Introduction: Methodology and Motivation

Nathan Smith, 1/26/2009

Economists love technology, yet find it elusive to understand. For example, the result of the "Solow model," the standard textbook model of economic growth, is (usually interpreted as) that technology is (a) the sole cause of long-run economic growth, but (b) "exogenous," that is, a result of noneconomic causes, not explainable by economic theory. Technology enters into the production function, causing more output to be made—somehow—from the same amount of labor and capital. Use (say) five units of capital and ten units of labor today, and you'll get 100 widgets, but tomorrow, you'll get 105. The model seems to say that productivity growth is a result of mere *time*, which is a violation of what is sometimes called the "law of science," the principle that if you do the same thing over again, it should have the same effect.

Is it possible to have an economic theory of technological change? One reason to think technological change may permanently elude systematic study is that any *particular invention* is, in a sense, unpredictable by definition. To illustrate, suppose that it is 10,000 B.C., and one caveman says two another:

"I predict that ten years from now, someone will invent the wheel."

"What's a wheel?" asks his companion.

"A wheel is a round object that can move along the ground by virtue of its shape, which has no sharp edges or corners. It is useful for moving and carrying things... But—wait a minute! No one can be *going* to invent the wheel, for I have just invented it!" (The example is taken from Alasdair MacIntyre's *After Virtue.*)

But a theory of technological change does not need to be able to predict particular inventions. Perhaps, indeed, it need not explain *invention* at all. Technological change in the economic sense is not about pure invention, but the widespread adoption of new products and processes that improve efficiency and expand society's productive capacity. Some accomplishments that are triumphs of human ingenuity and mastery of nature—the Pyramids of Egypt, the atomic bomb, or putting a man on the moon—are of little importance in the *economic* history of technology. By contrast, innovations such as the fast-food restaurant, the factory assembly line, or containerized shipping are important technological developments in the economic sense even though they involve no breakthroughs in science-based technology. There may be a lag of decades or centuries between the initial invention and the widespread, effective adoption of an invention such as the water-mill, printing, or gunpowder.

The crucial point is that *at any given time, there are unused technological possibilities*, whether they have been "invented" or not, and an economic theory of technology must explain how economies adopt and are transformed by them. Even if specific inventions are unpredictable, some generalizations about technological change can safely be made.

It is easy to observe that US labor productivity has generally grown by 1-2% per year or so for generations, and to project by simple induction that this trend is likely to continue. And we know some of the reasons why. Research and development spending by corporations and industrial research laboratories has often led to the introduction of new products and processes. Science as an independent enterprise, carried out partly for practical purposes and partly from the pure desire to know, and financed by both public and private sources, makes new discoveries about the laws of nature, some of which can then be exploited by engineers and inventors. Entrepreneurs and manufacturers, meanwhile, engage in experimentation, both conscious and accidental, which sometimes leads to better products and processes. Market competition both motivates entrepreneurs to try to adopt the best products and processes available, and selects among them by rewarding entrepreneurs with the most efficient processes and the most popular products, while their rivals lose money and disappear.

If we know the above by experience and intuition, why do we need a theory—that is, a model?

It's Saturday evening and you and your friend want to go out to a small Mexican restaurant. "Let's go late, maybe 9 p.m.," you say. "Then we probably won't have to wait for a table."

You may just have been engaged in social-scientific modeling.

First, you made some assumptions, such as the following:

1. There is a limited number of tables in the restaurant. At the busiest times, it is too few to sit all those who want to dine.
2. Customers come to the restaurant when they want to eat, then occupy the tables while they do so, usually for 1-2 hours.
3. Most people like to eat dinner around 7 p.m., give or take an hour.

From (1), (2), and (3), you infer that by 7 p.m., the restaurant will be crowded with customers. A queue will form, even as customers keep arriving. For some time after that, customers who arrived at 7 p.m. and later will be occupying the tables. But as those customers finish their meals and leave, and as the stream of new customers slows down, the queue will shrink and disappear. By 9 p.m., customers will be seated as soon as they show up.

Or maybe you weren't making a model. Maybe you've gone to that restaurant many times on Saturday nights, earlier and later, and you've noticed that when you arrive at 7 p.m., you usually have to wait, whereas if you arrive at 9 p.m., you don't, without ever thinking why. In that case, you used only *induction.* If you started by thinking about people's behavior and the physical constraints of the restaurant, then used *reasoning* to predict when there would be a wait, you were relying (partly) on *deduction.* Deduction might save you some time sitting in the entry of Chipotle. Of course, the premises from which you deduced your conclusion were presumably derived from induction. Induction and deduction are a bit like scissor blades: you need to use both to learn about the world.

Models are the deductive part of thinking. When the deduction you are engaged in is simple, you may be able to do it implicitly, without laying our your assumptions or your chain of inference. But if others don't understand and you need to explain it to them, or if you yourself get confused—and the more complex your deductive enterprise is, the more likely you or your interlocutors are to get confused—you may need to start from the beginning and take it step by step. That is what economic modelers do.

Formal economic models give definite and logical, if not always empirically correct, answers to important and interesting questions. In some cases, what a formal model shows may seem obvious intuitively (at least to some). In this case the model serves as a "mnemonic"—something to help remember it by—and a "heuristic"—something to help explain it to newcomers. For example, the supply-and-demand model taught in any introductory economics course can be explained, laboriously, through verbal arguments, but a simple chart can convey it far more quickly, both to students to whom it is a new insight, and among analysts and policymakers who are aware of it but often overlook or fail to apply it in their thinking and decision-making.

