of states. These conditions guarantee that the only steady states are the three absorbing states and that the economy ends up in one of these states, in finite time, with probability 1.

Clustered Network



Dispersed Network



B Visual Representations of our collectivist and Individualistic Networks

The connection matrix N of the collectivist network is

$$\begin{bmatrix} 0 & 1 & 1 & 0 & \dots & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & \dots & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & \dots \\ & & & \vdots & & & & \\ 1 & 1 & 0 & \dots & 0 & 1 & 1 & 0 \end{bmatrix}.$$

This matrix has zeros on the diagonal. (Typically, we don't consider one's relationship with oneself to be a connection.) It has two ones just to the left and right of the diagonal, indicating that each person is connected to the two people to their left and the two people to their right. The three entries in the top-right and bottom-left corners also have ones. This captures the connection between agents 1 and 2, who are located adjacent to agents $n - 1$ and n on the circle. The rest of the entries are zeros, indicating that these individuals are not directly connected in the network.

The connection matrix N of our example individualistic network is

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 \\ & & & \vdots & & & & & & & & \\ 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Again, there are zeros on the diagonal. There is one 1 entry just to the left and to the right of the diagonal. This represents each agent's connection with their immediate neighbor. There is also a 1 four columns to the left and four columns to the right of the diagonal, indicating the connection between agent j and $j + 4$, and between agent j and $j - 4$. As before, there are a handful of 1's in the top-left and bottom-right corners, indicating the connections between agents near n and those near 1, who are one or four spots away from each other on the circle. The rest of the entries are zeros, indicating that these individuals are not directly connected in the network.

C Data Details

Summary statistics for each of the variables we use are described in table 4.

Table 4: **Summary statistics**

	Obs	Mean	Std. Dev.	Min	Max
CHAT_tech	75	47.84	21.33	4	95
Hofstede	78	42.10	22.81	6	91
Pathogens	78	13.09	6,12	1	24
Diff_reservoir	78	7.45	4.67	-1	16
Diff_vector	78	2.76	1.30	0	5
Output per worker	67	.42	.28	.032	1
Solow Residual	67	.64	.30	.079	1.2
Density70	77	174.86	534.52	1	3587
Latitude	78	30.85	17.39	0	60
Lrgdpch70	74	8.72	1.031	6.33	11.49

Pathogen prevalence The pathogen prevalence measure used in these baseline regressions is from Murray and Schaller "Historical Prevalence of Infectious Diseases within 230 geopolitical regions: A Tool for investigating the origins of culture". They extended the work of Gangestad and Buss (1993) who employed old epidemiological atlases to rate the prevalence of seven different kinds of disease-causing pathogens and combined estimates into a single measure indicating the historical prevalence of pathogens in each of 29 countries. More recently, Murray and Schaller used a similar procedure to rate the prevalence of nine infectious diseases in each of 230 geopolitical regions world. The nine diseases coded were leishmanias, schistosomes, trypanosomes, leprosy, malaria, typhus, filariae, dengue, and tuberculosis. Epidemiological atlases were used to estimate the prevalence of each of these nine diseases in each region. For eight of them (excluding tuberculosis), prevalence of each disease was based primarily on epidemiological maps provided in Rodenwaldt and Bader's (1952-1961) *World-Atlas of Epidemic Diseases* and in Simmons and others (1944) *Global Epidemiology*. A 4-point coding scheme was employed: 0 = completely absent or never reported, 1 = rarely reported, 2 = sporadically or moderately reported, 3 = present at severe levels or epidemic levels at least once. The prevalence of tuberculosis was based on a map contained in the National Geographic Society's (2005) *Atlas of the World*, which provides incidence information in each region for every 100,000 people. Prevalence of tuberculosis was coded according to a 3-point scheme: 1 = 3 - 39, 2 = 50 - 99, 3 = 100 or more. For 160 political regions, they were able to estimate the prevalence of all nine diseases. The remaining 70 regions typically lacked historical data on the prevalence of either tuberculosis or leprosy; 6 of these regions lacked data on malaria as well. Therefore, in addition to create a 9 item index of disease prevalence (computed for 160), they also created a seven item index (excluding both leprosy and tuberculosis) for 224 regions and a six item index (excluding also malaria) for 230 regions. To ensure all different disease prevalence indices were computed on a common scale of measurement, all nine disease prevalence ratings were standardized by converting them to z scores. Each overall disease prevalence index was then computed as the mean of z scores of the items included in the index. Thus, for each index the mean is approximately 0, positive scores indicate disease prevalence that is higher than the mean and negative scores indicate disease prevalence that is lower than the mean. Figure 8 uses a color-coded map to summarize the data.

Classification of Infection Diseases Epidemiologists usually classify infectious diseases according to two important epidemiologic characteristics - their means of transmission and the reservoir of the organism. According to their means of transmission, diseases can be classified into five distinct categories: contact, foor- or water-borne, airborne, vector-borne (carried by some animal that comes in contact with a person) or perinatal (contact infection during pregnancy or delivery).

The second means for epidemiological classification of infectious diseases is according to their major reservoirs in nature. The reservoir is any person, animal, plant, soil or substance in which an infectious agent normally lives and multiplies. The reservoir typically harbors the infectious agent without injury to itself and serves as a source from

which other individuals can be infected. The infectious agent primarily depends on the reservoir for its survival. It is from the reservoir that the infectious substance is transmitted to a human or another susceptible host. Animals often serve as reservoirs for diseases that infect humans. The major reservoir for *Yersinia pestis*, the bacteria that causes plague, is wild rodents. There are also nonliving reservoirs. Soil is the reservoir for many pathogenic fungi as well as some pathogenic bacteria such as *Clostridium tetani*, which causes tetanus. Knowing the reservoir of the agent in addition to the means of transmission is usually needed to develop a strategy to prevent transmission. When organisms are classified according to their reservoirs in nature, four general categories are often considered: 1) Human 2) Animal (often called zoonoses) 3) Soil 4) Water.

