

2014

The Falcon's Gyre: Legal Foundations of Economic Innovation and Growth

Robert Cooter

Berkeley Law

Follow this and additional works at: <http://scholarship.law.berkeley.edu/books>



Part of the [Growth and Development Commons](#), and the [Law Commons](#)

Recommended Citation

Cooter, Robert, "The Falcon's Gyre: Legal Foundations of Economic Innovation and Growth" (2014). *Berkeley Law Books*. Book 1.
<http://scholarship.law.berkeley.edu/books/1>

This Book is brought to you for free and open access by Berkeley Law Scholarship Repository. It has been accepted for inclusion in Berkeley Law Books by an authorized administrator of Berkeley Law Scholarship Repository. For more information, please contact jcera@law.berkeley.edu.

The Falcon's Gyre

Legal Foundations of Economic Innovation and Growth

Version 1.4



The Falcon's Gyre

Legal Foundations of Economic Innovation and Growth

Robert Cooter
with Aaron Edlin

BerkeleyLaw

UNIVERSITY OF CALIFORNIA

Version 1.4 posted August 2014

Version 1.4, posted: August 2014

Copyright © 2014 by Robert Cooter

Creative Commons Attribution-ShareAlike 4.0 International License:



<http://creativecommons.org/licenses/by-sa/4.0/legalcode>

Version 1.0, originally posted: June 2013

Robert Cooter ISNI 0000 0001 1867 4300

http://works.bepress.com/robert_cooter/

This book available electronically through the Berkeley Law Scholarship Repository
<http://scholarship.law.berkeley.edu/books/1/>

Contents

viii **Acknowledgements**

ix **Preface**

Chapter 1. Overtaking

1.6 Innovation

1.9 Depletion

1.11 Freedom

1.13 Welfare

1.16 Conclusion

Chapter 2. Ventures

2.2 Life-Cycle of a Business Venture

2.4 Imitators or Enthusiasts?

2.8 Rate of Innovation

2.9 Conclusion

2.11 Appendix

Chapter 3. The Double Trust Dilemma

3.5 Relationships

3.7 Private Law

3.9 Public Law

3.11 Three Stages of Finance in Silicon Valley

3.13 Small and Large Firms

3.15 Ontogeny Recapitulates Phylogeny

3.17 Conclusion

Chapter 4. Separation

4.2 Paradox of Growth

4.4 Producer's Market Power

4.6 Innovator's Market Power

4.8 Conclusion

Chapter 5. Fertility

5.5 Applying the Fertility Principle

5.7 Bargaining and Transaction Costs

5.10 Conclusion

5.11 Appendix: Welfare Triangles

Chapter 6. Separation and Fertility in Intellectual Property Law

- 6.2 Strength
- 6.5 How Strong?
- 6.7 Applications
- 6.8 Reach: Invention or Discovery
- 6.10 Breadth
- 6.13 Duration
- 6.16 Optimal Remedy
- 6.20 Thickets and Holdup
- 6.23 Middlemen and Trolls
- 6.24 Conclusion

Acknowledgements

For the best feast, the host who supplies the main course should ask each guest to contribute a dish. I have supplied the main course to the table of my scholarship and many others have contributed their ideas. Most important, Aaron Edlin provided fundamental ideas and criticism, especially to the initial chapters of this book where he is credited. I would also like to thank Blair Dean Cooter, Robert Litan, Brian Broughman, Merritt Fox, Robert Atiah, George Geis, Steve Maurer, Patrick Goold, and Tal Niv. Technical support was provided by the Berkeley Law School library and the Berkeley Electronic Press. Financial support and encouragement were provided by The Kauffman Foundation, The Humboldt Prize, the Mercatus Center and the Law and Economics Program at George Mason University, and the Berkeley Law School

Preface

After years of parish work, a Catholic priest sat up in bed one morning and thought, “Maybe the Pope is wrong and Buddha is right.” After years of teaching and writing about economic efficiency and the law, I sat up in bed one morning and thought, “Maybe efficiency is wrong and innovation is right.” The effects on human welfare from inventing the tractor far exceed the effects from more efficient allocation of horses. Innovation causes compound growth that swamps static inefficiency like a tsunami swamps a scow. These facts compelled me to rethink previous work on law and economics, including my own.

The state provides infrastructure on which the economy runs. The material infrastructure includes roads and the institutional infrastructure includes laws. Smooth roads and good laws sustain the economy. Law and efficiency economics explains how law improves resource allocations. Its practical usefulness, systematic consistency, and even its intellectual beauty, have influenced America legal scholarship. By contrast, law and growth economics must explain how law increases economic innovation. Its potential usefulness potentially exceeds law and static economics, but, in its current state of underdevelopment, it lacks systematic consistency or intellectual beauty.

Law and growth economics is underdeveloped partly because growth economics is underdeveloped. Growth economics explains innovation’s effects, but not its causes. Mystery shrouds the causes of innovation because innovation is intensively legal and growth economics is not. Economics exports ideas to law and imports little, like contemporary China exports commodities to the USA imports little. Importing some legal ideas into economics, thus balancing trade in ideas, can dispel some of the mystery of growth.

Economic innovation usually requires combining new ideas and capital. They naturally repel each other because the investor distrusts the innovator with her money, and the innovator distrusts the investor with his ideas. The “double trust dilemma” refers to the problem of inducing the innovator to trust the investor with his ideas, and also inducing the investor to trust the innovator with her money. The solution to this problem depends on law, especially the law of property, contracts, corporations, finance, and bankruptcy. This book dispels some of the mystery of growth by explaining law’s contribution to solving the double trust dilemma.

Will law and growth economics fulfill its promise? Two shoe companies sent representatives to Hawaii in the 19th century. One of them telegraphed the message to the home office, “No opportunity. No one wears shoes.” The other telegraphed the message, “Great opportunity. No one wears shoes yet.” Instead of shoes, this book telegraphs the latter message about law and growth economics.

Chapter 1. Overtaking¹

Robert Cooter and Aaron Edlin

Robert Lucas, the economist who won the Nobel prize in 1995, famously commented, “Once one starts to think about economic growth, it is hard to think about anything else.” Compared to sustained growth, other sources of national wealth are insignificant. Compounded over a century, 2% annual growth (roughly the growth rate of the US economy in the last century) increases wealth by more than 7 times, and 10% annual growth (roughly the growth rate of the Chinese economy from 1980 to 2010) increases wealth by almost 14,000 times.² People systematically underestimate the effects of compound growth, like thinking that a 5% increase in height is the same for a teenager and a toddler.

Sustained growth causes wealth in one country to overtake another faster than most people can imagine. Assume that country x’s national income begins *twice* as large as country y’s, but the former stagnates and the latter grows. To overtake x’s national income, y will need 36 years if it grows at 2%, 15 years if it grows at 5%, and 8 years if it grows at 10%. To illustrate concretely, most people are surprised to learn that South Korea, Mexico, and Senegal had roughly equal income per capita in 1950, yet in 2008 South Korea was more than twice as high as Mexico and ten times higher than Senegal. Similarly, in 1870 Argentina’s per capita income was 33 percent higher

1 This chapter draws on Robert Cooter and Aaron Edlin, “Overtaking,” in *The American Illness: Essays on the Rule of Law*, ed. Frank Buckley (Yale University Press, forthcoming), and “Maximizing Growth vs. Static Efficiency or Redistribution,” working paper, Berkeley Economics Department.

2 Here’s a table of size reached by an economy that starts at 1 and grows at various rates and years.*

Growth rate	1 year	5 years	10 years	25 years	50 years	100 years
0%	1.00	1.00	1.00	1.00	1.00	1.00
0.5%	1.005	1.01	1.05	1.13	1.28	1.65
1%	1.01	1.05	1.10	1.25	1.64	2.70
2%	1.02	1.10	1.22	1.64	2.69	7.24
5%	1.05	1.28	1.63	3.39	11.47	131.50
10%	1.10	1.61	2.59	10.83	117.39	13,780.61

*size of economy = $(1+r/100)^t$, where r =percentage growth rate, and t =years of growth

than Sweden's, yet by 2004 Argentina's per capita income had dropped to 43 percent of Sweden's.

Angus Maddison estimated changes in wealth over millennia for regions in the world. He concluded that Egypt was the world's richest region two thousand years ago with per capita income 50 percent higher than elsewhere in the Roman Empire, China, or India. In the year 1000, Iran and Iraq under the Abbasids were the economically most advanced countries with a per capita income 50 percent higher than in Europe or Asia. By 1500 Italy had the lead with a per capita income 50 percent higher than in the rest of Western Europe, double that of Asia, and three times that of Africa. In 1820 Western Europe and the United States had the highest income per capita, twice as much as in eastern Europe, Latin America, or Asia, and three times as much as Africa. In the 19th century the rate of growth gradually accelerated, and accelerated more in the 20th century. Constant compound growth such as 2% implies accelerating absolute growth, just as a 2% gain in a teenager's weight is absolutely more than a 2% gain in a toddler's weight. From the perspective of two centuries, the compound growth rate in the industrializing countries accelerated modestly, which implies a large acceleration in the absolute growth rate.³ The successful countries are getting absolutely richer very fast,. However, economic growth in the world is uneven, so the economic gap between fast-growing countries and slow-growing countries is much larger today than ever before in history. The wealth of the fastest growers has risen above the slowest growers like the Himalaya Mountains rising above the Ganges Plain.

3

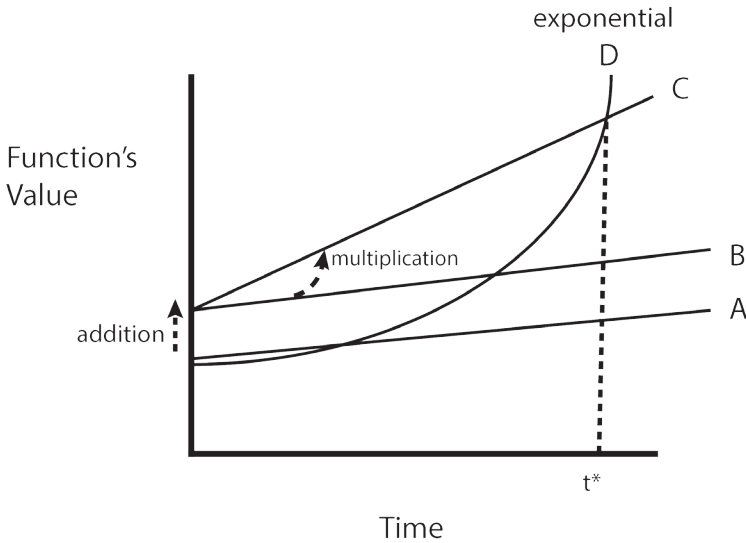
Figure 1.1. Ganges Plain and Himalaya Mountains



Behind these facts stands a mathematical generalization: an economy that increases at an exponential rate will overtake an economy that increases at a constant absolute rate. (“Exponential rate” is a mathematician’s term for “constant proportional rate”.) Figure 1.2 depicts this fact.⁴ First consider function A, whose value increases at a constant rate with time, as indicated by A’s constant slope. An addition to the value of A shifts it up and yields B. A multiplication of B’s value rotates it up and yields C. The additive and multiplicative increases make C larger than A or B at each point in time. Now contrast C to function D that increases at an exponential rate. D starts behind C at time 0, and overtakes it at time t^* . This book concerns exponential growth like function D.

⁴ Joke: “Economists need a graph to reach the wrong conclusions with certainty, whereas lawyer can do so immediately.”

Figure 1.2. Exponential Growth Overtakes Additive and Multiplicative Increases



An economy produces different goods like houses, clothes, books, and clean water. The usual measure of economic growth is the increase in wealth (a stock) or income (a flow). To measure wealth or income, economists combine different goods into comprehensive aggregates, such as gross national product (GNP). To combine heterogeneous goods, economists first weight quantities of goods by their market prices. Economists ideally adjust market prices for unpriced externalities such as the social cost of pollution. Comprehensive measures of wealth or income include externalities, and also public goods such as national security and air quality. To get closer to human welfare, economists measure consumption instead of wealth or income, and they divide the total by the number of people.⁵ Later we consider more controversial measures, specifically utility and social welfare.

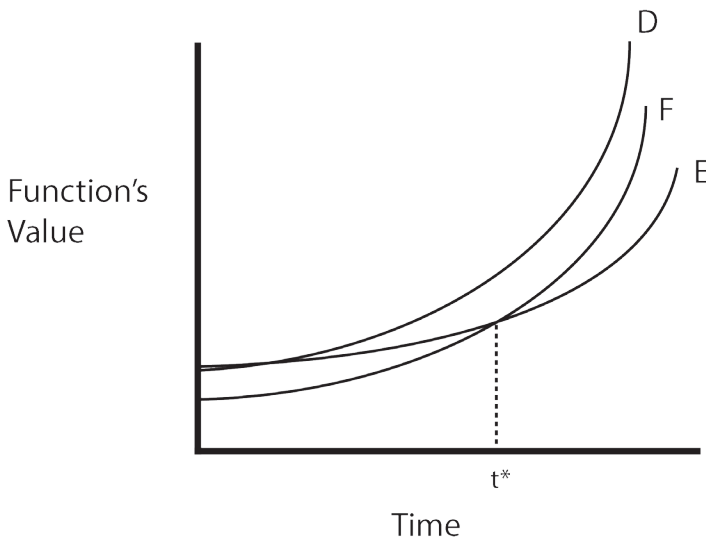
The vertical axis in Figure 1.2 is unlabeled because the measure of growth -- total or per capita wealth, income, consumption, utility, or welfare -- matters little to our conclusions. For all economic measures, exponential growth

⁵ In the national accounts, income can be “consumed” or “invested.” Consumption includes almost all the activities that people will pay to do, but not activities that people must be paid to do. Thus people “consume” an opera, a golf lesson, or a holiday. Note that in the technical language of economics, people “consume” many things without using them up, like playing a recording, whereas in ordinary speech “consumption” refers to using up something, like eating food or burning wood.

overtakes static effects. The *overtaking principle* in economics is my name for the proposition that rapid exponential economic growth quickly overtakes the direct effects of static efficiency and redistribution on wealth and welfare. Given fast growth, static efficiency and income redistribution effects are important for their contribution to growth, and unimportant in themselves.

Besides prioritizing dynamic over static effects, the overtaking principle prioritizes some dynamic paths over others. Thus in Figure 1.3, the growth-path of income or wealth represented by D is unambiguously better than the growth-path represented by E, because D starts higher and grows faster than E. The analysis is different, however, when comparing E and F. F starts lower than E, grows faster, and overtakes E at t^* . If overtaking occurs rapidly, then t^* is small and nothing matters except growth. If, however, overtaking occurs slowly, then t^* is large and static effects of efficiency and redistribution matter. Thus F is better or worse than E depending on the rate of growth. The measure of growth seldom changes these conclusions because different measures have similar overtaking times.

Figure 1.3. Different Exponential Growth Rates



In contrast, the overtaking principle loses much of its normative appeal when growth and overtaking is slow. Thus interpret E and F as two sequences of consumption in an infinite number of generations. In the first sequence E,

initial consumption is higher and growth is lower. In the second sequence F, initial consumption is lower and growth is higher. If overtaking occurs quickly (t^* is small), any reasonable economic measure will yield a larger value for F than E. Conversely, if overtaking occurs slowly (t^* is large), some reasonable economic measures will yield a larger value for F than E, and other reasonable measures will yield a large value for E than F. For example, utilitarians measure economic value by summing the utilities of different people. Theorists disagree over whether lawmakers should give similar weight to the utility of future generations as to the present generation. If growth is fast and t^* is small, the sum of utilities in F overtakes the sum of utilities in E for any reasonable discount rate.⁶ Conversely, if growth is slow and t^* is large, whether F or E yields the larger sum of utilities depends on the rate for discounting future utility.

Innovation

Economic theory ascribes long run growth to three broad factors: physical capital (buildings and machines), human capital (work adjusted for the skills people bring to their jobs), and “innovation”. An increase in these factors increases income and wealth. MIT economist Robert Solow won the Nobel Prize partly for showing that innovation was more important to growth in the 1950s in the United States than increased physical or human capital. Subsequent empirical work by Edward Denison, Robert Barro, and others confirmed this finding for other developed economies.⁷ In the last 100 years, innovation caused more economic growth than anything else, including using more resources.

Innovations use new ideas to produce goods cheaper or to make better goods. The supply curve shows the cost of making a good,⁸ so a cost-reducing inno-

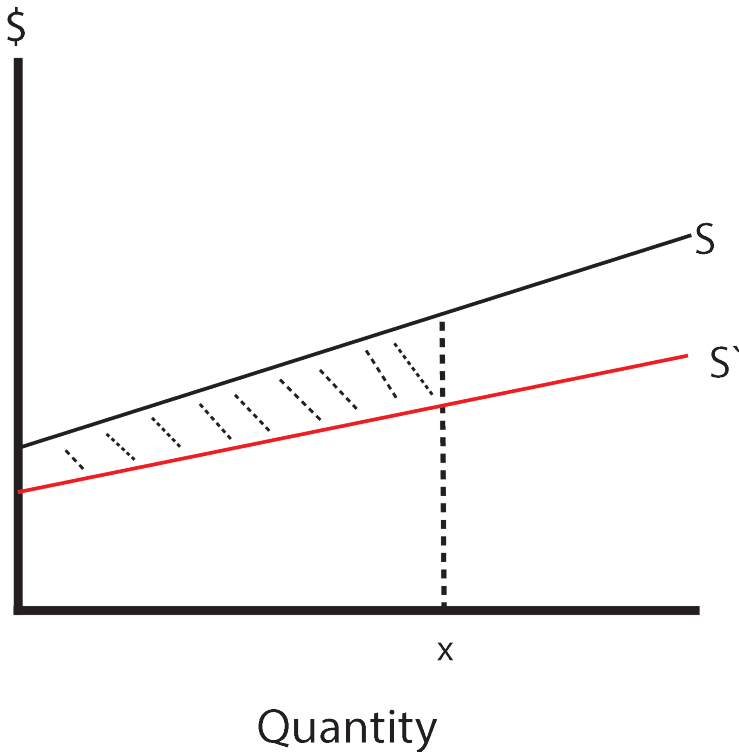
⁶ Utilitarians often hold that future utility should not be discounted, whereas future income should be discounted by its marginal utility. In this view, utility is what discounts money, but there is nothing to discount utility. For a justification of not discounting future utility or welfare, see Derek Parfit & Tyler Cowen, *Against the Social Discount Rate*, in JUSTICE BETWEEN AGE GROUPS AND GENERATIONS (Peter Laslett & Fishkin eds., 1992).

⁷ Add citations.

⁸ The supply curve is a monotonic relationship between price and quantity. Inferring from a price to a quantity, the supply curve shows the quantity that a firm or industry will supply at a given price. Conversely, inferring from a quantity to a price, the supply curve shows the cost for a firm or industry to supply a given quantity.

vation shifts the supply curve down, as when the tractor replaced the horse for plowing. In Figure 1.4, innovation shifts the supply curve from S to S' , and the shaded area indicates the social value of the savings in cost from producing the quantity x of the good.

Figure 1.4. Social Value of a Cost-Saving Innovation

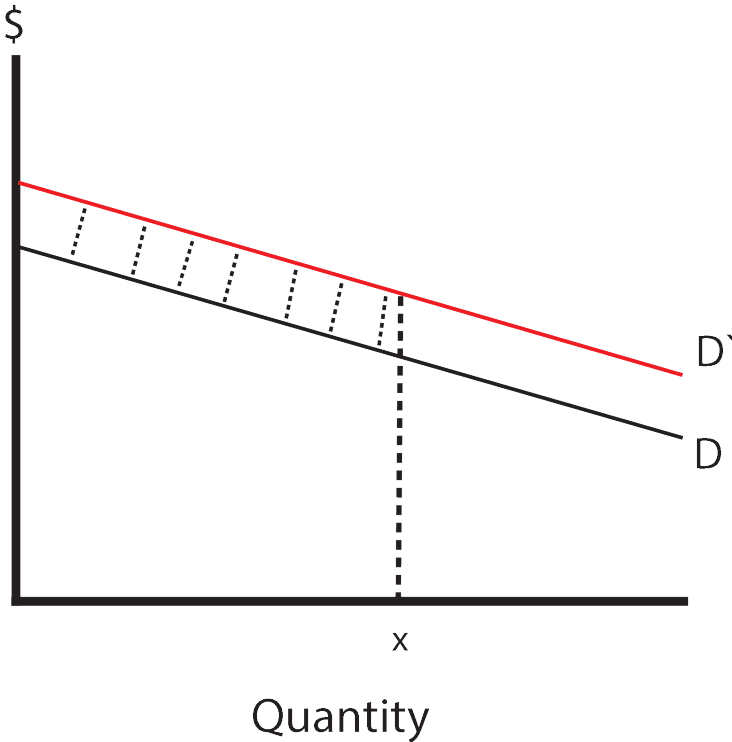


Similarly, producing a better good increases its value to consumers, as when waterproof fabrics were made to breathe (“Gor-Tex”). When quality improves, people are willing to pay more. The demand curve shows how much people are willing to pay for a good,⁹ so a quality-improving innovation shifts the demand curve up. In Figure 1.5, the innovation shifts the demand curve

⁹ Like the supply curve, the demand curve is a monotonic relationship between price and quantity. Inferring from a price to a quantity, the demand curve shows the quantity that consumers will buy at a given price. Conversely, inferring from a quantity to a price, the demand curve shows the price that consumers are willing to pay for a given quantity.

from D to D' , and the shaded area indicates the social value of the improvement in quality from producing the quantity x of the good.

Figure 1.5. Social Value of a Quality-Improving Innovation



When one person uses a scarce resource, others cannot use it, like taking a bite from a sandwich. Scarce resources like capital and labor have rival uses. In contrast, when one person uses an idea, just as much remains for someone else to use. Thus architects have used the Pythagorean Theorem for over two millennia and just as much remains as before. Using theorems, principles, natural laws, designs, discoveries, expressions, and compositions does not use them up. Economists call this property “nonrivalry”. Furthermore, when one person uses an idea, preventing others from doing so is difficult, more difficult than preventing someone from trespassing on your land or taking a bite from your sandwich. Ideas spread like gossip in Puddletown. Economists call this property “non-exclusion.”¹⁰

¹⁰ In economics, non-rivalry and non-exclusion define “public goods” like national security and air quality, which resemble ideas in these two traits.

“Nonrivalry” and “non-exclusion” are good news about ideas, and here is even better news: ideas are fertile like seeds, not dead like coal. 2,500 years after its proof, generalizations of the Pythagorean Theorem continue to expand geometry’s power.¹¹ Similarly, the number of transistors that can be placed inexpensively on an integrated circuit has doubled approximately every two years for the past fifty years (Moore’s law). Ideas about transistors are so fertile that innovation grows exponentially.

With a continuing stream of fertile innovations, growth can continue indefinitely and production can rise like the falcon’s gyre.¹² This hypothesis is a reasonable conclusion from the last century of experience. It is also a reasonable conclusion from recent technological and scientific progress, especially develops in computers, nanotechnology, robotics, genetics, and medicine. In spite of wars and recessions, the United States and other Western capitalist countries have averaged growth of 2 to 3 percent per year in per capita GDP for more than one hundred years. In 1930 after the stock market’s crash triggered the Great Depression, John Maynard Keynes wrote his essay, “Economic Possibilities for Our Children”, which argues that the economic problem of the future is how to use abundance. Similarly, in 2012 Diamandis and Kotler wrote Abundance: The Future Is Better Than You Think. Or perhaps absolute growth will accelerate so much that the future is unimaginable to us, as Kurzweil argued in THE SINGULARITY IS NEAR: WHEN HUMAN TRANSCEND BIOLOGY (2005).