In other cases, a formal model may yield highly non-obvious conclusions which then invite empirical testing. Even a model whose conclusions are clearly wrong can be useful, because, provided it is logically correct so that the conclusions follow from the premises, it forces us to re-examine the assumptions of the model to see which ones might be mistaken or incomplete. For example, the Solow model without technology would make the counter-factual prediction that economic growth stops. The falsity of this prediction is a stimulus to look for factors other than capital accumulation alone to explain ongoing economic growth.

It would be useful to know:

- whether and how national, global, and regional policy can influence the advance of the technological frontier;
- whether world poverty reflects a lack of technology in poor countries, and if so, whether and how their technological level can be raised;
- whether free-market capitalism is especially conducive to technological progress, and if so, how and why;
- whether technological change is a function of (among other things) *culture*, as is sometimes claimed; and
- whether and how technological change is related to other economic and social phenomena, such as business cycles, inequality, world trade, urbanization, and the spread of democracy and freedom.

A good formal theory of technological change would yield definite predictions about the answers to some or all of these questions. Some of its predictions would, hopefully, be testable empirically. In the absence of a formal theory, we can still try to look at the evidence. But empirical work is no substitute for good theory. Theory tells us what evidence to look for, and how to interpret it when we get it. If Sherlock Holmes has formulated a theory that a younger rival killed the elderly biology professor (whose job he wanted) by pouring into his drink a newly-developed, undetectable poison, he will carefully interview the witnesses to see if the suspect was ever alone with the wineglasses, and insist that the bright young scientist's recent purchases of arcane books on toxicology are relevant to the case. Watson, meanwhile, is bewildered by Holmes' lack of interest in the old man's angry dispute with his dissolute heir, which occurred on the balcony just minutes before his mysterious collapse. Economic Watsons with no theory, or an inadequate one, may not know evidence when they see it, and may end up filling fat volumes with statistical and historical spadework in support of vague claims that can be refuted with one sentence of sound theory.[1]

We already know—I believe—the "stylized facts" we need to illuminate the most fundamental and pervasive reasons why the economy grows. Indeed, these reasons were largely understood by Adam Smith, and expressed in the brilliant, though not rigorous by today's standards, verbal arguments of *The Wealth of Nations*, namely:

- the division of labor;
- that gains from trade and specialization are limited by the extent of the market and the scope for such gains increases as the market grows; and
- the way the 'invisible hand' of the market orchestrates self-interested (but subject to moral and/or legal side-constraints to respect property and refrain from coercion) individual effort for the common good.

Since then, economics at its best has reiterated, clarified, refined, formalized, demonstrated, applied and extended the insights of Adam Smith, greatly enhancing our understanding of the workings of (and the case for) market capitalism. However, beginning in the late 19th century, new ways of formalizing one of Adam Smith's great insights, the 'invisible hand', have tended to obscure his other great insight, the division of labor. To understand economic growth, we need to renew our understanding of the division of labor. To do that, we need new methods.

Let's look more closely at the Mexican restaurant model. What are its key elements? First, there are *agents*—the customers. Agents have *states*—hungry, not hungry, in the restaurant, not in the restaurant—*behaviors*—go to the restaurant, order, leave—and (most importantly: Marshall called these the 'motive-power' behind economic phenomena) *goals/purposes*—to get a meal, preferably when one is hungry for it, but without enduring hunger too long. Second, there are *constraints*—the number of tables in the restaurant. These two elements, *purposive agents* and *constraints,* are good starting places for models, because we know that they are real. We know by introspection that

[1] Famous example: the Marxist theory of 'exploitation' is refuted merely by the observation that workers will not agree to wages that do not make them better off than their next best alternative.

there are purposive agents because we *are* purposive agents: we want things, and we make choices in an effort to get them. And we know that physical laws constrain what can happen, and that in order to get some things we want we often have to sacrifice others.

To the extent that models are built on the foundations of purposive agents and constraints, *and nothing else*, they satisfy a desideratum that is regarded by many as the gold standard of economic theory: *methodological individualism.* The Marxist theory of class-struggle is an example of a theory that is *not* methodologically individualist:

> The history of all hitherto existing society [2] is the history of class struggles.
>
> Freeman and slave, patrician and plebian, lord and serf, guild-master [3] and journeyman, in a word, oppressor and oppressed, stood in constant opposition to one another, carried on an uninterrupted, now hidden, now open fight, a fight that each time ended, either in a revolutionary reconstitution of society at large, or in the common ruin of the contending classes.[2]

The entities that drive the action in Marx's class-struggle model are not individuals, but *classes.* Classes have purposes and fight with one another. But do we know whether, or in what sense, classes exist, let alone whether, or in what sense, they have purposes? The presence of these collective entities in Marx's model shows that the theory is, at best, incomplete. What classes are, and how the collective causes for which they fight emerge from the goals and behaviors of the individuals who comprise them, must be explained before Marxist class-struggle can be regarded as a coherent theory.[3]

The standard supply-and-demand model of microeconomics is not explicitly methodologically individualist. The demand curve shows how much 'customers,' a collective entity, will buy at any given price, while the supply curve shows how much 'suppliers,' another collective entity, will sell. To make it so, one must take a step back and begin by considering the purposive behavior of individual customers and suppliers.

Modern mathematical economists can do this by attributing a 'utility function' to agents and a 'production function' to suppliers. If we then given each agent a budget constraint, it is possible to write down his maximization problem, and calculate how much he would buy/sell at any given price—in other words, to calculate individual demand curves. By a similar process, we may (with more assumptions, and depending on the nature of the production function) be able to calculate individual supply curves for each supplier. Individual demand and supply curves can then be summed to get social demand and supply curves.

