Technological Change and Dynamic Equilibrium

Jeffrey Parker, Reed College

October 2016

The static nature of microeconomic general equilibrium

The standard microeconomic model of general equilibrium under conditions of perfect competition is an impressive piece of work. Even without understanding the complex mathematics that have transformed microeconomics since 1950, the wondrous mechanism through which the Walrasian market coordinates the independent decisions of thousands of products and consumers to engineer allocative efficiency is a worthy marvel. Notwithstanding the imperfections of competition in the real world, the basic model provides us a useful snapshot of economic equilibrium as it would look if we were able to stop time long enough for all markets to reach their equilibrium states.

It allows us to perform "comparative static" experiments—to see how changes in households' preferences or firms' production technologies would lead to changes in prices of goods and factors of production that would re-balance supply with demand in all markets. But these experiments are exactly what their name implies: comparing two static equilibria. They answer the question, "How would the economy be different if this underlying condition were different?"

The fundamental problem with static-equilibrium models is that they are indeed static. The world that the static model describes is one in which changes are discrete events and the expected normal state to which the economy moves is one in which prices and quantities remain constant at their equilibrium levels.

But we live in a dynamic economy that is constantly in flux, rather more like a flowing river than a placid pond. Economic change is not just movement between alternative static points of equilibrium but an ongoing process of change. The engine behind this change is incessant technological advancement. Instead of thinking of equilibrium as alternative stationary states, we seek a model in which the concept of "equilibrium" is defined not as a static point but as an ever-changing, dynamic path along which the economy moves.

Accounting for sustained technological progress changes the static general-equilibrium model in important ways. For example, while monopoly is anathema to static efficiency, in a dynamic setting some degree of monopoly power may lead to better outcomes over time than if all producers were perfectly competitive. Technological change introduces the importance of *dynamic efficiency*—the reduction in production costs associated with technological progress—alongside the familiar conditions of *static efficiency* such as reducing deadweight loss.

In this reading we shall analyze an economy in which firms must choose to invest in new technologies through research and development. It is a world in which discovery of economic knowledge is central to reducing costs and "competing" in dynamic industries. And if profit-maximizing firms are to invest in such discovery they must expect some reward in terms of economic profit or rent, such as might be provided by the temporary monopoly power afforded by exclusive rights to intellectual property.

The analysis owes much to the contributions of two authors: Joseph Schumpeter (writing in the 1940s) and William Baumol (writing 70 years later). Schumpeter (1950) describes a process of "creative destruction" in which existing markets are constantly being swept away and recreated by new technologies and new products. His analysis describes a very different concept of economic equilibrium than the static system taught in today's microeconomics courses. Schumpeter's ideas did not take root among microeconomists of the post-World-War-II era. Instead, the focus was on ever-more-mathematical elaboration of static competitive equilibrium. It wasn't until the 1980s that Schumpeter's ideas began to be described with more mathematical rigor with the advent of modern theories of "endogenous growth."

Baumol, in his 2010 book *The Microtheory of Innovative Entrepreneurship,* attempts to bring more microeconomic rigor to Schumpeter's conception of dynamic equilibrium with technological change. His model describes a "Red-Queen game" equilibrium in which firms in a dynamic industry must "run as fast as they can" by investing in research and development just to "stay standing still" with respect to their competitors. For many industries in today's advanced economies, this seems a more accurate depiction of their economic lives than the static story of maximizing current profit in an unchanging economic environment.

Schumpeter and notion of dynamic competition

Many economists trace the modern theory of innovation and technological progress to Schumpeter's disarmingly short seventh chapter (just a few lines beyond five pages) entitled "The Process of Creative Destruction." Writing at a time when microeconomic analysis was focused on allocative efficiency and the benefits of static price competition, Schumpeter sought to focus attention on a different kind of competition: competition in innovation among monopolistic firms.[1]

He begins by noting that the period beginning around 1890, when large monopoly "trusts" came to dominate many U.S. industries, was a period in which technological progress seemed to have accelerated, with rapid improvements in standards of living:

> If we list the items that enter the modern workman's budget and from 1899 on observe the course of their prices not in terms of money but in terms of the hours of labor that will buy them … we cannot fail to be struck by the rate of the advance which, considering the spectacular improvement in qualities, seems to have been greater and not smaller than it ever was before. … Nor is this all. As soon as we go into details and inquire into the individual items in which progress was most conspicuous, the trail leads not to the doors of those firms that work under conditions of comparatively free competition but precisely to the doors of the large concerns. [Schumpeter (1950, 81-82)]

This observation puts Schumpeter in direct confrontation with the view of standard microeconomic theory—both in his day and in our modern textbooks—that monopoly power leads to inefficiency and that perfect competition is the ideal market form.

