Are Nonconvexities Important for Understanding Growth?

By Paul M. Romer*

When economic reasoning is supplemented only by everyday experience, it leads to few definite conclusions. Systematic data collection and econometric analysis are usually needed to tell whether it is the one hand or the other hand that holds the right conclusion. The title of this note poses one of the few important questions that can be resolved (in the affirmative) using logic and the kind of evidence reported in the newspaper.

I. Are There Nonconvexities?

The *Wall Street Journal* reported last week that two former employees of Du Pont were convicted of extortion. They had threatened to reveal trade secrets about how to make Lycra. Earlier in the year, the paper reported a suit against a former employee of Intel over his alleged use of secrets learned there. It also described an investigation of a firm that used illicitly acquired mechanical drawings and metallurgical formulas to manufacture replacement blades for General Electric turbines.

These incidents have two things in common. First, the value of the economic good at stake is not small. The extortionists asked Du Pont for $10 million. Intel claims that it has invested $100s of millions on the design of its microprocessors. General Electric claims that its investment in the design of turbine blades is worth more than $200 million. Second, the notion of theft is unusual. The companies involved were not deprived of the use of the input that was allegedly stolen. Du Pont could still use its secret process for making Lycra. Intel could still use its chip designs. General Electric could still use its mechanical drawings and metallurgical formulas to make turbine blades.

These examples, and hundreds more like them, show that there are valuable inputs in production that can be used simultaneously in more than one activity. In public finance, consumption goods with this property are referred to as nonrival consumption goods. By analogy, I have referred to goods like a chemical process, a chip design, a mechanical drawing, or a metallurgical formula as nonrival inputs in production (see my 1990 paper).

There are (at least) two ways to think about nonrival inputs. One is to treat a good like a design or a list of instructions as something that is distinct from the medium on which it is stored, and to say that it can be used simultaneously by arbitrarily many different firms and people. A more literal way to describe a nonrival good is to treat the physical medium containing the design or instructions as the relevant good. Then a nonrival input has a high cost of producing the first unit and a zero cost of producing subsequent units. For General Electric, it may have taken millions of dollars of engineering work to produce the first mechanical drawing for its current generation turbine blade, but subsequent drawings can be made at virtually zero cost on a photocopy machine.

Rivalry is closely related to, but distinct from, the notion of excludability. A good is excludable if someone with a property right can exclude others from taking advantage of it. For a rival good, possession ensures excludability. The car I drive is both rival and excludable, since it is relatively easy to possess. The space in which I park my car at a shopping center is also a rival good, but it is one for which it is harder to enforce property rights. At best, it is partially excludable. Just as a good can be rival and nonexcludable, goods can be nonrival and excludable.

*University of Chicago and Center for Advanced Study in the Behavioral Sciences, 202 Junipero Serra Blvd., Stanford, CA 94305. Support from NSF grant no. SES-8821943 and BNS#87-00864 and a Sloan Foundation fellowship are gratefully acknowledged, as are the comments of Mancur Olson.

Computer software is a nonrival good, but it can be made excludable using copy protection or a copyright.

Conventional economic goods are rival and excludable. Public goods are nonrival and nonexcludable. Intermediate goods were described by Mancur Olson (1965) and labeled club goods by James Buchanan (1965). (See Todd Sandler and John Tschirhart, 1980, for a review of club theory.) Club goods like the right to swim laps in a pool are perfectly excludable and partially nonrival. Producing the second unit is cheaper than the first, but eventually, either the cost of subsequent swimming rights rises or their quality falls.

For growth theory, it is partial excludability not partial rivalry that is relevant. Goods like new mechanical designs, metallurgical formulas, chemical processes, and semiconductor designs are purely nonrival. They do not wear out or suffer from congestion. In this sense they are like basic science, for example, the advances in solid state physics that led to the discovery of the transistor. The difference is that advances in solid state physics are completely nonexcludable, whereas these goods are at least partially excludable. Otherwise, they would not be intentionally produced by profit-maximizing firms.

The distinction between rivalry and excludability is important because nonrivalry is inextricably linked to nonconvexities, whereas nonexcludability is not. Under the open field system, pasture land is a nonexcludable good that is used inefficiently. Efficiency can be restored by creating a legal system that assigns property rights and establishes excludability. No nonconvexities are involved. No difficulty for price taking arises.

