Notes

Introduction

1. Or, more accurately, if there was innovation, it was exogenous, not affected by what market participants did.

2. There are, of course, other exceptions. For instance, if risk markets are incomplete, free trade can lead to a Pareto inferior equilibrium; free trade increases risk, inducing investment away from high-risk/high-return sectors (see Newbery and Stiglitz 1982).

3. Note that this critique of the infant-industry argument is distinct from the usual criticisms based on political economy—that there is "capture" by vested interests. We deal with these issues below.

1. The Learning Revolution

1. How they make those choices will, of course, have profound effects on *measured* growth, since increasing leisure does not show up in conventionally measured GDP (see Stiglitz, Sen, and Fitoussi 2010).

2. This point was emphasized by Keynes (1930). See Stiglitz (2008d) as well as other chapters in Pecchi and Piga (2008).

3. The difference was referred to as the Solow residual. While technical change accounted for *most* of the residual, there were other factors, including the reallocation of labor from low productivity sectors to high productivity sectors (see Denison 1962).

4. Griliches and Jorgenson's work (1966, 1967), which entailed using alternative calculations of the value of capital, suggested a much smaller role for technical progress. Further problems were identified in the quantification of labor input, as economists attempted to assess the role of human capital

in economic growth (Klenow and Rodríguez-Clare 1997; de la Fuente and Doménech 2006).

5. There was a large literature describing how new technologies were "embodied" in capital goods. See, e.g., Solow (1962b) and the discussion and references cited in Stiglitz and Uzawa (1969).

6. For any doubters: Engage in a thought experiment in which primitive farmers accumulated more hoes, or even built more irrigation canals. If that primitive accumulation was *all* that had occurred during the past 200 years, standards of living would be incommensurately lower than they are today.

7. As Solow (1956) pointed out, an increase in the savings rate simply leads to an increase in per capita income, not to a (permanently) higher rate of growth. See the further discussion in the next chapter.

8. See Stiglitz (1998c), which describes development as a "transformation" into a society which recognizes that change is possible, and that learns how to effect such changes.

9. This work includes that of Kaldor (1957, 1961); Kaldor and Mirrlees (1962); Uzawa (1965); Nordhaus (1969a, 1969b); Atkinson and Stiglitz (1969); Inada (1963); and Shell (1967) and the papers contained in that volume. Not only did this early research address the question of the rate of technological progress, but also the direction (see Kennedy 1964; Samuelson 1965; Fellner 1961; Drandakis and Phelps 1966; Ahmad 1966; and others).

Of course, economic historians have long sought to explain the rate and direction of innovation (see, e.g., David 1975; and Salter 1966). Hicks (1932) made even earlier contributions to this field. More recent work building on these traditions includes Stiglitz (2006b).

A major stumbling block in the development of endogenous growth theory was that technological progress introduced a natural nonconvexity into the production function: if the economy was characterized by constant returns to scale in conventional factors, then, if technological progress were made endogenous, it would exhibit increasing returns in all factors. That is, if $Q = F(K, AL)$ is homogenous of degree one in K and L, where Q is output, A is the state of technology (technological progress is pure labor augmenting), K is capital, and L is labor, and if $A = A(L_R, K_R)$ describes the level of technology as a function of resources (capital and labor) devoted to research, then Q exhibits increasing returns overall. This posed problems both for normative (optimization) and descriptive models. Only under special values of the parameters of special formulations of the production function would a steady state exist. One could easily construct models that "worked," e.g., by assuming that $d(\ln A)/dt = \phi(L_R e^{-nt})$ where n is the rate of growth of population. Then, in a steady state with a fixed fraction of income saved, s, and with a fixed fraction of the labor force γ allocated to research, $d(\ln A)/dt = \phi(\gamma)$, and $d(\ln Q)/dt =$

$n + \phi(\gamma)$. But in the more general case, with say a fixed proportion of the labor force allocated to research, it is easy to see that in the limit $d(\ln A)/dt$ may either rise to some upper bound or fall to some lower bound. One could, alternatively, assume that there was a fixed factor, say land. In the absence of technological change, if the elasticity of substitution was less than unity, with increasing population, growth in income per capita would eventually halt. Technological change could offset the effects, say if it were "land" augmenting. But again it seemed a borderline case, where if a fixed proportion of the labor supply (or capital, or output) were allocated to land augmenting technological progress and similarly for labor augmenting progress, somehow the pace of innovation just offset the effects of the diminishing land per capita.

There was a second problem, discussed below: The nonconvexities naturally led to imperfections of competition.

10. This work includes the early work of Dasgupta and Stiglitz (1980a, 1980b) trying to endogenize both market structure and the rate of technological progress, subjecting some of Schumpeter's conjectures to more rigorous analysis. Other work includes that of Gilbert and Newbery (1982). Romer's (1986, 1990) work provided inspiration for much of the later work in this area. See Aghion and Howitt (1998) and Romer (1994) for surveys.

11. Arrow's 1962 papers (1962a, 1962b) are the classic references. Key properties of knowledge and its production (knowledge as a public good, nonconvexities associated with the production of knowledge, inherent capital market and risk market imperfections) are discussed at length below, with further references. See, in particular, Stiglitz (1987b; based on a 1978 lecture).

12. We do not, unfortunately, use the framework that we develop in this book to provide answers to two key historical questions: What happened to suddenly change the world, to initiate the process of becoming a "learning society"? And why did this process begin where it did and when it did? A few reflections on these questions are contained in later chapters.

13. The concept of "learning to learn" is developed by Stiglitz (1987a).

14. Gordon "suggests that it is useful to think of the innovative process as a series of discrete inventions followed by incremental improvements which ultimately tap the full potential of the initial invention" (2012, 2).

15. This idea gave rise to a vast literature on "embodied technological progress," closely related to work on "vintage capital," referred to earlier in note 5. The dynamics of economies with "putty clay" (where technology is fixed, once the machine is made) are markedly different from those of the "putty-putty" models that dominate in modern macroeconomics (see Solow 1959; Solow et al. 1966; Cass and Stiglitz 1969).

16. That is, taking into account the costs of creating markets or obtaining information (Greenwald and Stiglitz 1986, 1988).

17. Although Arrow did not frame the market failure in this way, it was clear from his analysis that knowledge was a public good, in the sense that Samuelson had defined public goods a few years earlier (Samuelson 1954). See the discussion below.

18. Of course, the principles are often contested, and the political influence of banks has, so far, been successful in resisting the adoption of many of the regulations around which there is broad consensus, at least among economists not associated with the financial sector.

19. Because industrial policies were often looked upon disparagingly in the years in which neoliberal economic doctrines predominated, some political leaders have looked for other terms to describe such policies, such as "proactive business policies." Alternatively, they have focused on particular categories of such policies (which typically meet with greater approval), such as export promoting policies. We will stick with the more conventional nomenclature.

20. This perspective was reflected in *Knowledge for Development*, the first World Bank Development Report done during Stiglitz's tenure as chief economist of the World Bank (World Bank 1999; see also Stiglitz 1998c, 1999b).

21. Most clearly articulated by Samuelson (1948).

22. For a more general discussion of leapfrogging (in the context of patent races), see Fudenberg et al. (1983).

2. On the Importance of Learning

1. Later discussion will explain why we also do not believe that these differences can be explained by the usual kinds of static inefficiencies, e.g., associated with distorted incentives.

2. To be sure, there may be problems of identification—when changes in capital are required to put into place changes in technology.

3. We recognize that the magnitude and sources of China's increase in productivity have been the subject of some controversy. For a contrary view, see Young (2003), who estimates productivity growth of only 1.4 percent for the nonagriculture sector from 1978 to 1998. There are many pieces of evidence collaborating the rapid increase in standards of living and output, e.g., trade statistics, and we suggest below why studies suggesting that total factor productivity growth was low are unconvincing.

4. Studies that suggest that total factor productivity growth has been low typically ignore the lag structure involved in human capita (see, e.g., Fleisher,

Li, and Zhao 2010). Moreover, many of these studies simply assume that the factor shares represent competitive returns; in the East Asian countries (and especially China), there is a presumption that that is not the case.

It is hard to reconcile a real return to investment in excess of 10 percent (or even 5 percent) with the patterns of investment, e.g., heavy investment in low-return infrastructure, or investments in U.S. government bonds, yielding real returns that are low or even negative (though the social returns to reserves may be somewhat higher, especially in the early years, when reserves were smaller; chapter 11 presents an alternative explanation for the accumulation of reserves). Studies that suggest low rates of factor productivity growth implicitly are assuming high rates of return to investment (or that is an implicit implication of their econometric analysis).

5. As we emphasize in Chapter 3, incentives are relevant not just for investment and labor supply, but also for learning.

6. For a recent discussion, see Zhu (2012), who argues that productivity growth has been central to economic growth since 1978. As Zhu notes, China's capital-output ratio has grown little since 1978. While in the initial period of "reform"—1978 to 1988—productivity growth came from agriculture, but between 1978 and 2007, non-state-sector "productivity growth contributed 2.27 percentage points per year to aggregate productivity growth" (p. 119). As we emphasize later in this chapter, improvements in allocative efficiency result in a one-time gain in productivity, not the persistent improvements that were observed, say, in China. (By contrast, Zhu argues, "Overall, gradual and persistent institutional change and policy reforms that have reduced distortions and improved economic incentives are the main reasons for the productivity growth" [p. 104].)

7. Though some studies do show high levels of TFP (total factor productivity growth) for the manufacturing sector in Eastern European countries, the small size of these sectors meant that there was relatively little impact on aggregate productivity. Brandt et al. (2012) find for 1998 to 2007, (p. 340) "firm-level TFP growth of manufacturing firms averaging 2.85 percent for a gross output production function and 7.96 percent for a value added production function."

8. In some countries, like the Czech Republic, multinationals did successfully bring in best practices. The highly educated labor force facilitated the requisite learning. In several of the seemingly successful countries of Eastern and Central Europe, the 2008 crisis made it evident that at least a significant part of this was related to a real estate boom/bubble.

9. Persistent differences across regions in many countries (such as Italy) are evidence of the deficiencies in the standard explanations focusing on artificial barriers to the movement of goods, services, or factors.

10. See, e.g., Foster, Haltiwanger, and Krizan 2001 for a survey. Hsieh and Klenow (2009) find very large gaps in the marginal products of capital and labor across plants within India and China.

11. This analysis does not explain the sources of the differences in learning ability. For example, is it due to differences in management/culture or differences in investments in learning?

12. The unemployment rate in the United Kingdom soared from 4.7 percent in 1979 to 11.2 percent in 1986 (OECD 2011), even as industry production, for instance, increased by nearly 5 percent over the same period (after recovering from an initial drop in the beginning of the decade). (World Bank data accessed through Google Public Data Explorer, GDO production, constant 2000 U.S.$, disaggregated by sector, http://www.google.com/publicdata/directory, accessed February 26, 2013).

13. One alternatively might have argued that the strike provided greater incentives for efficiency. But if that were the major explanation, why hadn't management adopted incentive structures to encourage these greater efficiencies, which would have saved enormous amounts of labor? The savings would have provided more than adequate compensation for the additional effort. Moreover, this and similar episodes exhibit hysteresis effects: once the organization has learned how to be more productive, productivity remains at relatively high levels even after the exigency which gave rise to the productivity rise is resolved.

14. A single episode of a productivity increase might be attributable to the removal of a static inefficiency; repeated increases should be seen much more as evidence of episodic learning—including possibly learning about how to remove certain static inefficiencies.

There are often periods of negative productivity change. Such periods reinforce the conclusion that much of the action in productivity occurs well *inside* the production possibilities curve.

15. Interestingly, some of the learning involved learning from foreign firms, e.g., about quality circles and just-in-time production (see, e.g., Nakamura, Sakakibara, and Schroeder 1998).

16. Total investment in the United States held steady at between 18.6 percent and 20.9 percent between 1995 and 2001, beginning at 18.6 percent and ending the period at 19.3 percent From 1981 to 1994 it ranged from 17.1 percent in 1991 to 22.3 percent in 1984, tending to decrease over the period (see the World Economic Outlook database of the International Monetary Fund, available at http://www.imf.org/external/pubs/ft/weo/2012/02/weodata/index.aspx, accessed February 26, 2012). Gross expenditures on R & D during the same period increased slightly from 2.5 percent to 2.7 percent; from 1981 to 1994 it ranged from 2.3 percent to 2.8 percent (see the indicators of the National

Science Board, available as table 4-19 at http://www.nsf.gov/statistics/seind12/c4/c4s8.htm#top, accessed February 26, 2013).

17. Some of the learning was related to computerization; some of the learning was learning how to exploit differences in costs between, say, the United States and China by constructing a global supply chain.

18. With monopolies in the consumer goods industries, the economy still operates along the production possibilities curve, but not at the point along that curve which maximizes societal welfare. With monopolies or imperfections in competition in inputs, however, the economy will not operate along the production possibilities curve.

19. The distinction is really artificial, once we take a broader account of production and value production correctly. That is, we need to take into account not only the good things that firms produce (steel) but the bad things (pollution) that are a byproduct of the good things that they produce. Once we do that, claims that controlling pollution may be good for consumers but bad for GDP make little sense. See the International Commission on the Measurement of Economic Performance and Social Progress (Stiglitz, Sen, and Fitoussi 2010).

20. We do not explore here those distortions or their interactions with learning and productivity growth. Our 2003 book lays out our interpretation of these macroeconomic disturbances. In Greenwald, Salinger, and Stiglitz (1990), we lay out the links between productivity growth and the business cycle (see also Stiglitz 1994c, 2006b; Greenwald, Levinson, and Stiglitz 1993).

21. With learning externalities (a major focus of this book), if learning is a function of investment, then individual decisions concerning learning will systematically result in too little investment.

22. Again, this was a major point made by the International Commission on the Measurement of Economic Performance and Social Progress (see Stiglitz, Sen, and Fitoussi 2010). There is now a large movement attempting to find and employ better metrics of economic performance and social progress.

23. For a developing country, facing a constantly moving frontier, there is a huge "one-time" gain in reducing the gap with the frontier, a gain which is an order of magnitude greater than the one-time gain from eliminating allocative inefficiencies. Once the gap is narrowed (or even better, eliminated), then productivity can only increase at the rate at which the frontier moves out.

24. We have noted that some of the disparities between what the economy or firm could produce and what it does can be related to distortions—allocative inefficiencies—within firms, but eliminating these inefficiencies can provide a one-time gain in productivity but not the kind of sustained increases in productivity with which we are concerned.

25. This formulation assumes that there are no learning spillovers across the firms that are at the frontier (or that there is a single firm on the frontier).

26. More generally, the nonconvexities associated with the production and acquisition of knowledge imply that the profit functions may not be single-peaked. Even if firms started with the same knowledge base, there might be no equilibrium in which all acquired knowledge grows at the same pace. The only equilibrium may entail some firms saving on investment in R & D and "poaching" off the knowledge acquired by others. The present discounted value profits for the two strategies might be the same. Later chapters will show that even if there exists a symmetric equilibrium, it is unstable. Chapter 11 develops an analogous model where some countries lag behind others in equilibrium.

3. A Learning Economy

1. In particular, beginning with the work of Stiglitz (1975a, 1975b), Rothschild and Stiglitz (1976), Akerlof (1970), and Spence (1973).

2. Hayek (1945) called attention to the problem of dispersed information—explaining it was precisely because information was so dispersed that central planning could never work. But somewhat inconsistently, he believed that the price system provided an efficient way of aggregating and transmitting information. As Stiglitz (1994c) has argued, if that were the case, then market socialism would have worked. Hayek never formalized his ideas. Later Chicago economists put forward the efficient markets hypothesis, but they too never created formal models to see whether markets efficiently aggregated and transmitted information. Grossman and Stiglitz (1976, 1980) showed that, in fact, they did not. The crisis of 2008 seems to have settled in most people's minds any lingering doubts about the efficient markets hypothesis.

3. Though, of course, there were other reasons for the failure of our regulatory system, e.g., related to special-interest politics.

4. The concept of learning to learn was developed in Stiglitz (1987a).

5. In 1977 Paul MacCready won the £50,000 Kremer prize offered by the Royal Aeronautical Society for a human-powered airplane with his Gossamer Condor (see http://aerosociety.com/About-Us/specgroups/Human-Powered/Kremer).

6. See Kanbur (1979) and Kihlstrom and Laffont (1979) for the canonical presentation of entrepreneurship in the setting of occupational choice. Emran and Stiglitz (2009) explain why competitive markets may do a bad job learning about who are good entrepreneurs.

7. To be sure, many developed countries showed in the 2008 crisis that they too had much to learn about risk management.

8. Supporting evidence includes Asher (1956) and Alchian (1963) on airframe production, Zimmerman (1982) on nuclear power technologies, Lieberman (1984) on production and investment in chemical process industries, and

Hollander (1965) on R & D. More recent studies include those focusing on learning by management, in rayon, in semiconductors, and in fuel-cell technology. See, e.g., Walters and Holling (1990); Jarmin (1994); Dick (1991), Gruber (1998); Argote, Beckman, and Epple (1990); Argote and Epple (1990); Barrios and Strobl (2004); and Schwoon (2008). Thompson (2010) provides a recent survey.

9. Arrow's work also gave rise to an extensive theoretical literature. See, in particular, Spence (1981); Fudenberg and Tirole (1982); Jovanovic and Lach (1989); Malerba (1992); Lieberman (1987); Leahy and Neary (1999); Ghemawat and Spence(1985); Young (1991, 1993); and Dasgupta and Stiglitz (1988a). As we note in chapter 5, some of these papers assume market structures that could not plausibly survive in the long run.