He goes on to sketch a dynamic model of innovating monopolistic firms who compete not by lowering prices of today's goods but instead by innovating tomorrow's goods, tomorrow's improved versions of goods, and tomorrow's improved production methods with lower costs. In doing so, they both create markets for new goods and (often) destroy the markets for existing ones—what he called creative destruction. This essentially dynamic process is at the core of Schumpeter's theory of capitalist economies:

> Capitalism, then, is by nature a form or method of economic change and not only never is but never can be stationary. … The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods of production or transportation, and the new forms of industrial organization that capitalist enterprise creates.
>
> … The opening up of new markets, foreign or domestic, and the organizational development from the craft shop and factory to such concerns as U.S. Steel illustrate the same process of industrial mutation … that incessantly revolutionizes the economic structure *from within*, incessantly destroying the old one, incessantly creating a new one. This process of Creative Destruction is the essential fact about capitalism.

---

[1] I use the term "monopolistic firm" to mean a firm with some market power, not necessarily a pure monopoly. This would also include most oligopolistic and monopolistically competitive firms, for example.

> It is what capitalism consists in and what every capitalist concern has got to live in. [Schumpeter (1950, 82-83), emphasis in original]

So central is the dynamic process of creative destruction to economic progress—and therefore to the ability of the economy to satisfy consumers' wants—that Schumpeter ends the chapter by arguing that static microeconomics ignores what is by far most important.

> Now a theoretical construction which neglects this essential element of the case neglects all that is most typically capitalist about it; even if correct in logic as well as in fact, it is like *Hamlet* without the Danish prince. [Schumpeter (1950, 86)]

In his view, fixation on minimizing dead-weight loss triangles at the expense of technological change is like focusing on the crumbs and ignoring the loaves of bread that really create economic benefit.

Why did microeconomics evolve away from Schumpeter and towards an emphasis on static equilibrium and price competition? Perhaps the application of advanced mathematical methods that was launched by Samuelson (1947) could proceed more easily in the analysis of static equilibrium under perfect competition. On the macroeconomic side, vivid memories of the Great Depression combined with steady productivity growth in the post-war period seemed to elevate Keynes's theories of macroeconomic fluctuations higher on the agenda than long-run growth analysis given center stage to technological progress. Perhaps Schumpeter's death early in 1950, just as postwar economics was finding its course, reduced the amount of attention devoted to his work.

For whatever reasons, mainstream economists did not build extensively on Schumpeter's model of dynamic competition for several decades after his death. It wasn't until the 1980s, after macroeconomists noted a pronounced slowing in the aggregate rate of productivity growth in advanced economies, that a substantive school of Schumpeterian economics took a central role in economic thinking. The "endogenous growth" literature, notably Romer (1986), Lucas (1988), and particularly Aghion and Howitt (1992), reached back to Schumpeter's ideas and developed them at a mathematical level commensurate with other modern economic models.

Role of technological progress in economic growth

Since the work of Robert Solow (1957), macroeconomists have used the method of "growth accounting" to break down overall growth in a country's GDP into components attributable to increased labor input, increased capital input, and a "Solow residual" that measures the increase in total factor productivity (TFP).

In terms of equations,

$$\frac{\Delta GDP}{GDP} = \alpha \frac{\Delta K}{K} + (1-\alpha)\frac{\Delta L}{L} + \frac{\Delta A}{A}, \tag{1}$$

where K is capital input, L is labor input, A is an index of TFP, and α is a coefficient that can be approximated by the share of GDP earned by capital (with share $1 - \alpha$ earned by labor). The expressions of the form $\Delta X/X$ are growth rates: the year-to-year change is divided by the level of the variable.

Growth accountants can obtain measures of all of the parts of equation (1) except the final term, so they solve to get an expression for TFP growth:

$$\frac{\Delta A}{A} = \frac{\Delta GDP}{GDP} - \alpha \frac{\Delta K}{K} - (1-\alpha)\frac{\Delta L}{L}. \tag{2}$$

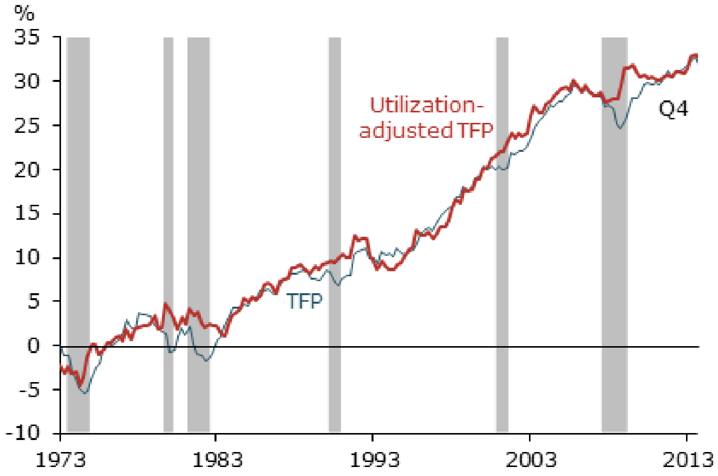
$\Delta A/A$ is the Solow residual and it is indeed a residual; it measures the part of output growth that cannot be explained by the growth in inputs. So what is it? Any kind of improvement in production efficiency would be captured in this term (along with any measurement errors in the variables on the right), but the biggest contributor is undoubtedly technological progress. New technological knowledge allows firms to produce more output per unit of input, so it's what TFP is all about.