Using information from GIDEON (Global Infectious Diseases and Epidemiology Network), we can classify the infectious diseases in the historical dataset according to agent, reservoir, vector and vehicle of transmission. GIDEON was founded in 1992 to provide health professionals with a diagnosis and reference tool for Infectious Diseases, Microbiology and Occupational Toxicology. The data in GIDEON are accessed and collated through a system of computer macros and dedicated source lists developed over the past 15 years. The sources for data included in GIDEON currently include all relevant health ministry publications (electronic and print), peer review journal publications and standard texts. A partial listing is available at <http://www.gideononline.com/resources.htm>. The quality and frequency of data input vary widely from source to source. The entire GIDEON database is updated weekly, and occasionally on a daily basis when major events occur (i.e., new case of avian influenza or an Ebola outbreak). According to these four categories we distinguish two diseases that are transmitted from human to human, through direct contact, and have a human reservoir. These are leprosy and tuberculosis. Then we have four diseases that reproduce in humans as well as animals and are transmitted through a zoonotic vector (mosquitoes or flies): malaria, trypanosoma, leishmaniasis and dengue. One disease that has a human reservoir but a zoonotic vector (filariasis). And lastly two diseases that are transmitted through water or food (schistosomiasis and typhoid fever). Figure 7 summarizes the classification of the pathogens that we collected data on.

Hofstede Index Hofstede defines individualism in the following way: "Individualism (IDV) on the one side versus its opposite, collectivism, that is the degree to which individuals are integrated into groups. On the individualist side we find societies in which the ties between individuals are loose: everyone is expected to look after him/herself and his/her immediate family. On the collectivist side, we find societies in which people from birth onwards are integrated into strong, cohesive in-groups, often extended families (with uncles, aunts and grandparents) which continue protecting them in exchange for unquestioning loyalty."

The original questions from the 1966-1973 Hermes (IBM) attitude survey questionnaires used for the international comparison of work-related values were listed in Hofstede (1980, Appendix 1). Appendix 4 of the same book presented the first Values Survey Module for future cross-cultural studies. It contained 27 content questions and 6 demographic questions. This VSM 80 was a selection from the IBM questionnaires, with a few questions added from other sources about issues missing in the IBM list and judged by the author to be of potential importance. In the 1984 abridged paperback edition of Hofstede (1980) the original IBM questions were not included, but the VSM 80 was.

A weakness of the VSM 80 was its dependence on the more or less accidental set of questions used in the IBM surveys. Therefore in 1981 Hofstede through the newly-founded Institute for Research on Intercultural Cooperation (IRIC) issued an experimental extended version of the VSM (VSM 81). On the basis of an analysis of its first results, a new version was issued in 1982, the VSM 82. This was widely used.

The VSM 82 questionnaire is too long to include in its entirety. However, factor analysis of 14 "work goals" questions from the survey produced 2 factors that together explained 46% of the variance. The first factor was demographic characteristics. The second set pertain to work goals. Here are those key work goals questions:

In this section, we have listed a number of factors which people might want in their work. We are asking you to indicate how important each of these is to you. Possible answers: of utmost importance to me (1), very important (2), of moderate importance (3), of little importance (4), of very little or no importance. "How important is to you":

1. Have challenging work to do
2. Live in an area desirable to you and your family
3. Have an opportunity for high earnings
4. Work with people who cooperate well with each other
5. Have training opportunities
6. Have good fringe benefits
7. Get the recognition you deserve when you do a good job

8. Have good physical working conditions
9. Have considerable freedom to adopt your own approach to the job
10. Have the security that you will be able to work for your company as long as you want to
11. Have an opportunity for advancement to higher level job
12. Have a good working relationship with your manager
13. Fully use your skills and abilities on the job
14. Have a job which leaves you sufficient time for your personal or family life

The answers to these questions were used to develop the Hofstede index of individualism for each country.

Figure 7: Classification of our nine infections diseases. Source: GIDEON database.

Diseases	Agent	Reservoir	Vector	Vehicle
Leprosy	BACTERIUM.	Human	None	Patient secretions
Tuberculosis	BACTERIUM	Human, Cattle	None	Air, Dairy products
Malaria	PARASITE - Protozoa	Human, Primate (Plasmodium knowlesi)	Mosquito (Anopheles)	Blood
Trypanosoma	PARASITE - Protozoa	Human, Deer, Wild carnivore, Cattle	Fly (Glossina or tsetse fly)	None
Leishmaniasis	PARASITE - Protozoa	Human, Rodent, Dog, Fox	Fly	Blood
Dengue	VIRUS - RNA	Human, Mosquito, Monkey	Mosquito (Stegomyia aegypti, S. albopictus)	Blood (rare)
Filariasis	PARASITE - Nematoda	Human	Mosquito (Anopheles, Aedes, Culex)	None
Schistosomiasis	PARASITE - Platyhelminthes, Trematoda	Snail (Bulinus, Planorbis) Rarely baboon or monkey	None	Water
Typhoid fever	BACTERIUM	Human	None	Water, Food

Figure 8: A world map of historical pathogen prevalence.