10. Warren Buffett, in his 2001 chairman's letter to Berkshire Hathaway investors, available at http://www.berkshirehathaway.com/2001ar/2001letter.html.

11. By the same token, societies (individuals) can develop capacities for learning how to use their leisure well, or they can enhance their capacities to enjoy consumption goods. (This can be thought of as improving individuals' capacities to translate inputs of time and goods into "enjoyment.") Thus, learning also can have large effects on consumption behavior (see Stiglitz 2008d).

12. Later in this chapter and in chapter 16, we will discuss societal characteristics that may facilitate learning and learning to learn.

13. See, in particular, the appendix to chapter 4.

14. An interesting aspect of a failure to adapt is that the school year in many countries is still related to the agricultural calendar—decades after that sector's decline to but a few percent of the labor force.

15. While we emphasize in this book technological knowledge, which enhances the ability to transform inputs into outputs, at every level there are other forms of "learning," e.g., changes in institutions or changes in beliefs, say about the way the economy or society functions. As we explain later, such changes in beliefs may not be based on an accurate analysis of the world and may in fact be counterproductive in terms of creating a learning society (see, e.g., Hoff and Stiglitz 2010, 2011; and the papers cited there).

16. Famously, Google has had a policy of allowing employees to dedicate 20 percent of their workweek to pursuing independent projects.

17. There is a small, but important, literature on an economy's innovation system (Nelson 2004, 1993; Freeman, 1987; Lundvall, 2010). There is also some writing on the "creative economy" (e.g., Florida 2002). Closer to what we have in mind is the work on, for example, agglomeration externalities. Moretti (2011), in his survey, groups the sources of agglomeration externalities in local markets into three broad bins: thick labor markets, thick markets for intermediate inputs, and knowledge spillovers. We might think of the latter as

(in our vocabulary) "learning spillovers." Moretti provides references for the current state of knowledge about learning.

18. Earlier, we noted another aspect of learning—learning the comparative advantages (skills) of different individuals. This is a central function of educational systems and is referred to as "education as a screening device." Because of marked differences between social and private return to such screening, market allocation of resources to such screening are not efficient (see Stiglitz 1975b, 2009).

19. There are other differences between young and old that may affect learning behavior. The major asset of the young is their human capital, and there is considerable uncertainty about the value of their human capital. One rational response (but not the only possible response) to this uncertainty is to increase investment in learning.

20. The new learning may, in fact, make the knowledge of those who are older obsolete. While, in the context of the standard competitive paradigm, individuals take the value of assets (including human capital) as given, in the small-scale microeconomics of the workplace, an increase in knowledge (learning) by one party can affect the value of human capital of others.

21. It is worth noting that similar considerations may have played an important role in explaining some of the differences in the transition from Communism to the market between, say, Russia on the one hand and Poland on the other. Russia had large, centralized firms, and these were, for the most part, retained as part of the transition. These firms were dominated by older managers. In contrast, Poland had more medium-sized firms and divided more of its large firms up in the process of the transition, providing a greater role for younger managers (see, e.g., Stiglitz 2002a, 2000c; Ellerman and Stiglitz 2000, 2001).

22. Letter from Isaac Newton to Robert Hooke, February 5, 1676.

23. See, e.g., the discussion in Henry and Stiglitz (2010). The far-reaching America Invents Act of 2011, which came into force in March 2013, has significantly deemphasized the need for disclosure in order to enforce patents. The law and related documents may be viewed at http://judiciary.house.gov/issues/issues_patentreformact2011.html (accessed on January 8, 2013).

24. Moreover, as we noted before, the patent system has encouraged a culture of secrecy, and some worry that the intrusion of IPR into the academic setting, through the Bayh-Dole Act, has had similar effects in universities, undermining the traditional academic norm of openness.

25. The General Assembly of WIPO adopted the Brazilian and Argentinean proposal for a developmentally oriented intellectual property regime on October 4, 2004.

26. And again, there is a worry that intellectual property creates incentives that undermine this traditional model.

27. Moreover, transportation systems are often centered around national hubs. It is easier to move within a country. Institutions and institutional knowledge are also likely to be local.

28. We again note that not all ideas are actually conducive to learning. As we argue extensively in this book, Washington Consensus policies which spread through globalization may have impeded learning.

29. This has provided one of the rationales for why advocates of trade liberalization, such as Grossman and Helpman (1991), suggest that enhanced trade will lead to more learning. As we explain later, there may be other, more than offsetting, effects.

30. But again, the effects can be ambiguous, as individuals are able to create on the Internet communities of like-minded people, reducing the exposure to new ideas (see Sunstein 2001).

31. The ideas in this paragraph are developed more extensively in Stiglitz (1998c). Chapter 16 (based on Hoff and Stiglitz 2010, 2011) elaborates on the role of cognitive frames and how they are shaped.

32. There were, of course, both technological and institutional changes, but they occurred very slowly. There were slow changes in farming technologies that evolved over time. The new world provided new crops, the use of which spread gradually over the entire world. In their time, feudalism and its end, and slavery and its end, represented important institutional changes. The enclosure movement was another institutional change with profound consequences. Many of the changes in technology and institutions were precipitated by exogenous events, such as the Black Plague (see, e.g., Ruttan and Hayami 1984).

Clearly, the Enlightenment helped create the cognitive mindsets that were conducive to innovation and change. Joel Mokyr (2009) suggests that the reason the Industrial Revolution began in England, rather than somewhere else in Europe, has a lot to do with social mindsets, e.g., the belief in the possibility of progress and social norms of honest dealing among businessmen.

33. This is, of course, a simplification. Conniff (2011) writes that this image of Luddites was due to particularly skillful branding, and that in reality the Luddites were not against machines, but "confined their attacks to manufacturers who used machines in what they called 'a fraudulent and deceitful manner' to get around standard labor practices."

34. According to a recent Pew Poll, a third of Americans do not believe there is solid evidence the earth is warming, and nearly 60 percent do not think that warming is mostly because of human activity (Pew Research, 2012, "More Say There Is Solid Evidence of Global Warming," October 15, http://www.people-press.org/2012/10/15/more-say-there-is-solid-evidence-of-global-warming/, accessed February 26, 2013). And according to a Gallup poll, nearly 50 percent of Americans do not believe in human evolution (http://www.gallup.com/poll/155003/hold-creationist-view-human-origins.aspx,

accessed February 26, 2013). If it is difficult to change people's beliefs about matters on which there is such overwhelming scientific evidence, it should be obvious that beliefs about our social and economic system may persist, even in the face of considerable evidence to the contrary. (Beliefs about markets being efficient and stable provide but one instance. While there was a wealth of theory, empirical evidence, and historical experiences suggesting otherwise before the 2008 crisis, it is remarkable how that crisis left the beliefs of so many adherents of "market fundamentalism" essentially unshaken.)

35. The ideas in this and the following paragraphs represent joint work with Karla Hoff and are elaborated in Hoff and Stiglitz (2010, 2011).

36. But that doesn't fully explain why these ideas are adopted by some individuals and groups and rejected by others. Chapter 16 provides a partial explanation.

37. While there are some immediate and important policy implications of these observations, we are not able to pursue them further in this book.

38. These ideas are developed more extensively in Stiglitz (1995b) and Sah and Stiglitz (1987a, 1987b).

39. They are also examples of ideas that are hard to protect with patents, though in some cases, America's business-process patents attempt to do so.

40. We do not comment here on whether their empirical approach really does capture fully the set of related capabilities. Since their work (see also Hidalgo and Hausmann 2009), alternative approaches to characterizing the product space have been explored (see, e.g., Jarvis 2013; Pietronero, Cristelli, and Tacchella 2013). The effects of an improvement in one sector on other sectors depends not just on the similarity of those sectors, but on the institutional arrangements, e.g., providing scope for exploiting linkages. Thus, the fact that natural resource sectors have traditionally not been closely linked to other sectors may be partly a result of the absence of effective industrial policies and the exploitive relationships often evidenced in that sector.

41. See chapters 4 and 12 for further discussions of why geography matters.

42. As the previous paragraph explained, even the notion of the production function is questionable, because without further research and development, we simply may not be able to produce with technologies markedly different from those currently in use. But putting that qualification aside, the standard formulations have $Q = F(\mu K, \lambda L)$, where Q is output, K is capital input, L is labor input, μ is the level of capital-augmente technological progress, and λ is the level of labor-augmenting progress, assuming that, in effect, all technologies are affected in a similar way. If labor is more productive in a capital-intensive technology, it is similarly more productive in a labor-intensive technology.

43. In chapter 6, we develop a theory of endogenous technological progress, including an analysis of the factor bias of technical change. I can be shown that, under plausible conditions, the economy evolves toward a steady state

in which the output capital ratio is equal to the rate of growth divided by the savings rate.

44. With perfect knowledge and in the absence of learning spillovers, a defender of the efficiency of the market economy might argue that the initial situation could not have been an equilibrium: Firms should have realized that would they shift to the capital-intensive technology there would have been more learning, from which they would have benefited. But knowledge of the learning benefits associated with alternative technologies is at best limited, and hence undertaking the capital-intensive strategy would have been risky—even if firms were sufficiently forward looking. Moreover, if learning spillovers are important (as we argue that they are), then the economy may be trapped in the low-learning equilibrium even without risk aversion. Moreover, if future benefits are discounted enough, the benefit of shifting to the high-learning technology may appear to be less than the costs; it is "optimal" to stay in the low-learning technology.

45. Of course, with modern globalization, global capital producing firms in the developed countries may have an incentive to develop these more labor-intensive technologies. But it is natural that they have developed capacities for research/learning of a particular kind—that which saves labor, at the expense of more capital.

46. This section draws heavily upon Hoff and Stiglitz (2010, 2011).

47. In contrast, field experiments have become, for example, a bigger part of the practice and study of development economics and increasingly a part of the learning strategies of some big businesses.

48. There is now a large literature in behavioral economics (with origins in psychology) based on these ideas, including the role of framing and biases in perceptions (see Ariely 2008; Thaler and Sunstein 2008; Kahneman 2011).

49. See Kindleberger and Aliber (2005). The most recent example is, of course, the real estate bubble that led to the Great Recession of 2008. As Stiglitz (2010b) explains, it is hard to reconcile behavior observed there with any notion of rationality (see also Holt 2009).

50. These ideas will be discussed further briefly in chapter 16 See also Hoff and Stiglitz (2010).

51. It is only, however, one determinant of learning, as evidenced by beliefs in evolution. While there is a correlation between beliefs in evolution and education and income, the United States stands out as a country with beliefs in evolution that correspond to those of far poorer and less-educated societies. For example, a survey by the British Council and the market research company Ipsos MORI reports that about 33 percent of Americans "agree the scientific evidence for evolution exists." This is a lower percentage than Argentina, China, India, Mexico, or Russia. See "God or Darwin? The World in Evolution Beliefs," *Guardian*, July 1, 2009, http://www.guardian.co.uk/news/datablog/2009/jul/01/evolution.

52. Bénabou (2008b) and Bénabou and Tirole (2006) see individuals as having the ability to choose their preferences (beliefs) so as to maximize their (meta-) utility. Our emphasis, in contrast, is on the social construction of preferences—where "outside" influences play a central role. The individual does not choose their preferences (beliefs) in isolation.

53. Some of what we are saying here can be expressed in terms of standard Bayesian inference: individuals often hold strong prior beliefs, so strong that new information has little impact on posteriors. The literature on confirmatory bias suggests that priors may be held with far greater conviction than can be justified.

54. This is what Hoff and Stiglitz (2010) refer to as "uber-ideologies." Gramsci argued that "The claim presented as an essential postulate of historical materialism, that every fluctuation of politics and ideology can be presented and expounded as an immediate expression of the [economic] structure, must be contested in theory as primitive infantilism . . ." (1971: 407).

55. As Hoff and Stiglitz (2011) note, prior to the seventeenth century, while slavery existed, it was not associated with race. Indeed, they suggest that the construct of race and racial differentiation may have evolved in part to reconcile Enlightenment beliefs in the equality of all men with the economic interests in slavery.

56. In this sense, our analysis is similar to that of Akerlof and Kranton (2010), who focus on identity and how it can be shaped.

57. As we emphasize below, the rewards do not have to be pecuniary. Even when societies pay lip service to learning, they often punish those who question received authority.

58. These beliefs are held in spite of overwhelming evidence to the contrary. For instance, the bipartisan National Commission on the Causes of the Financial and Economic Crisis in the United States, 2011, agreed, with one dissent, that government efforts to encourage housing among the poor were not responsible for the crisis. Stiglitz (2010b) presents further evidence: Not even the default rates on CRA lending (lending directed at poor communities) was higher than that on other lending. Another example is provided by the electricity shortages that developed in the early years of this century in California. Believers in free markets were quick to blame government regulations, particularly those associated with the environment. The real culprit, it turned out, was Enron's manipulation of the electricity market. When the market was re-regulated to prevent such manipulation, the shortage miraculously disappeared.

59. Many years ago, Tibor Scitovsky (see, e.g., Scitovsky 1986) described the drivers of human behavior—including the quest for excitement. See also Bénabou and Tirole (2003).

60. See, in particular, the important work by Hoff (1997). Later work has elaborated on her ideas (see, e.g., Hausmann and Rodrik 2003). Experiments are an important part of the learning strategies of some firms. Hal Varian (2011), chief economist at Google, reports in a letter to the *Economist*, "Last year at Google the search team ran about 6,000 experiments and implemented around 500 improvements based on those experiments. The ad side of the business had about the same number of experiments and changes. Any time you use Google, you are in many treatment and control groups. The learning from those experiments is fed back into production and the system continuously improves."

61. For an elaboration of this argument, see Emran and Stiglitz (2009).

62. Interestingly, the price system typically doesn't even work in the context of firms interacting with each other, simply because it is too difficult to value each individual patent. Often firms create patent pools, agreements to allow each other to make use of certain patents. (Such patent pools can often serve as effective barriers to entry, making it more difficult for firms that are not part of the agreement to enter the market.)

63. The parties to the contract obviously view these provisions in this way. But contract provisions which may be in the private interests of the contracting parties may not be socially desirable (see Greenwald and Stiglitz 1986).

64. However, the analysis below will show that markets can be so inefficient in allocating resources that sometimes growth can be increased even without a sacrifice in the short run.

65. That is, choosing a point on the production possibilities schedule where indifference curves are not tangent.

66. We note again, though, that our approach has questioned the relevance of the standard formulation of the production possibilities curve.

67. Some individuals may have an ability to learn quite generally, while others have developed more focused capacities. A well-structured learning society would recognize these differences.

68. The analysis of this section is based on Sah and Stiglitz (1989b) and Stiglitz (1995a).

69. There can be competitive effects going the other way, which we ignore for ease of exposition.

70. It should be obvious, given our emphasis throughout the book on externalities, that we do not believe that unfettered markets lead to efficient (let alone "fair") outcomes. It is even possible that welfare in the low-innovation economy might be higher than in the high-innovation equilibrium.

71. Sah and Stiglitz (1989b) provide a more elaborate model showing the existence of multiple equilibria.

4. Creating a Learning Firm and a Learning Environment

1. That is, it is hard to write good incentive-compatible innovation contracts. For instance, when a firm fails to produce a promised innovation, it is difficult to establish whether it was because of lack of effort or because of the intrinsic difficulty of the task. Cost-plus contracts, or other contracts designed to share the risk of the unknown costs required to make an innovation, have their own problems (see, e.g., Nalebuff and Stiglitz 1983a).

2. An alleged major disadvantage of firms is that transactions within firms are typically not mediated by prices, with all of the benefits that accrue from the use of a price system. But if the benefits of using prices exceeded the costs, firms presumably could use prices to guide internal resource allocations, and some enterprises do so, at least to some extent..

3. For instance, Sah and Stiglitz (1985, 1986) show that the hierarchical decision making which often characterizes large corporations leads to a greater likelihood of rejecting good projects but a smaller likelihood of accepting bad projects. But they go on to show how committees, "polyarchies," and more complex decision-making structures (e.g., polyarchies of hierarchies) can lead to improved decision making with fewer bad projects accepted and more good projects (see Sah and Stiglitz 1988a, 1988b). Large organizations may similarly encounter problems in choosing successor management teams. Sah and Stiglitz (1991) analyze the problem and show how it may be addressed.

4. For a discussion of convergence in productivity across industrial firms, see Rodrik 2013. His results stand in contrast to those presented in chapter 2.

5. As we noted earlier, it is these learning benefits that help explain an economy's industrial structure—the boundaries of what goes on inside firms. In general, the diseconomies of scale and scope (related, for instance, to oversight) are greater in agriculture than in industry. In the case of modern high-tech agriculture, there are increased benefits of learning, and that will affect the optimal size of establishments.

6. For a brief discussion of Xerox, see Wessel (2012).

7. The importance of these factors has clear implications for the conduct of macroeconomic policy, which we discuss later in this book.

8. This is consistent with earlier results on the inefficiency of competitive markets with labor mobility (see Greenwald and Stiglitz 1988; Arnott and Stiglitz 1985).

9. Exploitation by money lenders in the rural sector led to the development of rural cooperatives, e.g., in the United States and in Scandinavia.

10. Some learning, as we noted earlier, also gets at least partly embodied in capital goods.

11. In a model presented in chapter 11, the pace of learning in the industrial sector is related to the gap between the state of knowledge of the developing country and the developed country—there is some knowledge spillover across borders.