Depletion

Although the signs are good, the future remains in doubt until it arrives. A worrisome doubt concerns resource exhaustion and environmental degradation. Must depletion of resources eventually end growth? Some resources renew like a forest, river, or wheat. Use does not necessarily reduce their stock permanently because we replenish it. Instead of renewing, however, other resources deplete, like oil and iron. Use depletes their stock because we do not

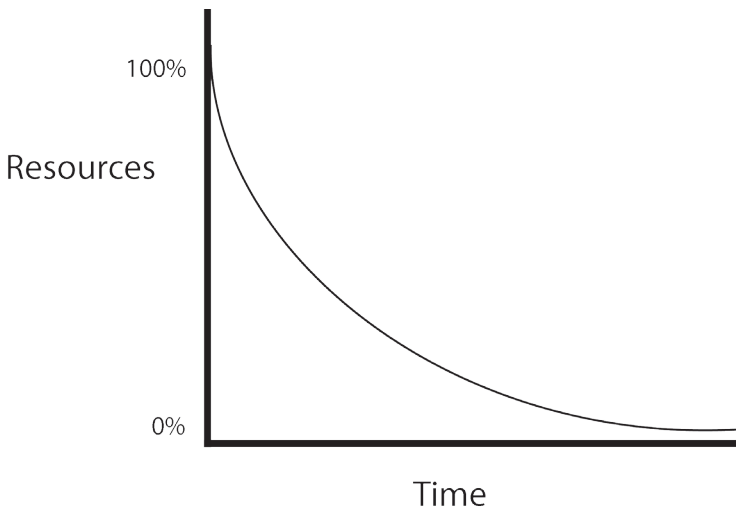
11 The theorem says, “The square of the hypotenuse of a right triangle is equal to the sum of the squares on the other two sides.” The fascinating story of the proof of its generalization in n-dimensions is told in Simon Singh’s book, Fermat’s Enigma: The Epic Quest to Solve the World’s Greatest Mathematical Problem (New York, etc.: Doubleday, Anchor Book, 1998).

12 For a survey of technologies where that great abundance is the future, see PETER H. DIAMANDIS & STEVEN KOTLER, ABUNDANCE: THE FUTURE IS BETTER THAN YOU THINK (2012).

know how to replenish it. Whereas fertile ideas increase with use, depletable resources diminish with use.

However, even with non-renewable resources, sustainable depletion is possible in principle. “Sustainable depletion” sounds oxymoronic like “unmarried husband”, but it makes sense. In the 5th century BC, Xeno posed his paradoxes of motion that resembled the following: “By traveling half of the remaining distance to your destination each day, you never arrive (except in the mathematical limit when time goes to infinity).” Similarly, if the stock of oil falls by 50 percent in every period, the stock never reaches zero in finite time. “Sustainable depletion” refers to a path of depletion that never reaches zero, such as Figure 1.6.

Figure 1.6. Sustainable Resource Depletion



Depletion is sustainable if the rate of depletion decreases forever. Instead of decreasing, the rate of depletion is increasing in the world for many non-renewable resources. This trend could be reduced by continually reducing consumption. However, laws and policies to reduce consumption encounter fierce political resistance. If sustainability requires reducing consumption, humanity might destroy its future rather than diminish its present.

Fortunately, sustainable depletion is possible in principle through innovation, without reducing consumption. Innovation slows depletion by using

resources more efficiently, as when new automobile engines economize on fuel. Innovation also slows depletion by substituting renewable resources for non-renewable ones, as when hydro replaces coal to generate electricity. To deplete sustainably without decreasing consumption, innovation must get more output from fewer inputs of non-renewable resources, and innovation must substitute renewable resources for non-renewable resources. Consumption can increase forever if the gains from innovation outpace the losses from depletion.

Freedom

Every youth who watches the Olympics fantasizes standing on the central platform with a gold medal while the national anthem plays. In wealth as in sports, individuals and nations hope to surpass others and fear being surpassed by them. Like the Olympics, the value of winning the race for wealth or income is self-evident to many people. If you are one of them, facts cited above provide ample motivation to study law and growth economics.

Other readers of this book, however, will want moral or political justification for focusing on the growth of wealth. Wealth and income are means for buying goods, not ends in themselves like beauty, love, holiness, knowledge, or self-fulfillment.¹³ Accumulating wealth only to misuse it is a labor of shame. Is the nation that wins the growth race like the winner of the pie-eating contest whose prize is another pie? Does studying growth turn wealth into a fetish like falling in love with a shoe?

13 The champion of this view in development economics is Amartya Sen, as two quotes suggest. “Economic growth cannot be sensibly treated as an end in itself. Development has to be more concerned with enhancing the lives we lead and the freedoms we enjoy.” *Development as Freedom* (New York: Knopf, 1999), 14. “The challenge of development . . . is to improve the quality of life. Especially in the world’s poor countries, a better quality of life generally calls for higher incomes—but it involves much more. It encompasses as ends in themselves better education, higher standards of health and nutrition, less poverty, a cleaner environment, more quality of opportunity, greater individual freedom, and a richer cultural life.” World Bank, *World Development Report 1991* (New York: Oxford University Press, 1991). Note that governments supply many non-market goods, and GDP measures their value by their cost (e.g. salaries paid to civil servants), not by their benefits to the citizens. Cost-benefit analysis can measure some of these non-market values more convincingly. To measure the value of non-market goods, economists try to find out how much people would pay for them if they had to pay, given that they don’t have to pay. This can be a measurement maelstrom, so national accounting limits its use of cost-benefit analysis. For the relationships between happiness and money, see B. S. Frey and A. Stutzer (2002), “What Can Economists Learn from Happiness Research?” *Journal of Economic Literature*.

One justification for focusing on wealth and income concerns freedom. Like driving a car, most people think that they are good at spending money, in spite of occasional lapses such as purchasing uncomfortable shoes. People know what they want to buy, and they criticize what others buy. A liberal education should help you to think critically about what is worth buying. To critique the uses of wealth, intellectuals draw on traditions in philosophy and religion concerning what is *really* good, as opposed to what *seems* good. Thus intellectuals in the university often carp about strip malls, popular movies, fried food, ostentatious furniture, pay-day loans, and television preachers.

Most economists, however, refrain from debate about what people ought to want. By doing so, economists cultivate neutrality in controversies about the values of consumers. Instead of embracing a particular conception of goodness, economists often accept “consumer sovereignty” – the right of consumers to buy what they want. Consumer sovereignty implies the right to buy things that are good or bad, just as free speech implies the right to say things that are true or false. I may disagree with what you buy and defend your right to buy it, just as I may disagree with what you say and defend your right to say it.¹⁴

Philosophers distinguish two kinds of freedom that apply to consumer sovereignty.¹⁵ First, the absence of prohibitions or “negative freedom” increases when people have fewer restrictions on what they can do, including spending their money. Economists promote negative freedom of consumers by defending free markets. Second, the presence of opportunities or “positive freedom” increases when people have more money to spend. Economists promote positive freedom of consumers by finding policies and laws that increase wealth.

Besides consumers, the two kinds of freedom apply to entrepreneurs. Economic innovation mostly occurs in business ventures that develop new ideas. Negative freedom gives entrepreneurs the legal right to experiment with new economic ideas. Positive economic freedom gives entrepreneurs the resources to launch innovative ventures. Negative and positive economic freedom is the legal condition for business creativity, as discussed in the next chapter.

¹⁴ That’s how I felt when I drove through Los Vegas.

¹⁵ The classical source is Isaiah Berlin, *Two Concepts of Liberty*, in *FOUR ESSAYS ON LIBERTY* 118–172 (1969).

Welfare

Increasing negative and positive freedom is an appropriate goal for economists in a democratic state. Increasing freedom, however, is not the usual justification given by economists for increasing wealth. Economists usually justify increasing wealth as the means to increase the welfare of people. The normative branch of economics connects wealth to welfare, rather than connecting wealth to freedom. In the history of economics, the philosophy of utilitarianism was central to connecting wealth to welfare, especially in 19th and early 20th century England.¹⁶ Utilitarians equate individual welfare with a person's utility, and they equate social welfare with the sum of individual utilities. In notation, $SW = u^1 + u^2 + \dots + u^n$. Instead of thinking of social welfare as a sum of individual utilities, however, modern economists think of it as an increasing *function* of individual utilities: $SW = f(u^1, u^2, \dots, u^n)$.¹⁷ The function may be additive as with utilitarianism, or it may be non-additive, such as multiplicative ($SW = u^1 \times u^2 \times \dots \times u^n$) or logarithmic ($SW = \log u^1 + \log u^2 + \dots + \log u^n$). From a mathematical viewpoint, the social welfare function generalizes utilitarianism,¹⁸ and different forms of the function have different ethical consequences for equality.¹⁹

16 Economic theory assumed its recognizably modern mathematical form when theorists understood that equating marginal benefits and marginal costs maximizes utility. This recognition joined Newton's calculus and the philosophy of utilitarianism and. See the "marginalist revolution" in economics as discussed in the classic by JOSEPH A. SCHUMPETER, *HISTORY OF ECONOMIC ANALYSIS* (1986).

17 "Utility" is a notoriously elusive concept that pervades economic theory. For a discussion of its various meanings, see...

18 The concept of social welfare as a function of individual utility was introduced by Abram Bergson, *A Reformulation of Certain Aspects of Welfare Economics*, *QUARTERLY JOURNAL OF ECONOMICS* 310–34 (1938). All forms of the social welfare function, additive or non-additive, assume that social welfare increase with individual utility. For a recent discussion of social welfare that is deep but challenging to read, see MATTHEW ADLER, *WELL BEING AND FAIR DISTRIBUTION* (2012).

19 Summing utilities gives the same weight to utility regardless of its distribution among persons. Thus, in a society of 5 people, social welfare is the same as measured by the sum of utilities if 5 people each enjoy 10 utils, or if one person enjoys 50 utils and the other 4 get 0 utils. In contrast, social welfare is higher for the multiplicative or logarithmic social welfare functions if 5 people enjoy 10 utils than if one person enjoys 50 utils and the other 4 get 0 utils. In general, the additive form of SWF is indifferent about the distribution of a fixed amount of utility across people, whereas the multiplicative and logarithmic forms favor its equal distribution across people. Note, however, that all three forms of SWF favor a more equal distribution of a fixed amount of income across people, so long as the marginal utility of money is decreasing for all individuals. The most sophisticated treatment of this problem, which is also challenging to read, is found in MATTHEW ADLER, *WELL BEING AND FAIR DISTRIBUTION* (2012).

People agree about some connections between wealth and welfare. Thus life expectancy at birth is 83 years in Japan and 66 years in Bangladesh,²⁰ and enrollment in secondary school is 98 percent among Japanese children of the appropriate age and 42 percent in Bangladesh.²¹ Facts like these make almost everyone agree that welfare is higher in Japan than Bangladesh. The poor in Bangladesh need extra money to buy basic health care and primary education more than the rich in Japan need extra money to buy cosmetic surgery and theater tickets. In general, the poor need extra money for necessities more than the rich need extra money for luxuries.

Economists often describe this fact as the “decreasing marginal utility of money with respect to income”.²² Progressive taxes and government expenditures can shift consumption from less to more basic needs, which presumably increases a nation’s welfare. Many economists believe that the poor need extra money more than the rich, but they disagree about how much more. After a century of trying, there is no generally accepted method for measuring the marginal utility of income like the thermometer measures heat. Thus economists have no uniquely correct measure of the difference in welfare between Japan and Bangladesh.

Perhaps comparing the marginal utility of income of rich and poor is like distinguishing between your face and the back of your head: the difference is real but the boundary is imprecise. In this respect, “social welfare” may be like other values that social scientists measure in controversial ways, such as “happiness”²³. If so, economists can talk meaningful about the declining marginal utility of income, but they will never find a uniquely correct measure of it. According to this view, economics must encompass controversial

20 Life expectancy at birth, total (years) in 2008, World Bank. <http://data.worldbank.org/indicator/SP.DYN.LE00.IN>. Note that life expectancy is in the 40s in many African countries.

21 Data & Statistics on Education for 2007 and 2008, World Bank, <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTEDUCATION/0,,contentMDK:20573961~menuPK:282404~pagePK:148956~piPK:216618~theSitePK:282386,00.html>.

22 The utility of a poor person increases more from an increase in wealth than the utility of a rich person (decreasing marginal utility of money. Social welfare increases in the utility of individuals. Therefore, redistribution of wealth from rich to poor (with no loss in the amount of wealth) increases social welfare. In Adler’s theory (see preceding footnote), social welfare increases with individual wealth, and social welfare increases with a more equal distribution of wealth or utility across individuals.

23 See Bruno S. Frey & Alois Stutzer, *What Can Economists Learn from Happiness Research?*, JOURNAL OF ECONOMIC LITERATURE (2002). Also see Daniel Kahneman & Richard Thaler, *Utility Maximization and Experienced Utility*, 20 JOURNAL OF ECONOMIC PERSPECTIVES 221–234 (2006).

measures of imprecise facts. Thus the “material welfare” school of the early 20th century held that additional money benefits the poor more than the rich, which can be shown by measures of the causes of welfare like health and education that are reasonable, pragmatic, and incomplete.²⁴

In contrast, another tradition in economics called “positivism” holds that the marginal utility of money has no *scientific* meaning. Given enough data, all scientific propositions can allegedly be proved true or false. In contrast, propositions about “social welfare” or the “marginal utility of money” are value judgments or expressions of political commitments, not testable assertions of fact. The question “How much higher is the marginal utility of income of the poor than the rich?” has no more scientific meaning according to the economic positivists than the question, “Is it morning or afternoon on the sun?” Since value judgments do not make testable claims, according to the economic positivists, “social welfare” cannot be part of scientific economics.

In any case, measuring social welfare and the declining marginal utility of money is less important to growth economics than to static economics. When progressive taxation and state spending shift consumption from luxuries to necessities, the static effects – whether measured in wealth, income, utility, or welfare for the poor or everyone in society – correspond to the shift in Figure 1.2 from A to B, or from B to C. The static effects are small relative to the dynamic effects represented by D, regardless of how they are measured. When growth is exponential, the mathematics of overtaking applies, regardless of the function’s interpretation in Figure 1.2. Measuring utility or welfare is unimportant to policy conclusions when fast growth causes rapid overtaking, whereas the measure is important when sluggish growth causes slow overtaking.

To illustrate, assume that redistributive policies improve the health and education of the poor, which directly increase their welfare. In addition, healthier workers with better education are more creative, so they may increase economic growth. Economists who believe that better education and health of poor Americans would increase U.S. growth point to Denmark and Korea as examples where good education and health have contributed to robust

24 For an account of the material welfare school and its tension with scarcity economics, see Robert Cooter & Peter Rappoport, *Were the Ordinalists Wrong About Welfare Economics?*, 22 J.ECONOMIC LITERATURE 507 (1984).

economic growth, whereas poor education and health of workers may partly explain the economic struggles of the Philippines.

In the preceding hypothetical, redistribution increases the growth rate by improving the welfare of workers. Redistributive policies can also decrease the growth rate by undermining economic incentives. China's Cultural Revolution, which commenced in the 1960s and expired by 1975, enforced strict economic equality and destroyed what was left of the private sector. Economic decline immiserized many people. Reversing policies after 1980, China dissolved agricultural communes, farmers were allowed to keep the surplus from selling their crops, and agricultural productivity soared. Spectacular success in agriculture prompted the Chinese leadership to allow the development of private business in the new export sectors, where entrepreneurs could keep much of what they earned. Improved incentives unleashed economic growth without historical precedent. Instead of trickling down to the Chinese workers, the benefits of growth cascaded down like the Yangtze River at the Three Gorges dam. Equality decreased, economic growth exploded, and poverty plummeted. The lowest wage earners in China benefited greatly from a faster growth rate in national income after 1980. Even unemployed poor people who depend on welfare payments and government subsidies benefitted from faster growth, which increased state revenues available for transfer payments.

Conclusion

The first question of law and growth economics is, "Which laws increase the pace of economic innovation?" Increasing the pace of innovation can lead to sustained growth in wealth, income, utility, and welfare. When growth is rapid, overtaking is the only normative standard required to choose among many laws and policies. Balancing growth against static efficiency or equality is unnecessary when growth is fast, which simplifies the analysis of many laws and policies such as patents. Isaac Newton invented calculus and discovered the laws of motion associated with gravity in 1666, which historians of science call the "miracle year" ("annus mirabilis"). Civil engineers still use Newtonian physics for ordinary objects, although it fails for very large objects (the cosmos) and very small objects (sub-atomic particles). Similarly, rapid growth is the domain of the overtaking principle, which is good for making

law and policy in most of growth economics. In contrast, sluggish growth is the domain of efficiency and equality.

The overtaking principle supplies the normative justification for prioritizing growth economics over static economics. It challenges conventional law and economics that treats static efficiency as a fundamental goal of law and policy. Which was more important to agricultural production, inventing a tractor or using horses more efficiently? A better allocation of horses for plowing the fields increases agricultural production marginally, whereas inventing the tractor caused a jump in production. Once you appreciate exponential growth, it's hard to care about static efficiency for its own sake.

The overtaking principle also challenges ethical theories concerning the redistribution of wealth. Controversies about fair distribution, social welfare, the marginal utility of income, and time-discounting do not matter when growth is rapid. In much of political philosophy, fairness concerns distributing shares of fixed income, and economic equality is an end in itself. With rapid growth, however, putative ends turn out to be only valuable as means. Economic equality mostly affects welfare through growth, not in its own right. When rapid growth is possible, progressive taxes and state expenditures increase the welfare of the poor if they cause faster growth in wages and subsidies.

Perhaps you think that static efficiency and growth align, with more efficiency causing faster growth. This view is roughly two-thirds right and one-third wrong. Efficiency and growth have a common cause: competition. To maximize innovation, the law must create a framework of open competition so innovators can develop their ideas. The core of economic growth comes from entrepreneurs. Only the possibility of extraordinary profits can induce entrepreneurs to bet big on risky ideas. However, extraordinary profits require market power, not competition. Thus patent law creates open competition to innovate and rewards the winners by giving them temporary market power. As told in this book, the story of law and growth economics is open competition to innovate and extra-ordinary profits for the winners.

Astronomers from the time of Aristotle saw the earth as the center of the universe and the sun revolving around it, until the Copernican Revolution put the sun at the center. In general, a revolution rearranges the central

elements of a scientific theory.²⁵ Such a revolution occurred in economics in the 1930s. Before the revolution, economics was defined as the science of material welfare, and then a famous essay proposed this alternative in 1932: “Economics is the science which studies human behavior as a relationship between given ends and scarce means which have alternative uses.”²⁶ The scarcity definition formulated a new scientific paradigm for economics that displaced its predecessor.²⁷ The scarcity definition characterizes contemporary law and economics, which can be called “law and scarcity economics.” This book replaces scarcity with growth as the paradigm, yielding “law and growth economics.” Once one starts to think about law and growth economics, it is hard to care about law and scarcity economics.

25 This is the thesis of the classic by THOMAS S. KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS* (1st ed. 1962).

26 Lionel Robbins, *An Essay on the Nature and Significance of Economic Science*. (London: McMillan, 1932), page 16.

27 For a discussion of the paradigm shift, see Robert Cooter and Peter Rappoport, “Were the Ordinalists Wrong About Welfare Economics?,” *J.Economic Literature* 22 (1984): 507.

Chapter 2. Ventures¹

Robert Cooter and Aaron Edlin

A bold ship's captain in seventeenth-century England proposes to investors that they finance a voyage to Asia for spices.² The voyage is inherently risky. Weather is uncertain, channels are uncharted, the Dutch prey on English ships, the English prey on Dutch ships, and other pirates prey on both of them. If the captain returns to the English port with a cargo of spices, however it will be worth a fortune. The ship's captain needs a large ship outfitted for two to five years of travel. He talks to investors and discloses secrets about how to get to Asia and what to do when he arrives. The investors decide to entrust him with a ship and supplies. Unlike so many other ships that sail for Asia, this one returns safely after two years. The townspeople spot the vessel sailing toward the harbor and the investors rush to the dock to divide the cargo. The success of the venture attracts imitators. As years pass, the spice trade eventually becomes a normal business with moderate risk and ordinary profits.

Similarly, an engineer in Silicon Valley in 1985 has an idea for a new computer technology. The engineer cannot patent the idea until he develops it. Developing it requires more money than the engineer can risk personally. He drafts a business plan, meets with a small group of investors, and explains his idea to them. Developing the idea is inherently risky—someone might steal the idea before it is patented, an unknown competitor might patent the invention first, the invention might not work, or it might work but not sell. If the plan succeeds, however, the innovators will make a fortune. The investors agree to form a company and develop the product. Unlike so many other start-ups, this one succeeds after two years and the firm acquires a valuable patent. The company sells the product to an established firm and divides the proceeds between the engineer and the investors. The success of the venture attracts imitators. Production and sale of the invention eventually becomes a normal business with moderate risk and ordinary profits.

1 This chapter draws on Chapters 4 and 5 of Cooter and Schaefer, 2011.

2 This characterization of the spice trade is based on Ron Harris, "Law, Finance and the First Corporations", in James J. Heckman, Robert L. Nelson & Lee Cabatingan, eds., *Global Perspectives on the Rule of Law* ((2009, Routledge).

Seventeenth-century spice voyages and twentieth-century technology start-ups involved up-front investment, high risk, and high return. Many business ventures have these characteristics. Discovering new ideas and developing them usually requires upfront investment. If development succeeds, the innovator has a temporary advantage over competitors until they catch up. While the temporary advantage lasts, the innovator enjoys extraordinary profits, which we call “venture profits.” In the end, the imitators catch up and profits fall to the normal level. The life-cycle of a business venture is this chapter’s subject.

Life-Cycle of a Business Venture

In efficiency economics, the production function and competitive equilibrium are fundamental tools of analysis. In growth economics, the business venture and its life cycle are fundamental tools of analysis.³ A successful business venture has numbers like Figure 2.1. The venture begins with the development of a new idea in period 1, which costs 8. When the product is developed, the innovator has a valuable secret or patent, or perhaps a cluster of secrets or a portfolio of patents (more on that later). After development, the innovation is launched and marketed to buyers. When launched in period 2, the innovation has no competitors, so the innovator is a monopolist who receives a payoff of 7. In period 3, imitators develop competing products that substitute for the innovation, which reduces the innovator’s payoff to 4. In period 4, imitations improve and competition intensifies. Taking competition to its logical extreme in period r , the imitations become perfect in period 4, so the market is perfectly competitive and the innovator’s payoff is zero. (“Zero profit” is economist’s code for “ordinary rate of profit”.)⁴ Summing over the life cycle, the venture’s net payoff equals +3.