Can trading begin? But at what price? Where do prices come from? It is at this step that traditional, neoclassical economic models take an ingenious shortcut: *the theorist sets the prices*. First, he deduces, by solving a system of equations, the vector of prices at which the quantity demanded exactly equals the quantity supplied, that is, the 'market-clearing'

---

[2] *The Communist Manifesto*

[3] Problematically for Marx, workers and capitalists do not typically cooperate, but compete with each other. For this and other reasons, Marx's theories are not influential among academic economists today.

price vector. He then assumes that agents and firms—somehow—know these prices, and only trade with each other at the market-clearing price. It is assumed that there are enough agents and firms in the economy that each one is 'small,' unable to influence the prevailing price, and must accept the price set by the 'market,' that is, by the theorist. This is called the 'price-taking' assumption.

This result is sometimes called 'the decentralized solution.' This is accurate in the sense that, given the price-taking assumption, prices established in this way are incentive-compatible. Agents will willingly engage in the trades that give rise to a Pareto-optimal allocation with no surpluses or shortages. The term 'decentralized' is ironic in the sense that prices are centrally imposed by the theorist. The theorist's role in equilibrating the model subtly compromises its methodologically individualist credentials.

For our purposes, the most serious problem with this type of equilibrium approach is that the ways theorists must rig the models mathematically in order to make them solvable blinds them to fixed costs and the division of labor. This is one of the cases where a useful theoretical advance has created a new blind spot. The neoclassical paradigm, which emerged from the 'marginalist revolution' of the late 19th century and has since constituted the mainstream of 20th-century economics,[4] was a great advance in the theory of value. It replaced the inadequate 'labor' or 'cost' theory of value of the classical economists—according to Adam Smith, "the real cost of everything... is the toil and trouble of acquiring it" (Smith, *Wealth of Nations*, 43)—with theory of value based at once on 'marginal utility' and 'marginal cost,' which, in equilibrium, are equal. But this innovation in value and price theory requires 'perfect competition,' and perfect competition, as the marginalists conceive it, turns out to requires a production function characterized by 'constant returns to scale' (CRS) production function.[5] That, in turn, has the double fault of (a) being unrealistic, and (b) of obscuring the way fixed costs motivate division of labor. After all, in the absence of fixed costs, for example in capital investment or acquisition of specialized skills, you could make everything in your backyard and you wouldn't need to trade with anybody![6] CRS production functions therefore cannot be integrated with Smith's fundamental insights about the link between the division of labor and the wealth of nations. As a result, division of labor has often been marginalized by economic theory in the neoclassical tradition.

---

[4] At least on questions of long-run growth. Keynesian economics, which has been very influential and even dominant in *macro*economics especially in the middle decades of the 20th century, departs from the neoclassical paradigm in its analysis of the *short run*, but tends to defer to the neoclassical paradigm on long-run questions like growth.

[5] Actually, it *is* possible to have fixed costs, but only if the production/cost function has the peculiar property that over some range of possible outputs it *mimics* a CRS function. Thus, average cost curves may be U-shaped as long as they have a 'flat bottom.' It is hard to specify mathematically what function would have this shape, or to argue intuitively why real production functions should have it.

[6] "But I don't know *how* to make everything," one might be tempted to object. But *acquiring the know-how* is also a fixed cost. Differences in natural propensity for different jobs—if they exist, which is hard to prove since training and natural propensity are difficult to distinguish empirically—are hardly ever considered in economic models, so I cannot say whether taking them into account would make constant returns compatible with division of labor.

There is a useful joke about scientists who end up studying irrelevant questions by following the methodological course of least resistance. Walking home one night, I see my next-door neighbor on his knees under the streetlamp, crawling and scanning the pavement.

“What’s the matter?” I ask.

“I lost my keys.”

“Did you lose them right here?”

“No, I lost them over there in the bushes. But the light’s better here.”

Economists from Solow onwards who try to study economic growth in model economies characterized by constant returns to scale and perfect competition may be like my neighbor looking for his key under the streetlamp. Or, to vary the metaphor, think of verbal theorists as pedestrians, while the mathematical modeling that has come to dominate the neoclassical tradition is like a road bike.

Pedestrians (verbal theorists) can go almost anywhere—over paved roads and dirt roads, bike paths and foot paths, even, with difficulty, straight through the forest. But feet are slow. The best hikers may find some beautiful places (important insights), but it will take them a long time (thousand-page volumes), their legs will get tired, and they might get lost in the forest (make mistakes) on the way home. When they do find worthwhile destinations, they often give bad directions (arguments) that cause others to get lost (to dismiss valid insights because of fallacious arguments used to support them).

Sightseers (theorists) who discover the road bike (mathematical economics) can go farther, faster. Not everyone knows how to ride a bike, and some who try take nasty spills (some verbal theorists like Schumpeter admired but could never master the mathy new tools) but for those who learn it, biking will easily become the preferred mode of tourism. Even pedestrians may prefer to go to bike-accessible destinations, because frequent traffic and good signposting makes the trails safer, and because of the lemonade stands (textbooks) that enterprising bike-messengers find it easy to supply at the end of trail.

But road bikes only work well on paved roads. Sightseeing cyclists who venture onto dirt roads will soon be stopped by flat tires (problems of mathematical intractability). Some curmudgeons (the ‘Austrian school’ of economics) will disdain the new fashion and insist that hiking is the only way to get to any good places, but cyclists will suspect, with some reason, that this is the sour-grapes reaction of those who never learned to ride. Meanwhile, the cyclists will begin to assume that, because the bike-accessible destinations are the only ones people *talk* about, they must be the best ones to see. Some formerly popular destinations, accessible only by footpath or gravel road, will be neglected.

One of the destinations that is not located along the bike paths of perfect competition is a famous yet rarely-visited place, Mount Division-of-Labor, once enthused about by the great early hiker, Adam Smith, who said that from there you could see the whole landscape.