12. There are exceptions, including the increase in productivity in the U.S. recession that began in 2008. While there are several explanations of this distinctive aspect of the downturn, one is the increasingly short-sighted behavior of firms which ignores the long-run costs of firing or laying off trained workers. In that case, it will still be true that there will be long-run adverse effects of the downturn on productivity. In the Great Depression productivity growth also appears to have been quite high, in part due to important investments made by government (including in transportation; Field 2011).

13. This is, of course, consistent with standard results on unit roots (see Dickey and Fuller 1979; Phillips and Perron 1988).

14. The previous discussion explained why long-lived stable firms provided an environment that was more conducive to learning.

15. For a more extensive discussion, see Stiglitz (2002a).

16. This can be put slightly differently: With capital (debt and equity rationing), the shadow price of capital often increases dramatically in recessions (see Greenwald, Stiglitz, and Weiss 1984; Greenwald and Stiglitz 2003).

17. To identify the effects of the reduced cash flow on firm investment, including investment in R & D, they focus on two situations where changes in cash flow or net worth might be uncorrelated, or negatively correlated, with future expectations. The first study focused on the automobile industry in the United States, particularly in the aftermath of the oil price shocks. Each of these shocks had strongly adverse effects on sales of American automobiles, particularly since they were not as fuel efficient as foreign cars. On the other hand, assuming that American firms could acquire the technological know-how to construct fuel-efficient cars, these oil price shocks should have increased the level of expenditures on R & D. For the unexpected changes in factor prices meant that, while the industry had gone far along the learning curve for large cars, they were still at the beginning of the learning curve for fuel-efficient cars. But the decrease in cash flow had an immediate and direct negative effect on those R & D expenditures, and those firms that were hit the hardest reduced their expenditures the most. Our econometric study corrected for the effect of future sales expectations, and even taking this into account, the effect of cash flow changes on R & D expenditures was significant.

A second study focused on the airline industry in the aftermath of deregulation, which increased competition, lowered prices, and adversely affected cash flows. But the increased output meant that the return to reducing the cost per passenger mile was increased. The evidence was consistent with a dominant

role played by cash constraints: The rate of productivity increases declined after deregulation, and those airlines whose cash positions were more adversely affected had the most marked effect on their rate of productivity increase.

Other studies have corroborated these findings. Hall (1992), based on earlier work (Hall 1990, 1991), shows in a large panel of U.S. manufacturing firms during the 1980s that firms that took on more debt subsequently reduced both investment and R & D. In the approximately 250 firms that increased their debt by at least one-half the book value of the capital stock during one year, the decreased R & D expenditures were large enough to account for a reduction in private industrial R & D spending in the United States of 2.5 percent, about one billion 1982 dollars.

18. These effects may be in evidence even in somewhat milder downturns (see Greenwald and Stiglitz 2003; Filippettia and Archibugia 2010; OECD 2009).

19. See chapter 14 below for an elaboration of these arguments and references.

20. Note, however, the complicated second-best nature of these problems: With more mobility, firms will invest less in training, and therefore the benefit of mobility will be reduced.

21. See, in particular, Arnott and Stiglitz (1985) and Greenwald and Stiglitz (1988). Arnott and Stiglitz focus on the case where there are turnover costs with costly search. Firms can affect the quit rate by raising wages. In the simplest model, where all workers and firms are identical, in the social optimum, there would be no labor turnover (except that caused by the death of a worker). But the market equilibrium will, in general, be characterized by a wage dispersion. There is excessive labor turnover (relative to the efficient level), as workers who are unfortunate enough to be hired by low-wage firms seek higher-wage firms. They also examine a model in which firms differ in nonpecuniary attributes, individuals only learn about these attributes by working at a firm, and as a result some level of labor turnover is socially desirable. Workers, in their mobility decision, do not pay sufficient attention to the costs that their mobility decisions impose on firms; and firms do not pay any attention to the costs that their actions to reduce mobility impose on other firms (e.g., if one firm raises wages, it reduces quits at the given firm, but may increase search at other firms.) In more complicated models, hiring and quitting decisions can impose externalities through impacts on the quality of labor (see Arnott, Hosios, and Stiglitz 1988).

5. Market Structure, Welfare, and Learning

1. In *Verizon v. Trinko*, the Supreme Court referred to collusion as the "supreme evil" of antitrust. Verizon Communications Inc. v. Law Offices of Curtis V. Trinko, LLP (02-682) 540 U.S. 398 (2004) 305 F.3d 89.

2. This discussion draws heavily upon Stiglitz (2010c).

3. The imperfections of capital markets go deeper: To get a loan, the innovator has to describe to the creditor his project; but the innovator worries that should he do that, the creditor may be able to steal his idea, or build on his idea to create a still better product. Thus, the struggle to appropriate returns from an idea runs into conflict with the necessity to get funding from others (whom one may not be able to trust).

4. It should be remembered too that the antitrust movement of the progressive era focused as much on the political consequences of trusts—the concentration of power—as on the economic consequences, which got fully explicated only with the development of the modern theory of monopoly.

5. As we noted in the previous chapter, Schumpeter even thought (incorrectly in our view) that recessions could have salutary consequences.

6. That is, of course, now changed. The large increases in inequality over the past quarter century mean that even significant increases in average incomes may not be accompanied by reductions in poverty (see Stiglitz 2012b).

7. This was the objective of Dasgupta and Stiglitz's two 1980 papers.

8. If we had assumed a constant elasticity demand curve, the first-order condition could have been written

$$p = c/(1 - 1/\eta), \tag{1'}$$

where η = the elasticity of demand. The assumption of linearity greatly simplifies the analysis, but the qualitative results are more general.

9. χ corresponds to what, in our later models, we identify as the learning elasticity, h.

10. If we had assumed a constant elasticity demand curve, the first-order condition would have been (using the symmetry conditions):

$$p = c/(1 - 2/\eta). \tag{3'}$$

11. This holds so long as $d\ln \{\chi c(a - c)/I\}/d\ln I < 0$, a sufficient condition for which is that $\chi' \leq 0$ and $c < a/2$. so long as $\chi' \leq 0$.

12. If $\chi (a - 2c)/(a - c) + 1 < 0$ and $\chi \geq 0$, then $I_i > I_m$.

13. It is straightforward from the above equation to derive sufficient conditions under which this is true.

14. The analysis of this chapter is not quite complete. In a multiperiod model, both the monopolist and the duopolist need to solve a multiperiod maximization problem, taking into account the lower costs associated with investments in R & D today on future production costs in all future periods. As we noted in Chapter 4 more-durable firms are likely to have in effect lower discount rates and therefore value future cost reductions more. This may further strengthen the advantages of monopoly.

15. In that case, with no spillovers, if $C_i^d = C^m$, there would be the same level of innovation and the same costs. The question is, at that level of output, is the marginal revenue of the Cournot duopolist less (or greater) than the marginal cost? For if it is less, then the duopolists will contract production, and that will mean that innovation will be lower. Marginal revenue at $C_i^d = C^m$ will be less than marginal cost if $p(2C^m) < c/(1 - 2/\eta)$, or equivalently if $p(2C)/p(C) < (1 - 1/\eta)/(1 - 2/\eta)$. Using second-order approximations, we can show that provided the elasticity of demand is large enough, the above condition will be satisfied.

16. And follow from an observation made by Arrow (1962b) in his classic paper.

17. Similar results hold for a large innovation, where if the firm is the only one to succeed, it lowers the price below c to the monopoly price (given c^*).

18. There is a little of the flavor of Sutton's *Sunk Costs and Market Structure* (1991) here. In Sutton's story, concentration levels are bounded from below in certain industries even as demand in those industries grows—tougher ex post price competition and the presence of fixed costs limit entry ex ante. Here the intuition is that as market demand grows, output of the incumbent grows. This increased learning lowers marginal costs, toughening ex post potential competition.

19. The optimal level of innovation, given the actual level of output, will be smaller with monopoly than with competition.

20. Capital market imperfections help explain why firms are risk averse. With risk aversion, firms may also undertake less innovation (see Greenwald and Stiglitz 1993, 2003). Nalebuff and Stiglitz (1983a) explain why risk-averse managers may undertake excessively correlated research strategies.

21. There may also exist asymmetric and mixed strategy equilibria.

22. An early attempt, using a rather different approach from that of this book, is Dasgupta and Stiglitz (1980a).

23. This section is based on joint work with Partha Dasgupta (1988a). Since Arrow's (1962a) original paper on learning by doing, there has developed a large literature on market structure and learning (see, for instance, Spence 1981).

24. This analysis provides a critique of Spence (1981), who provides the equations describing the market dynamics but does not detail their implications.

With product differentiation and diseconomies of scale and scope, equilibria in which multiple firms coexist are more stable and robust.

25. And eventually, Google's Chrome. In some ways, this experience is partially consistent with Schumpeter's view: While Microsoft has remained the dominant PC operating system now for more than three decades, its dominance in the browser market was much more short-lived.

26. See, e.g., Baumol (1982); Baumol, Panzar, and Willig (1982); and Martin (2000). The implication was that the "contestable equilibrium" was the same as a constrained Pareto optimum, where lump sum taxes and subsidies (or, more broadly, cross-sector subsidies) were not feasible and where each enterprise had to at least break even. But even if lump sum taxes and subsidies are not available, government interventions, even if restricted to commodity taxes and subsidies, are still desirable (see Sappington and Stiglitz 1987).

27. Similar results can arise even in the absence of irrationality. If each potential researcher draws randomly from a sample indicating the likelihood of success (the cost of achieving success), then those who get the most favorable draw will undertake the project. (This will be true even though they realize that it is likely that they have enjoyed a more favorable draw than others. The recognition that this is the case will lead them to have a higher threshold before undertaking the project. The reasoning here is parallel to that of the winners' curse in auctions.)

28. The distortions associated with monopolies in the context of imperfect information are more extensive, as they attempt to engage in price discrimination, extracting as much of the consumer surplus as they can (see Stiglitz 1977; Stiglitz 2009, intro. to part 2).

29. Recall the discussion of Chapter 4.

30. See, e.g., Dasgupta and Stiglitz (1980, 1988a); Fudenberg, Gilbert, Tirole, and Stiglitz (1983); Stiglitz (1987). These are not just theoretical niceties describing what possibly might happen: Microsoft took actions to discourage and suppress potential rivals, so as to maintain its dominant position. Their predatory behavior lowered the returns that these innovative rivals obtained on their inventive activities, serving notice on other potential rivals that Microsoft was able and willing to engage in activities to discourage entry—even if they were flagrant violations of competition laws, and even if they entailed significant short term losses in profits. In doing so, almost surely the pace of innovation was lowered both from what it would have been in the absence of this anticompetitive behavior and from the socially optimal level.

31. The literature suggesting an inverted U-shaped relationship between competition and innovation has to be treated with caution, in light of our perspective that market structure is *endogenous*. There may be common factors leading to both more competition and more innovation (e.g., a larger and more diversified opportunity set, drawing on a larger variety of skill sets).

Appendix C of the next chapter discusses briefly some of this literature on Schumpeterian competition. Most of this literature does not address many of the issues which have been the focus of this and the next chapter.

32. Chapter 15 explores a number of implications concerning the design of the patent system. The analysis of appendix C to the next chapter suggests that "utility patents," which allow the patenting of small innovations, by encouraging competition for "small" innovative steps may adversely affect the overall pace of innovation.

33. In particular, in models with monopolistic competition. See Dixit and Stiglitz (1977), Stiglitz (1986b, 1989). Dixit and Stiglitz do show that there is a "benchmark" case in which markets balance average costs and product variety perfectly. But the presumption is that that is not the case.

34. See, e.g., Dasgupta and Stiglitz (1980); Fudenberg, Gilbert, Tirole, and Stiglitz (1983).

6. The Welfare Economics of Schumpeterian Competition

1. See Greenwald and Stiglitz (1986), who develop the concept of *constrained efficiency*, i.e., taking into account the existing differences in information and the costs of acquiring and producing information. Beginning in the late 1960s, Stiglitz had explored the nature of the inefficiencies which arise when there is imperfect information and incomplete risk markets. Stiglitz (1975b) showed that there could exist Pareto inferior equilibria, and Newbery and Stiglitz (1982) showed that trade restrictions could make everyone in all countries better off. Stiglitz (1972, 1982) showed that stock market equilibria were in general not (constrained) Pareto efficient—Diamond's (1967) earlier result suggesting that they were rested on the special assumption that there was only one commodity, no bankruptcy, and highly restricted specifications of risk. Our 1986 paper in a sense provided a general formulation that embraced these and other earlier studies. (Arnott, Greenwald, and Stiglitz 1994 provides an alternative general formulation. See also Stiglitz 2009.) We extended our generic 1986 results to search and other models in 1988.

2. This was one of the central points made in Stiglitz's 1974 lecture before the Association of University Teachers of Economics in Manchester, U.K. (Stiglitz 1975a). In November 1978, he elaborated on the problems arising from the public-good nature of knowledge in a lecture to an InterAmerican Development Bank–CEPAL meeting in Buenos Aires (published later as Stiglitz 1987b). Knowledge is a special kind of public good—a global public good, the benefits of which could accrue to anyone in the world. After developing the concept of international public goods in an address to a UN meeting in Vienna (Stiglitz 1995b), Stiglitz (1999) applied that concept to knowledge.

3. For references on the topics of international public goods and knowledge as a public good, see the previous note.

4. As we commented earlier, and we elaborate on later in this chapter and elsewhere in the book, this is true so long as there are not other market failures. Pervasive rent seeking in some innovation sectors may lead to excessive expenditures on some forms of research.

5. One should, perhaps, not put too much emphasis on the fact that these individuals did not appropriate the full benefits of their innovations. There is little evidence that they would have worked any harder with fuller appropriability. Discussions among economists focus on economic incentives; these may be far from the most important determinants of learning and innovation, as we noted in chapter 3.

6. This result provides a telling criticism of aggregate endogenous growth models that have assumed competition. Only the limiting case of perfect spillovers is consistent with full competition, but then there will be little incentive for engaging in R & D or investing in learning (Romer 1994; Stiglitz 1990). Romer's use of Dixit and Stiglitz (1977) preferences provides a simple parameterization within which one can incorporate long-run dynamics in a model with imperfect competition. As Dixit and Stiglitz note, however, that utility function has some very special properties, and one should be careful about using that utility function, especially for making welfare assessments, e.g., on the optimality of the number of firms (diversity) in the market equilibrium. Alternative specifications can give markedly different results. See, e.g., Stiglitz (1986b).

7. Chapter 3 provided an extensive discussion of the nature of these spillovers.

8. If there are spillovers across sectors (products), but spillovers external to the firm are not full, there is a natural multiproduct monopoly (under our assumptions of linear technology) as a result of these natural economies of scope. These economies of scope and scale and offsetting diseconomies of scope and scale (e.g., arising from limits of the span of control and the benefits of managerial specialization) help define the boundaries of firms.

9. Given the lower level of production, however, the level of investment in learning/R & D may be optimal. When we say that there is less learning under monopoly, we mean *less learning than there would be in a first-best situation where the level of output was optimal and the level of learning reflected that higher level of output*. Later in this book, we will show that the level of learning is less than it would be in the second-best situations where there is government intervention, through subsidies, *even when there are costs to raising the taxes required to finance those subsidies*. Much of this chapter, however, focuses on the simpler question: Is the level of innovation higher with monopoly or competition?

10. Optimal learning may involve producing at a loss, necessitating borrowing (see Dasgupta and Stiglitz 1988).

11. This is an explanation of the high observed average returns to investment in technology (see Council of Economic Advisers 1995).

12. For a more extended discussion of these issues, see Stiglitz (2006, 2008, 2013).

13. This is, of course, a more general point, applicable as well to the entry of firms in monopolistically competitive markets. An entrant captures a fraction of the consumer surplus generated by its new product, but some of the entrant's profits are, in effect, profits diverted from other firms. Whether entry is socially desirable or not depends on the relative magnitude of these two effects. In some special cases, the market equilibrium may be efficient, but more generally, it is not (see, e.g., Dixit and Stiglitz 1977; Stiglitz 1986b).

14. More precisely, $S(n^* + 1)/(n^* + 1) < I$, and $S(n^*)/n^* \geq I$. In other words, profits would be negative with additional entry and are nonnegative at n^*.

15. See the discussion in chapter 15. Boyle (2003, 2008) refers to the patenting of knowledge that was previously in the public domain as the enclosure of the commons.

16. Strikes and wars represent similar inefficient breakdowns in bargaining (Farrell 1987). See also Shapiro (2010) for a discussion of holdup in the context of patents.

17. These problems are exacerbated by the fact that the "boundaries" of knowledge are often hard to define precisely. This and the more general problem of the patent thicket are discussed more extensively in chapter 15.

18. These problems can be exacerbated by other deficiencies in the market. Compensation schemes that reward individuals on the basis of relative performance encourage "herding" behavior, where individuals do what others are doing (see Nalebuff and Stiglitz 1983a).

19. In chapter 15, we describe in greater detail the problems that arise in the context of what has been called the patent thicket.

20. The discussion of this section borrows from Dasgupta and Stiglitz (1988a).

21. Inappropriately designed intellectual property regimes can actually inhibit innovation. (See the discussion in Chapter 15.)

22. Greenwald and Kahn (2009) have shown that most of the decrease in manufacturing employment, at least prior to 2000, was a result of improvements in technology (rather than globalization).

There is a large literature supporting the view that innovation in the United States has been "skill-biased." See, e.g., Greiner, Rubart, and Semmler (2003); Goldin and Katz (2008); Autor and Dorn (2013); Autor, Levy, and Murnane (2003); Autor, Katz, and Kearney (2008).

