# Stop Crying over Spilt Knowledge: A Critical Look at the Theory of Spillovers and Technical Change

Richard N. Langlois
University of Connecticut

Paul L. Robertson
University of New South Wales

## Abstract

This essay analyzes critically the idea of knowledge spillovers, especially as it enters the New Growth Theory. The conventional theory of spillovers, we argue, suffers from a thin and misleading account of the nature of productive knowledge. In this model, firms undersupply Ramp;D, which impedes economic growth and calls for research subsidies. We argue, by contrast, that a more subtle picture of the creation of knowledge, and the presence of network externalities (including true Marshallian external economies), tend to reverse the predictions of neoclassical theory: spillovers may actually lead to increases in the production of new knowledge.

# Introduction.

According to one popular story, technological knowledge is a peculiar kind of good. Because such knowledge is inherently slippery — it can be replicated and transmitted at essentially zero marginal cost — its generation contributes a positive externality to society. Knowledge "spills over" easily from its creator to benefit others, much as passers-by receive the esthetic and olfactory benefits spilling out from a well-tended private garden without having to pay for them. This means that the social benefit to the activity of research and development (R&D), which is what produces technological knowledge, must necessarily exceed the private benefit. As a result, R&D activity is undersupplied relative to the social optimum. And governments thus ought to subsidize or otherwise encourage R&D in order to promote social welfare.

This set of propositions has arguably constituted the central core of the neoclassical microeconomics of technological change for some decades. More recently, however, discussions of spillovers have shifted to a more aggregated venue, namely the so-called New Growth Theory (NGT) and New Trade Theory (NTT). As elaborated in several contexts, the basic argument of this theory is that technological change translates into growth. Technological change results from increases in knowledge, and those increase in turn result primarily (if not entirely) from research and development activities. As economic growth is assumed to be good, any factors that discourage investment in R&D are harmful. To the degree, therefore, that spillovers reduce the returns to innovation, they discourage the production of R&D and impair welfare. Echoing the traditional view in the microeconomics of technical change, Romer (1990, p. S89) and, more especially, Grossman and Helpman (1991, pp. 517-8) con-

tend that the returns to investment in R&D surely must be well above average returns in the economy and thus that *any* spillovers are potentially damaging.

In this essay we analyze critically the idea of knowledge spillovers. We call into question much of the web of syllogism involved in both the microeconomic and aggregate versions of the theory. We certainly do not deny the reality of knowledge spillovers, or that they may be important under certain circumstances. But we do dispute the policy conclusions of the mainstream view of these matters by questioning the plausibility of the assumptions on which the conclusions are laid. Our criticisms fall into three broad categories.

- **Knowledge and its production.** The conventional theory of spillovers, we argue, suffers from a thin and misleading account of the nature of productive knowledge (that is, as explicit, easily transferred "scientific" knowledge); a mechanistic reading of how such knowledge is produced (that is, through firm-level investment in explicit R&D activities); and an overly simple portrayal of how potential spillovers affect the incentives of innovators.

- **The complex role of spillovers**. Even if the neoclassical epistemology were correct, the role of spillovers in the economy would be far more complicated than the theory allows. If economic capabilities in the production of technological knowledge are not unbounded, then the R&D decision at the firm level is both "lumpy" and effectively supply constrained. This changes the analysis of the incentives to R&D in a world of spillovers. More importantly, however, where there are network externalities (including true Marshallian external economies), the prediction of neoclassical theory may actually be reversed. Spillovers may actually *increase* the firm's incentive to produce new knowledge.

- **Institutions and organizational forms.** Neither the conventional microeconomics of technical change nor (perhaps especially) growth theory concerns itself explicitly with underlying institutions or with the way production is organized. But institutions in the end determine the effects of knowledge spillovers as crucially as does the nature of knowledge itself. And organizational change — a neglected form of "technical" change — is

arguably a kind of learning not accounted for by the incentives to invest in R&D. Such change represents a true Marshallian external economy.

Our efforts, however, are not directed mainly towards intellectual (creative) destruction. We also make the affirmative case that spillovers may actually have a positive value for innovators and that a sensible economics of spillovers may be sought (at the microeconomic level) in the developing literature on strategic capabilities and (at the aggregate level) in institutional and evolutionary accounts of economic growth.

### *Knowledge, externality, and increasing returns.*

We may say that "spillovers occur when someone's actions affect anyone else in either a positive or negative way and this effect is not [fully] *paid for* (in the case of a benefit) or [fully] *compensated* (in the case of a cost)." (Bureau of Industry Economics 1994, p. 7, emphasis in original.) Spillovers are significant in the New Growth Theory because of the way in which that theory treats technological change. Whereas the Solow-Swan growth model regarded technological change as exogenous (Solow 1956; Swan 1956), Romer and other New Growth Theorists contend that economic motivations underpin substantial amounts of pure research as well as development. As a result, these theories are "endogenous," in that they explicitly model the production of technological change in terms of rational optimization.[1]

---

[1] There is a rhetorical tendency in the NGT to refer to the "old" (Solow-Swan) theory as "neoclassical growth theory"; the NGT is by implication not neoclassical or even perhaps "Schumpeterian" (Romer 1994, pp. 3 and 17; Grossman and Helpman 1994, pp. 27 and 30). In fact, however, it is hard to see what in the New Growth Theory is not neoclassical. As Grossman and Helpman (1994, p.34) point out, "[m]ost of the contributors to the new literature on innovation-based growth have adopted a general-equilibrium perspective." Moreover, from a methodological point of view, the NGT is arguably *more* neoclassical than the older Solow-Swan approach. One of the most characteristic aspects of neoclassical theory is precisely a positive heuristic that demands the progressive "endogenizing" of variables into the rational-

In a "generic" formulation of the NGT,[2] economic growth is postulated to be a function of technological change. Technological change, in turn, is a function of research and development activities, which are important because they lead to increases in knowledge. At its most basic, the NGT takes the following form:

R&D expenditure ⇒ increased knowledge ⇒ technological innovations
⇒ new products/processes ⇒ economic growth.

Growth may thus be viewed as essentially unbounded. The model is an economic sausage machine that assures us that economic growth will eventually ensue as long as firms invest in R&D. If we assume that economic growth is desirable, then we should try to minimize anything — such as spillovers — that may reduce incentives on the part of firms to invest in R&D.