As the population grows, so do labor input and real output, but that kind of "extensive growth" does not improve people's living standards. Increases due to capital input are important, but they may be self-limiting. In his seminal paper on economic growth, Solow (1956) showed that an economy with diminishing returns to capital and that saves a constant share of its income will eventually reach a steady state. The only source of growth in per-capita income in such a steady state is increases in TFP: technological progress. Increases in TFP translate directly into increases in the purchasing power of the economy's households; they make us richer and expand our options for consumption.

Growth accounting estimates the relative contribution of labor, capital, and TFP to growth. Specific estimates vary some across periods of time and countries but with only a few exceptions growth in TFP has proved an important contributor to growth ever since the Industrial Revolution. For example, a recent set of estimates for the United States by Jorgenson (2005) breaks down the overall 3.46% growth rate of real GDP over the period from 1948 to 2002 into 1.05% due to labor input, 1.75% due to capital input, and 0.67% due to growth in TFP.

Figure 1, from Fernald and Wang (2015), shows how total-factor productivity has grown in the United States since 1973. They estimate that real output is about 35% higher now than it would be if productivity were at the level of 1973. That means that the average household has one-third more spending power now as a result of 40 years of productivity growth. (And

TFP growth in this period has been considerably slower since 1973 than in the period from World War II to 1973.)



Note: Gray bars indicate NBER recession dates.

Figure 1. U.S. total factor productivity since 1973

To put the difference in lifestyles more vividly, consider a few technological advances that occurred during those four decades. The computer on which I worked in 1973 occupied an entire building; its user input was accomplished with a deck of punched cards with a maximum of 80 characters on each card. The computer had 60 kilobytes of usable memory (less than half the size of this document in Microsoft Word) and took several minutes to run a simple econometric analysis, which was delivered on a printout by the computer operator. Telephones were attached to wall outlets by cords and long-distance was expensive, in today's dollars an out-of-state call cost the equivalent of about $1 per minute. The Internet and electronic mail did not exist. Automobiles spewed lead-laden exhaust, broke down frequently, and 20 miles per gallon was considered outstanding economy. DVDs did not exist, nor did the video cassette recorders that preceded them: television options were whatever the three networks served you at any moment. Cameras used film, which had to be developed and printed before you could see your pictures. Interacting with your bank meant visiting a human teller during "banking hours," 10am to 2pm, Monday through Friday—no ATMs or 24-hour account access. Today's technology affords not only a 35% higher measured living standard, but a truly different lifestyle altogether.

The nature of innovation

From where do these life-changing technologies come? What is the mechanism by which new technologies are discovered and diffused through the economy? Does innovation just happen or is it an intentional act on the part of some innovator? Economists have begun to look much more closely at these questions in recent decades, using both descriptive and quantitative methods to examine the mechanisms behind technological progress.

At the center of technological change is the process of *innovation*. Economists distinguish innovation from invention, the latter of which is the more often heard in common speech. Invention is the discovery of something new: in the technology realm, a new product or a new production process. Innovation encompasses not only invention but also the ensuing sequence of events or by which the new product or new process becomes economically useful. Innovation includes turning the raw idea of the invention into an actual product or process that is successful in the market.

Innovation occurs at many different levels and through many different innovation processes. When we think of innovations we often think of high-tech innovations such as new computer chips or path-breaking pharmaceuticals that are usually the result of lengthy and intentional programs of research and development by giant firms. We also recall the often-apocryphal stories of the individual inventors slaving away for years in their garages to develop something new and wonderful, or programmers toiling in their parents' basements developing a spectacular new piece of software.

But most innovation is much more mundane. For every major breakthrough that gets headlines or makes billions of dollars, there are many thousands of small, incremental innovations that, taken together, have enormous impact on the economy. And not all innovations are the result of purposive research and development; some are "shop-floor" innovations that follow naturally as workers see opportunities to make things better or to make better things. An example of an impactful low-tech innovation occurred at a local recycling firm that was frustrated by their inability to separate small pieces of glass from strands of shredded paper, which fell through their standard sorting screens and went to the landfill together rather than being separately recycled. One day, a clever worker realized that by sending the stream of glass and paper up into the air and positioning a large fan sideways across the stream to blow the paper away from the glass, two separated and recyclable piles would accumulate next to each other on the floor. There were no headlines lauding this discovery and no fortune accrued to its inventor, but a useful improvement on the production process made it possible to increase output of recycled materials at a very low cost.

Without denying the importance of these almost-accidental innovations that occur without explicit R&D investment, we will now turn our attention to innovation that happens as a result of an explicit decision to invest in R&D. Why do firms make such investments? How do they expect to gain from them?