³ A production function determines outputs from inputs, such as $y=f(k,l)$ where the output is y and the inputs are capital k and labor l . In a competitive equilibrium, the price of output y equals its marginal social value. So the marginal value product of an input, say capital k , equals pf_k . A business venture at time t in its life-cycle has an expected profit function π_t determined by the probable payoffs as a function of future investment of k and l . As explain in Chapter 1, the social value of alternative business ventures is compared by the contributions to growth. According to the overtaking principle, one business venture that increases the sustained rate of growth relative to another business venture is much more socially valuable, perhaps infinitely so.

⁴ Perfect competition drives the prices of all goods to their cost of production. Profits are zero after including the cost of capital in the other costs of production. The cost of capital equals the ordinary rate of profit in alternative uses.

A successful innovation is like a horserace that pays handsomely to the winner, moderately to the near-winners, and nothing to the rest of the pack. A successful business venture earns extraordinary profits when launched, extraordinary profits when imitated, and ordinary profits when perfectly substituted. The industry eventually settles into equilibrium, like the crowd eventually stops yelling after horses cross the finish line.

Figure 2.1. Life Cycle of a Successful Venture

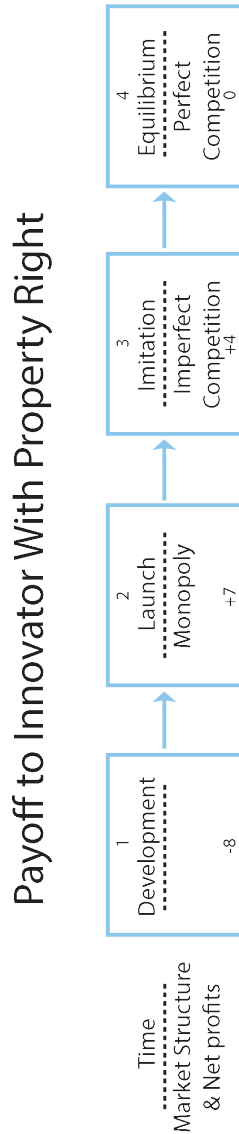


Figure 2.1 indicates the benefits and cost of the venture to its owners. What about its benefits and costs to society? As conventionally measured, the net social benefits from a successful business ventures equal the sum of innovator's profits, profits of other firms such as imitators', and the consumer's surplus.⁵ The innovator's profits are a small fraction of the net social benefits. The richest innovators get less wealth for themselves than the value that their innovations convey to consumers and other producers. Thus the wealth that Apple investors obtained from the iPhone is less than the value of the iPhone to consumers and other firms that imitate it or create applications for it. In the appendix to this chapter, Tables 2.1 and 2.1 use numbers to illustrate the venture depicted in Figure 2.1,

Production of the improved widget will continue beyond time 4 under conditions approximating perfect competition until the product becomes obsolete. The product becomes obsolete when a new innovation destroys the old one's value and the industry begins a new cycle of innovation. When ventures like the one in Figure 2.1 repeat themselves, one innovation follows another, and the path of net social benefits traces a gyre like the falcon on this book's cover. In the appendix, Figure 2.4 traces the gyre of net social benefits from a numerical example.

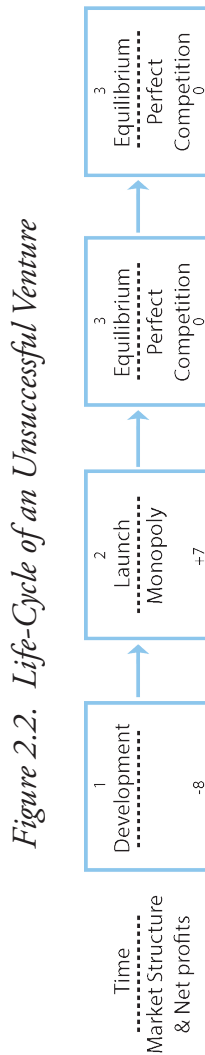
Imitators or Enthusiasts?

Figure 2.1 depicts a profitable venture, but most ventures fail and lose money. Recent U.S. data suggests that 40% of new businesses survive and 60% disappear within four years.⁶ Figure 2.3 depicts a losing venture. The innovator in Figure 2.2 spends 8 in period 1 to develop the product. Many innovations fail before completing development and beginning production, without recouping any development costs. The innovator in Figure 2.3, however, is a little more successful and brings the product to market. When the innovation is launched in period 2, the innovator has no competitors and enjoys profits of 7. The only difference between Figures 2.1 and 2.3 is in pe-

⁵ See appendix to Chapter 1. Recall that the consumer surplus equals the difference between what the good is worth to consumers and what they must pay for it, or the difference between the willingness-to-pay and the market price.

⁶ 44% of U.S. businesses started in the 1990s still existed 4 years afterwards. See Knap, A. E. (2005). "Survival and longevity in the Business Employment Dynamics data." [Bureau of Labor Statistics of the U.S. Department of Labor](#). The number given by Dun and Bradstreet is 37% as cited in "Some of the Reasons Why Business Fail and How to Avoid Them," Entrepreneur Weekly, Issue 36, 3-10-96.

riod 3. In Figure 2.1 the innovation yields a payoff of 4 in period 3 because it is better than the imitations. In Figure 2.2, however, the imitations in period 3 are just as good as the original – they are perfect substitutes. Perhaps the innovator cannot keep a secret or it has no patent. In Figure 2.2, the market reaches a competitive equilibrium in period 2, so the innovator's payoff is 0 in period 2, as well as in period 3. Summing over the life cycle, the venture's net payoff equals -1. The innovator in Figure 2.3 cannot recover the cost of developing the new idea.



The innovator sometimes has the advantage as in Figure 2.1, and the imitator sometimes has the advantage as in Figure 2.2. If the innovator in period 0 can foresee that the venture will have a net payoff of +3 as in Figure 2.1, it will launch the venture. The expectation of positive venture profits causes the development of innovations. If the innovator in period 0 can foresee that the venture will have a net payoff of -1 as in Figure 2.2, it will *not* launch the venture. The expectation of negative venture profits prevents the development of innovations.

The difference between a profitable venture in Figure 2.1 and an unprofitable venture in Figure 2.2 is the ease of imitation. Difficult imitation extends the innovator's period of extraordinary profits in Figure 2.1, and easy imitation reduces the innovator's period of extraordinary profits as in Figure 2.2. To slow imitators, an innovator often tries to keep the innovation secret. Some innovations reduce to explicit information that is easy to copy, like a recipe or a formula. Being easily copied, explicit information is intrinsically hard to keep secret. Thus developers of software need to understand parts of the core code in order to write new software Microsoft Windows. Microsoft will disclose parts of its core code to software developers, but not all of it.

Unlike a computer code, other innovations involve implicit information that is irreducible to simple communication. Implicit information is often imbedded in a practice or organization, like judgment in mixing chemicals, art in baking cakes, or methods to motivate salesmen. Information is implicit when someone knows how to do something that is hard to communicate. Because communication is hard, implicit information is easier to keep secret than explicit information. To steal another company's implicit information, you need to hire its employees rather than readings its documents. (See Chapter .)

The law of trade secrets helps innovators to keep their secrets private. A person who violates these laws by distributing proprietary information is liable for the harm done to the firm with the secret. When firms in Silicon Valley hire employees or discuss collaboration with other firms, they routinely sign non-disclosure agreements (NDAs) in which they promise not to reveal any secrets that they learn about the company. Sometimes trade secrets laws work -- the recipe for Coca Cola has remained a secret for decades. More

often, trade secret law is ineffective. Trade secrets laws are hard to enforce in Silicon Valley and they are unenforced in much of the world.⁷

Instead of secrets, another route to extraordinary profits is patenting. To patent an invention, the innovator must reduce it to explicit information disclosed in the patent application. Any member of the public can read the patent and obtain essential information for producing the invention. However, producing the invention infringes the patent, which is illegal without the consent of the patent holder. The creator of a patentable invention must decide whether secrecy or a patent is more profitable.

A patent applies to an invention, not to a market. However, some patent portfolios industry standards that dominate markets, like the Windows operating system currently dominates desktop computing in business firms. The owner of a portfolio of patents that becomes an industry standard can collect royalties from an entire industry. Standards are a kind of natural monopoly created by a coordination problem, as explained in Chapter 1. With industry standards as with pop stars, many volunteer and few are chosen.

An invention is likely to become a standard if its value for each user increases as the number of its users increases. The uncompensated benefit that one user conveys to another is a “positive externality.” The joint operation of the positive externalities is a “network effect.”

The three sources of extraordinary profits for innovators are secrecy, intellectual property, and natural monopoly. Some innovators enjoy extraordinary profits from all three sources. Thus Microsoft cloaks the core code of Win-

7 Yuval Feldman, *Confidential Know-how Sharing and Trade-secrets Laws: Studying the Interaction Between Legality, Social Norms and Justice Among High-tech Employees in Silicon Valley* (PhD Thesis, Law School, University of California at Berkeley, 2004).

In addition, patents and copyright give innovators a temporary monopoly by law. An effective patent gives the inventor monopoly profits until the patent expires or new inventions undermines its value. To illustrate numerically, consider how A's profits in Figure 1.5 might change if A holds an effective patent to its innovation that lasts for periods 1 and 2. If A refuses to license the innovation to anyone, it enjoys monopoly profits of 10 in time 1 as indicated in Figure 1.5. Assume that the patent extends through time 2 and that competitors make little progress circumventing the patent by their own innovations. In time 2, A enjoys profits much like in time 1, say, profits of 9. Thus the patent increases A's earnings net of development costs for all periods from 6 to 13. Later chapters contain more details about intellectual property, including patents and copyright.

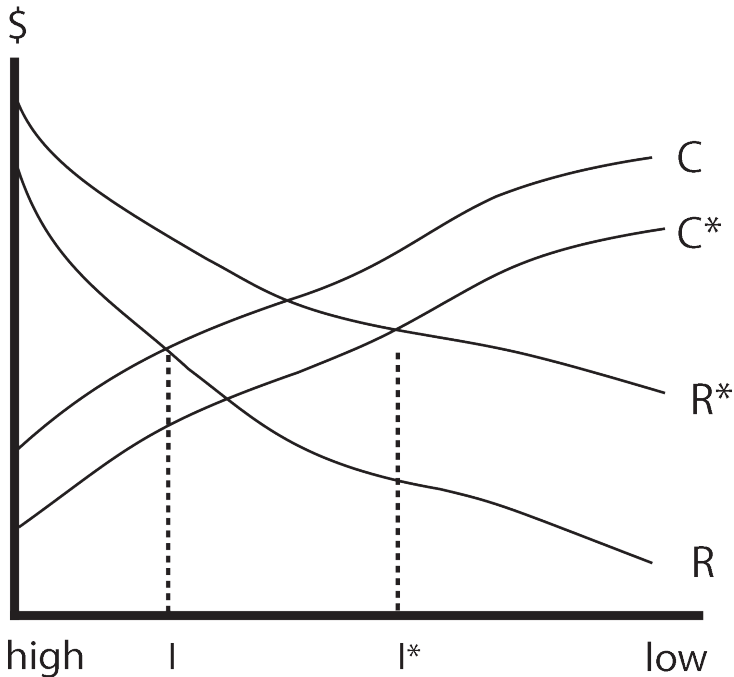
dows in secrecy, protects parts of it by patents and copyright, and enjoys natural monopoly from network effects.

Rate of Innovation

The expectation of positive profits launches innovative ventures, and the expectation of negative prevents innovative ventures. Law affects the profitability of all phases of a venture -- finance, development, marketing, and competition. When better law makes innovations more profitable, the number of innovations increases. To see why, think of an array of new ideas that differ according to the expected profitability of developing them. Innovators develop the ideas that they expect to yield positive profits, and they do not develop the ideas that they expect to yield negative profits. If better law increases the expected profits for innovations, some ventures will tip from negative to positive expected profits, so innovators will develop more ideas.

Figure 2.3 depicts this fact. The horizontal axis indicates an industry's array of possible innovations in order of profitability, with higher profit ventures farther to the left and lower profit ventures farther to the right. Figure 2.3 distinguishes between profits under bad and good law. An innovation's profitability under bad law equals the difference between its revenues R and costs C over the venture's life. To the left of I , revenues R exceed costs C , so the innovations will be developed. To the right of I , costs C exceed revenues R , so the innovations will not be developed. The revenue curve R intersects the cost curve C at point I where venture profits are zero (development costs before launch equal profits after launch). I is the tipping point that indicates the number of developed innovations under bad law.

Figure 2.3. Development of More Innovations with Better Law that Increases Venture Profits



Innovations Arrayed by Profitability From High on Left to Low on Right

Improving the law in this industry increases revenues from R to R^* and decreases costs from C to C^* . With improved law, the revenue curve R^* intersects the cost curve C^* at point I^* . I^* indicates the number of developed innovations under good law. I^* is larger than iI so better law causes higher venture profits, and higher venture profits cause more ventures.

Conclusion

Chapter 1 posed the first question of law and growth economics: “Which laws increase the pace of economic innovation?” Chapter 2 simply answers, “The laws that increase venture profits.” Fleshing out this answer requires the rest of this book. First we must look inside the business venture. The next

chapter explains how law enables individuals with different interests to give their ideas and money to innovative ventures with shared goals.

Appendix

Table 2.1 uses numbers to illustrate the business venture in Figure 2.1. Assume that the innovator develops an innovation that lowers the cost and improves the quality of a particular consumer good – say, a better widget. At time 0, before development of a better widget, widget makers earn zero profit and consumers enjoy a surplus of S_0 . At time 1, development of the improved widget costs the innovator -8. At time 2 when the improved widget is launched, the innovator has lower production costs, so the innovator enjoys extraordinary profits of 7 temporarily. Informed consumers buy the improved widget and enjoy its higher quality, so the consumer's surplus rises to S_0+10 . At time 3, some imitators partially succeed in replicating the innovation. This competition drives the price down, the innovator's profits fall to 4 at time 3, and the imitators enjoy profits of 2. The fall in price causes the consumer's surplus to increase by 30, and the rise in average quality of widgets causes the consumer's surplus to increase by 3. At time 4, competition drives the price down to the cost of production, so the innovator and imitators receive zero profits. The fall in price causes the consumer's surplus to increase by 70, and the rise in average quality causes the consumer's surplus to increase by 10. Summing over the life cycle in Table 2.1, the venture profits are 3 and the consumer's surplus increases by 123. Consequently, the innovator enjoys less than 3% of the innovation's total benefit to society.

Table 2.1. Net Benefits from Business Venture

Phase in Cycle of Innovation					
	time 0 equilibrium	time 1 development	time 2 market power	time 3 imitation	time 4 equilibrium
venture profits	0	-8	+7	+4	0
imitators' profits	+2	...
consumer's surplus	S_0	S_0	S_0+10	S_0+30+3	$S_0+70+10$
total net benefits	S_0	$-8+S_0$	S_0+17	S_0+39	S_0+80

The explanation of the following table shows that these numbers are reasonable.

Summary table of numbers underlying Figures 2.1 and 2.2.

	time or phase				
	perfect competition	development	monopoly	oligopoly (imitation)	perfect competition
price	1	1	1	.97	.93
ventures' production	100	100	100	100	100
ventures' revenues	100	100	100	97	93
ventures' cost	-100	-108	-93	93	-93
venture profits	0	-8	+7	+4	0
initiators' costs96	.93
initiators' profits	2x.01x100	0
consumer's surplus (100 x (value-price))	S_o	S_o	$S_o + .01 \times 100$	$S_o + .03 \times 1000 + .01 \times 300$	$S_o + .07 \times 1000 + .01 \times 1000$
total profits + consumer's surplus	S_o	$-8 + S_o$	$+7 + S_o$	$+6 + S_o + 30 + 3$	$S_o + 70 + 10$

The industry consists of, say, 10 firms and each of them supplies 100 units of the good for total production of 1,000. The unit cost of production under the original technology is 1, the market price of the good is originally 1 under perfect competition, and the consumer's surplus is S_0 . The innovator spends 8 to develop a new production method that reduces the cost to .93. Since the good's price is 1, the innovator earns .7 per item for 100 items in the monopoly stage when the innovation is launched. The innovation also improves the good's quality. 100 consumers know that the innovator's good has higher quality, so their surplus increases by $.1 \times 100$.

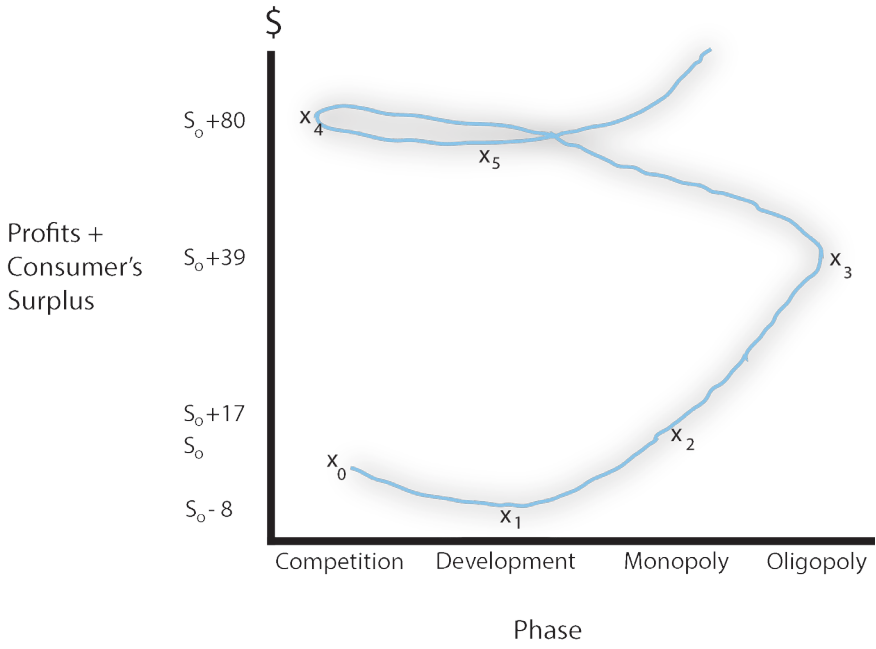
In the oligopoly phase, 2 imitators comprehend enough of the secret to lower their production costs from 1 to .96. Oligopolistic competition bids the market price down to .97. Each of the 2 imitators produces 100 units and earns profits of 1. When the price falls to .97, the increase in the consumer surplus from the fall in price equals $1,000 \times .03$. In addition, the 300 consumers who buy from the innovator or the imitators enjoy an increase in surplus by $.1 \times 300$.

In the final phase of perfect competition, all firms learn the new production technique, the price falls to the cost of production .93, and all firms earn zero profits. Consumers buy 100 units from each of 10 firms. The original consumer surplus is S_0 . When the price falls to .93, the consumer surplus increases by $1,000 \times .07$. In addition, the consumer surplus increases from the higher quality of the good by $1,000 \times .01$.

Notice that these calculations assume fixed market shares and perfectly inelastic demand, which eliminates the "welfare triangles" in a standard efficiency analysis. A more complete analysis that relaxed these assumptions would have larger total net benefits.

The product becomes obsolete when a new innovation destroys the old one's value and the industry begins a new cycle of innovation. The graph of the net social benefits forms a gyre as depicted in Figure 2.4. The first cycle of innovation begins at time 0 when social benefits equal S_0 and ends at time 4 when social benefits equal $S_0 + 80$. The curve beyond time 4 suggests what the next cycle of innovation looks like.

Figure 2.4. Growth Gyre



Chapter 3. The Double Trust Dilemma

Robert Cooter and Aaron Edlin

When someone discovers a better way to make something or something better to make, developing the new idea usually requires capital. One person often has ideas and someone else has money. To launch an innovative venture, they must combine their assets, which require the innovator to trust the financier with his idea and the financier to trust the innovator with her capital. We call the problem of uniting capital and ideas the *double trust dilemma of innovation* -- a new name for an idea with a rich economics pedigree.¹

An economist who worked at a Boston investment bank received a letter that read: “I know how your bank can make \$10 million. If you give me \$1 million, I will tell you.” The letter captures concisely the problem of buying information: The bank does not want to pay for information without first determining its worth, and the innovator fears to disclose valuable information without first getting paid. In general, a person cannot evaluate an idea until after its disclosure, and after its disclosure she may not need to pay for it.² If you know information then you have it, whereas you can know a car, cow, or coffee without having it.

Combining this characteristic of information with investment creates the double trust dilemma. A Berkeley mathematician named Richard Niles invented bibliographic software called EndNote that many professors use on

1 The phrase “double trust dilemma” was introduced by Cooter and Schaefer in Chapter 3, “The Double Trust Dilemma of Development,” *Solomon’s Knot: How Law Can End the Poverty of Nations*. Contemporary (Princeton University Press, 2012). Theories of finance often begin with the question, “How can an investor, who puts his money under the control of a manager, write a contract so that the manager profits most when the investor profits most?” This is the “principal-agent problem.” This is a single-trust problem because the investor must trust the manager with his money, but the manager need not trust the investor. This single-trust problem is the building block for analyzing double trust problems. A good introduction to this vast literature is Kenneth J. Arrow, “The Economics of Agency: An Overview,” in *Principals and Agents: The Structure of Business*, ed. John W. Pratt and Richard J. Zeckhauser, 1985. For a pioneering paper on secrecy and investment, see Edmund W. Kitch, “The Law and Economics of Rights in Valuable Information,” *J. Legal Studies* 9 (1980). Two-sided trust problems have been investigated in game theory, such as two-sided moral hazard or principals-agent with two principals. For an example of how modern financial institutions combine ideas and capital, see Bernard Black and Ronald Gilson, “Does Venture Capital Require an Active Stock Market?,” *11 Journal of Applied Corporate Finance* 36 – 38 (2005).

2 Economists call this fact “Arrow’s paradox of information.” For an early exploration of this problem of asymmetric information, see Arrow, Kenneth J. (1972) *The Value of and Demand for Information*, in C. B. McGuire and R. Radner (eds.), *Decision and Organization*, New York: North-Holland, Chapter 6.

their computers. In the early stage of development, he hoped and feared receiving a call from large firm asking for an explanation of EndNote. Once the large firm understood the product, it might buy Niles' company and make him rich, or it might develop its own version of his program and bankrupt him. Niles eventually got a call from a large publisher (Thompson) who bought the company.

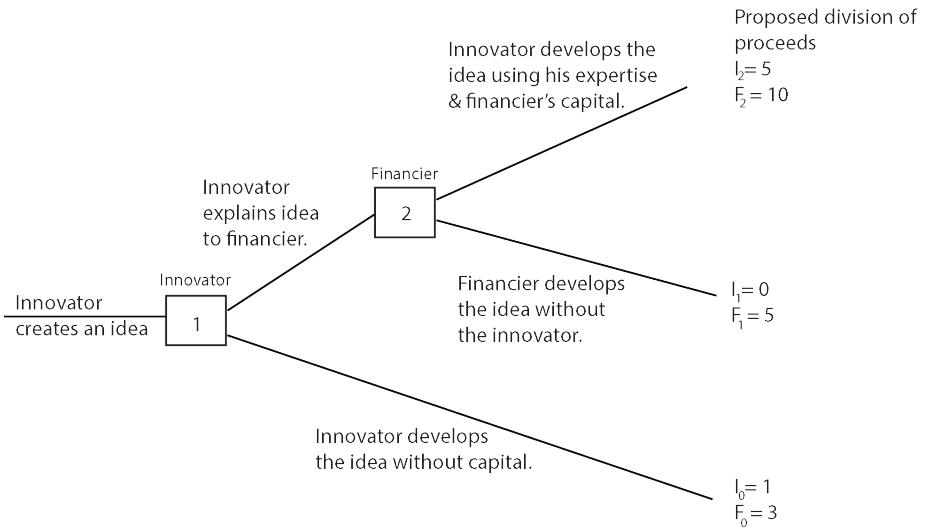
A new idea collides with capital in Silicon Valley when an innovator presents a business plan to potential investors. Here is a simple example of how such a presentation might go:

“I have a new idea. I can use my time to develop it and sell it for \$1 million, or I can explain it to you. I have no effective legal protection for my idea. Once you know it, you can use your capital of \$3 million to develop it yourself and sell it for \$5 million, and I will gain nothing. However, once you know the idea, I think that you will see that we can make a lot more money by working together. If you supply me with \$3 million in capital, I can use my expertise to develop the innovation and sell it for \$15 million. With \$15 million, we can give you \$5 million for not developing it yourself, and we can split the remaining \$10 million.”