The fact that nonrivalry creates a nonconvexity is apparent from the second definition of a nonrival good as one for which subsequent units have a lower unit cost of production than the first. For the analysis of technology and growth, it is more revealing to consider the definition of a purely nonrival good as one that can be used repeatedly. The most basic premise in our scientific reasoning about the physical world is that it possible to replicate any sequence of events by replicating the relevant initial conditions. (This both a statement of faith and a definition of relevant initial conditions.) For production theory, this means that it is possible to double the output of any production process by doubling all of the rival inputs. By definition, the nonrival inputs can be used again in the replica process. Hence, if there are any nonrival inputs that have productive value, output will increase more than proportionately with an increase in all of the inputs. If R denotes the set of rival inputs, N the set of nonrival inputs, and $F(R,N)$ denotes output, then for integer values of λ,

$$F(\lambda R, \lambda N) > F(\lambda R, N) = \lambda F(R, N).$$

This means that in the large, the elasticity of output with respect to inputs is greater than 1 and that the function $F(\cdot)$ is not concave.

The structure of this argument is very simple. Once you define the notion of a nonrival good and admit the principle of replication, it follows that nonconvexities are present. Newspaper reports remind us that nonrival goods exist and are important.

II. Previous Statements of This Argument

So far as I know, the statement of this argument in terms of nonrivalry is new, but the basic idea is old. Joseph Schumpeter suggested something like this in 1942. Richard Nelson and Sidney Winter have long recognized its importance for growth and industrial organization (1982). William Nordhaus (1969) noted this difficulty in his discussion of invention, and Karl Shell emphasized it in his models of growth (1966, 1967, 1973). Mancur Olson (1973) discussed the nonconvexity associated with public goods, as did David Starrett (1977).

In an influential discussion of these issues, Kenneth Arrow (1962) used "appropriability" in place of "excludability" and "indivisibility" in place of "nonrivalry." Appropriability is synonymous with excludability, but indivisibility is not quite precise enough to capture what is really different about goods like designs. Indivisibilities typically imply local nonconvexities. Nonrival goods create global nonconvexities.

Perhaps because the notion of rivalry remained obscure, or perhaps because property rights and excludability are easy to contemplate without undermining the notion of a price-taking equilibrium, what has come to be known as the "appropriability framework" (David Mowery and Nathan Rosenberg, 1989) has focused on excludability, and has tended to neglect the essential nonconvexities noted clearly by Arrow and others.

III. What About Fixed Factors and Aggregation?

The claim that nonconvexities are unavoidable has not gone unnoticed by defenders of price taking. At least three lines of attack on the argument outlined here can be discerned. They are: 1) that there are fixed factors of production; 2) that nonrival inputs are hard to measure; and 3) that the nonconvexities may be present, but they are really not very important for aggregate level analysis.

Because there are factors of production that are in fixed aggregate supply, it is certainly true that it is not possible to replicate all productive activities at the same time. This is a red herring. Nothing in the logic of the argument establishing the nonconvexities requires that it actually be possible to double all inputs. What matters for an equilibrium is what happens to output in any particular production activity if its inputs are doubled, for example, by taking inputs away from some other activity. In any activity, Euler's theorem implies that an elasticity of output with respect to all inputs that is greater than one makes it infeasible to pay each input a rental rate equal to its marginal productivity. Factors like land or managerial ability that are asserted to be in fixed supply have alternative uses in production (and consumption), and can be shifted between firms engaged in the same activity. How they are allocated depends on how they are compensated, and this depends crucially on convexity.

It is also true that the aggregate quantity of nonrival inputs like processes, formulas, instructions, and designs is hard to measure. But, in the years since the capital controversies, the assertion that difficulties of measurement and aggregation can be used to attack theoretical propositions has worn rather thin. These difficulties are hardly unique to nonrival goods. If you look carefully and worry about things like quality change and the introduction of entirely new goods, even concepts like GNP are plagued with ambiguity.