This book uses a relatively new method called *agent-based modeling*. Continuing the sightseeing metaphor, we may compare it to a *Jeep*, a rough, robust vehicle that can travel over rocky trails that the road bikes of mathematical-neoclassical theory can't handle. There are also certain bike paths through the woods (rational expectations, perfect competition) where bikes can go and the Jeep can't. And it can't necessarily go anywhere that pedestrians can't get to. But it can get to a lot of places faster—that is, more rigorously and persuasively.

As before, our analysis of a market start with *purposive agents*—consumers that want to maximize utility, firms that want to maximize profits—and *constraints*—consumers' budgets, firms' production functions, and we want to know how trading will develop. If we don't want to set prices centrally, we're back to the question, where do prices come from? In the real world, the omniscient theorist isn't there to set prices, so prices have to be set by firms or agents. In practice, it usually seems to be firms. Firms observe market conditions, imperfectly, and take account of them in their price-setting decisions, but they don't know whether or not the prices they set correspond to those that would equilibrate supply and demand in the market.

So we might proceed like this:

- Get sheets of paper, maybe a lot of them.
- On each sheet, write down an entity name (e.g., 'consumer' or 'firm') and some data about its states (e.g., bank account, inventory, price), behaviors (e.g., work, buy, consume, produce), and goals (e.g., a utility function).
- Regard the sheets as representing purposive agents.
- Think of decision rules for the agents and firms as they interact with each other.
- Taking each sheet in turn, let the agent or firm make decisions and implement behaviors in pursuit of its goals. Where necessary, let them interact with other sheets, e.g., to buy goods or hire workers. Let producers set prices at which they will sell.

A procedure along these lines would qualify as *agent-based modeling*. It is a simple and direct, almost crude, approach to studying the operation of markets. We can note immediately one advantage that it has over the neoclassical approach: it is *more methodologically individualist*, because the theorist doesn't intervene to equilibrate the market. Any equilibrium (or quasi-equilibrium) that occurs will emerge from the decisions and interactions of agents. Clearly, though, anyone who undertook this approach would be in for an ordeal of paper-shuffling and scribbling that would last for days, or even—if they wanted to model a large, many-agent economy, or if they wanted to “sweep the parameter space” to establish the generality of their results—years. No wonder the neoclassical economists came up with their ingenious mathematical shortcuts rather than trying to do things this way! But this cost barrier to agent-based modeling has been flattened by the IT revolution. Quite large and complex models, which would have taken, perhaps, months or years with pen and paper, can be executed on an average office computer with moderate run-times.

In the agent-based models in this book, finite but arbitrarily large numbers of agents (not a single ‘representative’ agent) engage in production and exchange, *over time* (an element omitted from equilibrium models), and some of them (producers, specifically) set prices (and wages) themselves, though while doing so they observe and take into account the prices applied in recently-executed transactions (‘market prices,’ if you will, but not ‘equilibrium’ market prices). Agents and entrepreneurs use utility and profit functions to make decisions, but they do not ‘maximize utility’ or ‘maximize profit’ in a strict sense, because they do not—nor do I, the theorist!—have the full information which would be needed to determine *ex ante* what decisions will prove to have been utility- or profit-maximizing *ex post.* Moreover, *randomization* is introduced into individual decisions such as setting prices or starting firms. Even a slight degree of randomization does important work in the model because it drives a process of (conscious or accidental) *experimentation*, by which the system may find its way to prices that roughly equilibrate the market, though it also ensures that strict and stable equilibrium will never occur. ‘Results’ may, nonetheless, take the form of *meta-stable* equilibria, in which prevailing levels or trends of the main variables of interest are clearly discernible despite continual micro-disturbances or recurrent/cyclical fluctuations. This allows for a modified form of what economists call ‘comparative statics,’ that is, analysis of the effects of exogenous assumptions or parameters by adjusting them and comparing the equilibrium ‘states’ that are the result.

Agent-based modeling requires both new technical skills and new concepts. A standard language to explicated models and characterize results is still emerging, and is far less well known than the mathematical language that has dominated the training of academic economicsts for decades. Fundamentally, however, agent-based methodology has much in common with mainstream economic modeling. It involves the same simplification and reductionism to make the complexity of the real world manageable. Its results are driven by, if not quite the ‘constrained maximization’ of the neoclassical paradigm (because maximization is usually unattainable in the more complex environments of agent-based models), then what may be called purposive agents subject to constraints. It has special pitfalls. Non-equilibrium results sometimes occur which are difficult to characterize. Sometimes results are too dependent on arbitrary parameter values. And agent-based modeling has yet to become as *transparent* as mathematical economics has become, now that it has thoroughly penetrated and transformed the discipline, and become part of every professional economist’s graduate training. But the weaknesses of the approach can be overcome, and the differences with mainstream theoretical methods are less fundamental than the similarities. Agent-based modeling is a world in which economists can learn to feel at home.

We are not here interested in agent-based modeling as an end in itself, but as a means for studying the economics of technology and the division of labor. We have seen how the Solow model fails to shed light on these issues, but what of other mainstream—that is, equations-based rather than agent-based—approaches?

A large 'endogenous growth' literature has taken shape in the past three decades, which took as its point of departure the same objections to the Solow model as outlined above. Indeed, it was a popular history of endogenous growth theory, David Warsh's *Knowledge and the Wealth of Nations,* that inspired the research leading to this book. What are the strengths and weaknesses of this literature? Is there still a puzzle to be solved?