23. Volcker's 2009 comments were widely reported in the news media. See, for example, Pedro da Costa and Kristina Cooke, 2009, "Crisis May Be Worse than Depression, Volcker Says," , Reuters, February 20. http://uk.reuters.com/

article/2009/02/20/usa-economy-volcker-idUKN2029103720090220 (accessed February 27, 2013).

24. For a more extensive discussion, see Stiglitz (2010b).

25. This analysis builds on Stiglitz (2006b), which itself builds off a large literature on factor-biased induced innovation, going back to Ahmad (1966), Drandakis and Phelps (1966), Fellner (1961), Kennedy (1964), and Samuelson (1965), with antecedents in the literature in economic history (e.g., Salter 1966; Habakkuk 1962).

26. Similar results obtain in a more dynamic, multiperiod model.

27. The share of labor is simply a function of λ, if K, L, and the unemployment rate are all fixed. In the standard models, the unemployment rate is zero. In the discussion below, we allow the unemployment rate to be endogenous.

28. The Shapiro-Stiglitz (1984) model is set in an infinite-period context. It is easy to set our choice of technique within such a context.

29. See J. Hicks (1932), for a discussion of a typology of innovation.

30. There are other interpretations, discussed briefly below. It seems, for instance, that at times wages did not fully adjust and that at times, as a result, firms had difficulty hiring workers. This will result in the shadow wage exceeding the market wage.

31. We also note that the dynamics can be markedly different from those of the standard Solow model, where convergence to equilibrium is monotonic. As we show in the appendix, convergence is oscillatory. We note, however, that the smooth convergence to the steady state in the Solow model is a function of its extreme simplifying assumptions. Other slight modifications (vintage capital, savings depending on the distribution of income) can also lead to more complicated dynamics. See, e.g., Akerlof and Stiglitz (1969); Cass and Stiglitz (1969); or Stiglitz (1967).

32. This is consistent with the evidence on the stagnation of median wages in the United States over a span of more than forty years and a decline in wages of unskilled workers (e.g., workers with only a high school education).

33. For a more extensive discussion of the issues raised, see Stiglitz (2010b).

34. For a more extensive discussion of these ideas, see Stiglitz (1975a, 1994c, 2010a).

35. We noted in chapter 5 that the firms that went bankrupt in Korea were not on average less productive than those that did not.

36. Even more so in countries, such as the United States, where those with money have disproportionate weight in the political process.

37. In the standard competitive model, where each firm faces a horizontal supply curve of labor of each type, these effects are not likely to rise. But in practice, labor mobility is imperfect. Firms are engaged in a bargaining process with their workers. The nature of technology—which they can shape—affects this bargaining process. Note that this analysis does not require that firms

coordinate their actions, to increase their bargaining power vis-à-vis workers (though under some circumstances they may in fact do so.) Rather, so long as there is imperfect mobility of workers, it pays each firm to take actions which increase its bargaining power vis-à-vis its workers.

38. See Braverman and Stiglitz (1986) for an analysis of these issues in the context of an agricultural economy.

39. In perfectly competitive labor markets, any nonpecuniary cost would lead to a demand for higher wages and thus would be taken into account by the firm, but this is not so in imperfectly competitive markets.

Moreover, in a world with uncertainty and imperfect information, management may come to believe that technologies that save on labor are profitable. There may be an "equilibrium fiction" in which the evidence, as they see it, confirms those beliefs (Hoff and Stiglitz 2010). This is especially so in managerial capitalism (with agency costs which enable managers to exercise considerable discretion for their own benefit), where managers may value their own time and trouble more than would be the case in an efficient market economy.

Interestingly, such beliefs will, in fact, serve the interests of the managerial/capitalist "class" as a whole, leading to outcomes that are consistent with what they might have wanted to do collusively, though they had no mechanism by which to do so.

40. There is, in this sense, a kind of increasing returns to scale. The more innovators think about how to improve labor efficiency, the better at it they get. This suggests that the innovation frontier, rather than being concave, may in fact be (at least in part) convex.

41. This is one of the points raised by Phelps (2013).

42. We should perhaps more accurately say that markets *never* exist in a vacuum. Society has to set the rules and regulations that govern them, e.g., what kinds of contracts can be enforced, and how they are to be enforced. Thus, the notion that there are "unfettered markets" is a chimera, an idea that is often used by those who are trying to shape markets in a particular way (*as if* there was a "right way" by which markets *should* be organized)—often in ways that are in their own interests.

43. More accurately, they result in less R & D or less learning than would occur in the first-best situation; given the lower output associated with monopoly, conditional on the monopoly power persisting, the optimal degree of investment in R & D is lower.

44. East Asia did this as a central part of its development strategy (see Stiglitz 1996; World Bank 1993).

45. We should emphasize, however, the difficulties of ascertaining well-being, highlighted by the work of the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz, Sen, and Fitoussi 2010).

For instance, while for most individuals improvements in life sciences have increased overall health, these gains have to be offset against increases in insecurity and decreases in the quality of the environment.

46. It is straightforward to show that $\sigma = -f''(f - Kf')/ff''$.

47. Equation (A.19) can be thought of as derived from an efficiency-wage model, where the efficiency wage is a function of the unemployment rate and the minimum wage, and the minimum wage rises with productivity. The wage an individual receives is $W = wa = G(U)W_{min}$.

48. Stiglitz (2006b) shows that the equilibrium is, in general, stable.

49. By contrast, in the text we noted the possibility of "learning to learn," implying that the process of skill-biased technological change may feed upon itself.

50. See, e.g., Dasgupta and Stiglitz (1980a, 1980b, 1988a); Fudenberg, Gilbert, Tirole, and Stiglitz (1983); Gilbert and Newbery (1982); Stiglitz (1987a); Greenwald and Stiglitz (forthcoming).

51. Stiglitz (2013b) describes how firms contribute to and take out of the pool of knowledge, affecting the innovation opportunity set. Dosi and Stiglitz (forthcoming) review the empirical literature and argue that the opportunity set is the most important determinant of the level of innovation. Successors thus benefit from this knowledge.

52. A similar argument applies in the case of product innovation.

53. Aghion and Howitt identify a similar effect, which they attribute to monopoly power. But the effect can be strong even when monopoly power is limited, e.g., because there is a sequence of rapid innovations, the effect of which is to lower consumer prices.

54. In the discussion below, we delineate carefully the effects of competition in innovation from those arising from competition in production. Schumpeter (implicitly) identified two sets of effects, that on the *marginal* returns to investment and that on the *ability* to invest. In a world with imperfect capital markets, there may be heavy reliance on internal finance, especially for R & D, which cannot be collateralized. We focus on the former effect, leaving to a later occasion an analysis of the latter.

55. In their work, and in much of the literature, a Poisson process is assumed. This is convenient for modeling steady states; but such processes have very special properties.

56. So too, the aggregative approach taken by Aghion and Howitt leaves unaddressed key issues about the sectoral direction of resource allocation and R & D, e.g., between industrial sectors and agricultural/craft sectors (the subject of Greenwald and Stiglitz forthcoming). Again, because of space limitations, we do not explore these issues here.

57. In Aghion and Howitt, there are a separate set of research firms who own the patents. In practice, most research is done by incumbent producing firms

because of the importance of the knowledge that is acquired in the process of production and the imperfect transferability of such tacit knowledge. This knowledge not only increases productivity directly (learning-by-doing) but also enhances research capacities. But even if this were not so, firms engaged in prior research on a product have a knowledge advantage over entrants, so there should be some tendency for perpetuation of monopoly power. As we have emphasized, knowledge spillovers are always imperfect.

58. For an early formulation of innovation with stochastic entry, see Futia 1980. Even with stochastic outcomes, in the absence of diseconomies of scope or differential knowledge on the part of potential entrants, any project that was worth undertaking for an entrant would be worth undertaking by the incumbent (see Dasgupta and Stiglitz 1980b; Sah and Stiglitz 1987).

59. With potential follow-on effects for subsequent innovation.

60. See, e.g., Aghion et al. (2005) and the references cited there.

61. We will say a little bit about a third aspect, the riskiness of the research strategy, and virtually nothing about a fourth, the extent of correlation of a given firm's research strategy with that of others. However, in the analysis, particularly in section 1, we show the importance of these.

62. Some of the more recent literature on Schumpeterian competition follows in this tradition. See, e.g., the excellent survey paper by Aghion, Akcigit, and Howitt (2013) and the papers cited there.

63. We can use (A.31) to derive an equivalent expression:

$$I = \phi'^{-1}\{-1/[c_0 Q(c_0)]\}.$$

64. Such nonconvexities arise naturally in the economics of information and knowledge. See Radner and Stiglitz (1984).

65. In this equation, the elasticity of demand is valuated at the point where $p = c_0$. At this point, we are not assuming constant elasticity demand functions.

66. This just highlights the fact that economies where R & D and learning are important are rife with convexities.

67. We should also note that the demand functions may be shifted from the earlier analysis as a result of income effects, the absence of profits from the exercise of monopoly power, and the imposition of lump sum taxes.

68. Not only are the two hypotheses questionable, the result on the efficiency of equilibrium is not general. In a market with multiple commodities, which the government can tax and subsidize differently, the contestability equilibrium is not efficient.

69. Highlighting the observation made earlier: if limitations on research are provided by access to capital, and if access to capital is affected by next period's profits, then there may be more investment in R & D under Cournot, even though marginal returns are lower.

70. Gross profits of the Cournot entrant (denoted R_C) are (using (A.45a)) $c_1(\alpha^2/\eta)Q_C/(1 - \alpha/\eta)$. Profits under Bertrand are $(c_0 - c_1)Q_B$. At $c_1 = c_0$, $R_C > R_B = 0$. At the boundary point between a small and large innovation, given by (A.47), $\alpha = 1$ and $R_B = R_C$. Using the facts that $Q_C > Q_B$, $Q_{1C} = \partial R_C/\partial c_1 < \partial R_B/\partial c_1 = Q_B$, it is clear that for all small innovations, $R_C > R_B$.

71. At Q_i equal to that of the myopic monopolist, Q_m, Marginal revenue $= p + p'Q_m < p(Q_m) + p'(Q_m)Q_m = c_1$.

72. Assume c_1 remains the same. Then

$$Q_1 = (a + c_0 - 2c_1)/3b < (a - c_1)/2b = Q_m$$

if $a + c_1 - 2c_0 > 0$, i.e., from (A.50b), if $Q_0 > 0$. Obviously, a higher Q leads to a lower c_1, leading in turn to a still higher Q.

73. $R = (p - c_1)Q = (a - c_1)Q - bQ^2$, which is maximized at $Q = (a - c_1)/2b$.

74. Cournot is 4/9 that of Bertrand, for small innovations; myopic monopoly is ½ that of Bertrand.

75. From the indirect utility function, $V_p = V_\gamma Q_1$, $dp/dc_1 = 1/3$ in Cournot equilibrium.

76. Most of the literature has taken the first approach, though some (such as Dasgupta and Stiglitz 1980a) have taken the second. In practice, there are often multiple entrants.

77. In contrast to the discussion of sections 8 and 9, we assume that there is no way that a potential entrant can, by allocating more resources to the speed of innovation as opposed to the depth of the innovation, increase the likelihood that it will be first. This is obviously a special case, but this assumption greatly simplifies the analysis. One way of thinking about the R & D process is in terms of a stochastic process as described in the next section. There is a small variance in the time of arrival, and the marginal costs of reducing the mean time of arrival are very large.

78. One can easily specify a game form for which this occurs.

79. This is the solution, provided the level of innovation is small. Alternatively, I^* is the solution to

$$I^* = c_0[1 - \phi(I^*)]Q[c_0(1 - \phi(I^*)/(1 - 1/\eta)]/(\eta - 1).$$

80. This follows from the assumption that $\phi'' > 0$. In the more general case, e.g., that depicted in figure 6.6b, there can exist more than one point at which the revenue and cost curves cross. So too in the case of large innovations, because both the revenue and cost curves are convex. The contestability equilibrium is that associated with the highest level of innovation, for a firm undertaking that level of investment would undercut potential rivals.

81. As c_1^1 diminishes, the slope of the revenue curve gets steeper. It is thus possible that there is more than one value of c_1^1 for which the revenue curve and cost curves have the same slope.

82. If $-c''_1(a - 2c_1^1 + c_1^2) + 2(c'^1_1)^2 > 0$, there is a unique value of c_1^1 corresponding to any value of c_1^2 in (A.69). Even if the above condition is not satisfied, there is always a value of I_1 that maximizes the first firm's profits, given I_2. Because of the nonconvexities, the reaction functions may not be as simple as depicted in the figure.

83. Which depends in turn on the elasticity of $\phi'(I_1)$.

84. For simplicity, we are assuming all firms have the same innovation function and the incumbent departs. Similar results hold if he does not.

85. More accurately, because n must be an integer, n^e is the largest value of n such that $\prod^i \geq 0$.

86. Recall our earlier caveat that these results depend on the likelihood of success being independent. If the different firms are engaged in identical research strategies, presumably the results would be highly correlated and ΔP would be zero or small.

87. Similarly, if firms are risk averse.

88. That was the central message of Dasgupta and Stiglitz (1980a).

89. For $n > n_{max}$, the value of profits in the symmetric equilibrium (where all firms engage in research) is negative.

90. In practice, there are likely to be large knowledge differentials, so that the cost of innovation is lower for some firms than for others. In this case, n_{max} is such that the expected profits of the next firm to enter (given the optimal response of all the other firms) is negative.

91. Except, as noted, in the case of a large innovation, where the innovator lowers the price below the marginal cost of production of the incumbent.

92. This is the essential insight of the "persistence" theorem (Dasgupta and Stiglitz 1980a). For a small innovation, the incumbent's optimal price may be above c_0 (its marginal cost the previous period), which would be the price confronting an entrant.

93. Entry can also occur because of the "irrational exuberance" of innovators—who may have more confidence in their own research project than it deserves. Thus, some of the dynamism of capitalism may be attributed to irrational expectations (see Knight 1921).

94. Note that, in general, the acceleration of research (in this model, the "sinking" of research expenditures) is not efficient, i.e., the threat of entry may actually lead to higher costs. It is even possible that it leads to a lower level of innovation.

95. This is highlighted by Dixit and Stiglitz (1977), who show that with the Dixit-Stiglitz utility function, the market for entry is "constrained Pareto efficient." The result is not general.

96. In moving to the analysis of a long-run maximization problem, it is convenient to change to constant-elasticity demand curves, which allow for the analysis of steady-state behavior.

97. While this assumption has become conventional in the literature, it is questionable. A bigger innovation today may draw down the pool of available ideas, and thus further improvements may be more difficult.

98. We do the calculations for the harder case of large innovations. The calculations for small innovations are similar. (See section 2.)

99. That is, the myopic monopolist sets $2\pi_1/2I_1 = 0$. From (A.88), since $2v_1/2I_1 > 0$, the non-myopic monopolist increases I relative to the myopic monopolist.

100. It follows from (A.88) that the magnitude of the deviation in I, between the myopic and non-myopic monopolist is increasing in δ.

101. For large innovations, the analysis is identical to that of the previous subsection, since at the price chosen by the monopolist, it does not pay for competitors (with the old technology) to produce.

102. Thus, the one-period output is smaller with the persistent monopolist, but the cost saving from more innovation relates to the (PDV) of all future outputs. With a small discount factor, that number will be very large.

103. In this formulation, we have assumed that there is no spillover of knowledge across sectors. It is easy to generalize the results to cases where there is some knowledge spillover.

104. This formulation assumes that the fruits of the research done this period become immediately available. A more plausible assumption is that there is a one-period lag. Nothing hinges on this assumption.

105. We assume that innovative activity each period builds on the benchmark of inherited knowledge, i.e., that an investment in innovation of I^t at time t leads to a percentage reduction in costs of $\phi(I^t)$, regardless of the level of previous innovation. This assumption greatly simplifies the analysis, but the qualitative results would be similar so long as research investments at time t don't fully dissipate opportunities in subsequent periods. At the other extreme, we could consider a case where the marginal cost of production at time $t + 2$ is a function only of total two-period investment: $c_{t+2} = c_t\Psi(I_{t+1} + I_{t+2})$, as opposed to the model investigated here, where $c_{t+2} = c_t\phi(I_{t+1})\phi(I_{t+2})$. Appendix D investigates a case where innovators at time t deplete the "knowledge pool," making further innovation more difficult until the knowledge pool has been replenished by basic research. The model of section 3 considers a similar situation, where faster innovation at time t makes innovation the next period more difficult.

106. It should be obvious that we require that $\delta H^a < 1$; otherwise, the PDV of utility becomes infinite, and the problem is not well-defined. We assume that the above condition is satisfied at all (relevant) values of I.

107. It is clear that there is a unique solution to (A.96) for $\delta = 0$ (and by continuity, δ near zero), so long as $h' \leq 0$. But if δ is large, the left-hand side of (20) may be increasing in I, even if $h' \leq 0$, since $d[\delta H^a/(1 - \delta H^a)]/d(\ln I) = \delta h H^a/(1 - \delta H^a)^2 > 0$.

108. We can think of all the investment being made up front, with the research maturing in T years, or, alternatively, that there is a steady flow of research. A flow of "i" for a period of T years has a (PDV) cost of $I = i(1 - e^{-rT})/r$. If a flow of i yields an innovation at time T, that is equivalent to a PDV of $I = i(T)(1 - e^{-rT})/r$ yielding an innovation at T.

109. If as T goes to zero, the required investment increases sufficiently, entry deterrence is still possible. And in that case, it is always desirable for the incumbent to deter the entrant. We assume that the entrant assumes that it will be the dominant firm for a period that does not increase (or increase too rapidly) as T decreases.