Concepts that are vague at the aggregate level are obvious at the industry or firm level. For the personal computer industry, it is clear what we mean by changes in output or inputs. Industry output (measured in millions of instructions per second of capacity produced per year) can double because rival inputs double; twice as much aluminum, silicon, copper, steel, and plastic can be used along with twice as many workers and machines to produce twice as many physical PCs. Or nonrival inputs can go up when Intel produces a new design for a microprocessor that lets each new PC run twice as fast.

IV. But Do the Nonconvexities Really Matter?

The fallback position to which some defenders of price taking retreat is this: Yes, we admit that nonrival inputs like new chemical reactions, new electronic devices, and advances in materials science will revolutionize living standards within each individual's lifetime. And yes, these inputs are intrinsically tied to nonconvexities. But no, we do not think that this poses a problem for aggregate equilibrium theories based on price taking.

The argument that nonconvexities might not matter at the aggregate level seems to be based on two inappropriate analogies. The first is a confusion of nonrivalry with excludability. If excludability were the only problem, policy towards research and development would be just like policy toward open fields and national defense: Establish property rights if this is possible; have the government pay for a good if it is not. If the problem was merely one of excludability, price taking would be observed in the markets for all of the excludable goods. Optimal provision of the public good by the government would then lead to a Pareto optimal equilibrium.

The deeper problem here is nonrivalry and its associated nonconvexity. Making a nonrival good excludable does not change the fact that it is nonrival. For a firm like Intel, selling its goods at marginal cost is not a viable strategy. Giving Intel better copyright or patent protection may be a good idea, but it will not cause it to behave like a price taker.

A related argument is that departures from the perfect market assumptions must be small if there is no evidence of spillovers or external effects. What the evidence is on spillovers is not entirely clear. In any case, the logic of the argument is wrong. An economy with perfect patent protection for ideas must exhibit departures from price taking. Patents are supposed to create monopolies. Even if patent protection is perfect and there are no external effects and no spillovers, the associated equilibria will generally not be Pareto optimal. (For an example of an economy of this type, see my 1987 paper.)

Economists who recognize the presence of nonconvexities sometimes rely on a second analogy to argue that they are not important at the aggregate level. The canonical example of a nonconvexity that can be convexified with large numbers is an indivisibility that gives firms a U-shaped average cost curve. As the number of identical firms become very large, price taking can be supported at a price equal to minimum average cost. This kind of analysis has been used in club theory to show that private clubs can provide the optimal level of an excludable, partially nonrival good.

The point that should be clear is that not all nonconvexities can be convexified this way. Large numbers of potential firms do not help if the average cost curve of each individual firm is globally decreasing. This is exactly the case that arises with nonrival goods. To see the problem from the production side, let $\bar{x}$ be the amount of an input x needed to create a nonrival good y. Then the production possibility set is the union of the sets $\{(x,0)\colon\ x \in [0,\bar{x})\}$ and $\{(x,y)\colon\ x \geq \bar{x}$ and $y \geq 0\}$ in $\mathbb{R}^2$. The closed convex hull of this set is the nonnegative orthant in $\mathbb{R}^2$. If the original production set is convexified, the implied technology says that any amount of the good y can be had for free. The only supporting price for y in terms of x is zero, and a zero price is inconsistent with an equilibrium in which firms earn nonnegative profits.

V. What Is the Alternative to Equilibrium Theory with Price Taking?

It is clear that we can do general equilibrium theory without assuming price taking. The relevant point was emphasized by Joseph Ostroy (1973, 1984): If an economy has many goods as well as many agents, markets for some goods may still be thin, even in the limit as the number of goods and agents goes to infinity. Market power can survive. But in the limit economy, individual agents are small relative to the economy as a whole, so these "large-square" economies avoid many of the ambiguities that arise in strategic situations with small numbers of agents.

It is not hard to construct large-square growth models that are directly comparable with conventional growth models. Examples are given in my 1987 and 1990 papers, by Philippe Aghion and Peter Howitt (1989), and by Gene Grossman and Elhanan Helpman (1989c). These models differ in the assumptions about spillovers (the last three have them, the first does not) and about the form of innovations (entirely new goods in my models; replacements for old goods in the other two). A product life cycle model that is closely related to the growth models cited here was described by Paul Segerstrom et al. (1987).