Much of the existing 'endogenous growth' literature relies on a model of 'monopolistic competition' developed in a 1977 paper by Avinash Dixit and Joseph Stiglitz. Monopolistic competition may sound like a contradiction in terms. It is not, because even quite different goods can 'imperfectly substitute' for one another. Dixit and Stiglitz were interested in people's taste for variety, and in the question of whether capitalism produced a superfluity of it. In the process, however, they developed modeling tricks that proved useful to other theorists with quite different interests. In particular, 'new Keynesian' macroeconomic theory, 'new trade theory,' 'new' or 'endogenous growth theory,' and the 'new economic geography,' relied heavily on the Dixit-Stiglitz monopolistic competition model.

The Dixit-Stiglitz utility/production function allows for the introduction of an indefinite, endogenously-determined number of goods within a formal model. For example, in Krugman (1987?)'s model of increasing returns and spatial agglomeration, there is one agricultural good and many manufactured goods, with manufacturers in monopolistic competition with each other. Krugman finds that, given positive transport costs, it makes sense for the manufacturers and their workers to agglomerate in one place, to take advantage of the concentrated demand. Perhaps the most significant paper that uses the monopolistic competition device is Romer's "Endogenous Technological Change" (1990), of which David Warsh writes:

> [W]ith the publication of "Endogenous Technological Change," Romer won a race of sorts, a race within the community of university-based research economists to make sense of the process of globalization at the end of the twentieth century, and to say something practical and new about how to encourage economic development in places where it had failed to occur...
>
> [T]he issues attending the post-World War II growth in the wealth of nations had been clarified and, if not resolved, at least reframed in the formal language of technical economics. The basic choices had become clearer than before. The contribution of the growth of knowledge had been broached in a way that permitted its analysis. A new emphasis had been placed on the role of institutions. And a secure role finally was assigned to that long-neglected figure (at least in economics classrooms), the entrepreneur.
>
> "Romer '90" (to use the article's citation shorthand) doesn't fit our conception of a classic, to be placed on the shelf alongside the works of other great worldly philosophers. But it *is*—for reasons that are relatively easy to explain. (Warsh, xii-xiii)

In Romer (1990), monopolistic competition is found in *intermediate goods* industries, which faced fixed costs of *researching* how to produce new kinds of capital goods... [describe Romer (1990)...]

Dixit-Stiglitz monopolistic competition, like many useful assumptions, has some unrealistic implications. Now, realism for realism's sake is not good practice in choosing the assumptions of economic models. Simplifications that imply implausible things are often useful and may be innocuous, and before rejecting a theoretical device that has proven its worth, one ought to have a clear idea of something fundamental or essential is lost from, or unwarrantedly added to, reality. For example, one unrealistic implication of the Dixit-Stiglitz utility/production function is that *every* consumer purchases some amount of *every* product available in the market, and, even more oddly, every producer purchases some of every intermediate good. This is clearly false (and the models developed here will *not* have that particular unrealistic implication) but it does not strike this author as necessarily a decisive impediment to understanding economic growth.

However, two other objections to the Dixit-Stiglitz monopolistic competition model, and the sub-fields built on it, do seem fundamental. First, the Dixit-Stiglitz model allows for only *horizontal,* not vertical, division of labor. Second, the model suggests that monopoly is a pervasive feature of modern market economies.

Horizontal versus vertical division of labor. Although it is not usually described this way, the Dixit-Stiglitz model may be regarded as a way of reintroducing the division of labor into technical economics. New goods (which can be interpreted as tasks) can be introduced into the economy without limit, which, as long as they cover their fixed costs, always contribute to productivity (whether of utility, e.g. in Krugman (1987), or of output of a final good, as in Romer (1990)). However, the division of labor thus generated is strictly *horizontal.* If there is a Dixit-Stiglitz utility function, new goods will always be *directly consumed*; if a production function, directly used in producing the consumption good.

What the Dixit-Stiglitz model does not shed light on is the *vertical* division of labor, what the Austrian school calls 'roundaboutness.' As Carl Menger compellingly explained in his *Principles of Economics* (1871), goods may be separated by any number of stages of production from the ultimate consumer. There are first-, second-, third-, and fourth-order goods, and so on, potentially *ad infinitum.* Roundaboutness is important, because an important feature of the history of technology is that many technological advances lack economic motivation until there is a *derived demand* for them that emerges from the production of a downstream good, or, alternatively, until an upstream good necessary for their production becomes available. Thus, movable-type printing only offers large efficiency gains for alphabetic languages, in which the number of distinct characters used is few. When movable-type printing appeared in Europe, which already possessed an alphabetic script, its impact was explosive. But in China, where movable-type printing was actually invented *earlier* than in Europe, it made little impact. In the absence of a downstream technology, the invention was stillborn.[7]

If all possible technologies were directly useful, we should expect the best advances to be developed first. It would be hard to explain why a technological feat like the Industrial Revolution should have waited 6,000 years while men tinkered around with comparative trivialities like stirrups and beer and [more examples of technologies earlier and less useful than the Industrial Revoluiton]. But the Industrial Revolution required steam power, which in turn required [get more of the technology tree behind the Industrial Revolution...] It is my contention, then, that a satisfactory theoretical model of technological change must have room for the indefinite growth of roundaboutness.

Monopoly and competition. When it comes to competition, economists tend to go to extremes: *either* perfect competition *or* monopoly. In the real world, by contrast, monopoly seems rare, and perfect competition, strictly speaking, never occurs, being an ideal case in which the number of firms competing is "infinite." Duopoly—two firms competing in an industry (e.g., Coke / Pepsi)—and oligopoly—several firms (e.g., McDonalds / Burger King / Wendy's / In-N-Out / Five Guys / etc.)—seem to be common, even typical.