110. How much investment the incumbent must make depends, obviously, on the information structure. If the incumbent can observe instantaneously what rivals are doing, then it need not invest ahead of time as much as it would if there were long lags in observation.

111. These are, obviously, extreme assumptions—highlighting the sensitivity of results to particular assumptions concerning the innovation process.

112. We model this as if the firm invests a flow for a period T, for a total PDV of investment of I, after which it receives a flow of profits $Q(1 - 1/H)$ for a period of T (until the next innovation arrives) of investment.

113. In fact, the entrant takes the interval over which it makes profits as given and maximizes

$$\lambda(1 - \lambda^*)Q(1)[1 - 1/H(I, T)]/r - I$$

with respect to T.

114. That is, $\{I^e, T^e\}$ do not in general maximize either social welfare or $H(I, T)/T$.

115. The analysis by Dasgupta and Stiglitz (1980a) is predicated on that assumption.

116. These knowledge spillovers are the focus of this book. Here, *intertemporal* knowledge spillovers have played a central role in our analysis.

117. In some cases, this is almost the explicit objective of R & D activity, e.g., in the case of "me-too" innovations in the pharmaceutical industry. Thus, while this appendix focuses on the overall pace of innovation, we should emphasize that the market has distorted incentives—excessive incentives for innovations which appropriate part of the rents of previous successful innovators and too few incentives for fundamental innovations. This appendix, exploring the intertemporal linkages and externalities associated with innovative processes,

has delineated the large number of forms that they can take, with different implications for the nature of distortions in the innovative process. While a persistent monopolist "internalizes" these externalities, and thus is not subject to these distortions, we have suggested that it faces an alternative set of incentive problems, for all—most importantly, potentially deficient incentives to innovate. But even with a sequence of monopolists, incentives are attenuated relative to a "first-best" since even the sequence of monopolists do not capture the consumer surplus generated over the long-term by their innovative activity.

In Appendix D, we show that similarly to other common-source problems, unbridled competition does not lead to the maximum production (of innovations, of fish, etc.). The analogy that has emerged in this appendix is even stronger. Just as in fisheries there is a danger from overfishing premature fish and not letting the stock of fish mature to the level where the steady-state flow of fish would be maximized, here too there is a tendency to capture (patent) ideas at too young a stage, impeding the subsequent development that would be associated with a faster rate of overall innovation.

118. As we noted, this is more generally true in the theory of imperfect and monopolistic competition.

119. And it is assumed that the incumbent firm cannot undertake some of the research projects that entrants can undertake, e.g., because of diseconomies of scope. If it can, it can be shown that any project that would be worthwhile for an entrant to undertake would be undertaken by the incumbent.

120. Dosi and Stiglitz (2014) argue that these opportunities may be the most important determinant of the pace of innovation.

121. These ideas are developed more fully in Appendix D.

122. This appendix is inspired in part by joint work with Giovanini Dosi, who, surveying the empirical evidence on innovation, has argued that the available set of opportunities plays a more important role in determining the pace of innovation than intellectual property rights (see Dosi and Stiglitz 2014). This appendix shows the endogenous determination of the opportunity set.

123. See, e.g., Bessen and Meurer (2008). That said, it should not be surprising that there have been a large number of attempts at finding correlations between some measure of the strength of a country's intellectual property regime and economic growth, or some variable purportedly related to economic growth, such as inward-bound foreign direct investment. As most scholars engaged in this research have recognized, these studies are bedeviled by a large number of econometric problems, e.g., of identification. With global harmonization, there is a dearth of natural experiments. Even when such experiments exist, long lags and the influence of a multiplicity of other factors affecting innovation and growth make it difficult to establish definitive, or even convincing, results.

Historical studies suggest a multiplicity of influences on the level of innovation: Several European countries with weak intellectual property rights had flourishing innovative sectors—more flourishing than others with stronger intellectual property rights (Chang 2001, 2002). See also David (1993, 2002).

More recent critiques of IPR regimes have focused on "flaws" in the IPR regime, arising from the patent thicket, the bias for excessive patenting (as opposed to the incentives for fighting patents), holdups, incentives for evergreening, etc. Whether it is possible to adequately "correct" these flaws, so that the net effect on the pace of innovation (as opposed to the level of investment in R & D, taking into account the distortions associated with these flaws) of the IPR regime is positive remains contentious. It is clear, however, that the relationship between IPR and innovation depends on fundamental rules of the IPR regime, governing what can be patented, the breadth and standards of patenting, how patents are enforced, and so forth.

More narrowly focused studies have identified areas where particular patents have had adverse effects on follow-on research. In particular, evidence presented in the Myriad BRAC gene patent litigation detailed adverse effects both on the development of tests and further research. (See *Association for Molecular Pathology v. Myriad Genetics*, 569 U.S. 12-398 [2013]; Huang and Murray 2008; Williams 2013).

The discussion of this appendix abstracts from the details of the patent system, which are discussed at length extensively in the large literature on intellectual property. So too, we ignore the details of the patent system which affect the welfare effects, e.g., the consequences of IPR regimes for access to life-saving medicines.

The state of the current debate around IPR is surveyed by the contributions to the Winter 2013 *Journal of Economic Perspectives* symposium (Boldrin and Levine; Moser; Hagiu and Yoffie; and Graham and Vishnubhakat).

124. Dasgupta and David (1994) also argue for the importance of other nonpecuniary motivations for research. See also David (2004a, 2004b).

125. See Dosi and Stiglitz (2013) and the references cited there.

126. To use Boyle's (2003) evocative phrase. See also Heller (1998); Heller and Eisenberg (1998).

127. We should emphasize the importance of the word *design*. There are many details to the IPR system, such as the breadth of the patent, which affect the extent to which knowledge is added to or "subtracted" from the pool. Knowledge that is not available for others to use is, in effect, subtracted from the pool. Broad patents subtract more from the available knowledge pool than more narrowly defined patents. Matters are, of course, far more complicated than this simple arithmetic analogy would suggest. Knowledge that can be

used, but at a price, is in a sense "partially" available, and even knowledge that cannot be used directly can trigger research. By the same token, some, perhaps much, of the investment in R & D in a poorly designed IPR regime is devoted to inventing around a patent or to increasing the rents that can be extracted out of a patent (e.g., by evergreening). In such circumstances, even if tighter IPR leads to more investment in R & D, it may not lead to faster real innovations, i.e., an increased pace of increases in standards of living.

128. $P_{\min}$ is defined by the lowest value of P for which $H(P) = 0$.

129. This would obviously not be the case if the number of firms is very small (say $n = 1$ or 2). In this simple formulation, the price of fish, p, is given exogenously.

130. It is easy to generalize (A.114) to include situations where there is imperfect competition among the firms (i.e., we generalize pQ to a more general revenue function R). The model is consistent with alternative interpretations of how additional investments in fishing fleets affect the marginal catch. For our purposes, these details are irrelevant.

131. The full welfare calculation is more complicated, because of the effects of externalities, even if markets are competitive. Ignoring the effect on P and the cross-firm externality associated with the increase in fishing costs for each firm, the marginal value of resources used by each firm is then equal to the marginal costs, and the value of the benefits of the marginal firm equals its marginal costs. The fact that the cross-firm externality effect (at fixed P) is negative implies that if the pool effect is small (P is only slightly greater than $\hat{P}$), it is still desirable to tax entry.

132. Matters are, of course, more complicated than this discussion might suggest. In the long term, the innovation may be "enabling," even if it takes away economic opportunities in the short term, because it provides the knowledge base on which further innovations can eventually be built. See, e.g., the various contributions to the Winter 2013 symposium on patents in the *Journal of Economic Perspectives*.

133. Notice that this formulation explicitly rejects the view that there is a fixed stock of knowledge to be discovered. If that were the case, then it is possible that a large value of "discovered" knowledge would diminish the set of knowledge to be discovered, and it is possible that $H_P < 0$, or even more, that $H_{PK} < 0$, i.e., the marginal return to research investments diminishes as the size of the knowledge pool increases. This could result in a figure looking more like that depicted earlier for fishing stocks.

134. All that the analysis below requires is that dH/dP not increase too much, for then there may not be a stable equilibrium pool of knowledge.

135. This is consistent with general results showing a fundamental nonconvexity in the value of information (see Radner and Stiglitz 1984).

136. We can think of I as the pace of, say, labor-augmenting technological change. For purposes of this analysis, however, we do not have to be specific about how we parameterize the level of innovation.

137. Throughout this section, we hold n constant.

138. Implicitly differentiating (A.116).

139. In the next section, we shall show how these results can easily be generalized.

140. We are explicitly assuming that knowledge is *not* industry specific.

141. This formulation allows us to avoid the more complicated intertemporal maximization problem that would arise if each firm's current innovation level depended not only on the current pool of publicly available knowledge, but on a pool of privately available knowledge. Qualitative results for this more general problem would, however, be similar to those described here.

142. $d(\ln I)/d\xi = [d(\ln H)/dP]dP/d\xi - d[\ln \gamma(\xi)]/d\xi < 0.$

143. These results can be seen directly in figure 6.14a, where, while the effects on P^* are unambiguously negative, those on i^* are indeterminate.

144. Notice that this formulation is consistent with there being many firms in each industry. The profits of any firm are a function of its own innovations and those of others in the industry.

145. More precisely, $dI/dn = H_P(\partial I/\partial n)_{|P}/[H_P(K, P) - \gamma(\xi)(I_i Z_P + I_P)].$

146. By analogy, in the first section, more fishing firms, *given a stock of fish*, will discourage investment by each of the fishing firms, as the marginal return to investment (at a given level of i and P) is reduced. In the context of innovation, see Aghion and Howitt (1992, 1998); Stiglitz (2013a, 2013b) and discussions elsewhere in this book.

147. But the analysis shows equally that if more competition should lead to lower innovation *at a fixed P*, then it will lead to more innovation in equilibrium. And that may well be the case. It is possible that more competition (large n) so lowers the marginal return to investments in R & D at any given level of research of others (and P), that investment by each firm is diminished enough that the depletion of the knowledge pool is actually reduced as competition increases.

148. There is a huge literature on the subject, some of which suggests an inverted U-shaped relationship, some of which suggests that innovation may decline monotonically with n. The empirical literature is bedeviled with the problems noted in earlier footnotes, and even the theoretical literature does not always separate clearly the effects of entry *given* a particular opportunity set, versus the effects with an endogenous opportunity set. The literature in which prior innovation affects returns to current investments in R & D typically emphasizes the benefits that arise from the increase in the "baseline" knowledge, from which current research efforts depart, rather than the

negative effects of the drawdown in the knowledge pool that has been the focus of this analysis. See, e.g., Aghion and Howitt (1992, 1998); Aghion et al. (2005); Aghion, Akcigit, and Howitt (2013); and Stiglitz (2013a, 2013b).

149. Thus, an increase in *n*, especially beyond a certain critical level, may lead to lower *I*, keeping *P* fixed, and thus to a higher level of innovation—taking into account the effect of entry on the size of the opportunity set (*P*).

150. There is, by now, a large literature discussing these and other similar reforms to the intellectual property regime. For a brief review, see below chap. 15.

151. As advocated, for instance, by Lewis and Reichman (2005).

7. Learning in a Closed Economy—the Basic Model

1. We reiterate that while we couch the results in this part of the book in terms of learning that results from doing, there are analogous results when firms make explicit investments in R & D.

2. Here, as elsewhere in the book, we use the terms *manufacturing* and *agriculture* in a quite broad way. *Manufacturing* includes other activities that typically occur within cities, including services like telecommunications and large-scale retail. *Agriculture* includes activities that are geographically dispersed and that occur in small production units, like crafts.

3. We are also assuming time-separable utility functions and utility functions which are separable in goods and leisure. Nothing essential depends on these assumptions.

4. The assumption that the spillovers are *perfect* is, of course, a polar case. In subsequent chapters, we will loosen this assumption. None of the qualitative results depends on this assumption.

5. $\psi(0) < 1$ implies that learning (cost reductions) occurs even in the absence of production. For simplicity, in much of the discussion below, we assume $\psi(0) = 1$.

6. In the next chapter, we will analyze the more general case where the labor supply is variable.

7. It should be emphasized, however, that the variables on the right-hand side of (21a) may themselves be functions of the taxes and subsidies and that there may be multiple values of t_M satisfying (21). For instance, higher subsidies for the industrial sector will normally be associated with a higher value of second-period output, and h too can depend on the level of industrial output.

8. Alternatively, we can write $\beta_M = \{d(\ln C_M)/d[\ln(p_M/p_A)]\}\{d[\ln(p_M/p_A)]/d(\ln \tau_M)\}$, where $d[\ln p_M/p_A)]/d(\ln \tau_M)$ can be calculated in a straightforward way, given in the next footnote. Similarly for β_A.

9. Relative consumption, C_M/C_A, is just a function of relative prices $p = (1 - \tau_M)/(1 + t_A)$: $C_M/C_A = Q[(1-\tau_I)/(1 + t_A)]$. Since $\tau_M C_M = t_A C_A$, $p = (1 - \tau_M)/(1 + \tau_M Q(p))$, which gives p as a function of τ_M. The elasticity of substitution, σ, is defined as $d[\ln (C_{MI}/C_A)]/d\{\ln[(1-\tau_M)/(1 + t_A)]\}$, where $dt_A/d\tau_M = (C_M/C_A) + \tau_M[d(C_M/C_A)/d\tau_M]$. Thus, $d[\ln(C_M/C_A)]/d(\ln \tau_M) = \sigma\, d(\ln p)/d(\ln \tau_M)$.

10. The analysis is somewhat more complicated than (25) suggests, because the terms in (25) are not themselves constant and may, in fact, depend on the tax rate.

11. We ignore here any effects on labor supply.

12. In the figure, U_m^t is the tth-period utility under monopoly, U_c^t, under competition. While U_m^t is slightly less than U_c^t, U_m^{t+1} is much greater than U_c^{t+1}.

13. v is itself an endogenous variable, depending on the magnitude of learning, ψ, the learning elasticity, h, the discount factor, and properties of the demand function. For instance, if there is a high level of learning, v will tend to be large; if there is a high elasticity of learning, v will be smaller than it otherwise would have been.

14. There are two other situations that deserve brief mention. The first is where there are not full spillovers, but competition is maintained within the learning sector as a result of diseconomies of scale. To take the extreme case, assume there were no within-sector learning spillovers but only cross-sector learning spillovers. Then each competitive firm would take full account of the benefits of its learning to itself—just as the monopolist would. In that case, the only distortion in the competitive equilibrium arises from the cross-sector externality. As long as there are spillovers, production in the learning sector will be suboptimal, and a first-period consumption subsidy would be desirable. The other case of interest is that where there is imperfect competition in both sectors. If the degree of monopoly is roughly the same, then both will raise prices relative to marginal costs, but relative prices will be unaffected. Hence, monopoly power won't distort (relative) production of the two goods; it will result in internalization of the within-sector externality, but not of the cross-sector externality. The results are much as discussed in the previous paragraph. Of course, the reduction in real wages not only has distributional effects, but will also affect labor supply, and thus learning. Some of these issues are discussed more extensively in the next chapter.

8. A Two-Period, N-Good Model with Endogenous Labor Supply

1. This is not an innocuous assumption but is done to simplify what is already a fairly complex analysis. Without separability, government may wish

to induce more consumption of the learning good(s) by exploiting complementarities and substitutability between these goods and consumption of various other goods at later dates.

2. H is the output per unit input; the function ψ introduced in the previous chapter is the input per unit output. Clearly $H = 1/\psi$.

3. While this is precisely true in the case of logarithmic utility functions discussed below, in the more general case, the analysis is somewhat more complicated because of the endogeneity of ξ_k, as discussed below.

4. Under normal circumstances, growth (an increase in H, productivity) will lead to an increase in consumption; but matters are slightly more complicated, as first-period consumption is subsidized because of the benefit of learning. If H_0, the increase in productivity with no subsidies, is large relative to δh_k (the value of the learning benefits), then $C_k^{t+1} > C_k^t$.

5. Recall that in terms of our previous notation, Ψ is the special case of H where there are no spillovers.

6. $U_i^t - v^{t\prime} = -\tau^* v' = -\delta U_i^{t+1} G'(L_i^t) L^{t+1}.$

7. These formulae provide a characterization of the equilibrium, but it is important to note that, in general, the elasticities can themselves depend on taxes and subsidies.

8. With full spillovers, the learning function is $G = G(L^t/n, L^t/n, \ldots)$ for all sectors. An increase in L^t thus increases learning by G_i. Denoting by G^* the function

$$G(L^t/n, L^t/n, \ldots), \text{ with } G^{*\prime} = G_i, \text{ we can write the first-order}$$
$$\text{condition with full spillovers as}$$

$$\Sigma\, U_i^t/n - v^{t\prime} = U_i^t - v^{t\prime} = -\delta\Sigma\, U_i^{t+1}\, G^{*\prime}(L_i^t) L^{t+1}/n =$$
$$-\delta U_i^{t+1} \Psi'^*(L^t) L^{t+1}.$$

Substituting G^* for $G\Psi$, we obtain exactly the same results, though with a slightly different interpretation. It is likely, because of learning spillovers, that normally $G^{*\prime}$ would be greater than it would be, implying a higher level of subsidy.

9. Since, in the first period (omitting the superscript t), $\Sigma\, p_i C_i = L + \Sigma\, \tau_i C_i$, at $\tau = 0$, $\Sigma\, dC_i/d\tau = dL/d\tau > 0$. Both the income and substitution effects lead to an increased labor supply.