Figure 3.1 depicts this proposal as a decision tree. In step 1 the innovator can develop his idea without capital and earn 1, in which case the financier keeps her capital of 3. Alternatively, in step 1 the innovator can explain his idea to the financier and go to step 2. The innovator has no effective legal protection of his idea. In step 2 the financier can use capital of 3 to develop the idea herself and sell it for 5. Alternatively, in step 2 the financier can supply capital to develop the innovation in cooperation with the innovator and they can earn 15, with the innovator getting 5 and the financier getting 10. Each receives what he could get without cooperation plus an equal share of the surplus from cooperation. (This is the “Nash bargaining solution” in game theory.³)

3 The Nash Bargaining Solution is named after its inventor, John Nash, whose genius and insanity inspired the popular Hollywood film, “A Beautiful Mind.” It is an axiomatic account of cooperation. The Nash bargaining solution corresponds to the outcome of a non-cooperative game repeated indefinitely. See Ariel Rubinstein, “Perfect Equilibrium in a Bargaining Model”. *Econometrica* 50 (1): 97–109. doi:10.2307/1912531..

Figure 3.1. Proposed Solution to Double Trust Dilemma



Many business ventures begin with plans resembling this one. The plan in Figure 3.1 *proposes* payoffs of 5 to the innovator and 10 to the financier. To implement this proposal, the innovator and financier must make 5 and 10 into the *actual* payoffs from cooperating. Successful implementation implies that the innovator and financier will get the payoffs in Figure 2.1 if they make the decisions depicted in the figure. If the plan can be implemented, self-interest will motivate the parties to cooperate as envisioned in the proposals. The financier will decide in stage 2 to supply capital to the innovator to develop the idea, because the payoff from cooperation exceeds the payoff from non-cooperation: $F_2 > F_1$. Foreseeing this fact, the innovator will decide to cooperate in stage 1, because the payoff from explaining his idea exceeds the payoff from developing it on his own: $I_2 > I_0$. These two inequalities express the conditions for cooperating in a venture structured like Figure 3.1. Successful implementation of the plan satisfies these two conditions.

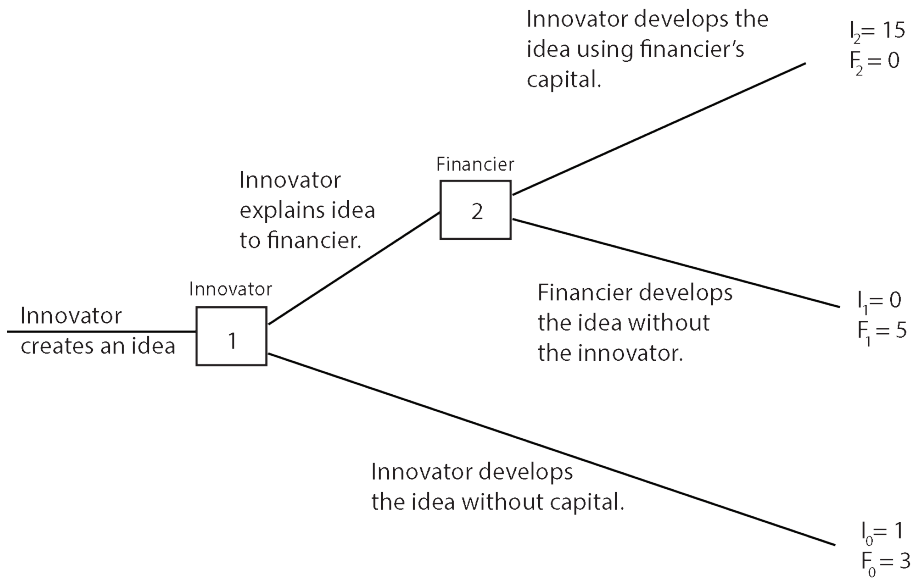
Implementing such a plan is remarkably difficult. Investing in a business venture requires confidence that strangers will not appropriate its profits. The innovator and financier will need protection against outsiders who might steal 15 from them. All ventures require protection from predators such as hoodlums, mafias, cheating accountants, Ponzi artists, conniving state regulators, oppressive tax collectors, and thieving politicians. Without effective property protection, people fear theft, so resources flow to protectors, not to

entrepreneurs. Families, clans, and gangs protected property historically, and they continue to do so today in some countries. However, an effective state is more reliable than private protectors. State protection of property is the legal foundation for investment in the future, including business ventures by families and friends.

If the innovator and financier are protected against outsiders, they still need to trust each other. At the final stage when the firm's has 15, the plan calls for the financier to receive 10 and the innovator to receive 5. What prevents the innovator from grabbing all 15? Perhaps the innovator will take all 15 as salary and leave nothing for paying dividends to shareholders. Many business ventures never launch because the financier cannot trust the innovator to distribute the profits as promised.

If the final stage of Figure 3.1 cannot be implemented, then the whole proposal collapses. Assume that if venture were to reach the final stage, the innovator would grab all 15, as depicted in Figure 3.2. Foreseeing this fact, if the venture reaches stage 2, the financier will develop the innovation herself, so the financier's payoff will be 5 and the innovator's payoff will be 0. Foreseeing this fact, in stage 1 the innovator will develop the idea himself and receive a payoff of 1, instead of explaining the idea to the financier. The payoffs in Figure 3.2 fail to implement the plan in Figure 3.1 for sharing the gains from cooperation, so cooperation unwinds and noncooperation stifles an innovative business venture. (Notice that the payoffs in Figure 3.2 violate the two conditions for cooperation: $F_2 \geq F_1$ and $I_2 \geq I_0$.)

Figure 3.2. *Unsolved Double Trust Dilemma*



Relationships

How can the innovator and financier implement a business plan as depicted in Figure 3.1, rather than collapsing as depicted in Figure 3.2? One way is to deal with trusted relatives. Loyalty within families is partly irrational and partly rational. Here's how rationality prompts family members to cooperate with each other. Kinship is ascribed, not chosen. Your uncle is your uncle until death. Ascription provides a framework for repeat dealings. Through repeat dealings, people come to know whom to trust and whom to distrust. One reason to be trustworthy is fact that your relatives will find out if you are not trustworthy.

The logic of relying on relatives is depicted in Figure 3.3 by repeating the game in Figure 3.2 indefinitely many times. Each repetition is called a "round" and the rounds are numbered consecutively. In each round, the innovator can cooperate or appropriate. The financier (possibly a rich aunt) decides to reward cooperation and punish appropriation. This strategy by the financier gives the payoffs to the innovator as depicted in Figure 3.3. If the innovator cooperates in round t , then the financier cooperates with him

in the next round, and likewise for subsequent rounds.⁴ With cooperation, the innovator receives 5 in each round of the game. Alternatively, if the innovator appropriates 15 in any round, then the financier refuses future dealings with him, so the innovator will have to develop his ideas without capital in each subsequent round and earn 1.

Figure 3.3. Innovator's Payoffs in Repeated Game

	round	t	t+1	t+2	t+3	t+...
innovator's choice	cooperate	5	5	5	5	5
	appropriate	15	1	1	1	1

Compare the payoffs from the two strategies. In round t, “appropriate” pays the innovator 15 and the innovator gains 1 in each subsequent round. In contrast, “cooperate” pays 5 in every round. Summing payoffs from round t to round t+3, “cooperate” pays 20 and “appropriate” pays 19. Thus “cooperate” overtakes “appropriate” in round t+3, and the gap widens with each subsequent round. A rational player who values future payoffs will cooperate and not appropriate.⁵ Conversely, a rational player who highly discounts future payoffs will appropriate rather than cooperate. Specifically, the innovator in Figure 3.3 will cooperate who discounts payoffs by less than 40% per round, and the innovator will appropriate who discounts payoffs by more than 40% per round.⁶

4 The financier's strategy is called “grim” or “grim trigger.” Once the innovator cheats the financier treats him as untrustworthy forever. Another possible strategy with similar implications for cooperation is a variant of “tit-for-tat, in which the financier punishes an innovator for cheating and then resumes cooperating. The payoff matrix looks like this:

round	t	t+1	t+2	t+3	t+3	t+5	t+6	t+...
appropriate	15	1	1	1	15	1	1	1

Experimental evidence indicates that tit for tat comes close to maximizing the party's payoffs in a variety of circumstances.

5 To be precise, the discounted sum across rounds is the game's “present value” to the player. The present value of “cooperate” exceeds the present value of “appropriate” if the game repeats itself enough times and the discount.

6 Cooperation yields $5 + 5\{\sum_1^\infty (\frac{1}{1+n})^i\}$

and appropriation yields $15 + 1\{\sum_1^\infty (\frac{1}{1+n})^i\}$

Setting them equal yields $5 + 5 = 15 + 5\{1/n\}$, whose solution is $n=4$.

Private Law

Family members can cooperate even if law cannot effectively enforce their promises to each other. However, entrepreneurs do not have enough family and friends to conduct business exclusively with them on the most profitable scale. Successful entrepreneurs look beyond relationships to strangers for money and ideas. Whereas kinship endures, strangers come and go. In business, friendships are often chosen for particular needs, not ascribed. The possibility that a business relationship will end poses an obstacle to cooperation.

In Figure 3.2, cooperation continues forever unless the innovator behaves badly and appropriates in one round. Even without bad behavior, the possibility that a relationship will end on its own can prevent cooperation. We already explained that solving the problem of cooperation by infinite repetition requires the actors not to discount future payoffs too much. If the game is likely to end, the players may discount future payoffs for uncertainty too much to cooperate in the present. Thus when communism began to collapse in Russia around 1990, production plummeted partly because producers could not cooperate when they lost confidence in future dealings with each other.

Whereas a possible end to a relationship increases the problem of cooperation, a definite end to a relationship makes cooperation even more problematic. Instead of assuming that the game in Figure 3.3 is repeated indefinitely, assume that $t+3$ is the game's last round in Figure 3.3. As before, the financier adopts the strategy of rewarding cooperation and punishing appropriation, but the financier has no opportunity to punish appropriation that occurs in the last round. Therefore, the rational innovator will appropriate in round $t+3$ without fear of retaliation. "Cooperate until the last round and then appropriate" pays the innovator more than "always cooperate," as shown in Figure 3.4.

Figure 3.4. Innovator's Payoffs in Repeated Game

Where Financier Cooperates Until the Innovator Appropriates

round	t	t+1	t+2	t+3
always cooperate	5	5	5	5
cooperate until the last round and then appropriate	5	5	5	15

The payoffs in Figure 3.4, however, do not reflect what each player will do if he foresees what the other player will do (“Nash equilibrium”). If the financier foresees that the innovator will appropriate in round $t+3$, she will change her strategy and not cooperate in round $t+3$. If the innovator foresees the financier’s behavior in round $t+3$, he will change his strategy and appropriate in round $t+2$. If the financier foresees the innovator’s behavior in round $t+2$, she will change her strategy and not cooperate in round $t+2$. By repeating this reasoning, cooperation unwinds all the way to the game’s beginning.⁷ With a definite end in sight, repeating the game provides no basis for cooperation. Every round looks like Figure 3.2, where the innovator and financier do not launch a joint venture. This is the “end-game” problem.

In a game with a definite end, cooperation may unwind back to the beginning. In a game with a likely end, cooperation may stop if uncertainty causes high discounting of future payoffs. Unlike relatives, strangers have more difficulty solving the problem of non-cooperation by repeating the double trust dilemma because they foresee an end to their relationship, or they discount future payoffs by the probability that the relationship will end. If the innovator and financier cannot cooperate, the business venture does not launch, the innovation is undeveloped, and the economy does not grow.

Where repetition fails, private law may succeed in solving the double trust dilemma. Here is one legal implementation often used in business ventures like Figure 3.1. The innovator and financier form a corporation to develop the idea. The financier loans the firm 3 for development (debt), and the innovator contributes his expertise to the firm. The firm issues 7 shares (equity) to the innovator and 5 shares (equity) to the financier. Shares are claims on profits left over after paying the firm’s debt. To implement the proposal, the law must effectively enforce the firm’s repayment of debt and an equal

⁷ Experimental data shows that people are not so rational. In a repeated game with a definite end, experiments often find that players cooperate in the beginning and they stop cooperating near the end. Apparently people are partly logical and partly psychological.

distribution of profits per share. If law is effective, the corporation earns 15 from developing the innovation, repays its debts of 3 to the financier, and distributes dividends of 5 to the innovator and 7 to the financier, as depicted in Figure 3.1.

This implementation requires effective contract law to enforce repayment of the firm's debts and effective corporate law to assure that profits get distributed equally to shares. The framework for finance by a small group of investors mostly comes from the law of contracts that belongs to "private law," and the law of business organizations that belongs to "business law." Ineffective contract or corporate law prompts chicanery, not cooperation.

Here is one way that weak corporate law could allow the innovator to appropriate all profits. Assume that the innovator is the corporation's chief executive officer (CEO) who controls the company without effective legal constraints. After the firm earns 15, the innovator/CEO repays the firm's debts of 3 to the financier, and then the innovator/CEO pays himself a salary of 12. Since all 15 go to pay debts and salary, no profits remain to distribute as dividends to shareholders. The financier gets nothing beyond the repayment of her debt of 3, and the innovator gets everything else. If the financier foresees this ploy by the entrepreneur, cooperation will unwind.⁸ (Similarly, another ploy transfers all profits to the financier instead of the innovator, which also causes cooperation to unwind.⁹)

Public Law

In private finance of a business venture, the innovator and financier negotiate with each other over price and non-price terms and then use contracts and corporate law to implement their plans. In contrast, traders in public markets like the New York Stock exchange buy and sell stocks and bonds at

8 In stage 2, the financier foresees that lending capital of 3 for the firm to develop the innovation will result in repayment of the loan and nothing more for her. Foreseeing this fact, the financier will choose to develop the idea without the innovator in stage 2, which pays the financier 5 and the innovator 0. Foreseeing this fact, the innovator will choose to develop the idea at stage 1 and earn 1, rather than explaining the idea to the financier and earning 0. Unless law can prevent the innovator's chicanery at the end of stage 2, cooperation will unwind.

9 Assume that the financier is the Chairman of the Board who controls the company without effective legal constraints. After the firm develops the idea, the financier/chairman fires the innovator/CEO and sells the innovation for 15 to another firm that she owns. The selling firm uses 3 to repay its debts, and the buying firm distributes its profits of 12 to the financier/owner.

prevailing prices, without bargaining or negotiating. The stock exchange approximates the economist's ideal of perfectly competitive markets. In perfectly competitive markets, the prices are public (the exchanges post the prices at which stocks are selling), everyone pays the same price, and the goods are standardized (one common share of Exxon is the same as another).

In developing innovations, relational and private transactions play the primary role, and public markets for stocks and bonds play a supplementary role. Supplementary by public finance especially occurs as a business venture matures. Thus assume that the innovator and financier in Figure 3.1 reach a private agreement to develop the innovation. To enjoy extraordinary profits, they must keep many details of their venture secret from competitors. Developing the innovation in stage 2 creates a temporary advantage over competitors, which might come from patenting the innovation or marketing it first. Patenting or marketing reveals the innovation to the public. At the venture's end, the firm gets 15 by selling its stock in an "initial public offering" or by allowing a public company to acquire it.

The ability to sell securities to the public presupposes compliance with securities regulations that ideally protect buyers from chicanery. Members of the public who buy stocks or bonds have no direct control over how the firm uses their money. Instead, the firm's managers and board of directors control it. Insiders have many opportunities to appropriate outsiders' investments. For example, insiders may use accounting tricks to convert profits into salaries, thus depriving stockholders of their dividends. Protecting outsiders from insiders in public companies requires more than securing property and enforcing contracts. For public finance, the additional protection comes especially from corporate law and the law of securities. These laws ideally assure that outside investors get their share of a firm's profits, rather than insiders appropriating all of it. Conversely, if these laws are ineffective, insiders will appropriate all of a firm's profits, and outsiders who foresee this fact will be unwilling to invest.

If the firm gets 15 by selling itself to another firm or by selling its stock in an initial public offering, the founders often take the cash from the initial public offering and exit the business soon thereafter. A timely exit by the founders makes a business venture more profitable, which increases innovative activ-

ity.¹⁰ Profits increase because innovating and producing commodities are different specialties requiring different talents. By exiting, the founders leave others to produce commodities while they pursue the next innovation.

As explained in Chapter 2, a successful business venture begins with a disequilibrium caused by developing a novel idea. To gain a decisive advantage over competitors, development requires secrecy in order, which precludes public markets. After development builds a decisive advantage, the need for secrecy diminishes. The innovator has fewer secrets to hide and the double trust dilemma attenuates. In these circumstances, the firm may conduct an initial public offering of its stock or sell itself to a public company. Public finance approaches the economist's ideal of a perfectly competitive equilibrium, where no one has valuable private information and everyone earns the same profit rate ("ordinary rate of return"). Like contentment, a perfectly competitive equilibrium is approached and never reached. Perfect competition is the end of an innovation, its irresistible and final fate, not the process of its creation.

Three Stages of Finance in Silicon Valley¹¹

Relational transactions require effective property law, private transactions require effective contract law and corporate law, and public transactions require securities regulations. The three stages of finance in Silicon Valley illustrate how these three types of laws support innovation. Relationships are especially important for many startup firms in Silicon Valley. First, someone has a new idea and obtains capital to develop it. At this point, the innovation's economic value has not been established. To establish its value, the innovator needs funds from a few "angel investors." Family and friends have confidence in the innovator invest, so they invest without fully understanding the innovation's market value. Perhaps a few people invest who think that they can evaluate the innovation without understanding it. According to a popular quip, initial funding comes from "the 3 Fs": Family, friends, and fools.

10 For a recent application, see Bernard Black and Ronald Gilson, "Does Venture Capital Require an Active Stock Market?" *Journal of Applied Corporate Finance* 11 (winter 1999): 35–38.

11 The three stages of finance in Silicon Valley are distinguished and related to stages in the development of financial markets in The phrase "double trust dilemma" was introduced by Cooter and Schaefer in Chapter 3, "The Double Trust Dilemma of Development," [Solomon's Knot: How Law Can End the Poverty of Nations](#). Contemporary (Princeton University Press, 2012).

In the first stage, the entrepreneur develops the innovation sufficiently to prove its value to knowledgeable strangers. Most innovators, however, have too few personal relationships with wealthy people to finance an innovation's full development. They must eventually turn to financiers. In Silicon Valley, the second stage of funding comes from "venture capitalists." They are experts in evaluating undeveloped innovations, but they are not family, friends, or fools. Venture capital is private finance because it comes from a small group of investors with information unknown to a broad public. Most countries impose different laws on finance by a small group of investors than on finance by many investors.

Innovators and entrepreneurs have good reasons to distrust each other. Silicon Valley innovators sometimes expropriate the investments of their financiers. John P. Rogers convinced some prominent California investors to give him \$330 million for a high-tech start-up named Pay By Touch that would "transform how America pays its bills" by using "biometric authentication technology" (e.g., fingerprints). In 2008 the company went bankrupt, and investors contend in lawsuits that Rogers burned through \$8 million per month without producing anything of value.¹²

Similarly, the creative people who found a company often manage it badly. When the founders prove to be bad managers, the venture capitalists must replace them with good managers. In these circumstances, the venture capitalists seize the firm to increase its profitability. Alternatively, where the founders prove to be competent managers, venture capitalists may seize the firm to avoid sharing profits with the founders. Venture capitalists sometimes want to remove *good* managers with large claims to the firm's future profits. The initials "v.c." stand for "venture capitalists" and also "vulture capitalists."

Innovators and venture capitalists use various legal devices to overcome their mutual distrust. The founders of the firm commit to performance goals implicitly or explicitly. If they fail to meet their goals, they may lose their investments and their jobs. Specifically, the venture capitalists often hold preferred shares of stock, unlike the common shares held by the founders. Preferred shareholders may have special powers of governance that common shareholders lack, as well as other advantages. The financing contract may say

¹² Lance Williams, "How a 'Visionary' Raised—and Lost—a Fortune," San Francisco Chronicle, December 7, 2008, available at <http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2008/12/07/MNIK137QU3.DTL>.

that preferred shareholders can demand repayment of their investment after three years. Such a contract reassures the venture capitalists that the founders will do their utmost to perform as promised. The contract also reassures the founders that the venture capitalists have an interest in keeping the firm's secrets.¹³

Corporate governance provides another device to solve the double trust problem in Silicon Valley. The firm's bylaws may stipulate that common shareholders (founders) and preferred shareholders (venture capitalists) appoint an equal number of directors to the company's board, plus an independent director accepted by both sides. If the founders and venture capitalists disagree, the independent director holds the decisive vote. Thus the independent director will decide whether or not the venture capitalists can replace the founders with new management.

In the third stage of finance in Silicon Valley, a successful start-up sells itself to the public, either directly through an initial public offering of its stock, or indirectly when a publicly traded company acquires it. In order to sell stock to the public in the United States, a firm must comply with disclosure rules of the Securities Exchange Commission. Brokers disseminate the firm's disclosed information to potential investors. After disclosures, many people understand the innovation sufficiently to decide whether or not to invest in its further development. The investors in stock markets are a large group of people – the “public.” Regulatory law, which belongs to “public law,” especially controls finance by a large group of investors.

Small and Large Firms

The double trust dilemma afflicts individual innovators and financiers. Does it also afflict large firms? In a large firm, the innovators and financiers are employed by one organization. They proceed something like this. Executives sift through proposals from the research divisions, choose the most

13 Instead of issuing preferred shares to venture capitalists, firms could create much the same result by selling them bonds and common shares. However, preferred shares have a tax advantage. The valuation of common shares affects the tax liability of managers and employees who are partly paid in common shares or options to buy them. Since preferred shares are different from common shares, selling preferred shares to venture capitalists does not necessarily affect the price at which the firm can value its common shares for tax purposes. The firm can raise money at the relatively high of preferred shares while claiming that common shares are worth much less.

promising ones, and order the finance division to pay the research division to develop the innovations. Perhaps you think that hierarchical authority solves the problems of trust that afflict negotiations between individual innovators and financiers.