Unfortunately, to model duopoly and oligopoly is complex and, worse, cannot be done with the parsimony that economists prefer. To close duopoly models, special assumptions about firm constraints and/or behavior are needed, which are (a) empirically questionable, and (b) difficult to integrate with the rest of economic theory. Since, in any case, what models there are (e.g. Bertrand, Cournot) tend to predict a more or less rapid convergence, as the number of firms increases, to the (perfectly) 'competitive equilibrium' where price equals marginal cost (P=MC), economists usually jump straight to the convenient assumptions of firm 'smallness' and 'price-taking' and 'zero profits' associated with perfect competition. Yet competitive equilibrium might not even exist! One case where it does not exist is when there are *fixed costs* of production. If there are fixed costs and marginal costs, the reason is fairly simple: economies of scale are inexhaustible, bigger is always better, the optimal number of firms in the industry is one, and if for some reason there are several firms, the biggest can undersell the others and drive them (or buy them) out.

If there are fixed costs and *rising* marginal costs—in other words, if there are U-shaped average cost curves—the reason for the non-existence of competitive equilibrium is subtle. In this case, bigger is not always better. There is an optimal efficient scale of production, and if it is small enough, and the market is big enough, several competing firms will be more efficient than one huge monopolist, and competition is sustainable. But there is an integer problem. Unless the production function has an unusual form (so that the 'U' shape of the average cost curve has a 'flat' bottom where average cost is constant), there will be a unique efficient scale. If P=MC=min AC,[8] the total quantity

[7] I owe this example to John Nye.
[8] That is, if price equals marginal cost equals minimum average cost, minimum average cost being the per unit cost of production (fixed and variable) at the efficient scale.

demanded in the market will not, except by chance, be an integer multiple of this efficient scale output. To satisfy demand, some firms will have to operate a bit above, or below, the efficient scale. But in that case, their average costs will be higher, and they can no longer afford to charge P=MC=min AC. Moreover, even if the total quantity demanded at P=MC=min AC *is* an integer multiple of the number of firms, it is not individually rational—to be more precise, it is not a Nash equilibrium—for firms to charge that price. *Given* that *(n-1)* firms are producing at efficient scale, charging P=MC=min AC, and satisfying a proportion *(n-1)/n* of the total demand at that price, the *n*th firm is faced, not with a flat demand curve, but with a *downward-sloping* demand curve in the residual market, within which it can (provided the behavior of other firms is held constant) act as a monopolist. Of course, if the firm *does* raise its price, competitors will respond by increasing production and invading its market, but the conclusion remains that there is no competitive equilibrium.

Despite these intricacies, perfect competition is probably a close approximation of many real-world markets, and it is appropriate that most of microeconomics uses it as a working assumption. If we want to model the division of labor, however, we have to take fixed costs into account, and then perfect competition breaks down. This is the point of departure for monopolistic competition, which goes to the other extreme, and assumes that, rather than an infinite number of firms in each industry, there is only one! It turns out that, because the goods these monopolists produce are (imperfectly) substitutable, and because there is no limit to the varieties that can be introduced, monopolistic competition actually resembles perfect competition in some ways, e.g., the 'zero profit' equilibrium condition. Nonetheless, monopolistic competition is a clear departure from standard micro theory, where P=MC is the prevailing assumption. The Keynesian revolution created a disjuncture between macroeconomics and microeconomics, since Keynesian models ignored, and on close inspection turned out to be inconsistent with, the "microfoundations" of individual rationality carefully developed in the microeconomic analysis of markets. The 'monopolistic competition revolution' in several sub-fields threatens to create a similar rift in economic theory.

It would be desirable, if possible, to develop an economics of the division of labor which was consistent, if not with idealized 'perfect competition,' then at least with something close enough to it that the dominance of this assumption in microeconomics could be justified—unless, of course, 'monopolistic competition' is the way the world really works. In that case, mainstream micro should make the necessary adjustments. But is 'monopolistic competition' realistic?

A fig-leaf of plausibility for monopolistic competition comes from *brands.* In the contemporary industrialized world, most products are branded: Coke, Pepsi, Hilton, Marriott, Snickers, Dole, Breyer's, Southwest Airlines, etc. Each branded product *is* produced (typically) by a single firm. Firms try to improve and differentiate their products. We can, then, define a Coke industry, a Pepsi industry, etc., call them all monopolies, and characterize their competition as 'monopolistic competition.' Individual utility functions could be assumed to include separate terms for consumption of Coke and Pepsi, Visa and MasterCard, etc. But do Coke and Pepsi, or Hilton and Hyatt, really enter separately into utility functions? My own introspection suggests that branded products are often, in effect, the same good, perfectly substitutable for one another. I'm not sure I can even tell the difference between Coke and Pepsi. Visa or MasterCard, a Burger King Whopper or a McDonald's Big-N-Tasty, Chiquita or Dole, Safeway or Giant—the differences are slight.

Why, then, do brands exist? Isn't the very existence of brands evidence that product differentiation has a place in people's preferences? But brands have another interpretation: people may prefer branded goods because they know the brand owner has an interest in maintaining the brand's reputation, so a familiar brand name is a credible signal of quality. It is possible for branded products to be *identical to each other*, yet still superior (expectationally) to generic products. Using introspection again, I have little "taste for variety" in brands of credit cards, burgers, bananas, or supermarkets. I may alternate among competitors by accident, or in response to price changes, but I would not count myself significantly benefited if I had twelve Cokes and you substituted six of them with Pepsis (or vice versa).

The question is not whether the economy is *really* competitive or Dixit-Stiglitz. One can always define each branded product as a separate good, and then monopolistic competition is pervasive. Surely there are *slight* differences between brands, at least for some consumers. There is probably even *some* "taste for variety" across very similar products such as Coke and Pepsi. Yet if neither the competitive nor the monopolistic competition paradigm is a *strictly* accurate description of most industries, I am inclined to think that the truth is much closer to the competitive paradigm. If so, while we should be grateful for the valuable theoretical developments based on the Dixit-Stiglitz model, we should try to find a different basis for these varied insights, then retire, for most purposes, the monopolistic competition assumption.