As a theory of aggregate growth, the NGT postulates an aggregate production function:

$$Y = f(K, L, H; A),$$

where *Y* is national income; *K* is physical capital; *L* is labor (*i.e.,* labor power minus its intellectual component); *H* is human capital; and *A* is knowledge (or the technological state of the art). But, despite its aggregate nature, this production function is really just the sum of the production functions of individual production units. Its plausibility rests, therefore, on the

---

choice framework (Field 1979, 1984). That the Solow-Swan models leave technological change as exogenous is a strike against them on neoclassical grounds.

[2] There are many types of New, or Endogenous, Growth Theories. Paul Romer (1986, 1990, 1993a, 1993b, 1994) is responsible for at least five. As a family, however, the theories present a broadly similar treatment of research and development and technological change.

plausibility of the micro behavior that it postulates; and at the micro-level, it is easy to demonstrate that the behavior postulated is, at best, a number of special cases in a more general formulation of firm behavior.

One animating idea of the New Growth Theory is actually an old idea, Marshall's notion of increasing returns — even if, as we will see, the modern variant is in the end rather different from Marshall's. Increases in knowledge, in this theory as in Marshall, are the source of overall increasing returns, which are the engine of economic growth.

There is, of course, a trivial sense in which R&D can lead to increasing returns. If R&D is a fixed cost, it, like all fixed costs, can yield (internal) economies of scale, as output can be doubled without replicating the fixed inputs, and therefore without doubling all inputs. These economies can affect other firms and other industries; but, in the modern lexicon, they do so only as a pecuniary externality that lowers the prices of the inputs those firms and industries purchase. Moreover, this source of increasing returns will not matter much at the aggregate scale of growth theory.

The real meat of the "spillover" idea is that R&D can also, and more importantly, affect other firms and industries as a *technological* externality, in much the way that the private flower garden, say, can benefit the general public in ways not taken into account by the price system. Why cannot the effects be taken into account? In the theory of knowledge spillovers, the answer is that knowledge is (some combination of) nonrivalrous and nonexcludable. Nonrivalry means that one person's use of the knowledge does not diminish the knowledge available to others. And nonexclusion means that others cannot be kept from using (and therefore not be made to pay for) the knowledge.

Most of theory and policy — including, for example, patent policy — has focused on the question of exclusion or, to put it differently, of appropriability. Indeed, the possibility that, because of the slipperiness of knowledge, inventors might not reap the benefits of their inventions is an idea of ancient vintage. It probably goes back at least to Francis Bacon, and among economists was first articulated by J.-B. Say (Langlois 1986). It was, however, A. C. Pigou, Marshall's successor at Cambridge, who articulated the problem of knowledge as an externality in the form that has most influenced modern neoclassical theory: namely, in terms of the divergence between private and social benefits.

> Among these examples we may set out first a number of instances in which marginal private net product falls short of marginal social net product, because incidental services are performed to third parties from whom it is technically difficult to exact payment ... Lastly, and most important of all, it is true of resources devoted alike to the fundamental problems of scientific research, out of which, in unexpected ways, discoveries of high practical utility often grow, and also to the perfecting of inventions and improvements in industrial processes to the general public in the form of reduced prices. The patent laws aim, in effect, at bringing marginal private net product and marginal social net product more closely together. (Pigou 1932, pp. 184-5.)

To pure theorists, however, it is the nonrivalry rather than the nonexcludability that is significant. For if a piece of information[3] really can be used by all without altering its value, then the marginal cost of its distribution is near zero (its production being entirely a matter of fixed cost); and zero is thus what its price should be in a first-best equilibrium.[4] To the extent, then, that exclusion becomes possible, by a patent or other institutional arrangement, the excluding

---

[3]  Neoclassical economics treats "knowledge" and "information" as essentially identical concepts. For an argument that knowledge is more than just information, even accumulated information, see Langlois (1983).

[4]  In technical terms, it is the nonrivalry that creates a nonconvexity in the allocation problem; excludability is really a matter of institutional arrangements, which do not enter into the theory except by *ad hoc* assumption. On this see below.

firm may recoup its investment, but it is unlikely to charge a zero price and therefore unlikely to do its part to achieve the first-best optimum. So exclusion is not by itself a solution to the problem of knowledge spillover, except perhaps in a second-, third-, or $n^{th}$-best world.[5]

The Pigovian approach became ensconced in the modern (i.e., formal) literature through Kenneth Arrow's influential 1962 article, which touched off a wave of interest in the "knowledge problem." One manifestation of this enthusiasm was empirical, as researchers, especially Mansfield *et al.* (1977), investigated econometrically the divergences between social and private rates of return to R&D, which they determined to be substantial. Their findings were cited more than once in calls for government subvention of R&D. By the 1980s, this microeconomic research agenda had mostly faded into the intellectual background of neoclassical economics, probably as a result of the natural ebb and flow of research trends, although the rise of Coasean economics (Coase 1960), which can be seen as a broad-based attack on Pigovian welfare economics, may have been partly responsible. But with the rise of the NGT, the fundamental idea of knowledge spillovers regained immediacy in the late 1980s.

### *Knowledge and spillovers.*

Our skepticism about a general problem of underinvestment because of spillovers begins when we examine carefully the assumptions about knowledge itself that are implicit in the neoclassical account. There is a growing body of literature on firm behavior and organization that starts from quite different presuppositions. In the evolutionary theory of the firm (Nelson and Winter 1982) and the related literature on dynamic capabilities (Teece and Pisano 1994;

---

[5] Like the one we live in?

Langlois and Robertson 1995), economic actors are not "profit maximizing and far-sighted entrepreneurs," as Grossman and Helpman (1991, pp. 517-8) put it, but are, in Herbert Simon's well-known phrase, "boundedly rational." The import of this limitation is not that entrepreneurs do not try to do the best they can with what they have; rather, the point is that what they have to work with is qualitatively different from what the neoclassical picture suggests.