All of the models have nonrival goods, so they will share a result about trade and growth. Suppose that two identical economies operate in total isolation from each other. Because they are identical, they allocate the same resources to the research sector and make the same discoveries in the same order. In the classical theory of trade with a convex technology, there are no gains from trade between identical economies. Here, there are large potential gains from interaction. Because there are no limits on the use of nonrival goods, there is no reason to have engineers in the different countries solve the same problem twice. It would be feasible for the engineers of one country to supply the whole world with the same level of nonrival goods as was produced using both sets of engineers under total isolation. The worldwide equilibrium with isolation must be Pareto suboptimal because the engineers in the other country could go on vacation without in any way reducing worldwide output. If increasing the inputs into research is good for growth and welfare, the second set of engineers could be set to work on different problems, thereby doubling the effective resources allocated to research.

The fact that isolation is suboptimal does not necessarily imply that opening trade will be welfare improving. Even if it is, free trade will not lead to an equilibrium that is first-best optimal because of the nonconvexities and departures from price taking that are present. Nonetheless, what seems to be a robust result is that trade in goods between similar countries will lead to a welfare-improving reallocation of resources used in research (see my paper 1990; Luis Rivera-Batiz and myself, 1989). Because the countries are similar, the main effect of opening them to trade will be to induce trade in goods developed through research, thereby reducing redundancy in research effort.

When the countries are different (for example, when one has a higher ratio of skilled human capital to raw labor) opening trade will also affect the relative price of human capital and labor. This can have an ambiguous effect on the market incentives for undertaking research. If an abundant supply of cheap labor reduces the incentive to produce nonrival inputs (as in the models described in my 1989 paper and in Grossman-Helpman, 1989c), then for a country like the United States, increased trade with Mexico might slow down the rate of growth even if increased trade with Europe or Canada would speed it up.

Without extensive (and implausible) government intervention, equilibria in this kind of economy will not be first-best Pareto optimal. As a result, economic reasoning alone can lead to a great variety of counterintuitive second-best results. The rate of growth and total research effort might be too high even if there are positive knowledge spillovers (Aghion-Howitt and Grossman-Helpman, 1989c). Tariffs on some of the goods produced in a country might speed up growth by driving human capital out of production of these goods and into research (Grossman-Helpman, 1989b). Allowing for reverse engineering and knockoffs of goods by less developed countries might speed up the rate of introduction of new goods in the developed country; even though the return to introducing a new good will be lower, the cost of inputs into research might be lower. Labor that is forced out of manufacturing of goods by the foreign goods will drive down wages, and some of this labor may shift into research (Grossman-Helpman, 1989a).

Whether these theoretical possibilities actually occur is the kind of question that does require systematic data collection and econometric analysis. I have argued (1989) that there is evidence for a negative effect of labor growth on technological change. Across countries, across sectors in the United States, and over time in the United States, higher rates of labor force growth are associated with lower rates of productivity growth. There is a large body of work that tries to measure the rate of return to research (Zvi Griliches, 1988). Both the private and the social rate of return to research seem high relative to other rates of return, so there is a presumption that too few rather then too many resources are devoted to research. This is at least a case in which the formal evidence jibes with the newspaper reports. As a fraction of GDP, Japan now spends considerably more on commercial (not basic) research that generates excludable benefits than does the United States, and the sentiment on both sides of the Pacific seems to be that they are better off as a result.

The other conjectures remain open. Whether increased trade generally increases growth rates is still a matter of some debate. Whether copying and product cycle effects speed up or slow down growth has to my knowledge not been studied in detail.

VI. Conclusion

The oldest question in economics is what causes growth. One of the oldest conjectures, built into Adam Smith's story of the pin factory, is that nonconvexities are important for growth and that they create an incentive for international trade that increases the extent of the market. (For a recent reminder of the fact that nonconvexities are essential in the pin factory story, see the discussion by Brian Edwards and Ross Starr, 1987.) We now know how to fit this kind of effect into an aggregate growth model, and we can already see that these models generate many theoretical possibilities. Whatever the verdict on the details, we have enough evidence in hand to come to a consensus about whether nonconvexities matter for growth. It seems clear that nonrival goods exist, that they are important for aggregate growth, and that they create nonconvexities that matter for aggregate level analysis. Moreover, there is some reason to believe that Smith's conjecture about the extent of the market is right. The very notion of nonrivalry suggests that there are large dynamic gains from trade between similar countries.