10. We make use of the fact that $V_{pi} = -V_I C_i$.

11. As we have noted before, one must take care in interpreting this and other optimal tax/subsidy formulas in this book, because the variables on the right-hand side are typically not constants but functions of the subsidy rate itself. Still, they provide insights into the determinants of the appropriate subsidies for optimally designed subsidies.

12. With a positive wage elasticity, ζ_j is normally greater than unity. If there is zero labor elasticity and a logarithmic utility function, then $\zeta_j = 1$ for all j. If a sector has elasticity of demand less than unity, ζ for that sector is less than unity.

For an explicit derivation of the scale parameter in the case of the symmetric model, see the discussion below in section 4.

13. Obviously, converse results hold if the demand elasticity is less than unity.

14. In this case of symmetry, $\rho = \delta V_I^{t+1}/V_I^t$.

15. That is, $(n - 1)h_{ij} >> h_{ii}$.

16. The techniques used in this section can be used to evaluate the scale parameter as a function of the tax rate in the nonsymmetric cases as well.

17. ε_{Lw} in turn can be related to underlying properties of the utility functions. In the case of separable utility functions, $\varepsilon_{Lw} = (1 - 1/\eta)/(v + 1/\eta)$, where η is the (absolute value of the) elasticity of marginal utility of consumption and v is the elasticity of the marginal disutility of labor.

18. It is worth noting that in the absence of the strong assumptions of separability, consumption the first period may depend on prices (taxes) the second period. This obviously complicates the calculations considerably.

19. Both of these are total derivatives, taking into account the direct effect of the change in τ and the indirect effect on the tax in the next period.

20. Again, since β^t and β^{t+1} are functions of the tax/subsidy rate, a fuller analysis requires expressing the optimal tax/subsidy in terms of underlying parameters, as in the earlier analysis.

9. Learning with Monopolistic Competition

1. In particular, the value of ζ_i (the relative scale parameter) is sensitive to the elasticity of demand, the amount of learning in this sector, the elasticity of labor supply, what happens in other sectors, and so on. As we have noted, $p_i^t = (1 - \delta\zeta_i h_i)/(1 - 1/\acute{\eta}_i)$. On the other hand, $p_i^{t+1} = 1/[H^i(1 - 1/\acute{\eta}_i)]$. If $H^i = 1/(1 - \delta\zeta_i h_i)$, the price would be the same both periods. The smaller the elasticity of demand, the higher the markup, so the larger the sensitivity of price to any differences between H^i and $1/(1 - \delta\zeta_i h_i)$.

2. More formally, we can use the techniques used in chapter 8 to calculate an approximation to ζ^m. Equations (2) and (3) allow us to write ζ^m as a function of h. For small h:

$$\zeta^m(h) = \zeta^m(0)[1 + [d(\ln \zeta^m)/d(\ln L^t)]\varepsilon_{Lw}\delta\zeta^m h].$$

Similarly, $\zeta^m(0) = 1 + \varepsilon_{Lw}H_0$. Hence,

$$\zeta^m(h) = (1 + \varepsilon_{Lw}H_0)[1 + d(\ln \zeta^m)/d(\ln L^t)]\varepsilon_{Lw} \, \delta\zeta^m h$$

$$= (1 + \varepsilon_{Lw}H_0)/[1 + (1 + \varepsilon_{Lw}H_0)(h\varepsilon_{Lw} - 1)\delta h\varepsilon_{Lw}],$$

which for small h can be approximated by

$$(1 + \varepsilon_{Lw}H_0)/[1 + (1 - \varepsilon_{Lw}^2 H_0 \delta h)].$$

ζ^m is greater than unity and increases with the elasticity of labor supply and the level and elasticity of learning.

3. For instance, in the absence of learning, by setting $\kappa_i(1 - \tau_i) = 1$, or $\tau_i = 1 - 1/\kappa_i = 1/\acute{\eta}_i$, we can correct the monopoly distortion.

4. Equations (A.1) and (A.2) can be derived explicitly from equations (2) and (3) in the text. It turns out simpler if we describe the equilibrium in terms of labor inputs each period rather than consumption levels.

$$p_i^t(L_i^t) = [1 - \delta(L_i^{t+1}/L_i^t)h_i](1 - 1/\acute{\eta}_i),$$

or

$$L_i^{t+1} = [1 - (1 - 1/\acute{\eta}_i)p_i^t(L_i^t)][L_i^t/(h_i\delta)]. \tag{A.5}$$

Equation (A.5) is a simple equation giving second-period labor input as a function of first-period labor input, where demand and technology elasticities may not be constant and may depend on first-period consumption (labor input). Equations (2) and (3) were the first-order conditions for the monopolist maximizing output (input), *given the output of all other firms*. This is the monopolistically competitive equilibrium. The demand elasticity, $\acute{\eta}_i$, is thus the partial elasticity—taking the output of other firms as given, how much the firm's price would have to change if it wished to increase sales by 1 percent. In the symmetric case, however, we can reinterpret (5) as a general equilibrium equation. In other words, we can drop the subscript i, and let $p^t(L/n)$ be the price level that would elicit the labor supply L, i.e., that solves (in the case of separable utility functions)

$$u_i'(L/n) = pv'(L). \tag{A.6}$$

Note that $d(\ln p)/d(\ln L)$ is no longer the inverse of the (partial) elasticity of demand, but is given by

$$d(\ln p)/d(\ln L) = 1/\{d(\ln u')/d[\ln (L/n)] - d(\ln v')/d(\ln L)\}. \tag{A.7}$$

We can then write

$$d(\ln L^{t+1})/d(\ln L^t) = 1 + d(\ln h_i)/d(\ln L_i)$$

$$- \gamma^t[d(\ln p^t)/d(\ln L^t) + d(\ln \acute\eta_i^t)/d(\ln L_i^t)]/(\eta - 1), \qquad (A.8)$$

where

$$\gamma = (1 - 1/\acute\eta_i)p_i^t(L_i^t)/[1 - (1 - 1/\acute\eta_i)p_i^t(L_i^t)]. \qquad (A.9)$$

Similarly, we can rewrite (A.3) to generate an equation giving second-period output (input) as a function of first-period input.

$$p_i^{t+1}(L_i^{t+1}H) = H/(1 - 1/\acute\eta_i) \qquad (A.10)$$

Again, we now focus on the symmetric equilibrium, with $L_i^{t+1} = L^t/n$. Totally differentiating (A.10), we obtain

$$\frac{d(\ln L^{(t+1)})}{(d \ln L^t)} = \frac{h(\eta'-1)}{1 + \frac{\eta'}{\eta'-1}\left(\frac{d \ln \eta'}{d \ln L^{t+1}}\right)}. \qquad (A.11)$$

Equations (A.5) and (A.10) define the equilibrium. It should be clear that they can take on quite complicated shapes, since h, $\acute\eta_i$, $d(\ln u')/d[\ln (L/n)]$, and $d(\ln v')/d(\ln L)$ can vary with the level of consumption and labor inputs.

It is clear that there can then exist multiple equilibria.

10. Long-Term Growth and Innovation

1. And typically, asymptotic labor supply goes to either its upper or lower bound.

2. Obviously, in the more general case, there is a transition to the steady state.

3. It is worth noting that the standard theory of monopoly requires a demand elasticity that is greater than unity.

4. $u^{t+n} = U_0 + ng$.

$$\Sigma\, ng\delta^n = \Sigma\, g\delta^n + \delta\, \Sigma\, g\delta^n + \delta^2\, \Sigma\, g\delta^n \ldots = g\, \Sigma\, \delta^n/(1 - \delta) = g/(1 - \delta)^2.$$

5. $\alpha_M/\upsilon_M - (1 - \alpha_M)/(1 - \upsilon_M) + (\alpha_M/\upsilon_M)h/(1 - \delta) = 0$, or

$$\alpha_M(1 - \upsilon_M)[1 + h/(1 - \delta)] = \upsilon_M(1 - \alpha_M), \text{ or}$$

$$\upsilon_M[1 + \alpha_M h/(1 - \delta)] = \alpha_M[1 + h/(1 - \delta)].$$

6. For convenience, we switch to continuous time. Analogous results hold in the discrete time version.

7. The full intertemporal maximization problem is somewhat more complicated and can be analyzed using standard techniques.

8. The simplest form is $\hat{H} = 1$ for $E < E^*$; that is, there is no learning.

9. As we noted earlier, Arrow's original model focused on learning through investment. But with the capital output ratio fixed, cumulative investment grows with output.

10. Long-run steady states require asymptotically constant elasticities.

11. The Infant-Economy Argument for Protection: Trade Policy in a Learning Environment

1. Indeed, in some circles, opposition to free trade would be grounds for taking away one's certification as an economist.

2. Actually, the circumstances in which free trade is welfare enhancing are more restricted than is justified by this widespread presumption. For instance, when there are imperfect risk markets, free trade can actually make all individuals worse off (see Newbery and Stiglitz 1982). For a broader discussion of these issues, see Charlton and Stiglitz (2006, 2012).

3. It is worth noting that in popular discussions, it is often argued that openness leads to more learning and that there are learning benefits associated with trade. While this may be true, this positive learning effect needs to be offset against the effects associated with the structure of production. For reasons set forth in earlier chapters, we believe that the latter effects predominate. For a contrasting view, see Grossman and Helpman (1991), who, however, essentially entirely ignore the effects upon which we focus.

4. There is a long history of the infant-industry argument for protection, dating at least back to the mid-nineteenth-century work of List (1841). For a more extensive discussion, see Chang (2002, 2003); Charlton and Stiglitz (2005); and Stiglitz (2006a). The discussion here borrows heavily from Dasgupta and Stiglitz (1988a).

5. Recall that we are using these terms as metaphors. *Agriculture* includes small-scale rural nonagricultural and craft activities. *Industry* may even include industrial agriculture.

6. As discussed at greater length in chapter 5 and formalized in earlier chapters in part 2.

7. Stiglitz and Weiss (1981) explain why there may be credit rationing with imperfect and asymmetric information.

8. In the case of East Asia, governments used rule-based systems, providing more finance to those firms that had demonstrated prowess in exporting, and especially in areas where there were significant potential technological spillovers.

9. A bank's granting of a loan, in this case, is little different from a government's decision about which researcher to support—except that in the latter case, the government can simultaneously evaluate different research applications, while the bank can only guess at what other researchers are receiving funding. Moreover, the government can assess the marginal social return associated with each project, while the private lender has to judge the expected (private) return of this particular project, i.e., the (average) probability of success times the profits that the firm will get if successful. As explained in earlier chapters, in the case of innovation, expected private returns are not closely aligned with (expected) marginal social returns. With government financing, projects get funded so long as expected marginal social returns are positive. With private financing, projects get funded so long as the bank's expected return is positive. In both cases, those with more learning potential get more access to funds. But the cutoff and the mix of projects can be markedly different. In particular, the government can take into account learning externalities.

10. One might argue that since patent protection is time limited (though firms have found clever ways of extending the effective life of patents), so too should protection be time limited. The analysis presented below shows that this may not be correct.

11. What is essential in this example is the unitary elasticity of substitution. There are problems in modeling long-term economic growth with nonunitary elasticity of substitution and differential rates of growth of productivity. With an elasticity of substitution less than unity, the high-productivity growth sector's share of global gross domestic product shrinks to zero; while with an elasticity of substitution greater than unity, it expands to unity. Both limits are uninteresting. At the same time, it is unsatisfactory simply to assume a unitary elasticity of substitution. A finite-period model of the kind presented below avoids this modeling dilemma.

12. As we noted in chapter 5, the existence of profits may not suffice to attract actual entry. Entrants care about what the market will look like after entry, and they may believe that after entry, competition would be sufficiently keen that they would make a loss. The incumbent can take entry-deterring actions which reinforce such beliefs. The incumbent can, in particular, undertake sufficient learning that it preempts entry of rivals.

13. It should be clear that in a more competitive marketplace, the sum of profits should be lower.

14. With constraints (and costs) associated with levying taxes, especially in developing countries, protectionism (a hidden tax) may seem preferable. But this lack of transparency is an important argument against protectionism.

15. Taxpayers in the country bear the costs of the infant-industry subsidy to help create an effective competitor. Government must judge whether such subsidies are worthwhile by assessing the value of the future profits it can glean from the eventual profits of the entrant plus the value of the consumer surplus that accrues to their citizens, ignoring the benefits to citizens of other countries. (If the monopolized good is an input into production, there can be further benefits from competition: higher profits to the firms that use the good as an input, higher consumer welfare from the lower prices on consumer goods that may result, higher tax revenues to the government. Such benefits are themselves global public goods, since all producers anywhere benefit. But there are likely to be localized benefits as well, and not just from knowledge spillovers; there can be, for instance, beneficial design interactions between the user and producer of the intermediate products.)

16. See chapters 3, 4, and 12 for a discussion for why geography matters for spillovers.

17. The only assumption that differs from that of earlier chapters is that concerning spillovers over space. We assume that there are perfect spillovers within a country, but no spillovers across borders. Obviously, this is a limiting case, and the more general case can be analyzed as in chapter 8.

18. Alternatively, it was argued that if it eventually should develop a comparative advantage in manufacturing, there was no point to anticipating the change. The critical assumption was that technology was exogenous.

19. As we noted in chapter 1, there are difficult problems in ascertaining a country's dynamic comparative advantage.

20. In the previous chapter, for simplicity, we normalized our units so that (in the first period) $Z = 1$ for both goods.

21. In a two-period model, nothing depends on this parameterization. In a longer-run model, we want there to be a steady state rate of growth. As the economy grows, if productivity growth depends on scale, then the rate of growth will increase.

22. It is easier to accommodate cases where η is greater than or equal to one.

23. We can use the framework of the previous chapters to give a precise calculation of U_0^a/U^f by using the indirect utility function, $U_0^a/U^f = V(p^L)/V(p^D)$.

24. That is, under our assumptions, using the notation of the previous footnote, p^L and p^D are fixed, and all that happens over time is that the scale of the economy changes.

25. As we noted earlier, the assumption of a fixed labor supply is, however, not innocuous.

26. $U^{t+1} = U^t + \ln(1 + g)$, and, using the same techniques employed earlier,

$$W = \Sigma\, U^0\{(1 + t[\ln(1 + g)]\}\delta^t,$$

from which (4) follows directly.

If U is not logarithmic but exhibits constant elasticity with respect to the scale of consumption (as before), with a fixed elasticity of marginal utility, there is a parallel analysis.

27. If at $\pi = 0$, $\partial U^*/\partial(\ln \pi) + g'\pi/[(1 - \delta)(1 + g)] < 0$, then there can be a corner solution at $\pi = 0$; and if at $\pi = 1$, $\partial U^*/\partial(\ln \pi) + g'\pi/[(1 - \delta)(1 + g)] > 0$, there can be a corner solution at $\pi = 1$. In particular, this means that if $g'(0)/\{(1 - \delta)[1 + g(0)]\} < 1 - k$, then $\pi = 0$.

28. So long as $g'' < 0$ and $g' > 0$, there is a unique solution to (7).

29. k is a measure of the difference in comparative advantage; κ is a measure of the difference in absolute advantage in the industrial sector.

30. Generalizing the results to the case where it is a trivial matter.

31. It would be an easy matter to generalize this to the case where the lagging country is nonnegligible in size and continues to import some industrial goods. We would then need an additional equation to solve simultaneously for π^{L*} and π^{D*}.

32. That is, along the first-order condition $W_\pi = 0$, $d\pi/d\kappa = -W_{\pi\kappa}/W_{\pi\pi}$. The denominator is always negative, so that the sign of $d\pi/d\kappa$ is the same as that of $W_{\pi\kappa}$, and the sign of $W_{\pi\kappa}$ is the same as that of $(f_{\pi\kappa}/f_\pi) - [f_\kappa/(1 + f)]$. We expect that an increase in κ reduces growth (when there is less catching up, there is less growth, at a given level of π), and that an increase in κ also reduces the marginal benefit of increasing π. In effect, we assume that the marginal effect dominates.

33. Notice the similarity between this result and the simple model in the appendix of chapter 2, where the backward firm never fully caught up.

34. Moreover, as we noted in an earlier footnote, even in the "normal" case, the result that the curve giving the optimal value of π as a function of κ is downward sloping depended on our assumption concerning the sign of $(f_{\pi\kappa}/f_\pi) - [f_\kappa/(1 + f)]$.

35. Assuming that the constraint (21) is binding.

36. See the discussion of dynamics with learning gaps earlier in this chapter.

37. See Korinek and Servén (2010). Other instruments, such as investment subsidies/tax credits and interest rates, are also relevant. For a discussion of the use of exchange rates (and changes in exchange rate) in the context of the East Asia miracle, see Stiglitz (1996).

38. Using the indirect utility function, it is easy to show that the effect of a change in the exchange rate depends on the magnitude of net exports or imports.

39. Though one should note there is some evidence of convergence within manufacturing. See Rodrik (2013).

40. But see Greenwald and Stiglitz (2010a, 2010b); and Stiglitz (2006a).

41. However, the same qualification exists as in note 39 (see Rodrik 2013).

41. An analysis of the full global general equilibrium effects of such policies, if pursued by enough developing countries to have systemic effects, is beyond the scope of this book.

12. The Role of Industrial and Trade Policy in Creating a Learning Society

1. For a survey, see, e.g., Acemoglu and Robinson (2012). For earlier discussions, see the 2001 World Development Report (World Bank 2001), Stiglitz (2000b), North (2005), and the references cited there.

2. For a broader discussion, see Stiglitz 2012b.

3. See Stiglitz (1998c), which describes development as a "transformation" into a society which recognizes that change is possible and that learns how to effect such.

4. See Stiglitz (2002), which lays out some of the reasons that matters turned out so differently from the way that the advocates of structural adjustment had anticipated.