If that is what you think, consider the following. When an employee has a new idea, he understands its worth better than others. If he gives a valuable idea to his employer, he may get a bonus and the firm may get richer. A typical employee, however, would rather enrich himself than his firm. Instead of giving the idea to his firm, the employee may quit, form a startup company, develop the idea on his own, and try to get rich. Various contracts and laws restrain employees from quitting firms and taking creative ideas with them.¹⁴ The more large firms can claim legal ownership of employees' discoveries, the less people who are creative will work for them. Large firms often cannot prevent employees from quitting and taking creative ideas with them, and large firms often cannot hire the most creative innovators to work for them. That is why many innovations in Silicon Valley begin in small, startup firms, where innovators own a significant share of the firm.

The double trust dilemma persists in large firms. The employee in the research division asks, "Can I trust the firm to reward me if I give it my innovative idea?" The employee in the finance division asks, "Can I trust the research division to develop this innovative idea as planned?" Each one fears that the other party is withholding crucial information. Perhaps the innovator-employee keeps secret a potentially fatal obstacle to development, or perhaps the financier-employee keeps secret his intention not to pay a bonus to the innovator-employee.

The problem of trust favors innovation in small firms. Why do large firms innovate? The amount of money required to develop an innovation often exceeds the innovator's wealth. The innovator is reluctant to risk all his wealth on an innovation with an uncertain future. In a startup firm, the innovator shifts much of the financial risk to the financier. If they form a business venture with the innovator in charge, the innovator risks another's money, not his own. This fact creates incentive problems that we discuss later. To improve incentives, the innovator usually retains substantial risk of failure.

¹⁴ See Chapter __ trade secrets law and Chapter __ on non-competition clauses in employment contracts.

Compared to a startup firm, the innovator-employee in a large firm shifts almost all of the risk of financial failure to the employer.

Creative people differ in their appetite for risk taking. Creative people who are risk-lover form startup firms, and creative people who are risk-averse find employment in the research division of large firms. The main advantage of developing innovations in large firms is risk spreading. Different tastes for risk contribute to the distribution of innovations by firm size.

Ontogeny Recapitulates Phylogeny

Firms in Silicon Valley often pass through all three stages of finance – relational, private, and public. Similarly, the three stages of finance for a start-up firm in Silicon Valley resemble three stages of historical evolution in capital markets for countries. The industrial revolution in England, which was the world's first, went through these stages. In the early eighteenth century, inventors mostly relied on their personal assets and loans from family and friends (relational finance). As industrialization proceeded, loans from wealthy investors and banks became available more readily to new industries. Finance of industrial companies by sales of stocks and bonds to the general public came later. Public financing of industrial companies originally concerned infrastructure like canals, docks, and railways, where private business and the state intertwine. As the law became more reliable, public finance spread to manufacturing firms.¹⁵

Like 18th century England, the poorest countries today have weak capital markets, so businessmen mostly borrow from family and friends. Starting from a condition of lawlessness, imposition of secure property rights can cause a spurt of growth based mostly on relational finance, as in China's new industries after the 1980s. Some peoples, notably the Chinese and the Jews, have family networks that extend business relationships beyond the usual boundaries. However, the conditions of trust among relatives do not reach the scale of modern businesses. Relational finance keeps business small and local. No modern country became wealthy by relying exclusively on relational finance.

¹⁵ See P. M. Deane, *The First Industrial Revolution* (Cambridge: Cambridge University Press, 1955), summary on 155–157. For financing of key eighteenth-century inventions by name, see 155.

To increase the scale of business, an economy must augment relational finance with private finance, especially bank loans. In countries where banks dominate, an elite of wealthy insiders often lend to business ventures based on private information. Thus bank finance in some developing countries performs a similar role to venture finance in Silicon Valley. As countries become affluent, they increasingly augment private finance with public finance, which means selling stocks and bonds to the general public. Stocks and bonds compete with banks and wealthy individuals to finance economic growth.

Compared to debt, equity financing requires a superior legal framework. Debt involves a fixed schedule of repayment specified in a contract, whereas equity involves sharing future profits. Enforcing a fixed schedule of repayment is easier for courts than enforcing profit sharing. Consequently, loans and bonds often flourish in countries where stock markets languish. When weak law shrinks local stock markets, businesses are deprived of equity capital needed to launch innovative ventures.

Biologists sometimes say, “Ontogeny recapitulates phylogeny,” which means that a single organism’s development from conception to maturity resembles an entire specie’s evolution^{16,17}. The same applies to solving the double trust dilemma. Startup companies in Silicon Valley and all companies in lawless countries tend to rely on relationships more than formal law to solve the double trust dilemma. With better law, finance expands from relational to private, and from private to public. The expansion supplements earlier forms without replacing them. All three forms of finance—relational, private and public—remain important in the richest countries. The extent of public finance varies significantly among countries, including rich countries. Japan and northern Italy have achieved affluence mostly through relational and private finance, with relatively little public finance, whereas the United

16 See Chapter 8, pages 115-120 of Cooter and Schaefer, “The Double Trust Dilemma of Development,” *Solomon’s Knot: How Law Can End the Poverty of Nations*. Contemporary (Princeton University Press, 2012)

17 Each animal begins life as a single cell containing genetic instructions for how to grow into a complex organism. For animals in different species with a common evolutionary ancestor, the path of individual growth suggests the older, evolutionary forms found in the fossil record. While the pattern of individual growth does not strictly recapitulate the evolution of the species, comparing them provides useful clues about the genes that control the development of individuals and species. For a book that finds the origin of the human fetus in fish, see Neil Shubin, *Your Inner Fish: A Journey into the 3.5-Billion-Year History of the Human Body* (New York: Pantheon Books, 2008).

States and Great Britain rely mostly on public finance for mature industries. Germany appears to be shifting from the former to the latter.¹⁸

Expanding the basis of finance requires effective law that controls behavior, not aspirational law that expresses lofty ideals. What makes a law effective? Not just writing it down. Written law in a poor country often resembles written law in a rich country. Property and contract law-on-the-books in India and Nigeria resemble English common law, and property and contract law-on-the-books in Peru resemble the Spanish civil code. Writing down a law, however, does not make it effective. The written laws are less effective in Nigeria or Peru than in England or Spain.

Conclusion

New business ventures begin with secrecy, risk, and high profit expectations. All three decrease as the venture matures. Sea routes from Europe to Asia were eventually mapped and secured, trade between them became commonplace, and middle-class Europeans could buy spices. In Silicon Valley, competitors work around patents and ferret out secrets, thus converting today's technological breakthroughs into tomorrow's commodities. However, an innovative economy never settles into a permanent condition without secrecy, risk, or extraordinary profits. For economic growth, new business ventures must repeatedly confront and solve the double trust dilemma. To solve the double trust dilemma, entrepreneurs, financiers, courts, and regulators have worked out contracts, laws, and institutions over years and centuries. One cannot understand these creations without studying them as legal scholars do. That is how law dispels some mysteries of growth that baffle economists.

Where law is weak, innovators and financiers cannot rely on formal law for much more than protecting property from predators. They must deal mostly through relationships. This is true for most business in some undeveloped countries, and it is true for startup firms in developed countries where no law can protect fragile secrets and implicit financial agreements. In developing countries and startup firms, good law cannot do all of the work of creating trust. When the effective law of contracts and corporations improves, finance expands from relational to private. This is true in undeveloped countries that improve their legal institutions, and it is true in startup firms after

18 For details on finance in different countries, see...

they can demonstrate an innovation's promise. When institutions improve further in underdeveloped countries, or when a startup firm develops an innovation sufficiently to acquire a decisive advantage against competitors, more effective securities laws allows, finance to expands from private to public. In general, improved law allows more ideas to combine with more capital so the economy grows faster.

Chapter 4. Separation

Robert Cooter and Aaron Edlin

When a successful business venture develops an innovation, its launch yields high profits temporarily, then profits fall towards zero as the innovation disseminates and competition intensifies. Over its lifetime, the venture's profits equal a fraction of the innovation's social value – usually a small fraction – and the rest goes to other firms as profits and to consumers as “surplus.”¹ Since the social value of an innovation far exceeds the innovator's profits, venture profits provide deficient incentives to innovate.

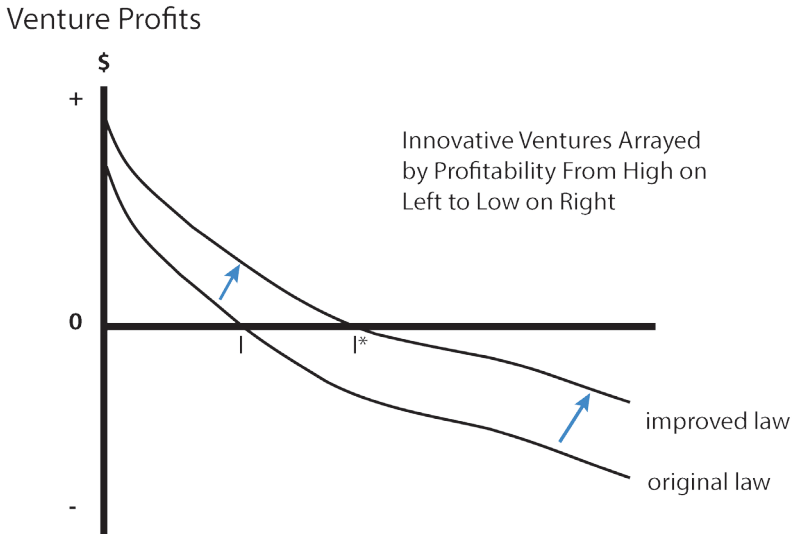
To increase innovation, law and policy should increase its profitability. Figure 4.1 expresses this idea. (It is essentially the same as Figure 2.4 in Chapter 2.). The vertical axis represents venture profits, which equal the discounted present value of the stream of revenues from the innovation's sales, minus the discounted present value of its development and production costs. The horizontal axis arrays ventures by profitability from high on the left to low on the right. With open competition, investors will finance ventures until profits fall to zero. Development stops when the present discounted value of revenues equals the present discounted value of development and production costs.

Figure 4.1 contrasts venture profits under two different legal regimes. Under the original law, venture profits reach 0 at innovation I. Improved law changes the situation by shifting the venture profit curve up as indicated by the arrows, so venture profits reach 0 at innovation I*. Thus improved law increases the number of innovations that get developed from I to I*. The figure depicts the increase in one period of time. If better law increases innovations in every future period of time, then the sustained growth rate increases. (Figure 4.1 depicts one of several ways that improved law increases growth by increasing the profitability of innovation.²)

1 By definition, the consumer's surplus equals the difference between the price that a consumer would be willing to pay for a good and the price that he actually pays.

2 Figure 4.1 does not depict the speed with which an innovation is developed. In a race to innovate, the expectation of higher profits may cause an entrepreneur to develop an innovation faster. Faster development affects infra-marginal innovations, not just the marginal innovation I*.

Figure 4.1. Laws that Increase Venture Profits Cause More Innovations



How much should law ideally increase the profits of innovative ventures? To maximize social welfare, according to the overtaking principle, law and policy should increase venture profits as much as possible. You might doubt the overtaking principle or the conditions that make it true, but when you consider the historical record you cannot doubt that sustained growth in a comprehensive measure of consumption dramatically improves social welfare. Nor can you doubt that faster sustained growth would increase social welfare faster. In any case, this book is playing the intellectual game of law and growth economics whose objective is to maximize sustained growth.

Paradox of Growth

Law and policy rests on a paradox, like “This statement is false.” The paradox of growth is that rapid innovation requires competition for extraordinary profits. This is a paradox because “perfect competition” in economic theory implies that all firms earn ordinary profits. If some firms earn extraordinary profits, competition must be imperfect.

How can law and policy combine intense competition and extraordinary profits for innovators? Patent law illustrates the solution. Chapter 2 distinguished a venture’s phases into development, launch, and imitation. Patent

law opens competition to develop patentable innovations, closes competition when development succeeds and the inventor gets a patent, and restores competition eventually when the patent expires.³ The formula for maximizing growth is open competition to innovate plus temporary market power for successful innovators.

With open competition to innovate, everyone is free to launch a venture and develop an innovation. Firms compete to launch business ventures and earn extraordinary profits. The development of innovations is very risky, and entrepreneurs take the risk in the hope of earning extraordinary profits. Monopoly is more profitable than competition. With higher profits, more ventures launch. Ventures launch until the marginal venture expects to earn zero profits, as depicted in Figure 4.1. Market power for innovators causes more competition to innovate.

Subsequent chapters in this book apply this formula to different bodies of law. Besides patents, many other kinds of law affect a business venture – property, contracts, torts, corporations, finance, regulations, antitrust, and taxation, to name a few. They should have the same aim as patents and use the same formula to combine competition and market power. The aim is to maximize the profitability of ventures in open competition to develop innovations. The gain from growth in the long run will overtake the loss from higher prices in the short run.

This is not the conventional approach in the economic analysis of law. The conventional approach first observes that a venture incurs losses in development, so overall profitability requires extra profits after development. How much extra? The usual words are “sufficient,” “adequate,” “cost-recovery,” or “break-even.” Second, the conventional approach notes that market power for a firm increases the price of its product to consumers. The words are “excessive,” “inefficient,” or “distorting.” Third, the conventional logic concludes that law affecting market power should balance the gain from faster innovation and the loss from higher prices for consumers. The words are “tradeoff” and “optimal.” Unlike the conventional approach, our words are “maximize innovation.”

³ A patent cycle begins with competition for a market, and ends with competition in a market. The phrase “competition for a market” was used by Bill Baxter to analyze antitrust law.

Producer's Market Power

In microeconomics, perfect competition and monopoly are extremes on a continuum of power to set a good's price. At the extreme of perfect competition, each seller must *take* the market price, not *set* a price. No seller can charge more than the market price because each buyer has so many alternative sellers. The demand curve is perfectly elastic. Furthermore, competition among sellers drives the market price down to the cost of production, so each seller receives ordinary profits. Moving along the continuum, as market structure becomes less competitive, a seller can set a price. Each buyer has imperfect alternatives among the sellers. The seller who charges a higher price will make fewer sales, not zero sales as in perfect competition. The demand curve is imperfectly elastic. Moving farther along the continuum, as the seller's market power increases, buyers have fewer choices. At the extreme of monopoly, there is only one seller. Everyone who wants the good must buy from the monopolist. The monopolist sets the price above the cost of production and earns extraordinary profits.

When a firm uses market power to increase its price, part of the buyers' losses transfers to the seller as higher profits, and part of the buyers' losses does not transfer to the seller. "Deadweight" describes a loss without an offsetting gain.⁴ To illustrate numerically, assume that a seller earns 0 profits under perfect competition and 100 under monopoly. A change from competition to monopoly transfers 100 from buyers to the seller. According to standard economics, the seller gains less from monopoly than the buyers lose. Thus if monopoly causes the seller to gain 100, buyers may lose, say, 130. The difference of 30 is the deadweight loss.⁵ (Behind this problem of monopoly lies a deeper theoretical problem of transaction costs.⁶) Firms with market

4 The deadweight loss occurs because the higher price causes buyers to purchase less of the good. When buyers forego some purchases, the buyers lose the gains that they would have enjoyed from these foregone goods, and the seller gains nothing from goods that buyers do not buy.

5 The standard analysis is somewhat more complicated than these remarks suggest because of the difference between long run and short run effects. In the long run, all of the deadweight loss falls on buyers, but in the short run sellers can also bear part of the deadweight loss. Note that in the standard graphical analysis, the transfer is a rectangle and the deadweight burden is a triangle.

6 If there were no obstacles to bargaining, the monopolist and the buyers could write contracts that eliminate the deadweight loss. The contracts would charge according to a price schedule, not a single price. The schedule discriminates according to how many units of the good the buyer buys, charging a higher price for the first unit and a lower price for the last unit. By definition, a perfectly discriminating monopoly charges the reservation price to each buyer, so the marginal rate of substitution equals the marginal rate of transformation as required for static efficiency. With zero transaction costs, perfect discrimination eliminates the deadweight loss of monopoly. The deadweight loss

power can raise prices to buyers and enjoy extraordinary profits. Static economics condemns market power as inefficient because the deadweight loss reduces wealth.

In contrast, growth economics condemns market power when it slows growth, and absolves market power when it increases growth. When does market power retard innovation, and when does it enhance innovation? As we explain, the identity of the buyers determines whether market power retards or enhances innovation.

Market power slows growth when a producer who does not innovate raises prices to innovators, so the cost of developing innovations increases and the profitability of innovation decreases. When profitability decreases, the marginal venture in Figure 4.1 shifts to the left as depicted by a move from I^* to I . To prevent this slowing of innovation and growth, law and policy should suppress market power by non-innovating producers against innovators.

Besides raising the cost of developing innovations, market power of non-innovating producers has another bad effect. Prices guide entrepreneurs in deciding which innovations to develop. Market power of non-innovating producers has the bad effect of aiming innovation at the wrong target. Innovators develop substitutes that are no better than the original good, instead of developing better goods. We call this distortion of inventive activity the “mistargetting effect”.

To illustrate, a cartel that controls diamond mines and raises diamond prices spurs the development of manufactured diamonds. Manufacturers will profit from developing diamonds so long as the cost does not exceed the cartel's price. Since mined and manufactured diamonds are virtually indistinguishable, the only gain to society from manufacturing diamonds is to undermine the cartel's price. The cartel's price could be undermined directly by laws and policies aimed at reducing the market power of the diamond monopoly, without the cost of developing manufactured diamonds.

To illustrate abstractly, if the competitive price of good α is more than good β , then society values α more than β . However, the producers of β might have more market power than the producers of α , so the price of β might occur because transaction costs make perfect price discrimination impossible. Chapter 5 discusses these theoretical claims at length.

exceed the price of α . Since the price of β exceeds the price of α , innovators might foresee higher profits from developing substitutes for β rather than α , even though society benefits more from the opposite.

When market power raises a good's price, more innovators try to invent substitutes for it. The invention of a perfect substitute is costly. However, a perfect substitute is no better than the original good. The only consequence of inventing the substitute is to reduce the distortion from monopoly pricing. Law and policy could reduce market power without the expense of developing the substitute.

Economists sometimes make the same point in different language. By definition, a "rent" is income from owning something that is scarce. Compared to perfect competition, market power makes a good relatively scarce. Thus monopoly profits are "rents" from having market power. High prices charged by producers with market power causes innovators to invent substitutes. Inventing substitutes for over-priced goods enables the inventors to get a share of the rents for themselves. Thus market power causes innovators to "seek rents" instead of inventing better products.

We have explained two reasons why market power of non-innovating producers against innovators retards growth. The first and most important reason is the transfer of profits from innovators to non-innovators. The second and less important reason is mistargetting innovation. This analysis in growth economics reinforces the usual condemnation of producer's monopoly in static economics.

Innovator's Market Power

If market power is condemned for non-innovating producers, what about market power for innovators? Does market power for innovators enhance or inhibit growth? Either effect is possible. This chapter describes the form of market power among innovators that usually enhances growth, and the next chapter describes the form that usually inhibits growth. The difference depends on who buys from the innovator. Divide buyers into non-innovating producers, consumers, and innovators. When the buyers are non-innovating producers and consumers, the innovator's market power increases the inno-

vations price, which transfers wealth from producers and consumers to the innovator. The transfer increases the average profits of innovative ventures, which causes the marginal venture in Figure 4.1 to shift to the right as depicted by the move from I to I*. More resources are devoted to innovation and fewer resources are devoted to production and consumption. Thus market power for innovators against producers and consumers enhances growth

Besides transferring wealth, the innovator's monopoly imposes a deadweight loss on producers and consumers. If maximizing growth is the aim, the deadweight loss does not matter in the long run. The benefits of faster growth from an innovator receiving monopoly profits overtake the costs of the deadweight loss to consumers and producers (the "overtaking principle" in Chapter 1).

To illustrate by our numerical example, assume that an innovator's monopoly imposes costs of 130 on non-innovating producers and consumers. 100 transfers to the innovator as profits and 30 is deadweight loss. The transfer of 100 increases the profitability and rate of innovation. Faster growth creates long run benefits for producers and consumers that overtake the temporary loss totaling 130.

Consider new pharmaceutical drugs. Stronger monopoly powers for pharmaceutical innovators immediately increase the prices of drugs paid by consumers. The higher profits of pharmaceutical innovators cause faster development of better drugs. Consumers lose from higher drug prices and gain more from better drugs. Thus relatively cheap ulcer drugs replaced relatively expensive ulcer surgery.

Similarly, cotton manufacturers buy ever-improving looms. Patents on looms increase the price of cotton textiles in the short run, but patents increase innovations that lower prices in the long run.

The three effects of market power are transfer, deadweight loss, and mistargeting. Analyzing these effects on growth implies a simple conclusion for law and policy. Growth economics *proscribes* market power for (non-innovating) production against innovation and consumption, and growth economics *prescribes* market power for innovations against consuming and (non-innovating) producing. In both cases, the aim is the same: to increase

the profitability of innovative ventures. We have arrived at the *separation principle*: law and policy should treat market power differently for producing and innovating, proscribing it for producing without innovating, and prescribing it for selling innovations for production and consumption.⁷

The separation principle can be explained in terms of externalities. Innovations build on each other. The value of each innovation spills over to other innovations, with large effects on growth. Each innovator contributes to economic growth that benefits others. Making an innovation increases growth whose benefits spill over from the innovator to others. In contrast, consuming an innovation or using it in production does not necessarily causes benefits to spill over to others. Thus innovation has larger external benefits than consuming or producing without innovating. The separation principle calls for transfers in favor of innovative activities with large external benefits.

Conclusion

In static economics, monopoly causes buyers to pay more than the marginal cost of goods, so too little is consumed and produced. The remedy is to abolish market power and to restore competition whenever possible. In growth economics, however, market power is necessary to increase the profitability of ventures competing to innovate. Crossing the boundary between static efficiency and growth changes economics, like crossing the border from France to Germany changes languages.

Market power of sellers has three consequences affecting growth: transfer, deadweight loss, and mistargeting. When (non-innovating) producers have market power in the sale of goods to innovators, all three effects reduce growth and human welfare. Conversely, when innovators have market power against non-innovating producers, the overall effect increases growth and human welfare. The transfer of profits from static to dynamic activities increases growth, in spite of mistargeting, and deadweight loss.

⁷ Note that the three effects of monopoly align in static and growth economics to reinforce the proscription of producer monopolies against innovators. The three effects do not align with respect to the prescription for innovator's monopolies against producers. The prescription for innovator's monopolies against producers rests on the transfers to innovators overtaking deadweight loss and mistargeting.

These facts imply the *separation principle*: to maximize growth, law and policy should separate innovation and (non-innovating) production. Growth economics proscribes market power for non-innovating production, and it prescribes market power for innovating against consuming and producing. The next chapter will analyze the effect of market power by innovators against each other.

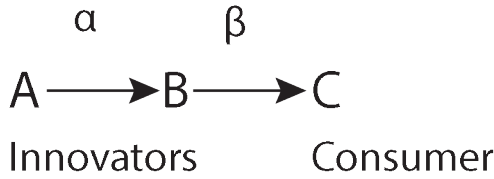
Chapter 5. Fertility

Robert Cooter and Aaron Edlin

The social value of an innovation far exceeds the innovator's profits. To increase innovation, law and policy should increase the profitability of business ventures in open competition with each other. Profitability increases when the venture that wins the competition enjoys market power. A successful venture obtains temporary market power by innovating first. Imitators eventually enter the industry and erode the venture's market power. Law and policy that slows imitators will extend the innovator's market power, thus increasing the venture's profits. Chapters 4 and 5 concern general principles for extending the innovator's temporary market power, and subsequent chapters discuss specific legal instruments such as intellectual property, secrecy, and collusion.