Technological information as a public good

The problem of public goods is a leading example of market failure. Every textbook tells us that profit-seeking firms will not produce public goods because there is no way to profit from them.

Public goods have two defining characteristics: they are *nonrival* and *nonexcludable*. Nonrivalry means that one person using the good does not limit the ability of others to use it. Street lights, radio broadcasts, and national defense are all examples of nonrival goods because two—or two hundred—can use them as easily as one. Nonexcludability occurs when it is impractical to provide the good to some users but to exclude others. Once the good exists, everyone can get access without paying. The three examples above are also nonexcludable, which makes them public goods.

When a good is nonrival (even if they are excludable), allocative efficiency requires that the good should be provided free of charge. The marginal social cost of providing the nonrival good to an additional consumer is zero, so it should be used up to the point where the marginal benefit is also zero, and consumers will only consume that much if the price is zero.

For nonexcludable goods it is impractical to charge individual consumers. Those who do not pay cannot be excluded from consuming it, so everyone has an incentive to be a *free rider*, consuming the good but letting others pay the cost. (Non-donors who listen to public radio, for example, are free riders.)

Both of these public-good characteristics apply to *knowledge* in general, and specifically to technological knowledge. Suppose that someone invents a revolutionary new way of producing widgets. The technological knowledge of this method can be shared with other firms (and anyone else) at negligible cost, so it is nonrival. And, as anyone who attempts to keep a secret knows, it is very difficult to keep knowledge from leaking out, especially if it is embodied in a product that can be "reverse-engineered" to figure out how it was made.

Suppose that WonderWidgets makes a homogeneous good and is initially in a zero-profit long-run equilibrium. In Figure 2, the long-run average-total-cost curve is at ATC_0, long-run marginal cost is MC_0, and the price of widgets is P_0.

Then WW comes up with a wonderful new cost-saving innovation that lowers its cost curves below those of its competitors, to ATC_1 and MC_1.

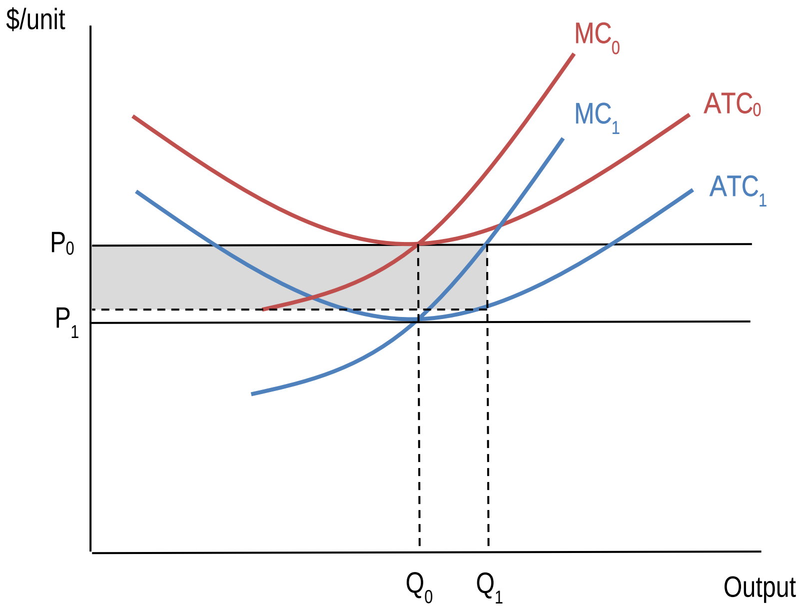


Figure 2. Innovation and appropriability

If WW were somehow able to prevent the rest of the industry from using the innovation, the price would remain at P_0 and WW could increase production to Q_1 and earn economic rent on its innovation equal to the area of the shaded rectangle. As long as rival firms' costs remain at the original, higher level, WW will continue to earn rents, perhaps repaying research and development costs incurred in the discovery.

If, however, the new innovation is copied by all of WW's rival firms, their cost curves will also fall and the widget price will drop to P_1, the minimum value of the new ATC_1 curve. WW makes no profit from its innovation once it has been imitated by the other firms in the industry.

This diffusion of knowledge is optimal from the prospective of society as a whole because knowledge is nonrival. It costs nothing to spread the knowledge to additional users once it has been discovered. And consumers benefit at the lower price of widgets, presumably buying more widgets and certainly getting additional consumer surplus. So immediate and complete diffusion of technological knowledge is in the social interest.

However, diffusion of the innovation is costly to WW. If rival firms are able to match the cost reduction immediately, WW sees no benefit at all from its innovation. Under such circumstances, it would not be in WW's interest to expend resources on research and development in order to innovate because there would be no gain to offset the cost.