Invoking Michael Polanyi (1958), Nelson and Winter argue that much of what economic agents know is actually tacit knowledge, an inarticulate kind of knowledge that cannot be easily codified or transmitted in explicit form. Such tacit knowledge is in the nature of a skill, something acquired only through a costly process of imitation, apprenticeship, and trial and error.[6] Clearly, knowledge of this type is far less "slippery" than explicit knowledge. To the extent, then, that resources devoted to the production of knowledge (R&D) are actually directed at producing knowledge that is tacit in this sense, spillovers are far less of a problem (Cohendet, Héraud, and Zuscovitch 1993). Knowledge becomes far more of a private good.

Far from being slippery, indeed, knowledge may be "sticky," both because of its tacit character and because there may be transaction costs of communicating even valuable (and relatively explicit) knowledge to others across markets and within organizations; and there may even be costs of persuading those others of the value of the knowledge once communicated (Silver 1984). Cohen and Levinthal (1990) argue, moreover, that, rather than receiving

---

[6]     This distinction between explicit and tacit knowledge also tracks very closely Gilbert Ryle's (1949) famous distinction between "knowledge that" and "knowledge how."

new knowledge as a free gift, firms must invest resources in the "absorptive capacity" neces-sary to understand the results of the R&D spilling in from others.

Furthermore, if knowledge is dispersed and sensitive to context, then firms will likely learn different things even when searching for the same things; that is, R&D effort will lead to relatively localized knowledge. This possibility militates not only against the likelihood of knowledge undersupply but also against the problem of the oversupply of R&D that arises in a game-theory context. In those latter models, agents are represented as all seeking the same holy grail of explicit knowledge. Typically, the one who finds it first receives a patent and thus all the rents of the knowledge. Under the right assumptions — including some made in NGT models (Grossman and Helpman 1994, p. 37) — the result will be a "patent race" in which firms engage in *more* R&D than is socially optimal[7] (Stoneman and Diederen 1994). Where knowledge is tacit and variegated, however, there will be unintended learning, and the R&D racers will discover *different things* even if they set out looking for the same thing.

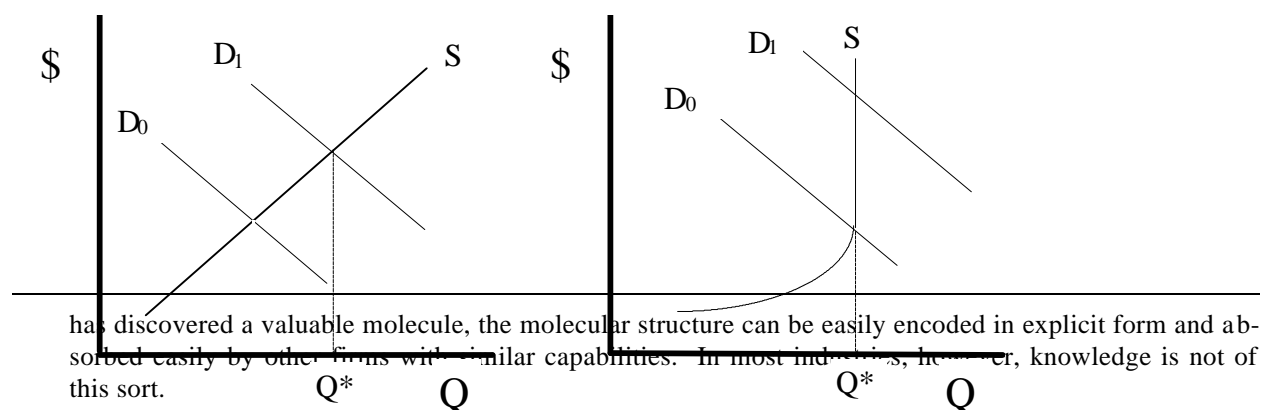### *Spillovers and the incentive to innovate.*

Clearly, of course, some aspects of all knowledge are at least potentially public; some aspects of the knowledge agents produce will have the character postulated in the neoclassical mod-els. For example, the success of otherwise private knowledge may generate certain kinds of "template" or "generic" knowledge that is easily transmitted.[8] In the early 1970s, the idea of

---

7    Alternatively, one may get no R&D at all, as each player realizes that the game is not privately worth the candle even if the resulting innovation would be publicly beneficial. (On this sort of problem, see gener-ally Richardson (1960).) All of these results, of course, are artifacts of the assumption of homogeneous and explicit knowledge.

8    As Levin *et al.* (1987) discovered in their survey of the appropriability of technology in various industries, knowledge seems to fit the neoclassical epistemology most closely in pharmaceuticals and (to a lesser ex-tent) in chemicals, where the knowledge transmitted is in fact typically of this template form. Once a firm

the personal computer was not considered valuable knowledge even by firms, like Hewlett-Packard and Atari, whom one would have thought to have had high absorptive capacity for such ideas (Langlois 1992). (These two turned away Steven Wozniak and Stephen Jobs, respectively.) But, with the success of the Apple II later in that decade, a number of firms tried to imitate the "free" idea of the personal computer, even if it cost them considerable resources to do so. This should not be surprising, as what constitutes information depends on what one already knows (Arrow 1974, Langlois 1983). And, over time, we might expect actors to share enough common structures of experience that bits of explicit information will be of value to them. To a firm not already skilled in a particular technology, providing even a blueprint may not be of much help absent a long and expensive learning process. But to a competitor who is up to speed, a small tidbit may be quite relevant.

It is our judgment that the economically important products of the conscious search for profitable information are mostly private, tacit, and expensive to transmit. Even if one adopts the opposite view, however, and takes seriously the neoclassical epistemology, the standard policy conclusions may still not follow. One reason for this is that the process of innovation may be "lumpy." Pigovian welfare economics diagnoses an "undersupply" of goods pro-



has discovered a valuable molecule, the molecular structure can be easily encoded in explicit form and absorbed easily by other firms with similar capabilities. In most industries, however, knowledge is not of this sort.