According to the separation principle in Chapter 4, law and policy should enhance the market power of the creator of an innovation who sells it for consumption or production. The aim is to transfer resources from consuming and producing (the "static economy") to innovating (the "dynamic economy"). As defined in microeconomics, market power is the power to set a good's price above the cost of production and to earn extraordinary profits. Thus innovator B develops innovation β and sells it to consumer C, as depicted in Figure 5.1. According to the separation principle, law and policy should enhance innovator B's market power against consumer C. If B has market power and raises the price of β , then the dynamic economy as represented by B extracts resources from the static economy as represented by C, and the rate of innovation increases. The separation principle commends lawmakers to enhance the market power of innovators against consumers and non-innovating producers.

Figure 5.1. Three-Party Straight-Line Transaction



What about market power of innovators against each other? Besides selling to consumers and producers, innovators sell to other innovators. Today's new pharmaceutical molecule is discovered from yesterday's new molecule; today's new operating system is discovered from yesterday's new operating system; today's new power cell is discovered from yesterday's new power cell.¹ The subject of this chapter is redistribution *within* the dynamic economy. The question is, "Should law and policy enhance market power of some innovators against other innovators?" Thus in Figure 5.1, A creates innovation α and sells it to innovator B, who uses α to create innovation β and to sell β to consumer C.

When one innovator has market power in the sale of an innovation to another innovator, the seller raises the price to the buyer and profits transfer from one innovator to another. In Figure 5.1, assume that A has market power against B, so A can capture some of B's profits from sales to C. Market power for A against B redistributes within the dynamic economic, whereas market power of B against C redistributes from the static economy to the dynamic economy.

According to the separation principle in Chapter 4, market power of A against B should have different legal treatment than market power of B against C. Market power of B against C transfers resources from consuming to innovating, which increases growth and improves human welfare. In contrast, market power of A against B transfers resources from one innovator to another innovator, which might increase or decrease growth.

To increase growth, A's innovation α must be more *fertile* than innovator B's innovation β . An innovation is *fertile* like a breeding horse if it can be used to create another innovation. In Figure 5.1, innovation α is used to create

¹ Brian Arthur describes technology as "self-creating" (autopoietic). "...if new technologies were constructed from existing ones, then considered collectively, technology created *itself*." W. BRIAN ARTHUR, THE NATURE OF TECHNOLOGY (2009), page 2.

innovation β , so α is fertile. Fertile innovations are often called “fundamental,” “basic,” “pioneering,” “general,” “first-generation,” “rising technology,” and “shoulders to stand on”.² Unlike innovation α in Figure 5.1, innovation β is *sterile* like a mule because it cannot be used to create another innovation. Sterile innovations are often called “applied,” “derivative,” “specific,” or “stand alone.” For purposes of economic analysis, one innovation is *more fertile* than another by definition if the costless transfer of profits from the latter to the former increases the growth rate. (Mathematical notation makes this characteristic more precise.³)

A costless transfer occurs without loss. Imagine that oasis Y in the desert fills a one liter container with ice cream and sends it across the hot sands to oasis X. Costless transfer implies that X receives one liter of ice cream. None of it melts. Conversely, a costly transfer between oases implies that some ice cream melts and X receives less than one liter. Similarly, a costless transfer from B to A in Figure 5.1 means that B’s loss from the transfer equals A’s gain. The burden of the transfer on B equals the benefit of the transfer to A. If law and policy can transfer profits costlessly, then maximizing growth requires redistributing profits from less to more fertile innovations. Redistribution should proceed until all developed innovations are equally fertile.

This fact can be explained in terms of spillovers. The value of a fertile innovation spills over to other innovations with large effects on growth, whereas the value of a sterile innovation spills over to consumers and non-innovating producers with small effects on growth. To maximize growth, some profits from selling sterile innovations to consumers and producers should pass through to fertile innovations. To pass profits through, the seller of a fertile innovation like α should have market power against the buyer who uses it for sterile innovation like β . (A footnote contains an example.⁴)

2 cite Tim Bresnahan, General Purpose Technologies).cite Scotchmer

3 In notation, let i denote an innovation. The winner of an open competition to make innovation i enjoys profits denoted π_i . The sustainable growth rate g is a function of the distribution of profits: $g=g(\pi_0, \pi_1, \pi_2, \dots, \pi_i)$. By definition, if j is “economically more fertile” than i , then a *costless* transfer of profits $\bar{\delta}$ from i to j increases the growth rate:

$$g(\pi_0, \pi_1, \pi_2, \dots, \pi_i) < g(\pi_0, \pi_1, \pi_2, \pi_i - \bar{\delta}, \dots, \pi_j + \bar{\delta}, \dots, \pi_l).$$

The characteristics of an innovation that make it economically fertile concern the way one idea prompts the discovery of other profitable ideas.

4 Innovator A in Figure 5.1 creates innovation α and sells it to B. Assume that A’s only income comes from selling α to B. If law and market structure deny A market power, then A must sell α

When innovators buy and sell innovations to each other, however, the transfer is usually costly, not costless. Market power imposes a deadweight loss on most transactions, including transactions between innovators. In Figure 5.1, market power for A might cause an increase in the price of α that benefits A by, say, 100 and burdens B by, say, 130. The difference of 30 is the “deadweight loss” as explained in Chapter 4. Whereas the transfer from less fertile innovations to more fertile innovations causes growth to increase, the deadweight loss causes growth to decrease. The net effect determines whether or not the overall growth rate increases. These considerations imply a general principle. *To maximize growth, law and policy should transfer profits from less to more fertile innovations until the increase in growth from higher fertility equals the decrease from a heavier deadweight loss.* This is the *fertility principle*. (The mathematical expression is in a footnote.⁵)

Figure 5.1 can illustrate this principle. Recall that A must develop α in order for B to develop β . If A develops α and sells it to B at the cost of production, then A suffers a loss equal to the cost of developing α . Foreseeing this fact, A will not develop α . Alternatively, if A develops α and uses market power to sell it to B at more than the cost of development and production, then A makes a profit. Foreseeing this fact, A will develop α . Thus market power of fertile innovators can benefit all innovators. (A footnote illustrates with numbers.⁶)

at its production cost, so A cannot recover α 's development costs. If A foresees this fact, it will not develop α , which B needs to develop β . To induce A to create α , B could pass some profits back to A obtained from selling β to C. To increase the rate of innovation, law should facilitate the pass-back of profits from sterile innovators like B to fertile innovators like A.

5 In footnote 2, if j is a “economically more fertile” innovation than i , then a *costless* transfer of profits δ from i to j increases the growth rate:

$$g(\pi_0, \pi_1, \pi_2, \dots, \pi_l) < g(\pi_0, \pi_1, \pi_2, \pi_i - \delta, \dots, \pi_j + \delta, \dots, \pi_l).$$

However, an actual transfer caused by law and policy creates a deadweight loss. The burden on i is δ , the transfer to j is $\gamma\delta$, and the deadweight loss is $(1-\gamma)\delta$. At the optimum as given by the fertility principle, the benefit from the transfer to a more fertile innovation exactly offsets the cost from the deadweight loss, so the growth rate does not increase or decrease:

$$g(\pi_0, \pi_1, \pi_2, \dots, \pi_l) = g(\pi_0, \pi_1, \pi_2, \pi_i - \delta, \dots, \pi_j + \gamma\delta, \dots, \pi_l).$$

6 Assume that A invests 75 to develop α and A spends 25 to produce α . If B buys α at 25, which is the cost of production, then A loses 75. Foreseeing these facts, A will not develop α . If A does not develop α , then B cannot develop β , and their joint profits are 0.

Instead of the fertility principle, court cases refer to other principles such as the principle of cost recovery (innovators should enjoy enough profits to recover their development costs) and the principle of proportionality (successful innovators should receive profits in proportion to their development costs). The next chapter discusses the connection between the fertility principle and the principles actually invoked by courts.

Applying the Fertility Principle

What determines the size of the deadweight loss caused by market power? In the standard economic analysis of market power, the deadweight loss increases with the elasticity of demand for the good in question, holding constant the monopolist's profits. The appendix to this chapter contains a graphical explanation of these facts. Here is a verbal explanation. When a seller uses market power to increase a good's price, buyers respond a lot if demand is elastic, and buyers respond a little if demand is inelastic. A larger response indicates that buyers are making more costly adjustments to cut back more on their purchases of the good. Thus if a patent on computer-driven printers causes their price to rise, people will print less, which causes productivity to fall. Productivity falls more when the elasticity of demand for printers is higher.

When demand is inelastic, creating monopoly profits for innovators imposes a small deadweight loss. The case for enhancing market power for fertile innovations is stronger when demand for the innovation by other innovators is less elastic. Conversely, the case for enhancing market power for fertile inno-

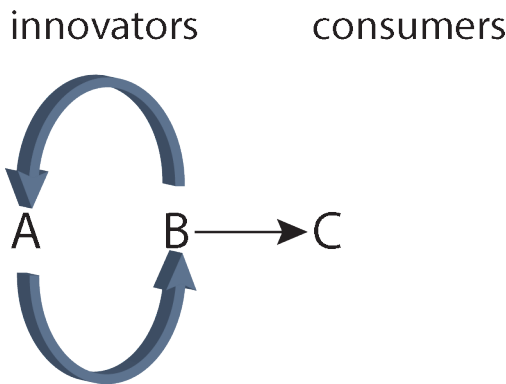
Alternatively, assume that market power enables A to sell α to B for 200. A will make a profit of 100, which equals the sales price of 200 minus the cost of development production 75+25. Foreseeing this fact, A will develop α . B buys α at a price of 200, develops β , and sells β for 300 to C, so B makes a profit of 100. The sum of the joint profits of A and B equal 200. Thus market power of A against B benefits both and causes their joint profits to rise from 0 to 200.

There is, however, a deadweight loss caused by A's market power against B, which we also illustrate. As noted, when B buys α for 200, A and B each earn profits of 100. We assume that the state pays a subsidy to A to develop α and then requires A to sell α to B at the cost of production. Specifically, the state imposes a tax of 175 on B, which the state pays to A for developing α , and the state requires A to sell α to B at 25, which is the cost of production. In sum, A receives a subsidy of 175, spend 75 to develop α , spends 25 to produce α , and sells α to B at a price of 25, for a net profit of 100 for A. A's profits remain constant at 100, and B's profits rise, say to 130, because B does not have to economize on the use of α when it develops β . The replacement of market power with a subsidy eliminates a deadweight loss of 30 and increases joint profits of A and B from 200 to 230.

variations is weaker when demand by other innovators is more elastic. *Market power of more fertile innovators should vary inversely with the elasticity of demand for innovations by less fertile innovators.*⁷ (The market power of innovators, however, should not vary with the elasticity of demand by consumers for innovations, according to the separation principle in Chapter 4.⁸)

Besides the elasticity, the pattern of transactions among innovators affects the deadweight loss. Circular transactions among innovators cause an especially large deadweight loss. Much that a venture buys to develop an innovation was originally an innovation by another venture. When innovators sell innovations to each other, some transactions circle within the dynamic economy, instead of proceeding straight from innovators to consumers as in Figure 5.1. In Figure 5.2 venture A makes innovation α and sells it to venture B, who uses it to develop innovation β and sells it to venture A for the next round of innovations, and so on. (In Figure 5.2, innovators A and B also sell innovations to consumers C.)

Figure 5.2 Circular Transactions



In circular transactions, the deadweight loss accumulates and becomes especially large. Thus assume that innovator A in Figure 5.2 uses market power

⁷ Ramsey worked out the inverse elasticity rule for optimal taxes in 1927. See analysis and footnote in the appendix to this chapter.

⁸ When demand by consumers for an innovation is elastic, market power for the innovator against consumer causes a large deadweight loss. This effect, however, primarily concerns static efficiency. Growth will overtake the effects on human welfare. So the market power of innovators should not vary much with the elasticity of demand for innovations by consumers.

to raise the price of the innovations that it sells to B. This use of market power gains profits of 100 for A and imposes costs of 130 on B. Similarly, assume that innovator B uses market power to raise the price of the innovations that it sells to A. This use of market power gains profits of 100 for B and imposes costs of 130 on A. The deadweight loss of monopoly in this cycle of transactions between A and B is 60. If market power were eliminated, the increase in profits of A and B would total 60, with each one gaining 30. Their profits would increase because each gains more from lower costs than it loses from lower revenues.

The more innovators deal in circular transactions with each other, the larger is the increase in growth from reducing their market power. Eliminating market power in circular transactions is relatively easy when everyone gains. For example, if patents were the source of market power between A and B, then both of them would enjoy higher profits if neither had patents that were good against the other. Later we explain how firms achieve this result by pooling their patents.

Circular transactions among innovators are sometimes described as an innovation “thicket.” In an innovation thicket, everyone needs the innovations of others in order to make their own innovations. If several innovators have the power to withhold innovations that are necessary to the success of another venture, then each of them can effectively veto the venture. As the number of innovators with veto power increases, the probability of a veto increases. Similarly, many barons built castles on the Rhine River in medieval times to extract tolls from passing ships. One baron who held out for a high toll could block most traffic on the river. If law allows holdouts, they sharply reduce innovation in thickets and trade along rivers. Consequently, growth economics proscribes market power in innovation thickets, as the next chapter explains for patents.

Bargaining and Transaction Costs

In perfect competition, everyone takes the market price. In contrast, market power gives the seller choice over the price. The seller names the price and the buyers respond to it. Sellers with market power name prices in two different ways. First, the seller may name a *firm* price, as when the vendor in

a football stadium sets the price of a glass of beer. In the standard model of monopoly, the monopolist maximizes profits by setting a firm price where the marginal revenue from a small increase in production equals the marginal cost.⁹

Second, instead of a firm price, the seller may name a flexible price, and each buyer may make a counter-offer. With price flexibility, the parties bargain with each other to reach an exact price. Bargaining takes time and effort. In addition to time and effort, bargaining sometimes fails to reach an agreement. Three major costs of transacting by bargaining are time, effort, and the loss from failure to agree.

To illustrate, assume that a football player is willing to play for 100 and a team values the player at 120. When they bargain over salary, the player will not accept less than 100 and the team will not pay more than 120. If bargaining succeeds, they will agree on a price in between 100 and 120, depending on their bargaining power.¹⁰ Say they agree on the price 112, so the player gains 12 and the team gains 8. Bargaining takes time and effort that costs, say, 1 for the team and 1 for the player. The gain net of bargaining costs thus equals 7 for the team and 11 for the player. In addition to the time and effort, bargaining involves uncertainty about the result. When the parties begin bargaining, they do not know whether or not they will reach an agreement. If the probability of success in our example is, say, 90%, then the probability of failure is 10%. Rational parties discount the expected value

⁹ A seller's revenue equals the price p multiplied by the quantity q that he sells. For a monopolist, the quantity is a decreasing function of the price. So the monopolist's revenue R can be written $R=pq(p)$ where $q' < 0$. The monopolist's cost of production C increase with the quantity, which decreases with the price, so $C=C(p)$ where $C' < 0$. Profits, which equal revenues $R - C$, are maximized when marginal revenues equal the marginal cost: $q+pq' = C'$.

¹⁰ When two parties bargain with each other, their bargaining power depends on their alternatives. In perfect competition, each party to a bargain has a perfect substitute. The buyer can buy the same good from another seller at the market price, and the seller can sell the good to another buyer at the market price. With perfect substitutes, neither party has any bargaining power against the other. That is why they are price takers. Alternatively, in monopoly, the buyer cannot buy the same good from anyone except the single seller, whereas the seller can sell to another buyer. With close substitutes for the buyer and no substitutes for the seller, the seller has all of the bargaining power and the buyer has none of it. Law and policy increases the market power of the seller of an innovation by reducing the substitutes available to buyers. For example, a patent that effectively stops infringements reduces the substitutes available to its buyers. A remarkably simple formulation of the relationship between bargaining power and negotiated prices is the Nash Bargaining Solution, which Chapter ___ discusses.

of the bargain by its probability of failure. The net gain discounted for the probability of failure is .9(7) for the team and .9(11) for the player, which sum to 16.2. The difference between the gross gain of 20 for the team and player, and the discounted net gain of 16.2, is 3.8.

This number, 3.8, can be called the bargain's "transaction costs," and it can also be called the bargain's "deadweight loss." The equivalence of "transaction costs" and "deadweight burden" yields two formulations of the fertility principle. According to the original formulation, to maximize growth, law and policy should transfer profits from less to more fertile innovations until the increase in growth from higher fertility equals the decrease from a heavier deadweight loss. Equivalently, law and policy should transfer profits from less to more fertile innovations until the increase in growth from higher fertility equals the decrease from higher transaction costs.

To apply this principle, contrast bargaining between two parties and many parties. When only two people must agree to a bargain (say, the buyer and the seller), negotiation often takes little time or effort, and the parties can usually reach an agreement. With two parties, transaction costs are low, especially when the law enforces contracts between them.¹¹ Thus the deadweight loss is negligible if A sells α to a single buyer B. Conversely, with many individuals involved, negotiation takes more time and effort, and the parties may not reach an agreement. Thus the deadweight loss is likely to be high if A sells α to B where "B" stands for many individuals who must agree.

"Frictionless plane" describes an engineer's theoretical ideal, not the real world. Without friction, objects move forever in the same direction at the same speed forever. In the real world, the amount of friction determines the speed and distance that an object will move. So a frictionless plane provides a baseline for calibrating and predicting actual movements. Similarly, "zero transaction costs" describes an economists' theoretical ideal, not the real world. Without transaction costs, exchange continues until all resources are owned by the parties who value them the most – a state of "allocative efficiency". Applied to law, this is the most famous proposition in law and economics, name the "Coase Theorem" after its discoverer. According to this

11 With good laws, many agreements can be implemented in enforceable contracts, but not all agreements. Recall the double trust dilemma in Chapter 3 where the innovator and financier cannot implement their agreement to develop an innovation because performance of the terms is unverifiable in court.

theorem, the allocation of legal entitlements does not matter to efficiency so long as private bargains can reallocate them without transaction costs. Thus intellectual property law can allocate ownership of an innovation to anyone (the inventor or someone else) in any form (strong or weak patents), and private bargains will re-allocate ownership to the party who values it the most, provided that transaction costs are zero. By referring to this intriguing and puzzling theorem, subsequent chapters will convey its meaning for intellectual property.

Conclusion

Maximizing growth requires an open competition among innovators with extraordinary profits as the prize. Innovators enjoy extraordinary profits when law and policy allows them to exercise the market power that they have or gives them market power that they lack. According to the separation principle in the preceding chapter, law and policy should strengthen the market power of innovators against consumers and producers who do not innovate. According to the principle of fertility in this chapter, law and policy should strengthen the market power of fertile innovators against sterile innovators.

Market power within the dynamic economy has two effects on growth: transfer and deadweight loss. Transfers from less to more fertile innovations increase growth, and deadweight loss in transactions among innovators decreases growth. Redistribution from less to more fertile innovators should proceed until the decrease in growth from deadweight loss equals the increase in growth from fertility. Intellectual property, which is the subject of the next chapter, is one bodies of law for implementing the separation and fertility principles.

Appendix: Welfare Triangles

Figures 5.3 and 5.4 depict the relationship between the burden, the deadweight loss, and the elasticity of demand when the seller has market power against the buyer.¹² The horizontal axis represents the quantity of innovation α produced, and the vertical axis represents prices and costs. After A develops α , c equals the cost of producing a unit of it. If A sells α to B at the cost of production c , then c equals the market price p_c . Since production cost equals price, A earns zero profits on production and cannot recover the cost of development. Alternatively, if A has market power against B, then A will raise the price to p_m and B will buy the quantity q_m . A's profit equals the difference between the price and cost of production multiplied by the quantity, or $(p_m - c)q_m$, which Figure 5.2 denotes T . The transfer T is the gain to A from receiving the high price p_m for the small quantity q_m , instead of receiving the low price p_c for the large quantity q_c .

In the numerical discussed in this chapter, T equals 100. However, the loss to B of paying the higher price p_m is more than the transfer T . Besides paying more for the q_m units of α that B buys, the higher price causes B to forego the purchase of $q_c - q_m$ units of α . The loss to B from foregoing this purchase equals the difference between the amount B was willing to pay for them, as given by B's demand curve, and their cost of production c . In Figure 5.2, L denotes this loss. B's loss from A's market power equals $T + L$, which is the burden on B. A's gain from its market power equals T . The difference L between B's burden and A's gain is the deadweight loss L .

¹² Ramsey first worked out the mathematics for taxation, and he concluded that optimal taxes are inversely proportional to the elasticity of demand. See Frank P. Ramsey, *A Contribution to the Theory of Taxation*, 37 *ECONOMIC JOURNAL* 47–61 (1927).

Figure 5.3. *Burden and Deadweight loss of Market Power*

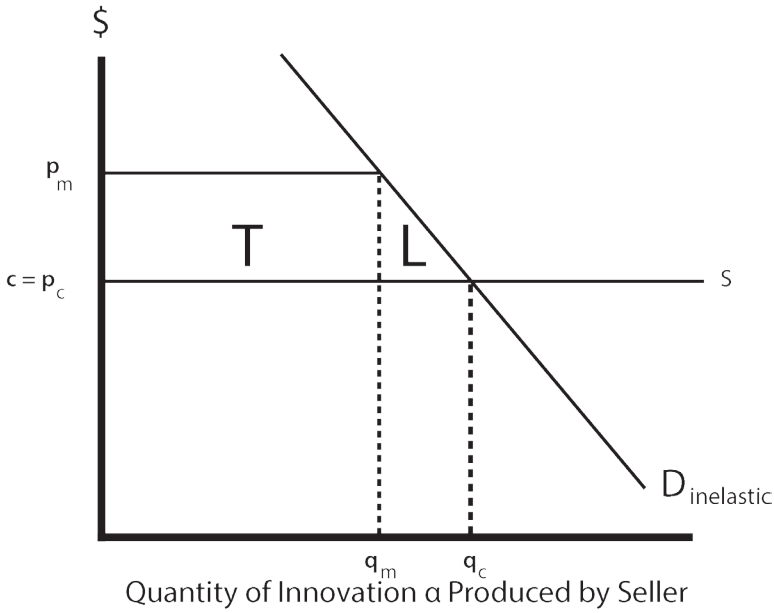
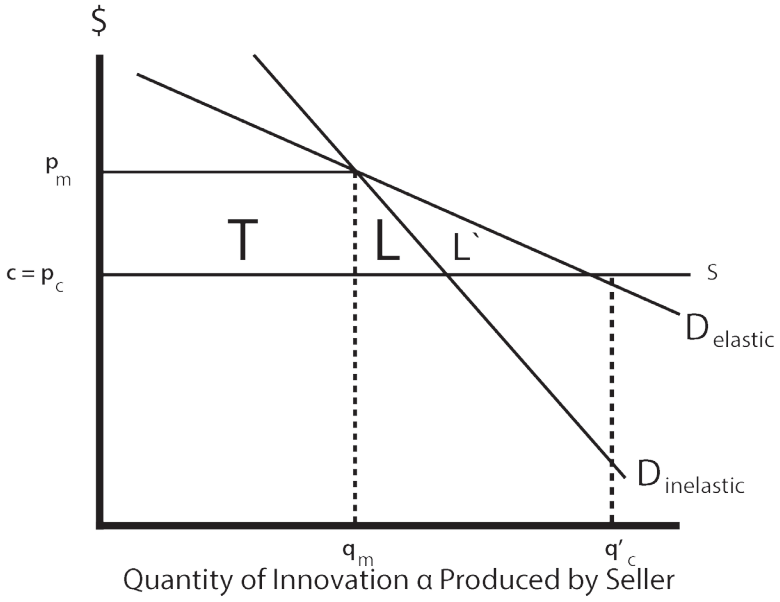


Figure 5.4 shows the increase in the deadweight loss when demand becomes more elastic. Figure 5.4 holds demand constant at the point (p_m, q_m) and rotates the demand curve from $D_{\text{inelastic}}$ up to D_{elastic} . This change holds A's gain of T constant, but the deadweight loss increase from L to $L+L'$. The deadweight loss increases because B responds to the price increase by making a larger reduction in the good's use. The larger reduction in use involves more costly adjustments to get by with less of the good.