R&D. The final problem with the 'endogenous growth' literature is the emphasis it places on research and development (R&D). No doubt R&D *does* play a role in economic growth. But it is difficult to quantify this link empirically. In R&D-based growth models, research produces useful knowledge according to a mathematical function whose form is quite arbitrary. Prospects for long-term economic growth depend heavily on the form of this 'knowledge production function.' Yet it is not even clear in principle how the real form of the knowledge production function could be researched empirically. [Attempts?...]

If R&D is costly, so much so that the costs of R&D are the major constraint on economic growth, how is it that many innovations seem to come for free, almost by accident? Only in a few industries, such as pharmaceuticals, is it clear that the competitiveness of firms and the dynamics of the industry are driven by R&D. A McKinsey study, *The Power of Productivity,* attributed much of the US productivity acceleration of the late 1990s to Wal-Mart. Where were the R&D laboratories that invented Sam Walton's business model? For that matter, what was the research budget dedicated to inventing the internet? Given how cheap it seems to have been, why didn't we invent it in the 1950s? If some highly fruitful inventions are so cheap, and if technological change is a result of deliberate, profit-motivated R&D investment, we should expect to have been invented much sooner.

Also, if economic growth is explained by advances in technology, how can we account for international differences in income? According to Romer (1990) and other contributions, knowledge is (designs are) an 'excludable' but 'non-rival' good which, once discovered (created), can be used by arbitrarily many at the same time. This suggests that today's poor countries could be raised to the living standards of today's rich countries by simply *not excluding* them from using state-of-the-art technology. Romer (1995) argues, *contra* Mankiw, that rich and poor countries differ because they do not have the same technology. But, in that case, why can't poor countries become rich just by having managers and machinists (or whoever) read the right literature—as the limited success of decades of foreign aid, if nothing else, strongly suggests they can't? And if *something else* is holding poor countries back from absorbing cutting-edge technology, might that same something else be preventing rich countries from adopting some useful technical advances whose costs of invention are negligible?

The models in this book assign *no role whatsoever* to R&D in the process of technological change: all technological knowledge is available *ab initio*, but derived demand is needed to make the introduction of a "new" (i.e., not previously produced) good economically viable. I do not, of course, mean to claim this extreme assumption is true. Right now, however, we have the opposite problem: R&D-based 'endogenous growth' is almost the only game in town for explaining long-run growth (supplemented by institutional stories). Rather, by raising a competitor—*Marshallian external economies*—to the R&D-based explanations of long-run economic growth, I hope to stimulate a useful theoretical and empirical debate. My own feeling is that the external-economies story of technological change is probably more fundamental, but that R&D and invention are needed as well, and are sometimes the limiting factor for economic growth. But a theoretical demonstration of how long-run growth and ongoing technological change might occur *without any increase in knowledge at all* may be a useful corrective to the vaguely grandiose claims about 'knowledge and the wealth of nations' which the endogenous growth literature seems inevitably to suggest.

The 'wealth of nations' is surely, today as when Adam Smith wrote about it, one of the most important questions in the world. As Robert Lucas famously said:

> I do not see how one can look at figures like these without seeing them representing possibilities. Is there some action a government of India could take that would lead the Indian economy to grow like Indonesia's or Egypt's? If so, what exactly? If not, what is it about the "nature of India" that makes it so? The consequences for human welfare involved in questions like these are simply staggering: once one starts to think about them, it is hard to think about anything else.

Indeed. Today, amidst unprecedented abundance, billions of people live still live on the edge of subsistence. In 2007, according to IMF data, GDP per capita in the United States was $45,725. In the Democratic Republic of the Congo, it was $312. Malawi, a country in southeastern Africa where I lived for a couple of months in 2004 on a mission with the World Bank, ranks 169 of 179, with a GDP per capita of $793. I spent most of my time in the capital, Lilongwe, a city of 870,000 now, less then. What was strange was that the *capital* in the capital—capital in the economic sense: buildings, roads, etc.—seemed to be (to make an unscientific guess) roughly the same total volume as that in the town I grew up in, Louisville, CO, with a population less than one-twentieth as large. I wouldn't have understood how Lilongwe could hold so many but for one day when I tried to walk from the "Old Town" to the "City Center" and wandered into what seemed like the countryside. Only I could tell it wasn't by the *people,* who were everywhere: women, running children. Where did they all *live,* I wondered?—and then I started noticing the shanties amidst the corn, too small and makeshift to communicate the thought *dwelling* to an untrained American mind. I would have called the place cornfields in America, but there they were, I suppose, suburbs of sorts, and these extended, perhaps, in all directions. There seemed to be only a handful of restaurants in Lilongwe, and only a couple of bookstores, with a pitiful selection, and most of that in *English,* not Chichewa, the language most Malawians spoke. In the marketplace in Old Town, though, the morning crowds were thick. The strange fact that I had to bear in mind, though, was that *85 percent of the population was rural.*

The country's motto was "the warm heart of Africa," and the days were warm and the people warm-hearted—indeed, when they had food in their bellies they took a relaxed attitude to life that belied what, from a foreign point of view, must be called the desperate state of the country. Life expectancy in Malawi is now 43 years; I think it was lower then. I noticed the *youth* of the population. I was twenty-five at the time, quite a young man still in America, but here the median age seemed to be about fifteen. On the minibuses that relayed between the "City Center" and "Old Town"—less than ten cents for a seat on the crowded benches, but you had to wait till they filled up—I felt, not merely tall, which I'm used to, but also *old,* surrounded by teenagers. AIDS was the main culprit in that, but malaria and hunger, too, played a part. In Lilongwe, one frequently saw advertisements for coffins, in macabre juxtaposition with the African sunshine. In Malawian schools, it was common to have, say, over 200 pupils in the first and second grades, *twenty-five* in eighth grade: a bit of that was population growth, but mostly it reflects the dropout rate. Many who reached 8th grade stopped then, because only primary education was free. The number who reached university was a minuscule share of the population. (When I visited the college the students were on strike because the professors had refused to cancel an exam for the funeral of one of their classmates—an anecdote that seems to capture both the ubiquity of death and the unbearable lightness of educational ambition.) It was my job to try to change that, but I came to sense the logic of it: why spend so much of your life in school, when it is so likely to be cut short?