**Figure 1**
The Pigovian picture.

**Figure 2**
The capabilities picture.

duced under conditions of positive externality; but this presupposes that there is a smooth and continuous supply curve for those goods: because I am not properly compensated for the external benefits of my flower garden, I do not plant as large a garden as is optimal. (See figure 1. $D_0$ represents the demand for flowers felt by the gardener — or the demand for R&D felt by the firm — with uninternalized spillovers; $D_1$ represents the demand with spillovers fully internalized.) In fact, however, because of constraints of time or land or gardening ability (or all three), my costs of planting a garden may rise sharply beyond what I have already planted. In that case, the fact of spillovers may be irrelevant to my decision, and I will plant the optimal amount anyway. (Figure 2.) The literature on economic capabilities deriving from Penrose (1959), Richardson (1972), and others suggests that this may in fact be a general phenomenon from which R&D is not exempt. Because R&D projects are quite varied, each requiring capabilities slightly different from the next, a particular organization's or individual's menu of R&D projects may be constrained by its capabilities for carrying out projects (the supply-of-R&D side) rather than driven by the potential quasi-rents to be derived from projects (the demand-for-R&D side).

From this point of view, then, we do not need to ask whether there are incentives to supply the "optimal" amount of R&D but rather whether the expected returns are sufficient to induce the entrepreneur to undertake a particular (lumpy) R&D project. (In terms of figure 2, the question becomes whether the expected returns are sufficient to induce a demand for R&D of $D_0$ rather than whether returns are enough to yield $D_1$.) And, we argue, there are good reasons to think that the typical answer is yes.

Spillovers may take three forms. First, they may result from increases in consumer surplus if buyers do not have to pay for the full benefit that they receive from an innovation embodied in a good or service that they have purchased. Second, they may result from competitors of the innovator acquiring the new knowledge at less than the full cost of R&D that the originator had to pay. Finally, spillovers may result from firms in other industries acquiring the knowledge at less than the full cost of R&D (Bureau of Industry Economics 1994).

Any assessment of the importance of spillovers requires an assumption concerning the investment behavior of firms. The task is made simple if one accepts, along with Grossman and Helpman (1991, pp. 517-8), that "profit maximizing and far-sighted entrepreneurs … invest in research and development in order to capture monopoly rents from innovative products." In this case, governments must clearly internalize fully or otherwise compensate for *all* spillovers. Romer (1990, p. S89), however, has contended that, "profit-maximizing agents make investments in the creation of new knowledge and … they earn a return on these investments by charging a price for the resulting goods that is greater than the marginal cost of producing the goods." The issues Romer raises are more problematical because, although he states that firms must be able to collect some economic profit, he also implies that some unspecified degree of spillover may be permissible. As most of the economy is organized quite differently from the picture of perfect competition in neoclassical economics, such profits may be expected to transpire frequently even in the absence of government intervention (although the desirability of intervention is one of Romer's main messages). In fact, both of these attempts to define the extent of returns required to generate R&D are improbable as a general guide to the behavior of investors in a world of rapid technological change.

To be sure, some risk-averse managers will insist on a guarantee of monopoly profits before proceeding; but they will be balanced to a degree (or even overbalanced) by less calculating investors who rank the excitement of engaging in a novel pursuit above the prospect of profits in their list of preferences.[9] In general, though, we should expect that considerations of perceived opportunity cost will prevail and that firms will invest in R&D as long as the return that they expect to receive is at least as great as the return from alternative investments. Since there is usually a high degree of uncertainty associated with R&D, the returns expected from successful projects will, of course, have to be great enough to compensate for the losses incurred on failures; but it is improbable that firms would in general insist on significantly higher expected profits from each individual item of R&D investment, or on such investment as a whole, than they expect to receive from other potential destinations for their funds.[10]

If this analysis is indeed accurate and perceived opportunity costs provide a good guide to the volume of investment in R&D, then the numerator (the amount earned) can only be judged sensibly in relation to the denominator (the delete amount invested). Furthermore, this standard should hold regardless of whether one is attempting a neoclassical or an evolutionary analysis. It is obvious that anticipated spillovers are important in the case of marginal investments which may be rendered submarginal if the investor expects that a share of the

---

[9]    See, for example, Freear *et al*. (1994) on the behavior of "angels" who help fund high-technology firms in their earliest stages.

[10]   Because the investment decision occurs *ex ante* but the profits, if any, are collected *ex post*, there is a high degree of subjectivity in investing. This raises the prospect of variations in behavior that extend beyond the presence or absence of animal spirits. The low level of R&D expenditure in many mature industries, for example, may stem from an expectation that nothing of importance remains to be discovered. As an expectation is not necessarily an accurate portrait of reality, super profits (or significant losses) may be the reward of firms in such industries that defy the conventional wisdom and invest in R&D when their competitors are more complacent.

return will be lost to others, but this does not mean that spillovers will, as a rule, deter investment in R&D. Such spillovers may be irrelevant in determining investment behavior, or, in many cases, they may even encourage investment.

Consider, for example, the third category of spillovers listed above: returns earned by firms in other industries (non-competitors with the innovator) who can acquire the new knowledge for less than the full cost invested in R&D by the innovator. As the free-riding firms are not taking business away from the innovator, there may be no change to either the numerator or the denominator of the innovator's rate of return.[11] For instance, it is entirely possible that, perhaps out of ignorance, an innovator does not realize, or does not feel it relevant, that producers in other industries might also benefit from an improvement intended to achieve a competitive advantage in the innovator's own industry. Thus the *ex ante* expected rate of return to R&D would not be adversely affected by adoptions elsewhere in the economy. Furthermore, these spillovers may lead to serendipities in the form of unanticipated licensing fees or other benefits that raise the rate of return above the expected level and therefore give greater incentives for *future* investment in research and development.

Two important points follow from this discussion. The first is that investment behavior in regard to R&D is necessarily subjective. The factors that the investor includes in calculating expected returns depend of the investor's own focus and expertise. Some firms may base their behavior entirely on returns expected from their own operations, while other firms

---

[11]   Paul David (1992, p. 242) has made much the same point in relation to developing nations obtaining innovations without paying. "[T]he important point remained that even though [developing countries] were not expected to pay, it would not much matter from an efficiency standpoint. The world's supply of such innovations would not be much diminished, if at all, by these countries' adoption of a free-riding policy."

may take a broader perspective. Both their decisions and their perceptions of the adequacy of the returns that they gain eventually will vary accordingly. Second, as the denominator is sunk by the time that the numerator is generated, any extra amount collected can only *raise the rate of return*. Although rent seekers may try to give the impression that spillovers *per se* discourage investment in research and development, the important consideration is not the amount of return that others get but the amount that the investor receives. As long there is good reason to expect that the amount received by the investor will satisfy opportunity-cost criteria, the benefits that flow to others are irrelevant to decisions to invest in R&D.