Figure 5.4. Burden, Deadweight loss, and Elasticity of Demand



Chapter 6. Separation and Fertility in Intellectual Property Law

Robert Cooter

To speed innovation and increase human welfare, law should provide a framework for open competition among innovators with high profits for the winners. High profits especially come from the innovators' market power, which is the power to affect prices. The first firm to develop and market an innovation often enjoys market power temporarily (the "first mover's advantage"), which is a form of natural monopoly. Secrecy often extends the innovator's market power by slowing the discovery's dissemination to competitors. Patents and copyright that shield innovators from competition create market power by legal fiat. These three sources of market power -- natural monopoly, private information, and legal fiat -- especially relate to three bodies of law: antitrust, trade secrets, and intellectual property. This chapter concerns the third source of market power for innovators -- intellectual property law.

Article 1 Section 8 of the U.S. constitution reads:

"The Congress shall have power...To promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries..."

The "useful arts" mentioned in Article 1 Section 8 form a heterogeneous collection of activities. To measure "progress" requires combining the activities into an index. A comprehensive measure of economic growth is often the best index of progress in the useful arts. Thus Article 1 section 8 authorizes Congress to promote economic growth through intellectual property law.

Congress asserted its constitutional power by enacting Title 35 Sections 100-105 of the United States Code.¹ The constitution and much of the code is general and abstract. Regulations and court decisions increase the precision and certainty of these laws. In the U.S. where courts justify their interpretations by reference to prior court interpretations, much intellectual prop-

¹ Section 101 [2] sets out subject matter that can be patented; section 102 [3] defines novelty and loss of right to patent; section 103 [4] lists what constitutes non-obvious subject matter.

erty law is “judge-made”. Economic growth provides a useful standard for interpreting and critiquing intellectual property law made by Congress and courts. The separation from Chapter 4 and fertility principle from Chapter 5 contribute to interpretation and critique, especially where innovations can grow exponentially.

This chapter applies the separation and fertility principles to innovations that can grow exponentially, so the effects of growth on human welfare overtake other effects. Patented inventions can grow exponentially. So can some copyrighted expressions such as computer programs, as well as some innovations covered by special statutes.² In contrast, cultural expressions (novels, songs, plays) evolve and develop, but these changes do not constitute exponential growth on a generally accepted measure such as cost-benefit analysis or market value. Compared to economic growth, any attempt to measure progress in artistic expression immediately encounters deep disagreements over cultural values.

The patent or copyright for inventing a machine, creating a molecule, extracting a vaccine, or writing a computer program can be weak or strong. The separation principle distinguishes infringement by innovators from infringement by consumers or producers. Patent protection should be strong against infringers who consume the invention or produce with it. According to the fertility principle, patent protection should be weak or absent against users who innovate with it in the usual cases. In exceptional cases, however, patents should be strong against other innovators where the patented invention is more fertile than the innovation for which it is used.

Strength

We will develop the theory and application of the optimal strength of intellectual property law. The *strength* of intellectual property rights depends on three attributes: duration, breadth, and remedy. The *duration* of a patent can be measured in years. Thus U.S. patent duration is 20 years from the date of filing the patent application.³ When filing, the applicant and her patent attorney make claims about the innovation’s *breadth*. The patent office and

2 The Semiconductor Chip Protection Act of 1984, which legally protects the layouts of integrated circuits, exemplifies such special statutes.

3 Before 1995, U.S. patent duration was 17 years from the date the patent issued.

the courts ultimately decide the validity of these claims. Breadth is easier to order than to measure. Thus a patent covering rain gear is broader than a patent covering umbrellas, and a patent covering umbrellas is broader than a patent covering automatically opening umbrellas. While one patent can be broader or narrower than another, there is no natural measure of the difference in breadth.⁴

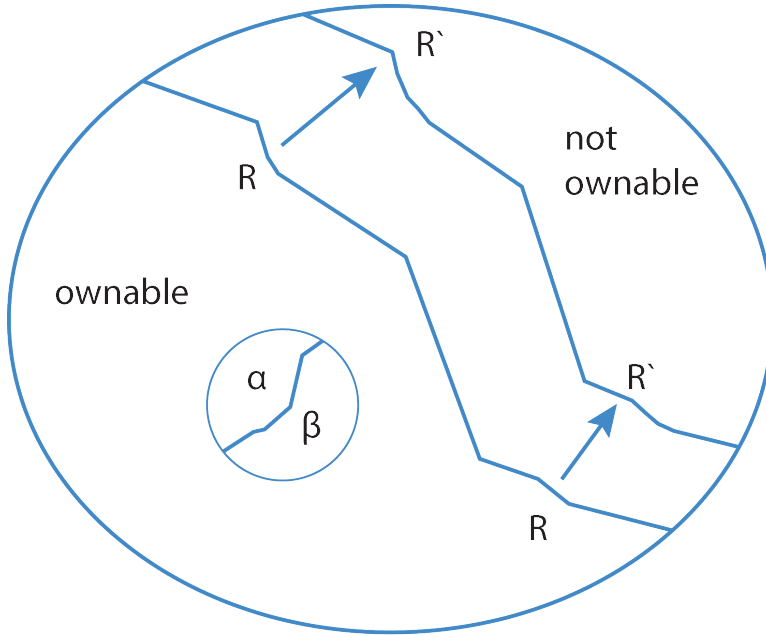
Taken together, the duration and breadth of an intellectual property right define its *scope*. Acting within the scope of a patent or copyright without a license from the owner infringes it. Law usually provides a remedy for infringement. For past infringements, the owner can usually sue to recover compensatory damages. Higher damages (“super-compensatory”) are allowed sometimes. For future infringement, the owner can usually enjoy the infringing conduct.⁵

4 Patent breadth resembles the alphabet. Letters A through N encompasses more letters than A through M, but there is no natural measure of the distance between M and N. Similarly, no natural measure exists to answer the question, “Is the difference in breadth between a patent on rain gear and a patent on umbrellas larger or smaller the difference in breadth between a patent on umbrellas and a patent on automatically opening umbrellas?” In this respect, patent breadth contrasts with height. One person is taller than another and the difference can be measured in centimeters.

Fortunately, an ordering without a measure can support the mathematics of maximization. This fact makes modern utility theory possible in economics. Economic models often maximize utility functions that order states of the world by preference (“ordinal utility”), but do not measure differences in utility levels. Similarly, finding the breadth of patent that maximizes growth only requires an ordering by breadth, not a measure.

5 Note that money *measures* damages like years measure duration, whereas breadth *orders* (but does not measure) injunctions like patent claims order their breadth.

Figure 6.1 Future Innovations



In addition to increasing the strength of individual patents, the patent system as a whole can strengthen against unpatented activities. Figure 6.1 divides future innovations into “ownable” and “unownable.”⁶ A new computer chip is ownable (patentable) and a new metaphor in the English language is unownable (no patent, no copyright). The strength of all patents indicates the patent system’s *reach*, as illustrated by the boundary between ownable and unownable innovations in Figure 6.1. Thus shifting the boundary in Figure 6.1 from R to R' increases ownable innovations and decreases unownable innovations. To illustrate, U.S. patent law increased its reach when inventors were allowed to patent business processes for the first time, such as Amazon’s “one-click” ordering from its online catalogue. Expanding patent law’s reach fences in more of the common land of innovation.⁷

6 This is a picture, not a graph, because the breadth of innovations has no natural measure. See the discussion in this chapter of the difference between an ordering and an index.

7 Increases in the scope of a particular patent do not necessarily increase the scope of patents as a whole. Rather, the scope of a particular patent can increase by decreasing the scope of another patent. Instead of fencing more of the common land of innovation, the fence between patent holders moves to the advantage of one of them and to the disadvantage of the other, as with the boundary between α and β in Figure 6.1.

Besides strengthening the patent system against unpatented activities, patent law can strengthen one patent by weakening another. The space of innovations represented by the small circle in Figure 6.1 is subdivided into α and β . Assume that firm A invents α and firm B subsequently invents β . A claims that α and β are a single invention and A files for a patent covering both of them. Later B files for a patent for β and claims that β is a different invention from α . The patent office must decide whether to award a broad patent to A that covers α and β , or to award a narrow patent to A that covers α and a narrow patent to B that covers β . For a given space of inventions, the state can generally give few broad patents or many narrow patents, rather like a city planner can divide undeveloped land into a few large lots or many small lots.⁸

How Strong?

What reach and strength (breadth, duration, and remedy) of intellectual property rights maximizes innovation? The answer is remarkably simple. Maximizing innovation requires transferring resources from consumption and production to innovating. According to the separation principle (Chapter 4), an innovator should have a strong patent against others copying the innovation to sell to consumers, such as a smart phone application. An innovator should also have a strong patent against others producing consumer goods by using the innovation, such as industrial robots that manufacture cars. Applied to the strength of intellectual property, *the separation principle asserts that the innovator's rights against others consuming it or producing with it should be strong -- broad and long with a generous remedy for infringement.*⁹

Besides producing and consuming, some innovations need prior innovations like a carriage needs wheels. When a prior innovation is owned, the owner usually has market power against a subsequent innovator who needs a license

⁸ Once lots are surveyed in a new town, owners can sell a lot but not half of a lot. The owner who cannot sell half of a lot can lease it for use by someone else. Much the same is true for selling patents. An owner cannot sell half of a patent, but an owner can contract for someone else to use half of it. Some theorists think that divisibility is the essential difference between property rights and contract rights. See Thomas A. Merrill & Henry Smith, "Optimal Standardization in the Law of Property: The Numerus Clausus Principle," 110 Yale Law Journal (2000).

⁹ Economists may wish to compare broad patents and broad taxes. Economists favor broad taxes because the demand for broad categories of goods is less elastic than the demand for narrow categories of good, so the deadweight loss is smaller. Similarly, the demand for broad patents is less elastic the demand for narrow patents, so the deadweight loss is smaller and the transfer from buyers to sellers is larger.

to use it. Market power of innovators against each other imposes a deadweight loss on all of them, which slows innovation and growth. This fact argues that innovators should have weak intellectual property rights against each other, or none at all.

Chapter 4 described four forms of deadweight loss and their likely extent. Creating market power for an innovator causes a large deadweight loss when the patented good has close substitutes (high elasticity of demand),¹⁰ few obstacles to resale (no price discrimination),¹¹ the innovation is complex and opaque (incomplete information for buyers),¹² and many buyers purchase small amounts of the good.¹³ These conditions argue that innovators should have weak intellectual property rights against each other, or none at all.

To apply the separation principle, lawmakers and courts must distinguish the uses of an innovation into consuming, producing, and innovating. In many cases, the distinction requires non-technical judgments of the kind that courts routinely make without statistics or mathematical theories. Common sense usually suffices, but difficulties arise at the boundaries of these three activities. In hard cases, economists can assist by drawing on their long history of distinguishing these three activities. Economists can also assist by predicting the effect of different distinctions on the goal underlying the separation principle: maximizing innovation.

Legal application of the separation principle can mimic licensing practices by some firms owning fundamental inventions like computer platforms. Apple allows developers of applications to use the iPhone platform for free, whereas consumers must pay a fee for using many of the resulting applications. Thus Apple separates innovation from consuming, and asserts weak rights innovat-

10 Goods are substitutes when having more of one makes a person demand less of the other. When a good has close substitutes, raising its price (while keeping other prices constant) causes buyers to switch to the substitute. Switching costs are a deadweight loss.

11 With few obstacles to resale, the owner cannot charge different prices to different buyers for the same innovation. Price discrimination is impossible. The more the seller can discriminate among buyers, the less the buyers will switch in response to a price increase, so the deadweight loss of switching falls. This is an application of the proposition found in microeconomics textbooks that a perfectly discriminating monopolist is efficient.

12 As a good becomes more complicated and opaque, the amount of bargaining increases, and so do the search costs of determining whether an innovation is already patented.

13 Transactions costs of purchases often have economies of scale. Consequently, the transaction cost of market exchange is higher with many buyers make small purchases than when few buyers make large purchases.

ing and strong rights against consuming. Similarly, Sun Microsystems developed the Java programming language in the 1990s and made it available for free to developers under a general-purpose license.¹⁴ Developers use Java to write programs that they sell to consumers.¹⁵ Thus Java is free for innovators and costly for consumers of its applications.

As explained, the overtaking principle and the deadweight loss from monopoly argue for weak intellectual property rights of innovators against each other. However, an argument in the opposite direction applies when innovations differ significantly in fertility. By definition, *costless* transfers (no deadweight loss) from less to more fertile innovations increase the rate of innovation and growth. Stronger intellectual property rights for more fertile innovations transfer wealth away from less fertile innovations. In transactions among innovators, relatively greater fertility favors relatively stronger intellectual property rights.

As explained, deadweight losses argue against market power for innovators against each other, and greater fertility argues in favor of market power for some innovators against others. Maximum growth requires balancing these considerations. *The fertility principle asserts that the ideal balance is struck when further strengthening of property rights of fertile innovations causes growth to decrease from greater deadweight loss by the same amount as it increases from greater fertility.* Equivalently, the innovator's rights are ideally strong against others using it for less fertile innovations when further weakening increases growth from lower deadweight loss by the same amount as it decreases growth from lower fertility.

Applications

To maximize progress in the useful arts, intellectual property law should apply the separation and fertility principles. The separation principle favors giving innovators strong rights – broad, long, and with a generous remedy for infringement -- against others consuming the innovation or producing

14 A reverse of this pattern sometimes occurs. Thus when the owner of a platform for running computer games allows consumers to use it for free and charges developers who want to use it to create new games to sell to consumers.

15 Legal disputes subsequently arose when Microsoft modified java to run exclusively on its Windows program, thus necessitating the use of Windows to run programs written in the modified java language.

with it. The fertility principle usually favors giving innovators weak rights or no rights against others using the innovation to make subsequent innovations. In unusual case of especially fertile innovations, an innovator's rights against others using the innovation to make subsequent innovations should balance incentives for fertile innovations and the deadweight loss from stronger property rights. We will apply this general theory to intellectual property's strength as indicated by reach, breadth, duration, and remedy

Reach: Invention or Discovery

The boundary between owned and unowned innovations defines the intellectual property system's reach as depicted in Figure 6.1. Courts and regulators have attempted to draw the boundary in various ways. One line of cases distinguishes inventing and discovering. To be patentable in U.S. law, an innovation must be an invention, but inventions often involve discoveries about nature. A fault-line of legal dispute concerns the difference between patentable inventions and unpatentable discoveries about nature.

To illustrate, the inventors of the telegraph in the 19th century attempted to patent the use of the electromagnetic spectrum to transmit messages. In *O'Reilly v. Morse*, the U.S. Supreme Court decided that the telegraph's inventors could patent a particular machine to transmit messages by using the electromagnetic spectrum, but not all ways of doing so. The electromagnetic spectrum is unpatentable, like the laws of nature and observations of physical phenomena.¹⁶ A similar conclusion concerns DNA sequencing. Does the first person to observe a DNA sequence acquire the exclusive right to observe it? In *Myriad Genetics* (2013) the U.S. Supreme Court unanimously decided "No". Inventions can be patented, but not discoveries of natural laws and facts.

How does this legal distinction square with the separation and fertility principles? The separation principle favors strong patent protection over static activities (consuming and producing) and weak or no patent protection over dynamic activities (innovating). The telegraph was the first device to use the electromagnetic spectrum to transmit messages, and many others followed, including telephones and computers. The patent for the telegraph gave its inventor's temporary market power over transmitting messages by using the

¹⁶ The court recently sustained this view in *Bilski* (2010).

electromagnetic spectrum. The patent was strong against telegraph transmissions, which mostly concern consumption and production. However, the patent was absent against subsequent inventions that transmit message by using the electromagnetic spectrum like telephones. So the breadth of the patent for the telegraph was consistent with the separation principle and it created strong incentives to innovate, not obstacles to it.

The separation and fertility principles broadly favor the patentability of inventions and the non-patentability of discoveries of natural laws and facts. Easy cases can follow common-sense distinctions, and hard cases can look to evidence about locating the boundary to maximize economic growth. An example of a hard case concerns whether diagnostic devices in medicine are inventions or observations of nature. Diagnostic devices are used to treat patients (consumption) and to conduct medical research (innovation). The separation and fertility principles commend a different standard of “infringement” for medical treatment and medical research.

To speed innovation, the separation and fertility principles argue for patent coverage of patient treatments, which implies that unlicensed use of a diagnostic device to treat a patient infringes the patent. The patent will increase the cost of treatment, and patients will lose from higher prices in the short run, but, according to the overtaking theorem in Chapter 1, patients will gain more from faster innovation in diagnostic devices. The conclusion is different for research as opposed to treatment. To speed innovation, the fertility principle argues against patent coverage for research, which implies that unlicensed use of a diagnostic device to make medical innovations does not infringe the patent. The law follows these prescriptions imperfectly. In *Prometheus* (2012) the U.S. Supreme Court disallowed the patenting of diagnostic devices (and also surgical methods) without distinguishing between treatment and research. In pharmaceuticals, however, the law follows the separation principle and allows a “research exemption,” as we will discuss shortly.

The preceding discussion of “reach” mostly concerns the difference between inventing and discovering. To be patentable in U.S. law, an innovation must be an invention, not a discovery of a natural phenomenon. In addition, it must be useful. For most patent applications, the authorities presume that the invention is useful without weighing the arguments. In unusual cases

that weigh the arguments, some commentators doubt that an invention has utility if its only use is to generate more inventions. This reasoning is confused. A fertile innovation promotes “progress” in the “useful arts” even if it is not used in consumption or production. Progress in the useful arts is the constitutionally recognized purpose of intellectual property law. Thus fertility is a constitutionally recognized form of utility, not a suspect category.

Besides being useful, U.S. patent law requires an invention to be new. An invention is new if it is unknown to the public and not obvious. The most important disputes about novelty concern whether one innovation is sufficiently different from another to justify a separate patent. The breadth of patents is our next topic.

Breadth

In Figure 6.1 the patent authorities could issue one broad patent to A that encompasses both α and β , or a narrow patent to A for α and a narrow patent to B for β . Are α and β sufficiently similar to justify one patent, or are they sufficiently different to require two patents? Common sense understanding of it and engineering facts give the answer in most cases, but not in the hard ones. Resolving hard cases requires connecting legal rules to their purposes. The constitutional purpose of intellectual property law in the U.S. is promoting progress in the useful arts, which economic innovation measures.

Connecting patent breadth and growth requires understanding how economists measure similarity in goods. Economists regard two goods as similar if they are substitutes in markets, which implies that people will buy whichever one is cheaper, and not buy the dearer one (high cross-elasticity of demand). The breadth of patents affects innovation partly through the substitutability of goods. If α and β are close substitutes for each other, and if no close substitutes exist for them, then one broad patent will create a monopoly and two narrow patents will create a duopoly. Compared to duopoly, a monopoly can charge at least as high prices and extract at least as much wealth from buyers.¹⁷ So one broad patent for A covering α and β transfers at least as

¹⁷ Duopolists sometimes cooperate perfectly with each other, or they merge. In that case, they maximize their joint profits and earn the same profits as a monopolist. More often, however, duopolists cooperate imperfectly with each other and do not monopolize their joint profits. So the profits of a monopolist are at least as high as a duopolist for the same product.

much wealth from buyers to the owner A as two narrow patents transfer to the owners A and B.

When goods are substitutes, their use determines whether transferring wealth from buyers to sellers increases or decreases incentives for innovation. If innovations α and β mostly have *static uses* (consuming and producing), then one broad patent will transfer more wealth from static to dynamic activities than two narrow patents, which increases the incentive to innovate. Thus the separation principle favors broad patents for innovations that substitute in consumption or production. Conversely, if innovations α and β are substitutes and they mostly have *dynamic uses* (innovating), issuing two narrow patents will impose less deadweight loss on innovating, which increases the incentive to innovate.

In hard cases, the connection between similarity as measured by substitutability and incentives for innovation can often provide definite answers about the optimal patent breadth. The separation principle favors broad patents for substitutes in consumption and production. The fertility principle usually favors narrow patents for substitutes in innovation. However, in the exceptional case the fertility principle favors broad patents for very fertile substitutes in innovation.

Demand for two goods is independent if the price of one does not affect how much people will pay for the other (zero cross price elasticity). Independent goods are dissimilar economically. Intellectual property law cannot increase the market power or profitability of innovating by issuing broad patents encompassing independent inventions. Thus when demand for α is independent of demand for β , owning the patent on β does not increase the profitability of owning the patent on α . Since independent goods are dissimilar, broad patents do not increase the incentive to innovate. The breadth of intellectual property makes little difference to market power created by patents for independent innovations. The patent authorities should issue narrow patents for inventions of independent goods.

Table 6.1 summarizes these conclusions about use, substitution, and preferred patent breadth.

Table 6.1. Preferred Patent Breadth

Use	Substitutability		
		close substitutes	independent
consuming or producing		strong patent (broad preferred to narrow)	weak patent (narrow preferred to broad)
innovating, usual case		weak patent (narrow preferred to broad)	weak patent (narrow preferred to broad)
innovating, very fertile case		strong patent (broad preferred to narrow)	weak patent (narrow preferred to broad)

This reasoning applies to a series of suits that Apple brought against Samsung beginning in 2011 for infringement of design patents for the iPhone and iPad. For smart phones and tablets, Apple was the original innovator and Samsung was the imitator. Samsung’s strategy was to build close substitutes to the originals and to make small improvements. By imitating, Samsung saved development costs, so Samsung built substitutes at lower cost and sold them to consumers at lower prices. Consequently, Samsung’s share of the smartphone market increased at Apple’s expense.¹⁸

The legal question is whether Samsung infringed Apple’s patents. Table 6.1 provides an answer by relating preferred patent breadth of substitutes to their use. In so far as Samsung mostly used Apple’s innovations to make consumer goods, the separation principle favors strong protection. Strong protection will reward innovation and speed progress in the useful arts. In so far as Apple’s innovations were fundamental and Samsung’s innovations were derivative, the fertility principle also favors strong protection of Apple’s patents against Samsung. Again, strong protection will reward fertile innovation and speed progress in the useful arts. However, in so far as Samsung mostly used Apple’s innovation to make its own innovations of similar fertility, the fertility principle favors weak protection of Apple’s patents against Samsung. Weak protection will reduce the deadweight loss on innovation and speed progress in the useful arts.