We can think of this problem as a classic public-good problem: private firms that achieve no gain from innovation are unlikely to incur costs in order to innovate. In standard public-good settings we expect that government intervention would be required to provide the optimal amount level of provision. (And indeed government agencies are very involved in subsidizes research and development.) We can also think of innovation as a situation of externality. WW's innovation creates external benefits for widget consumers—and perhaps little or none at all for itself. In such a world, the only innovations that we would expect to see emanating from firms would be those that are costless or accidental: perhaps the fan in the recycling plant but surely not new pharmaceuticals, computer chips, or airliners.

Appropriability mechanisms

To motivate profit-seeking firms to invest resources in innovation, they must either be subsidized in some way or the economic system must provide some means by which they can *appropriate* some of the gains from their investment. Most often, the appropriability mechanism limits the ability of rival firms to imitate the innovation; it prevents—at least temporarily—the diffusion of the new technology.

The simplest appropriability mechanism is *secrecy*. If the firm can keep rivals from discovering the new technology, then they may earn the rents shown in Figure 2 for a sustained period of time. Trade secrets have some legal protections, for example against corporate espionage or revelation by an employee who has signed a nondisclosure agreement, but if a rival discovers the innovation independently there is no protection.

Some trade secrets survive for a long time; perhaps the best-known example is the formula for Coca-Cola. But in many cases secrecy is impossible. Process innovations can be kept behind the closed doors of a factory, but new or improved products are available on the market for anyone, including rivals, to see and use. Once the product is revealed it is often feasible for other firms to reverse-engineer how to build it and enter the market with a rival product, sometimes relatively quickly.

Intellectual property protection through patents and copyrights provides an alternative appropriability mechanism by legally restricting imitation. The inventor of a machine, a manufactured product, a composition made from two or more substances, or a process for manufacturing objects can apply for a patent, which will be granted provided it is found by the patent examiner to be novel, useful, and non-obvious to someone experienced in the field. If a patent is granted, it gives the inventor the exclusive right to make, use, sell, offer, import, or offer to import the innovation for a period of twenty years from the date of application. Effective patent protection prevents industry rivals from using the patented technology to lower their cost curves until the patent expires, potentially preserving the innovator's rents. Alternatively, the patent-holder can earn royalties on the patent by licensing its use to other firms, which is ideal for a firm that is better at innovating than at using its innovations to produce goods.

But patent protection is not always strong. The social quid-pro-quo for patent protection is publication of the invention by the patent office, which eliminates secrecy as an option. It is often possible for rivals who understand the patented innovation clearly to “invent around” the patent, designing a product or process that is equally effective but just enough different that it technically does not infringe the patent. Moreover, patents cannot always be enforced. It is the responsibility of the patent-holder to enforce its patent by bringing suit against those it thinks are infringing. This is costly whether the suit is successful or not, and it can be impractical to enforce patents when the alleged infringer is a deep-pocketed firm.

Another potential mechanism for appropriability is simply a sufficient *head start*. Even if rivals know about the innovation and even if patents are not effective in preventing imitation, it is likely to take many months or years for the imitator to bring rival products to market. In fast-changing industries (smart phones?) a period of one to two years may be sufficient for the innovator to earn enough rents to make R&D profitable. And by the end of that period, the innovator may have the next-generation product ready to release, setting the process in motion once more. Staying one product generation ahead of its rivals worked very well for Intel for several decades in the production of microprocessors.

Finally, in some situations it may be possible for the innovator to prevent imitation by gaining exclusive control over strategic complementary resources. If the innovation requires a particular rare input, it may be possible to arrange exclusive contracts for the input prior to releasing the innovation, making it difficult for potential imitators to get access.

Alternatives to appropriability

Sometimes no appropriability mechanism will be effective. In the case of *basic research*—the kind of pure science that does not lead directly to a salable product—there may be great benefits for downstream research and development but little potential for revenue for the researcher. Some profit-seeking firms do perform extensive basic research as part of an integrated R&D effort. Perhaps the best example is the Bell Labs affiliate of former national telephone monopoly AT&T, where scientists have won eight Nobel Prizes for inventions such as transistors and for basic discoveries in chemistry and physics.

But given the very long time frame associated with harvesting commercial fruit from the seeds of basic research, few firms are large enough or far-sighted enough to find potential profits in large-scale basic research. (And disciplinary norms require that discoveries in pure science be published, which makes any potential for appropriability that much more difficult.) Because its rewards are more public than private, much basic research is conducted in universities. Such activity is often funded by government agencies such as the National Science Foundation and the National Institutes of Health. Government support of the production of technological information follows the government-provision model that is common for other public goods. Even when government funding is incomplete, universities and colleges may provide basic research as a by-product of their educational product. Students fund the research through their tuition payments and society receives the external benefit of their discoveries.