5. We should be somewhat more careful. Williamson (1990) articulated the Washington Consensus in the context of the policies that were pushed in Latin America, and Africa's structural adjustment policies began well before his formulation of these ideas. But the underlying beliefs about economics and economic policy, which Williamson put so clearly, had long dominated development thinking in the international economic institutions, and the ideas were applied not just in Latin America. Williamson actually distanced himself from some of the ideas that subsequently get identified with the Washington Consensus. He was, for instance, always cautious about capital market liberalization, and capital market liberalization was *not* part of the set of policies he identified as the Washington Consensus (see Stiglitz 2008c; Williamson 2008; as well as other papers in Serra and Stiglitz 2008). For further discussion of the Washington Consensus, see Stiglitz (1998a, 1999c, 2002a).

6. In chapter 14 we will explain why even this prescription may have been misguided, at least from our learning perspective; what matters is *real* stability, as much or even more than *price* stability, and the excessive focus on the latter may have undermined the former (see Stiglitz et al. 2006).

7. It is, perhaps, worth noting that what is viewed as corruption in one society may not be viewed in that way by others. Many outsiders look at the American system of large campaign contributions and revolving doors,

which seems to "buy" favorable legislation, as a form of corruption, even if there isn't money stuffed into brown paper envelopes for the politicians themselves.

8. Dixit (2012) has argued that firms from developing countries may have a knowledge advantage in dealing with governments of other developing countries.

9. See chapters 3 and 4 for more extensive discussions of why geography matters and why learning (and learning policies) that may be relevant in one locale may be less so in others.

10. Herbert Simon emphasized that if there are differences in the performance of public and private enterprises, the differences could not be explained just by differences in incentives, since in both, typically, most individuals work for others and have to be incentivized (see, e.g., Simon 1991). "This examination of authority and organizational identification should help explain how organizations can be highly productive even though the relation between their goals and the material rewards received by employees, if it exists at all, is extremely indirect and tenuous. In particular, it helps explain why careful comparative studies have generally found it hard to identify systematic differences in productivity and efficiency between profit-making, nonprofit, and publicly controlled organizations" (Simon 1991, 39).

11. Rashid (2011, 2012) and Detragiache, Tressel, and Gupta (2008) provide data and an empirical analysis strongly supporting this conclusion. Greenwald and Stiglitz (2003) present the general theory.

12. See, e.g., Shapiro and Stiglitz (1984); Stiglitz (2002b); and the references cited there.

13. Episodic periods of labor scarcity, where firms cannot easily fill vacancies, may provide further impetus for highly risk-averse firms to focus on saving labor. See, e.g., the historical discussion of Salter (1960) and the literature to which it gave rise, including Kennedy (1964); Fellner (1961); Atkinson and Stiglitz (1969); Ahmad (1966); Stiglitz (2006b); Samuelson (1965); and Drandakis and Phelps (1966). For a recent discussion, see Acemoglu (2010), who shows how labor scarcity can encourage innovation if the technology is "strongly labor saving."

14. It is difficult to track inequality because of data limitations. The *Africa Progress Report* states that 24 countries in Africa have Gini coefficients in excess of 42, the level in China. It also points out that in a number of cases, recent growth has not been matched by falling poverty—which they attribute to inequality: "In many countries, the pattern of economic growth is reinforcing these inequalities" (Africa Progress Panel 2012, 16).

15. See, e.g., Newbery and Stiglitz (1982), who show that free trade can make everyone worse off (that is, it can be Pareto inferior) when there are imperfect risk markets, because it increases risk.

16. There is a long-standing theoretical presumption that this would be so for advanced countries (Stolper and Samuelson 1941), but there is also evidence that this is so even for developing countries (see Stiglitz 2006a). With trade liberalization often associated with an increase in unemployment, it is not surprising that there are adverse distributional consequences: Those at the bottom are most likely to be laid off, and higher unemployment puts downward pressure on wages (see Furman and Stiglitz 1999). The adverse effects of trade liberalization were often exacerbated by simultaneous measures liberalizing financial and capital markets, which contributed to economic volatility (see, e.g., Stiglitz 2008a, 2010e, 2012a). For further discussions of the possible adverse effects of liberalization on inequality, see World Bank (2005); and Topalova (2010).

17. See, e.g., Rodriguez and Rodrik (2001). Wacziarg and Welch (2003) found that roughly half of the countries in their survey experienced zero or even negative changes in growth post-liberalization.

18. A few econometric studies (cross-country regressions) have been particularly influential (see, e.g., Dollar 1992; Sachs and Warner 1995). But while these authors were careful to qualify their results, others have not been.

19. See Charlton and Stiglitz (2006, 2013). They note that sub-Saharan Africa's share of world exports decreased from 3.9 percent in 1980 to 1.9 percent in 2006, and the least-developed countries did even worse, with their average share falling from .06 percent to .02 percent over the period. Part of the explanation, as they point out, is that there are other nontariff barriers to trade, including supply constraints and infrastructure deficiencies, providing the rationale for "aid for trade."

20. See, in particular, Chang (2002, 2003). Moreover, developing countries that have reduced their tariffs have not been able to make up for the resulting shortage of revenues, e.g., through value-added taxes.

21. And indeed, with constraints on taxation (or subsidies), differential taxation of traded goods (as compared to domestically produced goods) is in general desirable (Dasgupta and Stiglitz 1971, 1974, 2000; Emran and Stiglitz 2005). These results are consistent with those that show that certain types of liberalization (e.g., of intermediate goods and capital goods) may have beneficial effects (see Estevadeordal and Taylor 2008). The effects of liberalization may depend too on the economy's situation and structure: When there already is a high level of unemployment, liberalization may have adverse effects, even if it has more positive effects in other circumstances (see Charlton and Stiglitz 2005).

22. See the brief discussion in chapter 4 above and the more extensive discussion in chapter 14 below.

23. Where tariffs are much higher on produced goods than on raw materials (see, e.g., Charlton and Stiglitz 2005).

24. Recall our discussion in chapter 3 on the observed distance in product space between the production of different commodities, with mining and

hydrocarbons being more distant from other commodities (see, e.g., Hidalgo et al. 2007).

25. Irwin and Kroszner (1999) outline the conversion of the Republican Party away from its long-time support for industrialization behind high tariff walls, beginning in the 1940s.

26. Of course, trade interventions have sometimes not worked out well. (They have been used as protectionist tools by special interests, rather than to redirect society's resources toward creating a learning society.) But the history of successful interventions suggests that failure is not inevitable. And hopefully, countries will learn from the failures (and successes) of the past, so that the returns from future interventions will presumably be greater than those from past interventions.

27. In chapter 2, we noted similarly that the existence of large unexploited potential productivity was confirmed by special historical circumstances where there was a sudden need to increase output.

28. Moreover, the circumstances confronting Latin America in the 1960s and East Asia in the 1980s and 1990s were markedly different. It is not obvious that an export growth strategy would have worked in the 1960s.

29. See Rodrik and Subramanian (2005) for the case of India. Rodrik (2001) shows that growth relative to all developing countries actually increased from 1975 to the 1980s, even though import duties increased.

30. As we noted earlier, U.S. public investments in research have had enormously high returns (Stiglitz and Wallsten 1999; Council of Economic Advisers 1995).

31. See the references cited earlier in this chapter.

32. The returns on U.S. government investments in technology and science are even higher than those of the private sector (which in turn are far higher than private sector returns elsewhere; see Council of Economic Advisers 1995).

33. Though some conservatives do argue, on this basis, that there should be a return to the gold standard, and that there should be no role for discretionary monetary policy. However, since the failure of monetarism, these extreme positions have garnered little support among economists.

13. Financial Policy and Creating a Learning Society

1. The existence of these externalities provides the rationale for financial sector regulation, and the failure to adequately take into account these externalities provides an important part of the explanation of the 2008 and other financial crises (see Stiglitz 2010b).

2. At the center of lending activities are issues of information: assessing credit worthiness and monitoring fund usage. Markets characterized by imperfect and asymmetric information—features which are central to financial

markets—are inherently characterized by externalities, resulting in market allocations not being (constrained) Pareto efficient (Greenwald and Stiglitz 1986). There are, in fact, a number of distinct categories of externalities, besides those associated with the macro-instability upon which the next chapter focuses. Actions (investments) affect credit constraints, self-selection constraints, incentive compatibility constraints, and price distributions.

3. By the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994.

4. They may also have learned bad risk management practice or a variety of other bad practices that have come to light in the aftermath of the financial crisis.

5. See also Beck, Demirgüç-Kunt, and Martinez Peria (2010). We note that there are some empirical studies that claim the opposite (see Clarke, Cull, Martinez Peria, and Sanchez 2006).

6. Because of differential information, there is likely to be more subjective risk associated with a project in the developing country than in the home country, so the expected return required to induce a loan will have to be correspondingly greater. Matters may be even worse: The foreign lender may know that in competing for domestic borrowers, it faces a winner's curse. If there are local lenders with better information, the foreign lender only succeeds in "winning" if it offers a loan at too low of an interest rate—at an interest rate below the rate at which the (better) informed domestic lender is willing to lend.

7. Even borrowers will care about their lender's life expectancy. Lending is informationally intensive; borrowers develop a relationship with the lender, which makes the market for loans particularly imperfect. If a lender goes into bankruptcy in a downturn, borrowers are especially likely to find it difficult to find an alternative source of funds (see Jaffee and Stiglitz 1990; Greenwald and Stiglitz 2003).

8. The fact is that backing any country's banking system is its government; credit default swap spreads for banks and for the sovereigns of those banks are highly correlated (IMF 2012a). Argentina and the events of the global financial crisis of 2008 showed that depositor beliefs are not always fully rational: Governments did not always come to the rescue in the way hoped.

9. Iceland provides an interesting case, because depositors in the U.K. and the Netherlands evidently felt that their assets were secure, though any "rational" analysis would have made clear the severe limitations in this small country's ability to protect them. The governments of the Netherlands and the U.K. put enormous pressure on Iceland, but in the end, largely failed.

10. A result that is particular startling, given the destruction of the war and the impediments posed to trade.

11. While the United States did not fully create "national banking" until the 1990s, the country's national banking system was created in 1863, with the

National Currency Act, which created a system of regulation for nationally chartered banks (the Office of the Comptroller of the Currency, under the U.S. Department of Treasury). These newly established nationally chartered banks were able to attract funds from outside the state, and though funds didn't flow as freely as they might have with banks that could operate freely across state boundaries, funds flowed more freely than in the previous regime.

12. More recently, some foreign banks have engaged in extensive consumer lending, taking advantage of their "learning" about how to better exploit uninformed consumers, replacing in some cases even more exploitative local money lenders.

13. Thus, *financial market liberalization* refers to opening up a country's markets to foreign financial institutions and the deregulation of financial markets more generally, while *capital market liberalization* focuses on the movement of capital itself into and out of a country. Discussions of capital market liberalization usually focus on short-term capital flows (bank lending, portfolio investments) rather than foreign direct investment.

14. See, e.g., Prasad et al. (2003) and Kose et al. (2006). For these authors, the fact that volatility did not decrease in many of the countries which became more financially integrated into the global economy was a puzzle. As Stiglitz (2008a) pointed out, in models with imperfect and asymmetric information (endogenous capital market imperfections), pro-cyclical capital flows could easily be explained. More generally, in models with finite-lived individuals, capital market liberalization could lead to more volatility of consumption. (The standard models assumed infinitely lived individuals.)

15. See, e.g., Stiglitz (2000a, 2002a, 2006a, 2008a); Stiglitz et al. (2006, 2008); and the references cited there.

16. Thailand provides an example. There is an important distinction between short-run flows, the major effect of which may be an increase in the exchange rate, thus discouraging export sectors, and foreign direct investment, which *may* go into sectors associated with more learning and learning externalities.

17. See Stiglitz (2002) for a fuller discussion of the East Asia crisis and its impact on the structure of these economies.

18. It is worth noting that in Arrow's (1962a) original analysis of learning by doing, learning was related not to output, but to investment.

19. Earlier discussions also noted other impediments to financing research. The borrower had to disclose enough information to make the lender willing to provide money, but then the lender could "steal" the idea, appropriating the returns for himself.

20. For a fuller discussion, see Stiglitz (2010c).

21. This section borrows heavily from Greenwald and Stiglitz (2003); Hellman, Murdock, and Stiglitz (1996, 1997, 1998, 2000, 2002); Honohan

and Stiglitz (2001); Murdock and Stiglitz (1993); and Stiglitz and Uy (1996).

22. Part of the reason is that information (like knowledge more generally) is a public good. If capital markets were really informationally efficient, as its advocates claim, there would be no incentive to gather information. Everyone would try to be a free rider on the investments in information of others. While securitization may lead to improved risk diversification, it had adverse effects on incentives for assessing credit worthiness and monitoring, and these effects played out disastrously. For the general theory, see Grossman and Stiglitz (1976, 1980). For a discussion of the problems posed by securitization, and an explanation why the contention that it improves risk diversification may be incorrect, see Stiglitz (1992, 2010b).

23. Financial restraint needs to be distinguished from financial repression, which typically entailed large negative real interest rates. One of the standard arguments against financial restraint was that the lower (real) interest rates associated with it led to less savings; but interestingly, the East Asian countries all had very high savings rate. This may be partly because the interest elasticity of savings may be very low, but it also may be partly because government policies enhanced both the safety and convenience of financial savings.

24. The lower lending rates in turn helped increase firm equity, enabling them to engage in more risky investment (see Greenwald and Stiglitz 1993).

25. Moreover, it has also been more widely recognized that private banks also engage in "connected" lending, and, especially when the private bank looms large in the economy, taxpayers wind up picking up the tab. The distinction, in this sense, between public and private institutions is somewhat blurred. Monitoring public institutions may, in fact, be easier than monitoring private institutions.

26. We are suggesting, in other words, the creation of Arrow-Debreu securities related to the macro-state of the economy. Though it should be relatively easy to create these risk products, neither government nor the private sector has done so.

Governments and international organizations have, however, created instruments designed to eliminate, or at least reduce, risks associated with expropriation.

27. Moreover, there are limits to the interest rate that banks can charge in the initial period, because of adverse selection and incentive effects, described by Stiglitz and Weiss (1981).

28. This argument is quite apart from the question of the role that small firms play in generating new ideas.

29. As the prime minister of one developing country argued, this was the most important example of a taking by one country of another country's intellectual property (see Stiglitz 2006a).

14. Macroeconomic and Investment
Policies for a Learning Society

1. As we noted in chapter 4, there are exceptions, including the increase in productivity in the 2008 U.S. recession. In footnote 12 of chapter 4, we discussed alternative explanations. Note, however, that while the above analysis suggests that typically, foreign investors are more sensitive to adverse signals, domestic investors have access to inside information. Thus, there will be some cases where that inside information indicates to them to leave before there is an adverse public signal to which foreign investors react. In the Tequila crisis of 1994/1995, it appears that Mexicans attempted to take their money out of the country first (see Lederman et al. 2003).

2. There are, of course, other explanations. For example, they have greater incentives to become tax havens, because the loss in domestic revenues from increased tax avoidance is more than offset by the increase in foreign revenues.

3. Since weaknesses in the local banking system would not be highly correlated with weaknesses in the large international banks.

4. An exception occurs if banks become so undercapitalized that they "gamble on resurrection."

5. Moreover, as we noted above and in chapter 4, learning benefits from having a stable environment.

6. This is the case in most developing countries. Some critics have suggested that a low exchange rate exposes a country to more inflation. Two responses are in order: First, that would only be the case if the central bank did not take offsetting actions. When the economy is already at full employment, the exchange rate affects the composition of output, and it may still be the case that it shifts it toward the learning sectors. Second, inflation is affected not by the level of the exchange rate (which affects relative prices), but by changes in the exchange rate.

7. As we noted earlier, there are still costs of such an intervention—the opportunity cost of forgone consumption or investment. As we showed, the learning benefits may well exceed these opportunity costs.

8. See chapter 13 for a discussion of some of the adverse learning consequences of such policies.

9. In some cases, learning can be related to the choice of technology by the firm, and government policies can affect such choices.

10. Some investment treaties provide foreign investors greater protections than they do domestic firms (see, e.g., Stiglitz 2006a, 2008e).

11. We omit the inclusion of public goods the second period. It would complicate the analysis, without adding any insights into the particular questions at hand.

12. In this simplified formulation, because there is no second-period production of public goods, learning-by-doing in the production of public goods is not relevant. In a more general model, it would be.

13. We have made use of the standard results: $-V_{pi}^t/V_I^t = C_i^t = L_i^t$.

14. Moreover, individuals who are absorbed with ensuring their basic survival have less ability to learn. Good systems of social protection thus enhance individuals' learning capacities.

15. Intellectual Property

1. For a slightly longer discussion, see Stiglitz (2006a). (Stiglitz was the member of the Council of Economic Advisers responsible for innovation and intellectual property at the time the Uruguay Round was being discussed within the Clinton Administration.)

2. Conditional on the level of production, however, the level of expenditure on innovation may be optimal, as we noted earlier.

3. For a broader discussion of this issue, including empirical evidence, see Dosi, Marengo, and Pasquali (2006); and Dosi and Stiglitz (forthcoming).

4. We should reiterate our cautionary note: More competition does not necessarily lead to more innovation. As we noted in chapters 5 and 6, however, especially because of agency problems (managerial capitalism), a monopoly may have little to spur it into innovation. This effect may dominate in markets with only one or two firms.

5. Moreover, as we noted earlier in the volume, monopoly innovators fail to take into account any consumer surplus that results from large innovations — or the consumer surplus that accrues from higher levels of innovation.