## *The strategic use of spillovers.*

Not only are monopoly returns, or even economic profits, not a *sine qua non* to induce R&D expenditures, but spillovers may even be an important strategic incentive for investment in R&D in a disequilibrium situation arising from the introduction of an innovation. This is true of the first type of spillovers, which accrue to customers in the form of increased consumer surplus. In the neoclassical model, such consumer surplus is a free gift to buyers and, in it-self, creates no incentive for additional R&D. However, in a dynamic world — in which both firms and buyers are still learning about a technology — a "gift" of consumer surplus can pay dividends, and can induce further R&D, to the extent that it furthers the wide diffusion of buyer capabilities on which consumption is based.[12] Development activities on personal computers and their components may have generated some consumer surplus that was "free" in the short run. But, in the longer run (that is, with social learning about the relative merits of new products), consumers may tend to learn more about, and thus to take greater advantage,

---

[12] On the idea of consumer capabilities, see (Langlois and Cosgel 1997).

of those products, like personal computers, offering higher potential consumer surplus.[13] This will have the effect of shifting out demand, which may well increase the return to the original development activities

If firms are to gain economic profits or even monopoly rents from investments in research and development, as the NGT suggests they should, then they must absorb at least some proportion of the consumer surplus that would result in a more competitive situation. But this is not a zero-sum game. As will be shown, reducing consumer surplus does not necessarily increase the returns to the innovator, which are based on what the innovator
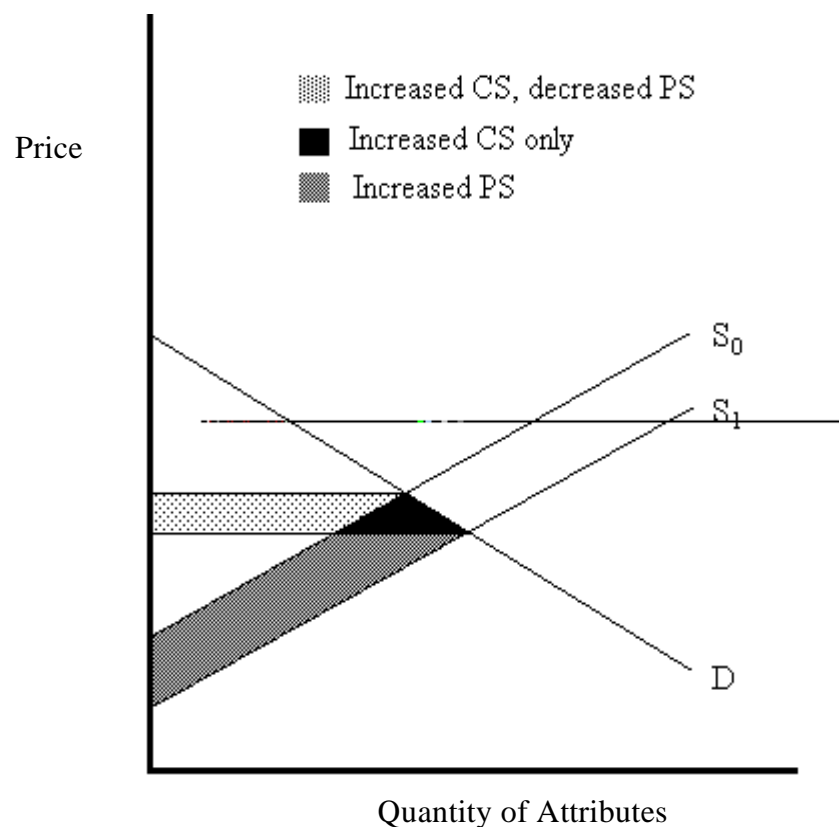


**Figure 3**
Effects on consumer and producer surplus of a shift in the supply curve resulting from an innovation, showing net increases in both types of surplus.

---

[13]    In neoclassical demand theory, the fact of higher consumer surplus in one market does not induce consumers to switch in the long run to that product. (This is unlike the case of firms in neoclassical production theory, which enter a market in the long run when producer surplus is higher there than elsewhere.) However, if we think of consumption as an activity actually analogous to production (Lancaster 1971; Stigler and Becker 1977), or perhaps even as an activity requiring capabilities and learning (Langlois and Cosgel 1997), then consumers may well respond in the long run to the potential for consumer surplus in a way more analogous to the way in which producers respond to producer surplus.

collects and not on the benefits that others gain.

As the investment costs in R&D are sunk by the time production begins, innovators should be willing to produce up to the point at which they maximize the extra returns that flow from an innovation in order to maximize their rates of return. However, the maximum rate of return does not necessarily occur at the level of sales at which consumer surplus is minimized. In fact, producer surplus is a better guide to the returns received by an investor, and consumer surplus and producer surplus may increase simultaneously. Assume, not unreasonably, that the demand curve for a first-mover producing an innovative product slopes downward, although if there are substitute products available the slope will be less steep than that of the industry demand curve. As in Figure 3, let the effects of an innovation be represented by an outward shift of the firm's supply curve. Following Lancaster (1971), we may say that consumers are purchasing collections of attributes. The term "Quantity of Attributes" on the horizontal axis refers, therefore, to the product of the number of units of a particular good or service which are being demanded or supplied, and to the number of attributes embodied in each unit. In the case of a process innovation, we should expect the outward shift of the supply curve to represent an increase in the number of units with unchanged attributes that a firm is willing to supply at a given price. When there is a product innovation, however, the rightward shift would indicate an increase in the quantity of desired attributes per unit.[14]

---

[14] Of course, an outward shift in the supply curve may also represent a simultaneous increase in both types of attributes, or an increase in one type and a decrease in the other, as long as the net effect is a positive change in the utility of the attributes supplied at any given price.