Since passage of the Bayh-Dole Act in 1980, researchers have been allowed to seek patents on products they develop using funding from government grants. Bayh-Dole is controversial because it might seem like "double-dipping" to give royalties to the innovator (or, more often, to the innovator's university) given that the government paid for the product's development in the first place. The Bayh-Dole Act has been instrumental in turning the attention of academic researchers toward products with potential markets. This model of R&D has been exploited extensively in emerging fields such as biotechnology and software, where key patents developed at universities have led to many successful (and many unsuccessful) startup firms. The most successful universities earn significant revenue from their portfolio of patents. For example, the patent rights to Google's "PageRank" algorithm were assigned to Stanford when it was developed there. Stanford sold its rights to the algorithm (to its discoverers, then at Google) about ten years ago for $336 million—a sum comparable to Reed's entire endowment at the time.

Beyond grants, governments or foundations can reward successful research through *prizes* offered for specific advances. Although prizes are sometimes effective, there are several problems with them. On a practical level, the winner does not get paid until the discovery is made and validated. In the meantime, all of those pursuing the prize must fund their own research activities and most will *not* win the prize. It is difficult for small-scale scientists to pay for the research expenses and even the most intrepid venture capitalists are wary of funding such projects.

Another important problem with prizes is that the government (or whoever offers it) must decide that a problem is of sufficient social benefit to warrant establishing the prize. Sometimes this is obvious (as with a cure for cancer, for example), but governments are not always successful at identifying the most promising research questions. (This criticism, of course, applies to government grants as well.)

Even when the government finds an obvious target for research and establishes a prize, there is often controversy about whether or not the prize has been won. The conditions for winning the prize must be very specific in order to rule out innovations that satisfy the prize criteria but do not solve the problem in a useful way. A famous historical example is the Longitude Prize, established in 1714 by the British government as a reward to the first person to discover a method for ships to measure their longitude. The government had in mind a method that would be nonrival knowledge, so that once discovered it could be used freely by all navigators. John Harrison was eventually awarded the prize. Harrison's invention was a sea clock whose mechanism would remain accurate despite the heaving of the ship in the waves.[2] The government resisted awarding the prize, however, because Harrison's method required the navigator to have possession of Harrison's clock, a rare and expensive instrument. Although the method was far from nonrival, Harrison eventually succeeded in claiming the prize money because nonrivalry was not part of the detailed specification of the prize.

How much innovation is optimal?

How much research and development is desirable? The principle of this question is similar to the question that we would ask for other kinds of investment: Does the rate of return, adjusted for risk, compare favorably with other uses of funds?

All real and most financial investment is risky. When a firm builds a factory, the rate of return it will achieve will depend on the demand for the factory's product and on the cost of producing there. The firm may have a pretty good idea about these variables but neither is fully predictable, so the prospective rate of return on investment in plant and equipment must be balanced against the risk. Higher-risk projects will require a higher expected rate of return.

Investment in R&D is particularly risky. Even if a research project is focused on a highly specific outcome it may take much longer than expected, it may require unanticipated expenditures, or it may even fail entirely. Those research agendas that are more "exploratory" in nature carry even more uncertainty about costs, duration, and outcome.

Economic and financial models can incorporate risk as long as investors have some idea of the probabilities involved. The central problem with R&D investment is not risk but externalities. Private investors make decisions based on their *private rate of return* whereas the socially optimal amount of R&D depends on the *social rate of return*. Absent an iron-clad appropriability regime, the innovating firm will be subject to imitation and much of the benefit from innovation will go to rival firms and consumers. This externality means that investment in R&D is likely to be too small.

Figure 3 shows a hypothetical example in which the expected return to the firm on non-R&D projects with similar risk is 15%. The higher, blue curve is the expected social return to

---

[2] Knowing the time at the ship's home port (by reading the clock) and the time at its location at sea (by measuring the angle of the sun), the ship's navigator could easily work out the ship's east-west position relative to the port.

R&D: the expected flow of gains per dollar invested both to the firm and to broader society, which includes spillovers to rivals and consumers. The lower, orange curve is the private return that the firm expects to appropriate in terms of economic rents. It is lower because of imperfect appropriability. Both curves slope downward because firms will undertake the most promising R&D opportunities (those with the highest potential returns) first.[3] The vertical distance between the two curves is the positive externality.