6. With perfect information, presumably the owner of intellectual property could act as a perfectly discriminating monopolist and extract from potential users the surplus associated with the use of knowledge — so that there would be no distortion. But information is imperfect, and owners of intellectual property are far from perfectly discriminating monopolists. (For a discussion of imperfect information and monopoly distortion, see Stiglitz [1977].)

7. For a more extensive discussion of this point and related issues, see Mowery et al. (2001); David (2004a, 2004b); and Dosi and Stiglitz (forthcoming).

8. He in fact used his patent to try to organize an automobile cartel. Had the patent not been challenged by Henry Ford, who wanted to create a low-priced car, the development of the automobile would have been greatly impeded. For a discussion of this and other problems with the patent system, see Stiglitz, 2006a.

9. For an early discussion of the importance of the scope of the patent, see Merges and Nelson (1994).

10. There is now a large literature on this subject. See, e.g., Farrell and Shapiro (2008); Lemley and Shapiro (2007); and Shapiro (2001, 2010).

11. For a discussion of this story, see, e.g., Crouch (1989). Fighting their patent claims may also have diverted the attention of the Wright brothers from making further development in their own design, contributing to the United States falling behind Europe in the development of the airplane. The irony is that it appears that the critical insight concerning the control of the airplane had been patented decades earlier, in 1868, by British inventor Matthew Piers Watt Boulton. Had the patent examiners known of this earlier patent, they might not have granted the Wright brothers their patent. The limitations of the U.S. patent system may be further highlighted by the fact that the Wright brothers' original patent application in 1903 was rejected. It was only when they reapplied, using a patent attorney, that they were granted the patent.

12. Michael Heller and his coauthors have provided other examples of how patents can deter innovation, invoking the term *anticommons*. See Heller (1998, 2008); and Heller and Eisenberg (1998).

13. It used to be the case that once granted a patent, the owner could exclude others from using that intellectual property *until the patent was overturned*. This has become a source of special concern, given the large number of bad patents—patents which should not have been issued, some of which are eventually overturned. Those who have such patents could impose extortionary demands on those who wish to make use of their patents. These patent owners can even insist that those to whom they grant license not sue—eliminating a major source of challenge to patents.

14. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388 (2006). Case documents can be found at http://www.supremecourt.gov/opinions/05pdf/05–130.pdf. (Accessed January 15, 2013.)

15. The court decision went some way to creating what intellectual property lawyers like Reichmann had long called for, a "liability system," under which those who use another's intellectual property have to pay compensation, but the owner of the intellectual property cannot exclude someone from using the property.

16. In 2012, a small company named X2Y sued Intel, Apple, and HP to exclude from the American market all of Intel's advanced microprocessors, all of Apple's computers (which employ these microprocessors), and those HP computers that do so. The claim was that these microprocessors infringed, in their "packaging," on an X2Y patent. X2Y had offered to sell this and a bundle of other patents for a few million dollars. Intel viewed it as a holdup and refused. The cost to Intel, Apple, and HP—let alone to the U.S. economy—of the exclusion would have been the order of billions of dollars. The law providing for the exclusion had a narrow exception—the exclusion order was not to

be issued if it was against the public interest. But the International Trade Court (ITC) had so narrowly defined the exception that it had been used only four times in forty years. The irony, of course, was that a law designed to protect American firms against foreign firms who violated the intellectual property rights of Americans was being used by a small American firm that had spent a miniscule amount on research—and far more on lawyers—to hold up some of America's leading IT companies, who were spending billions on research. Those who argued against the exclusion order contended not only that exclusion would have a large negative effect on the economy in the short run, but also that it would be counterproductive, disincentivizing research.

17. In practice, there is usually some value to a me-too innovation—for instance, there may be some patients for whom the side-effects are less—but still, the social return to such innovations is very limited and less than the private returns.

18. See chapter 6 for a formal model demonstrating this.

19. There is a large literature on the subject. For a review, see, e.g., Gallini (2002).

20. For instance, in the case of "orphan drugs," the life of the patent was extended, because it was thought that the benefits from greater incentives to innovate exceeded the costs. A still better way of creating incentives for such innovation, however, could have been provided through the prize system.

21. In April 2010, the U.S. District Court for the Southern District of New York invalidated patents on a pair of genes linked to breast and ovarian cancer held by Myriad. But in July 2011, the Court of Appeals for the Federal Circuit overturned this decision (Pollack 2011). In 2013, the Supreme Court supported the District Court decision that one could not obtain a patent for isolating a naturally occurring gene. *Association for Molecular Pathology v. Myriad Genetics,* 569 U.S. 12-398 (2013)

22. See Brunswick, 36 Fed. Cl. at 207; cited in Love (2004, 13).

23. Differences in politics—including the influence of the pharmaceutical and entertainment industries—may, however, be the predominant explanation of the differences.

24. *United States v. W. Elec. Co.*, 1956 Trade Cas. (CCH) ¶ 68,246, at 71,139 (D.N.J. 1956).

25. Earlier, we noted the airplane patent pool that helped resolve conflicting claims and that allowed progress to go forward on further developing the airplane.

26. Under a U.S. law called the Tunney Act (Antitrust Procedures and Penalties Act, 15 U.S.C. §16), members of the public have an opportunity to comment on a proposed settlement of a civil antitrust suit before it is accepted by a court. At the time of the proposed Microsoft settlement, I filed an affidavit together with Jason Furman (later the deputy head of the U.S. National

Economic Council and chairman of the Council of Economic Advisers under President Obama) explaining why limiting the length of the patent would be a preferable way to address the anticompetitive abuses.

27. The discussion of this section is adapted from Stiglitz (2008b); Stiglitz (2013a); and Dosi and Stiglitz (forthcoming. It draws heavily upon Freeman (1987); Lundvall (2010); and Nelson (2004).

28. The general theory of prizes is set forth in Nalebuff and Stiglitz (1983a, 1983b). Subsequently, there has developed a large literature on the use of prizes as an incentive system (including Love and Hubbard 2007; Davis and Davis 2004; and the papers cited there). Also, a bill has been introduced into the U.S. Senate to use prizes as a way of incentivizing medical research. And in 2012 the WHO Consultative Expert Working Group on Research and Development: Financing and Coordination (CEWG), linked with its Commission on Intellectual Property Rights, Innovation and Public Health (set up by WHO in 2003 to look at the relationship between intellectual property, innovation, and public health), recommended establishing a prize system, as well as other measures to make medical research more "open." The CEWG also recommended creating patent pools and putting research outputs that address the health needs of developing countries into the public domain or making them available through open licensing.

29. For a recent discussion, see Kremer and Williams (2010). For an earlier discussion, see Stiglitz (2006a) and the references cited in earlier footnotes.

30. The Royal Agricultural Society in England also provided prizes. Brunt, Lerner, and Nicholas (2011) show that these prizes provided effective inducements to research.

31. Some of the benefits from using competitive markets to disseminate the knowledge can be obtained if the government buys out patents, i.e., giving the patent holder what the monopoly profits would have been (see Kremer 1998).

32. This chapter is devoted to IPR and its impact on creating a learning society. There are also questions associated with producing knowledge, e.g., whether knowledge production is best carried on in public, private for-profit, or nonprofit institutions. The issues of production and finance largely can be separated. Production can be undertaken privately or publicly; finance can be undertaken privately or publicly. At one extreme are government research laboratories—publicly financed research that is also publicly "produced." The IPR system is often described as the polar opposite, a private-sector solution combining private funding and private finance. But this description is misleading in two respects that we have already noted: First, much of the innovation is based on basic research that is publicly funded and often publicly produced or at least produced by not-for-profit entities, such as universities. And second, in the case of both health and defense, even the seemingly "private" funding under

an IPR regime is really public funding, since all defense expenditures are from the public purse and since the government provides most of the funding for health care expenditures in most countries. Even in the most market-oriented country, the United States, much of the funding comes from government: The National Institutes of Health represent publicly financed and publicly produced research; and government spending on health care, both through its program for poor people, Medicaid, and its program for elderly people, Medicare, represents a large share of total health care spending.

33. We say "excessive" because it may in fact be optimal to have several independent, parallel research efforts.

34. The evidence is that capital markets do not fully spread risks faced by firms, because of imperfections of information. See, e.g., Greenwald and Stiglitz (1990), who discuss the effect of information imperfections on firm behavior and argue that informational problems in the capital market cause firms to act in a risk-averse manner. (See also Stiglitz, 1982c). There is also considerable empirical evidence that markets do not efficiently distribute risk; i.e., firms act in a risk-averse manner, even when risks are uncorrelated with the market (see, e.g., Stiglitz 1982b).

35. For broader discussions of these issues, see Cimoli et al. (2013); Lewis and Reichman (2005); Nelson (2004); and Odagiri et al. (2010).

36. In this sense, as we have repeatedly noted, knowledge is a public good. Indeed, it is a *global* public good (see Stiglitz 1999a).

37. There were two critical ideas in the resolution. It recognized that intellectual property "is not an end in itself" (statement by Brazil on September 30, 2004, before the WIPO General Assembly at the introduction of the proposal for a development agenda), and it reiterated WIPO's mission to "promote creative intellectual activity" and "the transfer of technology to developing countries." The new development agenda calls for ascertaining how different intellectual property regimes affect developing countries.

38. Statement by Brazil on September 30, 2004, before the WIPO General Assembly at the introduction of the proposal for a development agenda.

39. See, e.g., Stiglitz (2006a). Indeed, it was not even clear that the IPR regime that was foisted on the world through TRIPS was well designed for the United States, as we suggested in the beginning of the chapter. It reflected the interests of the entertainment and pharmaceutical industries, and not of the scientific community.

40. For instance, in granting pharmaceutical patents, developing countries should reserve the right to grant a compulsory license for any lifesaving or life-extending drug. To be exempt from this provision, the patent applicant would have to state that the patent does not cover any such medicinal use; and if subsequently such a use were established, the government would have the right to issue a compulsory license, limited, of course, to sales for such usages.

In the context of trade agreements, see Charlton and Stiglitz (2012) and Ismail (2007) for a discussion of the "right to development."

41. See also Odagiri et al. (2010) and the various chapters of Cimoli et al. (2013).

16. Social Transformation and the Creation of a Learning Society

1. As we also noted in chapter 3, in some quarters and in some countries, it appears that the notion that policies ought to be based on the principles of the Enlightenment has to be constantly relitigated.

2. For instance, a classic experiment in psychology by Bruner and Potter (1964) suggests that preconceived ideas serve as unconscious filters of sensory impressions.

3. *Confirmatory bias* is the tendency to search for, interpret, and remember information in a way that supports one's initial beliefs. For a survey, see Rabin and Schrag (1999).

4. On the other hand, it provides a more disciplined approach to the formation of beliefs than that based on "animal spirits," which suggests that any set of beliefs is possible.

5. In this sense, our analysis goes beyond standard behavioral economics, which has used insights from psychology to modify economists' traditional reliance on hypotheses concerning individual rationality.

6. Of course, sociologists have long recognized the importance of social constructions (see, e.g., Douglas 1986), but they have not focused on modeling "equilibrium," where there is some correspondence between beliefs and perceptions and what the individuals observe.

Some economic historians have also emphasized ideas similar to those articulated here (see, e.g., North 2005).

7. Standard theory treats the categories as if they were objectively determined. Standard rational expectations theory assumes that individuals use all the relevant information, updating prior beliefs through a Bayesian process. There are no biases.

This approach is also markedly different from the very interesting models of Piketty (1995); Bénabou (2008a, 2008b); and Bénabou and Tirole (2002), who assume that individuals *strategically* choose the probability that they will remember certain signals.

8. There are an infinite number of possible correlations between observables. Individuals have to choose which among these they study. They do not gather information about many of these possible correlations because the way

we see the world suggests that they are irrelevant. If we came to believe that they were relevant, they possibly would be. This is called preconfirmatory bias. Fryer and Jackson (2008) analyze bias that emerges from categorization. See also Loury (2002).

9. Again, there is a large literature in both psychology and economics consistent with this hypothesis. Smith et al. (2008) showed that invoking in experimental subjects the feeling that they have little power impairs their performance in complex cognitive tasks. Steele (2010) provides a survey of the literature demonstrating that cuing an identity associated with a stereotype, or cuing a condition that could confirm a negative stereotype, shifts an individual's performance in the direction of the stereotype. See also Hoff and Pandey (2011); and Afridi, Li, and Ren (2011). Experiments summarized by Compte and Postlewaite (2004) demonstrate that psychological states can affect performance. Among the earliest examples are the efficiency wage theories in economics, which noted that perceptions of unfairness can affect morale, which can in turn affect performance (see Stiglitz 1974; and Akerlof and Yellen 1986).

10. Efficiency wage theory (referred to in an earlier footnote) provided early examples of this. Perceptions of fairness can affect morale, morale can affect behavior, and this can explain the persistence of dysfunctional inequality.

11. The possibility of multiple equilibria of this sort is enhanced once it is recognized (as in chapter 11) that a more dynamic society enhances the returns associated with innovative skills and attributes, while in a less dynamic society the relative returns of bureaucrats may be higher. A high-learning society creates an ecology which is self-supporting.

12. See in particular the report of the National Commission on the Causes of the Financial and Economic Crisis in the United States (2011).

13. The complexity of the issues is illustrated by vicissitudes in attitudes toward government policies to restimulate the economy. In the aftermath of the collapse of Lehman Brothers, there was a moment in which all the world adhered to Keynesian ideas. But within two years, there was a shift toward "Hooverite" fiscal austerity policies—even though the empirical (scientific) evidence that such policies would lead to slower growth with disappointing results on deficit reduction had actually mounted in the interim.

14. But it is not as if those who believe in that institution—or even the smaller group that benefits from it—got together and figured out a set of beliefs that would accomplish what they sought. As we have noted, the theory we have presented is incomplete, in that it does not adequately explain when beliefs change and when they do not. But we believe it is a step forward to break out of the mold of rational expectations, in which the variables described above play no role.

15. Even the way we perceive institutions is affected by the prisms through which we look at the world, by our ideology. At one time, some economists suggested that institutions have a simple role in society—to fill in the "holes" in markets, to remedy market failures (see North 1973). Arnott and Stiglitz (1991); and Hoff and Sen (2006), as well as others, showed that nonmarket institutions purportedly resolving a market failure (like incomplete insurance markets) could, in this sense, be dysfunctional—they could lead to Pareto inferior outcomes.

More recent literature has highlighted the role of institutions in preserving inequalities—in the context of repeated games, equilibria in which one group is exploited by others may be sustained (see, among others, Dasgupta 2005; and Mookherjee and Ray 2003).

16. In the United States, those wanting to insulate the Fed from scrutiny as it provided massive subsidies to certain banks opposed Fed transparency (see Stiglitz 2010b).

17. The ideas in this paragraph are developed more fully in Stiglitz (2012b).

18. We are deeply indebted to Tim Besley for discussions on the ideas in this section (see Besley and Torsten 2009, 2010, forthcoming; and Besley, Persson, and Sturm 2010). Hoff and Stiglitz (2004a, 2004b, 2007) modeled the political economy of transition from communism to a market economy, employing analogous ideas. See also Acemoglu and Robinson (2000).

17. Concluding Remarks

1. See Stiglitz (2011) and the references cited there.

2. As we noted earlier, there can be some knife-edge circumstances, where firms begin in precisely identical situations, and remain so. But these are highly unstable. Any perturbation that leads one firm to have any advantage over others will have cumulative effects, until that firm becomes dominant.

3. We emphasized in earlier chapters that the market is *constrained* Pareto inefficient—even taking into account the costs of gathering information or removing information asymmetries.

20. Commentary: The Case for Industrial Policy

1. Aghion is also the author of a leading textbook in that area, *Endogenous Growth Theory* with Peter Howitt, and coeditor of the *Handbook of Economic Growth* with Steve Durlauf.

2. The choice by governments as different from one another as those of Germany, the U.K., the United States, or France, to protect the automobile

industries located on their territories, illustrates that point. The policy mix of providing credit support for the refinancing of automobile conglomerates, the "cash for clunkers" programs to speed up the renewal of domestic car fleets, R & D supports to accelerate the conversion of the car fleet to electric vehicles, the support given to first- and second-order subcontractors, the implementation of foreign aid policies whereby recipient domestic industries can also operate in the countries that provide the foreign aid, all this shows the extent to which countries which experience a crisis can mobilize the instruments available to them in order to avoid the collapse of a sector which is decisive for economic activity.

Afterword: Rethinking Industrial Policy

1. Thus, Frankel and Romer (1999) and Wacziarg (2001) point to a positive effect of trade liberalization on growth. In particular, Wacziarg showed that increasing trade restrictions by one standard deviation would reduce productivity growth by 0.264 percent annually. Similarly, Keller (2002, 2004) showed that 70 percent of international R & D spillovers are due to cross-country trade flows. More recently Aghion et al. (2008) pointed to large growth-enhancing effects of the trade liberalization and delicensing reforms introduced in India in the early 1990s, particularly in more advanced sectors or in Indian states with more flexible labor market regulations. And several studies summarized in Aghion and Griffith (2006) point to a positive effect of liberalizing product market competition and entry on innovation and productivity growth by incumbent firms, particularly those that are more advanced in their sector.

2. See also Young (1991).

3. However, Harrison (1994) questions these findings.

4. More specifically, relatedness between products i and j is measured by:

$$\phi_{i,j} = \min\{P(x_i / x_j), P(x_j / x_i)\},$$

where $P(x_i/x_j)$ is the probability that a country export (enough of) good i conditional upon exporting (enough of) good j.