REFERENCES

Aghion, Philippe and Howitt, Peter, "A Model of Growth Through Creative Destruction," MIT working paper, 1989.

Arrow, Kenneth, "Economic Welfare and the Allocation of Resources for Invention," in Richard Nelson, ed., *The Rate and Direction of Inventive Activity*, Princeton: Princeton University Press, 1962.

Buchanan, James A., "An Economic Theory of Clubs," *Economica*, February 1965, *32*, 1–14.

Edwards, Brian K. and Starr, Ross M., "A Note on Indivisibilities, Specialization, and Economies of Scale," *American Economic Review*, March 1987, *77*, 192–95.

Griliches, Zvi, *Technology, Education and Productivity*, New York: Basil Blackwell, 1988.

Grossman, Gene and Helpman, Elhanan, (1989a) "Endogenous Product Cycles," NBER Working Paper No. 2913, 1989.

______ **and** ______, (1989b) "Growth and Welfare in a Small Open Economy," NBER Working Paper No. 2970, 1989.

______ **and** ______, (1989c) "Quality Ladders in the Theory of Growth," NBER Working Paper No. 3099, 1989.

Mowery, David C. and Rosenberg, Nathan, *Technology and the Pursuit of Economic Growth*, Cambridge: Cambridge University Press, 1989.

Nelson, Richard R. and Winter, Sidney G., *An Evolutionary Theory of Economic Change*, Cambridge: Harvard University Press, 1982.

Nordhaus, William, *Invention, Growth, and Welfare*. Cambridge: MIT Press, 1969.

Ostroy, Joseph, "Representations of Large Economies: The Equivalence Theorem," presented at the winter meetings of the Econometric Society, 1973.

______, "A Reformulation of the Marginal Productivity Theory of Distribution," *Econometrica*, May 1984, *52*, 599–630.

Olson, Mancur, *The Logic of Collective Action: Public Goods and the Theory of Collective Action*, Cambridge: Harvard University Press, 1965.

______, "The Measurement of Economic and Social Performance," in Milton Moss, ed., in *Evaluating Performance in the Public Sector*, New York: Columbia University Press, 1973.

Rivera-Batiz, Luis and Romer, Paul M., "International Trade with Endogenous Technological Change," University of Chicago working paper, 1989.

Romer, Paul M., "Growth Based on Increasing Returns Due to Specialization," *American Economic Review Proceedings*, May 1987, *77*, 56–62.

______, "Endogenous Technological Change," *Journal of Political Economy*, forthcoming 1990.

______, "Capital, Labor, and Productivity," prepared for the December meeting of the Brookings Microeconomics Panel, 1989.

Sandler, Todd and Tschirhart, John, "The Economic Theory of Clubs: An Evaluative Survey," *Journal of Economic Literature*, December 1980, *18*, 1481–521.

Schumpeter, Joseph, *Capitalism, Socialism, and Democracy*, New York: Harper, 1942.

Segerstrom, Paul S., Anant, T. C. A. and Dinopoulos, Elias, "A Schumpeterian Model of the Product Life Cycle," Econometrics and Economic Theory Paper no. 8606, Michigan State University, 1987.

Shell, Karl, "Toward a Theory of Inventive Activity and Capital Accumulation," *American Economic Review Proceedings*, May 1966, *56*, 62–68.

______, "A Model of Inventive Activity and Capital Accumulation," in his *Essays in the Theory of Optimal Economic Growth*, Cambridge: MIT Press, 1967.

______, "Inventive Activity, Industrial Organization, and Economic Growth," in J. Mirrlees and N. Stern, eds., *Models of Economic Growth*, New York: Wiley & Sons, 1973.

Starrett, David A., "Measuring Returns to Scale in the Aggregate and the Scale Effect of Public Goods," *Econometrica*, September 1977, *45*, 1439–55.