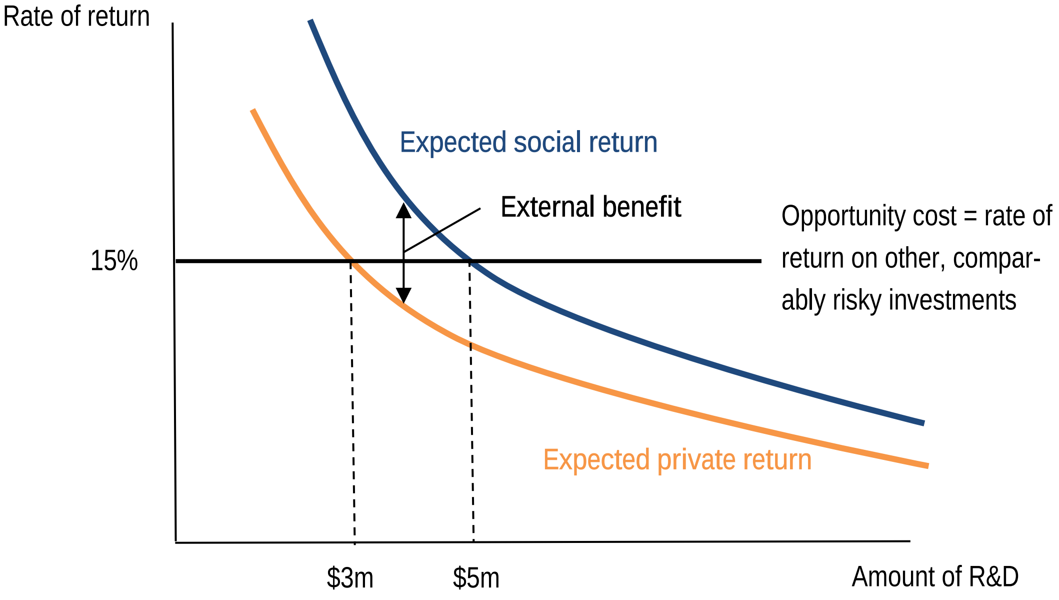


Figure 3. Optimal R&D investment

A profit-maximizing firm will choose to perform $3m of R&D, but based on the social return the optimal amount would be $5m. This underinvestment in R&D is exactly what we would expect in a situation of positive externalities: by not considering the external social benefits of its action the firm undertakes too little.

One standard response to externalities is to impose a tax on—or, as in this case, grant a subsidy to—the firm to reflect the external costs or benefits. If the firm were granted a subsidy on R&D in the amount of the black vertical arrow in Figure 3, then it would choose the socially desirable amount of R&D investment, $5m. But such a policy intervention asks a lot of policymakers. Those deciding on the subsidy never know as much about the true private and

---

[3] We assume, though this need not be the case, that the R&D opportunities with the highest social returns also have the highest private returns.

social returns to R&D as the firm does. And the firm has every incentive to portray the blue curve reflecting social benefits as being very high and the orange curve showing private benefits as being very low. Any sort of precise targeting of R&D subsidies to individual firms and projects is likely to be impractical.

While it may be infeasible to exactly align private and social R&D incentives, policymakers surely know that external benefits exist and may be able to identify some project areas in which they are likely to be large. Thus, general R&D subsidies through tax credits and targeted grants are common policies to attempt to move toward the optimal amount of investment in new technologies.

Baumol's Red-Queen equilibrium

Now let's try to tie together the various ideas we have explored and think about a microeconomic model that incorporates R&D investment and technological progress as an ongoing, dynamic phenomenon. In such a model, there must be some degree of appropriability to give profit-seeking firms incentive to invest in R&D. This appropriability gives firms some effective temporary monopoly power following successful innovation, which enables it to recover the costs of its R&D.

We will not attempt to describe the model in rigorous mathematical detail, but its basic outlines are as follows:

- Firms are not price takers, they had some monopoly power over their current selection of products.
- Firms earn economic rents on their past innovations.
- They use these rents to perform research and development in search of additional innovations.
- These additional innovations, if successful, then perpetuate the dynamic cycle by providing a new generation of products yielding some degree of monopoly rents.

To summarize, in dynamic equilibrium firms invest continuously in R&D to develop new technologies. Firms earn economic rents on their innovations, but they use those rents to do further R&D.

To see that continuous innovation can be a situation of equilibrium, consider what would happen to a firm that chose to halt its R&D program. Initially, it would be better off because it could save the R&D costs and instead distribute the rents from previous innovations to shareholders. But the firm's technological capability would eventually fall behind the industry, leaving it with high costs as rivals' innovations pushed prices down and eventually forcing it to leave the industry. So breaking the cycle of innovation by halting R&D is not in the interest of a firm that wants to remain in the industry.

Baumol (2010) has characterized this kind of dynamic equilibrium as a "Red-Queen game," citing the passage in Lewis Carroll's *Through the Looking Glass* where the Red Queen asserts:

> Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!

Just as the Red Queen claims, in this dynamic equilibrium firms must continue to invest in R&D and discover new technologies just in order to survive and maintain their current market position.

Can a typical firm in such an equilibrium earn economic profits? Perhaps not. Baumol characterizes the "innovation industry" as having low barriers to entry. If any firm were to achieve large economic profits, its rivals and potential entrants would begin developing alternative new technologies that would chip away at its profits and return the industry to a "normal profit" in a process similar to Schumpeterian creative destruction.

If economic profit cannot be sustained, then in long-run equilibrium the firm's rents from previous innovations can only be just enough to support its ongoing research and development. If they were any higher, new innovators would enter until equilibrium was restored. We therefore have a paradoxical situation in the Red-Queen equilibrium where firms' revenues exceed their operating costs, creating what looks like economic profit, but where this profit is entirely dissipated on R&D investment, leaving firms with a normal rate of return and zero economic profit.