Malawi has a subsistence economy and hunger is only one bad harvest away. The hunger was so bad in 2002 that a colleague told me people from the city visiting parents in the village with food had to put clothes on top of it, so that people wouldn't know they had food and rob them. A Peace Corps volunteer I met told me that shortly after arrived, she saw a person dead of hunger lying on a village path. It is difficult for the imagination to grasp that one is surrounded by a population forever gnawed by hunger, and, once you internalize this, difficult again to convey it in words. I learned it when I ventured outside the capital the first weekend after my arrival. As I drove further from the capital, cars grew fewer, until the large SUV rented for me by the World Bank was the only one on the road. I was not alone though: up and down the country dirt roads were people walking, especially women with loads balanced on their heads—and they looked at me as they passed. There came a time later in the day when I stopped in a village to ask directions. I walked up and down the village's one street (no cars here; but a few goats and chickens) drawing the eyes of everyone, and the children in particular. I found no store, no one to ask (in English) the way back to the capital. But by the time I got back into my car a crowd of twenty children stood around me, looking with silent, sad eyes. I can't remember if they begged—you got so used to pleading hands reaching up to you that it is forgotten—but their eyes said enough. I longed to give them something but had nothing appropriate, nothing but a single apple. And what use was that among so many? Guilt-stricken, I gave it to one of them, and was horrified to see that I had given the proverbial "apple of discord," for the favored child ran away and the rest began to chase him. I could have told a dozen other episodes as well as that one. One knew that the begging hands had a juster claim to one's substance than any beggar in the West ever has, that it was really no fault of their own that they were hungry, but one felt at the same time the futility of it, amidst the hunger of so many millions. It came to feel like a law of nature, this hunger, the mouths to feed always exceeding the means to feed them. And indeed it almost is, yet it is one the West (among other places, now) has somehow managed to revoke.

Minibuses and street vendors seemed to be the only businesses run by Africans. All others—guesthouses, photo shops, restaurants, clothing stores, gas stations, mechanic shops and car rental agencies, tourist travel lodges, plant nurseries, bookstores—were run by what Amy Chua calls "market-dominant minorities," of which I recall four: Arabs (they were probably Lebanese); whites (mostly what an American would call "British," though often they called themselves "South African" or "Rhodesian"—not "Zimbabwean"; I recall one German); Indians; and Chinese/East Asian (e.g., the Japanese mail-order bride of an African-born Briton who worked as a consultant). Perhaps there were exceptions to this rule that I have forgotten; anyway, I found the dominance of these minorities—numerically tiny; surely less than 1% of the population—amazing. Here, too, Malawi and not the modern West is typical: in much of the world, for much of history and even today, trade has been/is the special province of ethnic minorities, living insecurely amidst populations that are, for some reason, more or less inert economically, lacking some ineffable capacity or inclination. There were African checkout clerks in the Indian supermarket, African waiters in the Chinese restaurant, an African chef at our guesthouse, and I sensed that these menial but steady and paying jobs, which would have been associated with low social status in the United States, made them a breadwinning elite amidst the hungry, hand-to-mouth masses of the city streets, let alone the far more numerous subsistence farmers of the villages who perhaps scarcely saw money. It was high as Africans seemed to rise in their own country—except in the government, and the fact that the government was exclusively African was proof enough that the market-dominant minorities prospered by their own merit and not through political favors. I never knew quite lesson to take from the economic levitation market-dominant minorities, but it seems to be a hopeful one, of some sort.

Meanwhile in the villages, I was told, the people "eat maize while it's still green in the fields"—that is, they wasted this precious resource, food, by eating it before its nutritional maturity. I never learned how widespread this practice was, or what motivated it. Did people eat green maize because they were short-sighted, because they discounted the future? Or were they afraid that if they let it ripen their neighbors would steal it? Or did they have to eat it early simply to survive until harvest time? Institutional and cultural stories are probably important *determinants* of economic growth, and the moral imperative is not just to relieve human suffering but to unlock human possibilities, but the substance of economic growth lies, above all, in physical production, which provides the means both to fill bellies and to pursue dreams. And while acquisition of physical capital—shoes, agricultural tools that could increase the yield of the fields, or factory machines that could diversify the economy—and human capital—the know-how, or discipline, that allowed the Arabs and whites and Asians to succeed amidst destitution—are critical, these processes and the incentives that drive them are inseparable from technology and division of labor. What role policy has to play—development cannot be planned from above, as the experience of communism showed, but policy can assist in a variety of ways, from providing market space to protecting property to offsetting externalities to knowing when to get out of the way—depends on one's vision of the development process as a whole. Great effort has been devoted to getting this vision right, resulting often in statements of the obvious, sometimes to a deepening of mysteries, sometimes to errors, fallacies, and missteps, but also to major insights which, however, do not always cohere. My hope in presenting this economics of technology and the division of labor is to refresh the insights of *The Wealth of Nations*, and to provide a theoretical nexus to unify what we know about the wealth of nations.