Samsung’s liability to Apple ideally depends on the extent to which Samsung’s smart phones and tablets embody its own fertile innovations. Samsung’s de-

¹⁸ Sang-Hun Choe, *After Verdict, Assessing the Samsung Strategy in South Korea* By CHOE SANG-HUN ‘’, NEW YORK TIMES, Summer, 2012, http://www.nytimes.com/2012/09/03/technology/companies/south-korea-reassesses-samsung-after-battle-with-apple.html?_r=1.

vices appear to use Apple's designs mostly to increase consumer sales, and Samsung's innovations appear to be derivative. If these appearances are the deeper truth, then Apple's patent protection should be strong against Samsung. Liability of Samsung to Apple will promote faster growth in the future by causing Samsung to devote more of its formidable technological prowess to fundamental inventions and less to imitations.

We have discussed the alignment between patent breath and the separation principle. For the pharmaceutical industry, an important exception to U.S. patent law aligns them closely. Firms that develop a new drug need regulatory approval before marketing it in the U.S. To gain regulatory approval, the drug's owner must prove that it is effective and safe. Proof often requires research on other chemically related drugs, including rivals for the same market. The owner of a drug has reason to refuse to license its use by a rival. A legal exemption ("Hatch-Waxman") allows firms to use patented drugs for research on regulatory compliance without a license from the patent's owner. Instead of obtaining permission to license the patented drug at the price demanded by the owner, the innovator can use the drug and pay a royalty set by legal authorities. In the language of property rights, the innovator can "take" the patentee's property right by paying an "objective price" set by the court, rather than having to pay the "subjective price" demanded by the patent's owner.

The *research exemption* for pharmaceuticals discriminates between using an innovation for innovating and using it for consumption or production, as prescribed by the separation principle. The research exemption for pharmaceuticals should ideally generalize and extend to other industries. Unfortunately, a recent case on research by universities went in the opposite direction.¹⁹

Duration

Next, apply the separation and fertility principles to the duration of a patent. The horizontal axis in Figure 6.2 indicates the length of patents in years, and the vertical axis indicates the rate of innovation. The curves in Figure 6.2 depict some possible relationships between the rate of innovation and patent length. For curve A, the rate of innovation is an increasing function of

¹⁹ Universities were traditionally exempted from the legal requirement to license patented inventions for use in research in laboratories. This traditional research exemption was eliminated in ...

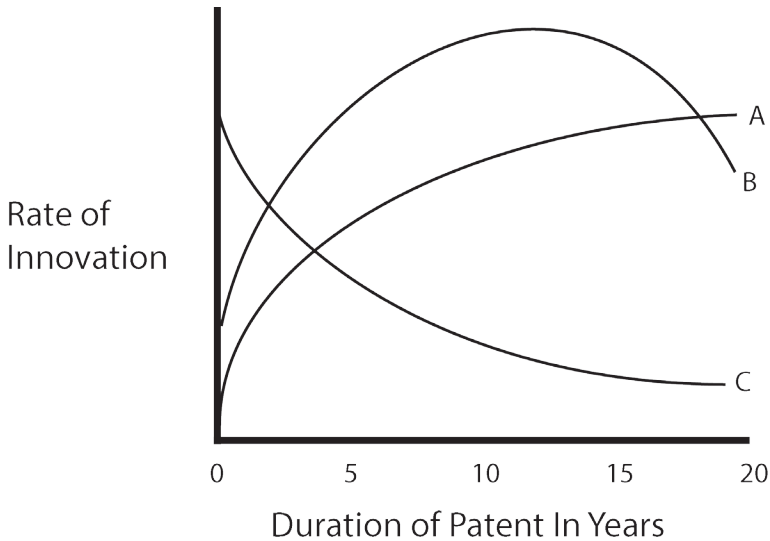
patent length. Curve A depicts an industry where innovators sell mostly to consumers and producers, such as pharmaceutical drugs. For a curve like A, optimal patent length is infinite for patents against consuming and producing an innovation.²⁰

For curve C, the rate of innovation is a decreasing function of patent length. Curve C depicts an industry where innovators of similar fertility sell mostly to each other. In Curve C the dominant effect of longer patents is to transfer wealth among dynamic users without regard to fertility. The deadweight loss from the transfers slows innovation.

Curve B depicts the intermediate case in between the pure cases of Curve A and Curve C. In the intermediate case, strengthening weak intellectual property rights increase the rate of innovation at first, but a point is reached where further strengthening decreases innovation. In Figure 6.2, the optimal patent duration of patents is internal, not infinite as for curve A or zero as for curve C. In the intermediate case, moderately strong intellectual property rights maximize the rate of innovation.

²⁰ Landes and Posner advocate patents of infinite length, which we favor for patents against consuming and producing, but not for patents against innovating. CITE;

Figure 6.2. Rate of Innovation and Patent Length



Curves A, B, and C in Figure 6.2 depict three possible relationships that often exist between the rate of innovation and the duration of intellectual property rights. (We will not discuss other possible shapes.²¹) For curves with these shapes, a simple rule leads to the maximum: Starting from 0, increase the strength of intellectual property rights so long as venture profits increase. The maximum occurs at infinity for curve A, zero for curve C, and an interior point for curve B.

The ideal patent system grants patents of different length according to the product's growth profile as in Figure 6.2. German patent authorities issue "petty" patents of short duration for minor inventions, and "full" patents of longer duration for major inventions. In contrast, American patent law issues full patents for all patented inventions. Does the German or American patent regime come closer to this ideal? For inventions mostly sold for producing and consuming, full patents will transfer more wealth from static to the dynamic uses. Consequently, American patent law is closer to the ideal than German patent law for innovations mostly used for consuming and producing. In contrast, for innovations mostly sold for innovating, the ideal law issues strong patents to fertile inventions and weak patents to sterile

²¹ Curve B is a concave function with a unique maximum. Another possibility is a convex function with two local maxima, one for weak intellectual property rights and one for strong intellectual property rights. Another possibility is multiple equilibria. For analyzing many activities, economists assume a concave function with a unique maximum.

inventions. A stronger patent for fertile innovations and a weaker patent for sterile innovations transfer wealth from the latter to the former in accordance with the fertility principle. In so far as German patent authorities issue “full” patents to fertile inventions and “petty” patents to sterile inventions, German patent law is closer to the ideal than American law for innovations mostly used for innovating.

The international default standard for patent duration, which the U.S. has adopted, is twenty years from the date of the application for a patent. Quantitative research on the optimal duration of intellectual property rights has barely begun. No compelling evidence indicates whether the pace of innovation would increase from lengthening or shortening the duration, or from distinguishing between petty patents and full patents. According to U.S. law, copyright endures for the creator’s life plus 70 years.²² Many scholars believe that this is far too long. The shadow of dead writings apparently inhibits creativity. A U.S. statute on semiconductor protection possibly has a better rule. The Semiconductor Chip Protection Act of 1984 (or SCPA) legally protects the layouts of integrated circuits for 10 years from registration.²³

Optimal Remedy

Patent law provides damages for past infringements. Three benchmarks for damages are compensation, disgorgement, and treble damages. Compensation ideally puts the *victim* in the same position as if the infringement had not occurred. For firms, this requires transferring profits from the injurer to the victim until the *victim’s* profits equal what they would have been but for the infringement. Disgorgement ideally puts the *injurer* in the same position as if the infringement had not occurred. For firms this requires transfer-

²² In October, 1998, Congress passed the Sonny Bono Copyright Term Extension Act, which lengthens copy- right protection for works created on or after January 1, 1978, to the life of the author plus 70 years, and extends existing copyrights “created for hire and owned by corporations” to 95 years. Before the change, the 1976 Copyright Act had given protection for the author’s life plus 50 years. Whatever other reasons there may be for the Copyright Term Extension Act, one justification is that it brings U.S. practice into conformity with Western European practice.

²³ The three dimensions of strength of intellectual property protection for the layouts of integrated circuits, which the law calls a chip’s “mask”, mix standards from patents and copyright. Breadth is determined by the substantial similarity” test of copyright law, duration is 10 years from registration, and the available remedies correspond generally to copyright and patent law. Protection covers all embodiments of the chip’s layout, but does not cover the chip’s functionality that patents law protects or the information on chips that copyright law protects (if they are protected at all). If the defendant asserts reverse engineering as a defense and provides “probative evidence” of it, then the standard shifts from “substantially similar” to “substantially identical.”

ring profits from the injurer to the victim until the *injurer's* profits equal what they would have been but for the infringement. Treble damages simply mean three times compensation.

For infringement of intellectual property rights, compensation is the usual remedy in U.S. law. Disgorgement is seldom available except when infringement of an intellectual property right involves breach of a fiduciary duty or unjust enrichment. When the law allows the disgorgement remedy, the victim can choose between compensation and disgorgement. Treble damages are the remedy in U.S. patent law for *willful* infringement of intellectual property rights.

The following inequalities order these remedies by the usual size of damages, or, equivalently, by their strength:

treble damages \geq choice of compensation or disgorgement \geq compensation.

In addition to damages for past harm, the owner can often obtain an injunction against future infringements.²⁴ This possibility gives the owner a choice: either prevents future infringements by enjoining them before they occur, or deter future infringements by obtaining damages after they occur. If the law allows the owner to defend his patent with an effective injunction, then others must desist from using it or pay the owner's *asking fee* for a license.

However, the law does not always give the owner a choice between injunction and damages. Sometimes the law allows damages but not injunction. If the owner defends his patent by suing for damages, then, instead of licensing, others may infringe and pay a price set by the court, not the owner's asking price.²⁵ Thus in *eBay v. Merc*, the U.S. Supreme Court held that when a

24 In addition to patent law, U.S. firms have an additional remedy against infringing imports. The U.S. firm can bring a case before the U.S. International Trade Commission asking it to ban the import of all infringing goods. This powerful remedy against foreign infringement allegedly leads to abuse by firms shielding themselves from foreign competition. See K. William Watson, *Still a Protectionist Trade Remedy: The Case for Repealing Section 337*, No. 708 POLICY ANALYSIS (CATO INSTITUTE) (2012).

25 Legal scholars say that property rights cannot be taken without paying the owner's *subjective price*, whereas the law allows other rights to be taken by paying an *objective price*. An implication of this fact is that sellers will not sell property for less than it is worth to them, and buyers will not buy property for more than it is worth them. Consequently, voluntary sales move property from people who value it less to people who value it more, which yields a surplus that can benefit buyer and seller. By creating surpluses, sales allocate resources to their most productive uses and reward innovators.

patented invention is a small component in a large product, compensatory damages can compensate sufficiently for future use of the patented good, and no injunction will be given by the courts. Consequently, the owner receives compensation instead of getting paid whatever she asks. The decision effectively compels the owner of a patent on a small component in a large product to license its use at a price close to compensatory damages.

The owner's right to choose between damages and an injunction is at least as strong a remedy as the right to damages without any choice of an injunction. The following inequality orders these two remedies by strength:

$$\text{compensation or injunction} > \text{compensation.}$$

The preceding inequalities characterize different remedies for infringement by their strength. To maximize growth, how strong should the remedy be? The separation principle distinguishes between static (consumption and production) and dynamic (innovation) activities. Infringements for static uses should have strong remedies, in order to transfer more resources from static to dynamic activities and increase the pace of innovation. For past infringements by consumers and producers, the separation principle favors treble damages instead of a choice between disgorgement or compensation, and it favors a choice between disgorgement or compensation instead of compensation without a choice. For future infringements by consumers and producers, the separation principle favors the choice of compensation or an injunction instead of compensation without a choice.

The fertility principle distinguishes between the usual innovations and very fertile innovations. Infringements for dynamic uses involving the usual innovations should have weak remedies in order to reduce the excess burden on innovators. Consequently, the fertility principle usually ranks remedies the opposite of the separation principle. For past infringements by innovators, compensation without a choice is preferred to a choice between disgorgement or compensation, and a choice between disgorgement or compensation is preferred to treble damages. For future infringements by innovators, the fertility principle favors compensation without a choice instead of the choice of compensation or an injunction. In contrast to the usual cases, unusual cases concern infringements of patents for fundamental innovations that are

very fertile. Here the fertility principle ranks remedies the same as the separation principle.

Table 6.2 summarizes these conclusions about the preferred remedy for economic growth.

Table 6.2. Preferred Strength of Remedy for Infringement (“P”)

	Time of Infringements		
		past infringements	future infringements
Infringing Use	consuming or producing	strong treble P choice P compensation	strong injunction P compensation
	innovating, usual case	weak compensation P choice P treble	weak compensation P injunction
	innovating, very fertile case	strong treble P choice P compensation	strong injunction P compensation

Table 6.2 compares compensatory damages to the strength of two other remedies. Compensatory damages also differ in strength according to their basis of computation. Production costs and development costs often figure as elements in compensation for infringement. Compensation can include the cost of the invention’s production and development, which suggests this ordering of compensatory damage measures for patent infringement:

production and development costs \geq development costs \geq no compensation.

The same reasoning as in Table 6.2 motivates the relationship between the infringing use and the preferred measure of compensatory damages that Table 6.3 depicts.

Table 6.3. Preferred (“P”) Damages for Infringement

Infringing Use		Damages for infringement
	consuming or producing	high production & development P development P nil
innovating, usual case	low nil P development P production & development	
innovating, very fertile case	high production & development P development P nil	

To illustrate the use of the tables, return to the case of *eBay v. Merc*, where the court held that compensatory damages can compensate sufficiently for infringing a patent on a small component in a large product. This decision is apparently consistent with Table 6.2, which indicates that a patent owner should receive compensatory damages for infringement by innovators in the usual case. The “usual case” involves an innovator’s infringement of a patent with low or average fertility, and a “small component in a large product” presumably has low or average fertility. Furthermore, Table 6.3 indicates that compensatory damages should preferably exclude development costs or extraordinary profits. (Notice that the rule of compensatory damages replicates the cost structure in a patent pool, which eliminates market power and deadweight loss.²⁶) In the exceptional case where one patented innovation is much more fertile than others, however, Table 6.2 requires a strong remedy – injunction or high damages.

Thickets and Holdup

Bell Labs developed the computer operating system Unix in the 1960s. Subsequently, many different programmers created variants of Unix, including the version powering Apple computers (BSD developed at the University of California at Berkeley). Chapter 5 described circular transactions among innovators as an “innovation thicket.” In an innovation ticket, everyone needs the innovations of others in order to make their own innovations, as with variants of Unix. We will apply Tables 6.1, 6.2, and 6.3 to patent thickets.

Unix is “unowned” or “open access” or “in the public domain.” Its users are unrestricted by patent or copyright, although it is trademarked.²⁷ Since no one privately owns Unix, using it does not require negotiating with anyone. If someone owned Unix, however, anyone who wanted to improve it would have to license its use. Negotiating a license takes time and effort, which reduces profits and slows innovation. An innovation thicket contains so many circular transactions that bargaining over property rights slows growth. Innovations in a thicket ideally remain in the public domain where anyone can use them freely, like Unix. If not in the public domain, then innovations in a

²⁶ In a patent pool, anyone is free to produce and use another’s patented good, without paying a license fee. So anyone selling its patented good to another member of the pool must charge a price that does not exceed the good’s production cost. If courts award compensation for production and not development among innovators in a patent pool, costs replicate a patent pool.

²⁷ Only users conforming to certain specifications can apply the name “Unix”.

thicket should, according to the fertility principle, receive weak patent rights (narrow breadth in Table 6.1, weak remedy in Table 6.2, low damages in Table 6.3). In the unusual case of a pioneering patent in a thicket, however, its high fertility warrants strong patent rights (wide breadth in Table 6.1, strong remedy in Table 6.2, high damages in Table 6.3).

Three different kinds of private transactions among innovators can sometimes overcome the patent thicket problem. Large firms often accumulate a rich patent portfolio. When firm A needs to use a patent in B's portfolio, sometimes B needs to use a patent in A's portfolio, so they can negotiate *cross licenses*. Or, instead of waiting until the need arises, cross licensing can be negotiated in advance. Each member of a *patent pool* agrees in advance to cross-license its patents for a given technology to everyone else.²⁸ Or, instead of forming a pool, owners of complimentary patents can *merge* into a single firm that holds all of the patents required to develop an innovation.²⁹

Each of these solutions – cross licensing, pools, and mergers -- has limited success in solving the patent hold out problem.³⁰ In practice, firms cross license or pool a small fraction of patents, and small firms often resist merging for fear that size will impede creativity.³¹ Private transactions among individuals often fail to solve the challenge of innovation thickets because of a problem that game theory explains. If innovations in a thicket are patented, then each venture to develop an innovation needs licenses from several other innovators. When several patent owners have the power to withhold licenses that are necessary to develop an innovation, then each of them can effectively veto the venture. Given multiple vetoes over a venture, every veto holder must consent for the venture to proceed. If everyone with veto power consents except for one, the last hold out can demand most of the venture's value in exchange for his consent. Since a patent holder gains by consenting last, no one wants to consent first. As the number of patent-holders with veto power increases, the probability increases that someone will hold out for a higher license fee.

28 A subsequent chapter on contracts discusses cross licensing and pools.

29 A subsequent chapter on antitrust discusses mergers among innovators.

30 Each firm wants to give the pool its least valuable patents, and each firm wants to use the most valuable innovations of others in the pool. This is the problem of adverse selection.

31 Merging dilutes the concentration of ownership needed to solve the double trust problem, as mentioned in Chapter 3 and examined further in a subsequent chapter on antitrust.

Behind this bargaining story lies a technical fact. When distributing the value of a joint product, the usual rule in economics is for each participant to receive his marginal product. The marginal product of a participant in a cooperative venture equals the increase in the venture's value when he participates in it (holding the participation of others constant).³² In a patent thicket, many people own essential patents for further innovation. Consequently, the holdout's marginal product is large, possibly equal to the venture's entire value or nearly so. Furthermore, anyone can hold out who owns an essential patent for the venture. In these circumstances, the sum of the marginal products of the contributors to the venture exceeds the venture's total value. The venture does not have enough value for each contributor to receive his marginal product. Demands that are reasonable individually are unreasonable collectively. If the participants negotiate and each one demands his marginal product, the venture is doomed.³³ (For the same reason, marginal product provides no basis for computing compensatory damages or reasonable fees.³⁴)

Given its intractability, private actors need law's help to solve the hold out problem. Law can help by weakening patents among innovators in a thicket. Narrowing a patent's breadth weakens its market power and lowers the price that a license commands. In *Apple v. Motorola*, Judge Posner recently followed this prescription by dismissing Apple's infringement claims. Law can also help to solve the holdout problem by weakening the remedy for patent infringement. As mentioned above, in *eBay v. Merc*, the U.S. Supreme Court held that the weaker remedy – damages instead of injunctions – applies to small components in a complex innovation.

32 This is usually equivalent to the decrease in the venture's value when he stops participating in it (holding the participation of others constant).

33 In more technical language, increasing returns to the scale of a coalition implies that the sum of the marginal contributions of each member exceeds the coalition's total product, and the game's "core is empty".

34 Damages for using another's property often equal the marginal product that the owner lost when the infringer used his property (compensatory damages), or else damages equal the marginal product that the trespasser got from using another's property (disgorgement damages). In a thicket, "compensatory damages" or a "reasonable fee" for using a patented good without a license cannot refer to its marginal product, because the sum of the marginal products exceeds the total product. Marginal productivity theory is unreasonable for computing damages from infringement or negotiating shares in cross-licenses.

Middlemen and Trolls

“Middlemen” refers to intermediaries who buy goods for resale. Throughout history, people have viewed middlemen with suspicion because they make money without making goods. Confusion abounds about their economic role. Do they increase or decrease prices to consumers? Do they create cartels or match buyers and sellers?³⁵ The same confusion infects discussions about patent middlemen, who buy patents to enforce them, not to produce or innovate. According to the Norwegian fairy tale of the “Three Billy Goats Gruff”, a troll living under a bridge gobbled up pedestrians who tried to cross. Similarly, some law firms gobble up patents and sue others who need them for production or innovation. “Patent troll” is the derogatory term for “patent middlemen.”

Name calling aside, law requires analysis of patent middlemen. After a business venture develops and patents an innovation, the innovation must be marketed to users and enforced against infringers. Innovating, marketing, and enforcing are different specialties that different people can perform. With respect to enforcement, the innovator can directly defend her patent or she can indirectly defend by selling the right to collect damages to someone else. Thus firm A invents and patents α . A can enforce the patent on α , or A can sell α to B, or A can sell to B only the right to sue and collect damages from infringers of α .

By facilitating middlemen, the law gives patent owners more enforcement choices. Innovators tend to choose the most profitable form of enforcing patents, whether by themselves or others. More enforcement choices strengthen the patent owner’s rights *de facto*, but not *de jure*. Stronger patent rights can stimulate or stifle innovation, as previous discussions show. According to the separation principle, increasing a patent’s strength against consumers and producers stimulates innovation. By prosecuting infringers who use an invention to produce consumer goods, patent middlemen transfer resour-

³⁵ Middlemen often save buyers and sellers the cost of finding each other, but they sometimes form cartels to separate buyers and sellers, and to extract higher prices. Thus middlemen try to prevent buyers from obtaining information about sellers from any source but themselves. To illustrate, the American Bar Association historically prohibited its members from advertising their specialties, including printing a specialty on a business card. Consequently, a person who needed, say, a patent attorney often had to ask the family attorney for a reference. This restriction on information generated referral fees for lawyers. Such predatory practices by middlemen cause hostility and resentment among those who must pay for it. The U.S. courts eventually struck down such legal rules as violations of “commercial free speech”.

es from consuming to innovating, which speeds progress in the useful arts. To increase progress in the useful arts, law should facilitate sales of patent rights to middlemen for innovations used in consumption or production. Conversely, increasing a patent's strength imposes a deadweight loss on innovators in the usual case. By prosecuting infringers who use an invention to innovate, patent middlemen impose a deadweight loss on innovators, which slows progress in the useful arts. To increase progress in the useful arts, law should inhibit sales of patent rights to middlemen for innovations used in innovation. In the usual case of a very fertile patent, however, law should facilitate sales of patent rights to middleman. Perhaps the derogatory word "troll" should be reserved for patent middlemen who stifle innovation.

Conclusion

Intellectual property law, which includes patents and copyright law, establishes the ownership of innovations by people. It conveys a bundle of rights to creators as determined by rules. The normative question of growth economics is, "Which ownership rules maximize innovation?" In order to increase the pace of innovation, ownership rules should open competition to innovate and increase the winner's profits. This chapter poses the question, "Which ownership rules maximize profits among ventures in open competition to innovate?" Answering the normative question of growth economics requires comparing two opposite effects of stronger intellectual property rights. The first effect is a transfer from static activities (consumption and production) to dynamic activities (innovation), which increases the profitability and pace of innovation. According to the overtaking theorem explained in Chapter 1, transfers from static activities to dynamic activities increase the welfare of people. So the first effect favors strong intellectual property against infringement by consumers and producers. The second effect is the dead-weight loss of licensing dynamic activities, which decrease the profitability and pace of innovation. So the second effect favors weak intellectual property rights against infringement by innovators.