The nature of competition in the red-queen equilibrium is quite different than that modeled in static perfect competition. Instead of firms competing to undercut one another on price, they compete to out-innovate one another, reducing costs and improving their products to try to stay ahead of their rivals. Although this kind of competition is difficult to find in microeconomic textbooks, it is representative of a kind of inter-firm interaction we commonly read about in the business pages of newspapers. Price competition is surely important in some markets, but in those industries characterized by rapid technological progress the Red-Queen game is a more accurate model of the nature of competition.

Static and dynamic efficiency: Coping with scarcity vs. reducing scarcity

We began by considering Schumpeter's argument that "monopoly trusts" had provided most of the gains in living standards during the early 20th century. We then considered the importance of technological innovation and the incentives for the research and development investment that often underlies it, ending by sketching a microeconomic model of innovating firms in dynamic equilibrium.

The dynamic model is very different from the general equilibrium model of perfect competition we find in textbooks. The standard competitive model *takes scarcity as a given* and discusses how we can allocate resources, at any moment in time, in a way that maximizes social benefit *given that degree of scarcity*. The dynamic Red-Queen model describes an equilibrium in which technological progress gradually *reduces the severity of scarcity* by allowing more goods to be produced using less resources.

We can think of this difference in terms of two kinds of efficiency. Pursuing *static efficiency* in the general equilibrium competitive model amounts to reducing the size of dead-weight-loss triangles in static models. Among other things, this means eliminating or reducing monopoly power whenever possible. Pursuing *dynamic efficiency* means achieving the level of investment in research and development that matches the social rate of return with the opportunity cost. This leads to a steady reduction in production costs in the long run and/or a steady introduction of new and better products.

Each of these is surely desirable! The problem is that achieving dynamic efficiency through R&D *inevitably* leads to static inefficiency. In the Red-Queen equilibrium, firms earn monopoly rents on their previous innovations, leading to distortions and dead-weight losses in the markets for their products. Nor is government provision a panacea: to the extent that R&D is publicly supported, the government must impose taxes to pay for subsidies and grants, which themselves introduce distortions.

Deadweight losses are important, but most of the gains in living standards we have achieved since the Industrial Revolution have been associated with technological advance, so surely that is important, too.

The tradeoff between static and dynamic efficiency is most clearly seen in debates over appropriability. Stronger protection of intellectual property—for example, broader and longer-lasting patent protection—should allow firms to appropriate more of the gains from their innovation, increasing research and development toward the optimal level. But it does so precisely by restricting rivals' access to innovations, creating artificial monopoly power that leads to dead-weight losses in the static model. If technological knowledge is nonrival, then static efficiency requires that everyone have immediate, costless access to it. But if all innovations diffuse immediately and completely, no one will have economic incentive to pursue them. Put in these terms, the choice is between less-widely-shared innovation and less innovation, period.

There is no "right answer" to this choice. But a brief thought-experiment should convince you that the case for the importance of dynamic efficiency is strong. We noted earlier that technological progress had increase average incomes in the United States by about 35% since 1973, and that much of that increase had come in the form of wonderful new and improved products that we all enjoy every day. Suppose that instead of providing incentives for these innovations, policymakers had focused successfully on making the 1973 economy perfectly competitive, somehow eliminating all static inefficiency from the economy but in the process eliminating the incentives for innovating new and improved products. Would you rather have more 1973 products at lower prices due to more competition in their production or would you prefer to pay somewhat higher "monopoly prices" for the new products that we now enjoy?

References

Aghion, Philippe, and Peter Howitt. 1992. A Model of Growth through Creative Destruction. *Econometrica* 60 (2):323-51.

Baumol, William J. 2010. *The Microtheory of Innovative Entrepreneurship*. Princeton, N.J.: Princeton University Press.

Fernald, John, and Bing Wang. 2015. The Recent Rise and Fall of Rapid Productivity Growth. *Federal Reserve Bank of San Francisco Economic Letter* 2015-04.

Jorgenson, Dale W. 2005. Accounting for Growth in the Information Age. In *Handbook of Economic Growth*, edited by P. Aghion and S. N. Durlauf. Amsterdam: North-Holland.

Lucas, Robert E., Jr. 1988. On the Mechanics of Economic Development. *Journal of Monetary Economics* 22 (1):3-42.

Romer, Paul M. 1986. Increasing Returns and Long-run Growth. *Journal of Political Economy* 94 (5):1002-37.

Samuelson, Paul A. 1947. *Foundations of Economic Analysis*. Cambridge, Mass.: Harvard University Press.

Schumpeter, Joseph A. 1950. *Capitalism, Socialism, and Democracy*. 3rd ed. New York: Harper & Brothers.

Solow, Robert M. 1956. A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics* 70 (1):65-94.

———. 1957. Technical Change and the Aggregate Production Function. *Review of Economics and Statistics* 39 (3):312-320.