In this case, as long as the price elasticity of demand is greater than unity, both consumer and net producer surplus increase. As producer surplus is the sum of economic profit (B) plus fixed costs (F),

$$R = B + F,$$

any change that increases the producer surplus of a firm by more than its fixed costs yields an increase in economic profit, *i.e.,*

$$\mathbf{D}B > 0 \text{ whenever } \mathbf{D}R > \mathbf{D}F.$$

In a static model, a firm will adopt a pricing strategy that maximizes its return on an investment in research and development by allowing consumer surplus from an innovative product to rise to the point at which producer surplus from the product is maximized, because this also maximizes the rate of return from the investment in R&D. Therefore, in a static analysis, a firm will be willing to invest in R&D whenever

$$E(\mathbf{D}R) > E(\mathbf{D}F).$$

In a dynamic model, the strategic options open to a firm are broader. Given that returns would not be generated simultaneously or immediately, a firm needs to estimate the present value of its returns before investing in R&D. Therefore, a firm will undertake investment in R&D whenever the expected present value of the increase in producer surplus is greater than the expected present value of the associated increase in fixed costs:

$$E[PV(\mathbf{D}R)] > E[PV(\mathbf{D}F)].$$

In order to maximize the present value of the return flowing from an investment in R&D, a firm may be willing, for example, to engage in forward pricing by accepting an initial increase in consumer surplus that is so great that it erodes producer surplus. In this way, the firm could gain greater market share at the onset which might yield subsequent benefits by allowing it to take advantage of economies of scale in production, or raise prices (and producer surplus) by a larger margin in the future because of greater market power, etc.[15]

The point is perhaps even clearer if we consider spillovers to other firms and to other industries. Clearly, the phenomenon of "clones" in the personal computer industry reflects a spilling over of knowledge from (relative) first movers like Apple and IBM to cloners.[16] But the cloning in fact spurred the development of the computer industry in a way that has vastly increased the private returns to what little R&D was originally undertaken, even by the more research-intensive contributors like Intel and Motorola. The reason, of course, is that the microcomputer is an instance of a network industry (Katz and Shapiro 1985; Farrell and Saloner 1985). In such industries, one person's demand is a function of how many others have already purchased the good. Thus, increases in supply shift the demand curve rightward; and, by increasing supply, knowledge spillovers help shift demand and increase the size of the network.

As Liebowitz and Margolis (1994) have cautioned, however, such network effects are often in the nature of pecuniary rather than technological externalities. If, by taking advan-

---

[15]  For a fuller list of the first-mover advantages that might be strategically tapped in this way, see Lieberman and Montgomery (1988).

[16]  Or from Xerox's Palo Alto Research Center, which originated many of the features now found on personal computers, to Apple and Microsoft.

tage of internal economies of scale available to me (including capitalized R&D), I lower the cost of (what to you are) your inputs, I have merely passed on my internal economies through the price system. True technological externalities occur only in networks like telephone systems, in which there are non-price benefits accruing to users from network size (in this case the enhanced probability of contacting the party you want). Personal computers may also fall into this category to the extent that there are non-price benefits in software availability and support proportional to the size of the installed base.

In Marshall's concept of external economies, however, the line between pecuniary and technological externalities was not as sharp as subsequent writers have drawn it.[17] Marshall recognized that even seemingly pecuniary network externalities could have non-price effects: if my internal economies are transmitted to users as lower prices, the increased demand by those users will lead to additional learning and the creation of complementary capabilities among consumers that effectively shift the demand curve rightward. This may in turn lead to further increases in quantity supplied, which partake of further economies of scale, and which lead to more learning, etc., in the kind of self-reinforcing process popularized lately by Brian Arthur (1990). Spillovers obviously play an important — and positive — role in this Marshallian account of economic growth. Knowledge spilled in one part of a linked industrial system increases output and spurs additional learning, which in turn spills over and increases output, etc.

---

[17] For an excellent discussion of Marshall's concept, see Loasby (1989, 1990).

This is a process somewhat in the spirit of the New Growth Theory, in the sense that knowledge generation creates increasing returns and economic growth; but, because it sees social learning as essentially a network phenomenon, the Marshallian version generates a much different verdict on the benefits of spillovers. In the neoclassical process, spillovers appear undesirable because most of the burden of social learning is placed on the act of private R&D. In the Marshallian process, spillovers appear quite beneficial because most of social learning *takes place in the spillover process itself*.

There is considerable evidence in industrial history to support the idea that spillovers can key the process of social learning that leads to economic growth. There is also evidence, moreover, that innovating firms often recognize this fact and use it to their advantage.[18] When AT&T invented the transistor after World War II, it consciously chose to license the device widely and inexpensively, and held seminars to tutor licensees in its fabrication. Although there may have been a number of motives for this, including (before 1956) antitrust concerns, one of those motives was clearly to reap the benefits of technological development that would come from widespread diffusion of the technology. AT&T was still primarily concerned with the usefulness of the transistor to its own line of business, telephone switching. Although Bell had begun using the device in telephone devices and circuits, it was still an extremely immature technology. AT&T believed that if it allowed access to the transistor, telephony would reap the benefits of spillovers from the external capabilities of the market to an extent that would far outweigh the foregone revenues of proprietary development

---

[18]   Economides (1996) provides an equilibrium analysis that demonstrates in theory that firms can benefit from sharing innovations with competitors in a network context. Innovative situations, however, generally demonstrate non-equilibrium properties, which in our view makes the case even stronger.

(McHugh 1949; Bello 1953; Braun and Macdonald 1978, p. 54; Levin 1981, p. 76-77). An AT&T vice president put it this way. "We realized that if this thing [the transistor] was as big as we thought, we couldn't keep it to ourselves and we couldn't make all the technical contributions. It was to our interest to spread it around. If you cast your bread on the water, sometimes it comes back angel food cake."

Now, one might argue that AT&T was moved to such largess in this case only because it was a (domestic) monopolist in telephony, and could therefore appropriate a significant part of the external benefits. There is evidence, however, that even firms in extremely competitive industries recognize and pursue the benefits of spillovers. Columbia deliberately encouraged other firms to produce long-playing records, while Microsoft and Apple may be positive and negative lessons from the computer industry (Langlois and Robertson 1992). And Garud and Kumaraswamy (1993) have documented the conscious and successful policy at Sun of pursuing a high-spillover policy in the workstation market.

### *Institutions, spillovers, and economic growth.*

As we saw, one pillar of the conventional theory of technology spillovers is the assumption that knowledge is "nonexcludable," that is, that it is technically difficult or impossible to exclude others from gaining the benefit of knowledge once produced and therefore difficult or impossible to get them to pay for their use of the knowledge. In the Pigovian approach, this condition is taken to be epistemological: the assumption of nonexcludability is justified on the grounds of the inherent slipperiness of knowledge. In fact, however, the excludability of knowledge is an empirical question and an institutional one. In a sense, indeed, it was pre-

cisely Coase's (1960) contribution to make clear this point.  What are and are not externalities depends on actually available structures of property rights and transaction costs.

Coase and his followers (Coase 1974; Cheung 1973) set about demonstrating that some of the canonical cases of nonexcludable public goods in economic theory were in fact no such thing:  both lighthouses and beekeepers found it perfectly easy to charge people for their services.  More generally, these writers argued that the diagnosis of "market failure" could not be rendered in the abstract but only against the backdrop of actually existing institutional alternatives.  This Coasean approach is what Demsetz (1969) — in a famous and well-aimed attack on Arrow (1962) — called "comparative institutional analysis."  Most criticisms of the conventional (Pigovian) theory of technology spillovers have followed Demsetz's lead.  Indeed, one recent critique of the policy conclusions of the New Growth Theory (Weder and Grubel 1993) argued precisely that excludability is far less of a problem in R&D than the NGT imagines, as there are in fact a plethora of institutions, including trade associations and industrial districts, that serve to internalize spillovers.[19]

We agree with this line of reasoning, but wish to go even further.  Institutions matter to a theory of economic growth for reasons that go well beyond their role in internalizing knowledge spillovers.  As Douglass North (1981, 1990) has persistently argued, theories of economic growth that rely on movements in aggregate stocks like capital, labor, or even knowledge "beg all of the interesting questions" (North 1981, p. 5). Such theories either assume institutions to be unimportant or, as is usually the case, tacitly assume that institutional ques-

---

[19]    On this point see also Eliasson (1996), who also makes a number of other arguments similar in sprit to our own.

tions reduce to questions of relative price: institutions are just a matter of "getting the prices right."

Much could be said on this point. But we would like to focus on the role of institutions in the creation of the knowledge that — all are agreed — is conducive to economic growth. At one level, institutions are important for growth because they create incentives; and, if those incentives channel behavior into the creation of new wealth rather than into the redistribution of existing wealth, they are favorable to growth. This is an idea dating back at least to Smith. At another level, however, institutions are important to growth because they are themselves embodiments of knowledge. Thus one might "produce" knowledge not only about how to produce goods and services more effectively (holding institutional structure constant) but also about how to change institutions themselves in a manner more conducive to economic growth. Obviously, institutional change at the most abstract level — that is, change in basic institutions of government, law, and culture — is an extremely complex process of "coevolution" that is not well described as a progressive search for more efficient forms (North 1981). However, if we consider institutions at a more concrete level, including the kinds of organizational forms and contractual structures that emerge within an abstract system of law and policy, then we might reasonably think that more-or-less cumulative productive knowledge might well be embodied in such institutions in much the same way that it can be embodied in technological innovations.

North (1990) makes a sharp distinction between "institutions," which are the abstract background rules of society, and "organizations," which are the actors upon that background. Langlois and Robertson (1995), however, consider an intermediate set of organizational forms

and structures they call "business institutions."[20]  The term is intended to refer both to the pattern of capabilities within an economy and to the systems according to which those capabilities are organized, with the latter running the gamut from decentralized markets (broadly understood) to highly integrated firms.  Rosenberg and Birdzell (1986, p. 115) have noted the importance for the economic growth of Western Europe of a number of innovations in business institutions, including bills of exchange and banking; insurance; economic association without kinship; double-entry bookkeeping; and the mercantilist partnership.  One could easily name some more recent ones, including those managerial innovations chronicled by Alfred Chandler (1977).

The productive knowledge embodied in such innovations has very much the properties hoped for in the New Growth Theory: it is template knowledge that is widely applicable and reusable without replication.[21]  But such knowledge is not typically, if at all, "produced" in the way R&D is supposed to produce knowledge; it often enters an economy in something much more like the serendipitous way Schumpeter (1934) described[22] — which means that its contribution to growth is thus exogenous, not endogenous.[23]  "Ironically," writes Howard Pack (1994, p. 60), "the new generation of growth models relies on externalities and R&D at

---

[20]  On the blurriness of the line between institutions and organizations, see also Langlois (1995).

[21]  That is to say, one can double output without having to double the input of this knowledge.  To form another Renaissance business, one may have had to have more labor and capital, but one didn't need to reinvent double-entry bookkeeping.  Notice that an innovation of this sort is a true Marshallian external economy.  It lowers everyone's costs of doing business, but is not the result of the internal economies of any firm.

[22]  That Schumpeter did not see innovation as endogenous is, of course, another reason why the "Schumpeterian" label ill fits the New Growth Theory.

[23]  On the other hand, Alfred Chandler (1962) has shown that the development of organizational forms may also be endogenous and proceed by a process analogous to industrial development projects.

precisely the time a sense is emerging that one of the important factors determining interme-diate and perhaps long-term productivity growth is organizational."

But it is not merely that institutions represent a kind of knowledge in addition to strictly technological knowledge. It is that the structure of background and concrete business institu-tions in a society largely determines the way in which the society generates and uses more purely technological knowledge. This is in part what Nelson (1993) and Lundvall (1992) mean when they speak of national systems of innovation.

More recently, Nelson and Pack (1996) have made a distinction between "accumulationist" and "assimilationist" theories of growth. Accumulationists are those who stress the role of "moving along the production function," that is, of accumulating ever greater stocks of inputs in order to generate higher output. The institutional assumptions in such theories are tacit and minimalist: so long as an economy gets relative prices right, accumula-tion should translate into higher output. By contrast, assimilationists see the role of institu-tions, including business institutions and capabilities, as more subtle, more complex, and more important. A country cannot simply accumulate inputs, even knowledge inputs as those are conceptualized in the New Growth Theory, without the proper set of background and business institutions to permit learning — to permit the assimilation of the knowledge others have spilled over onto them. Countries as well as firms need absorptive capacity.

**Figure 4**

New Growth Theorists are still largely accumulationists. (See figure 4.) Like assimilationists, they stress the importance of "knowledge capital," even if understood in a homogenized sort of way; and, by pointing to the importance of the generation of new knowledge for economic growth, they do arguably go beyond the "old" theory. But New Growth Theorists do not see growth as an evolutionary process. And they do not see the crucial knowledge issue in economic growth as essentially a qualitative and institutional one.

## References.

Arthur, W. B. (1990) "Positive Feedbacks in the Economy," *Scientific American* 262:2, February, pp. 92-99.

Arrow, K. J. (1962) "Economic Welfare and the Allocation of Resources to Invention," in R. R. Nelson, ed., *The Rate and Direction of Inventive Activity: Economic and Social Factors* (Princeton: Princeton University Press).

Arrow, K. J. (1974) *The Limits of Organization* (New York: Norton).

Bello, F. (1953) "The Year of the Transistor," *Fortune*, March.

Braun, E., and Macdonald, S. (1978) *Revolution in Miniature* (Cambridge: Cambridge University Press).

Bureau of Industry Economics (1994) *Beyond the Innovator: Spillovers from Australian R&D* (Canberra: Australian Government Printing Service).

Chandler, A. D., Jr. (1962) *Strategy and Structure: Chapters in the History of Industry Enterprise* (Cambridge, Ma: MIT Press).

Chandler, A. D., Jr. (1977) *The Visible Hand: the Managerial Revolution in American Business* (Cambridge, Ma.: the Belknap Press of Harvard University Press).

Cheung, S. N. S. "The Fable of the Bees: An Economic Investigation," *Journal of Law and Economics* 16:1, April, pages 11-33.

Coase, R. H. (1960) "The Problem of Social Cost," *Journal of Law and Economics* 1, October, pp. 3-44.

Coase, R. H. (1974) "The Lighthouse in Economics," *Journal of Law and Economics* 17:22, October pages 357-76.

Cohen, W. M., and Levinthal, D. A.. (1990) "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly* 35, pp. 128-152.

Cohendet, P., Héraud, J.-A., and Zuscovitch, E. (1993) "Technological Learning, Economic Networks and Innovation Appropriability," in Dominique Foray and Christopher Freeman, eds., *Technology and the Wealth of Nations: The Dynamics of Constructed Advantage* (London: Pinter), pp. 66-76.

David, Paul A. (1992) "Knowledge, Property, and the System Dynamics of Technological Change," *Proceedings of the World Bank Annual Conference on Development Economics 1992*, supplement to *The World Bank Economic Review* and *The World Bank Research Observer*, pp. 215-248.

Demsetz, H. (1969) "Information and Efficiency: Another Viewpoint," *Journal of Law and Economics* 12, pp. 1-22.

Economides, N. (1996) "Network Externalities, Complementarities, and Invitations to Enter," *European Journal of Political Economy,* forthcoming.

Eliasson, G. (1996) "Spillovers, Integrated Production, and the Theory of the Firm," *Journal of Evolutionary Economics* 6:2, pp. 125-140.

Farrell, J,. and Saloner, G. (1985) "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, Spring, pp. 70-83.

Field, A. J. (1979) "On the Explanation of Rules Using Rational Choice Models," *Journal of Economic Issues* 13:1, pp. 49-72.

Field, A. J. (1984) "Microeconomics, Norms, and Rationality," *Economic Development and Cultural Change* 32, pp. 683-711.

Freear, J., Sohl, J. E., and Wetzel, W, E., Jr. (1994) "Angels and Non-Angels: Are There Differences?" *Journal of Business Venturing* 9, pp. 109-123.

Garud, R., and Kumaraswamy, A. (1993) "Changing Competitive Dynamics in Network Industries: An Exploration of Sun Microsystems' Open Systems Strategy," *Strategic Management Journal* 14, pp. 351-369.

Grossman, G. M. and Helpman, E. (1991) "Trade, Knowledge Spillovers and Growth," *European Economic Review* 35:3, May, pp. 517-526.

Grossman, G. M. and Helpman, E. (1994) "Endogenous Innovation in the Theory of Growth," *Journal of Economic Perspectives* 8:1, Winter, pp. 23-44.

Katz, M. and Shapiro, C. (1985) "Network Externalities, Competition, and Compatibility," *American Economic Review* 75, pp. 424-440.

Lancaster, K. (1971) *Consumer Demand: A New Approach* (New York: Columbia University Press).

Langlois, R. N. (1983) "Systems Theory, Knowledge, and the Social Sciences," in Fritz Machlup and Úna Mansfield, eds., *The Study of Information: Interdisciplinary Messages* (New York: John Wiley), pp. 581-600.

Langlois, R. N. (1986) "Science, Technology, and Public Policy: Lessons from the Classicals," paper presented at the American Economic Association annual meeting, December 28, New Orleans.

Langlois, R. N. (1992) "External Economies and Economic Progress: The Case of the Microcomputer Industry," *Business History Review* 66:1, Spring, pp. 1-52.

Langlois, R. N. (1995) "Do Firms Plan?" *Constitutional Political Economy* 6**:**3, pp. 247-261.

Langlois, R. N., and Cosgel, M. M. (1997) "The Capabilities of the Consumer," in Maria Bianchi, ed., *The Active Consumer* (London: Routledge).

Langlois, R. N., and Robertson, P. L. (1992) "Networks and Innovation in a Modular System: Lessons from the Microcomputer and Stereo Component Industries," *Research Policy* 21:4, pp. 297-313.

Langlois, R. N., and Robertson, P. L. (1995) *Firms, Markets, and Economic Change: A Dynamic Theory of Business Institutions* (London: Routledge).

Levin, R. C. (1982) "The Semiconductor Industry," in Richard R. Nelson, ed., *Government and Technical Progress: A Cross-Industry Analysis* (New York: Pergamon Press).

Levin, R. C., Klevorick, A., Nelson, R. R., and Winter, S. G. (1987) "Appropriating the Returns from Industrial R&D," *Brookings Papers on Economic Activity* 3, pp. 783-831.

Lieberman, M. B. and Montgomery, D. B. (1988) "First-Mover Advantages," *Strategic Management Journal* 9, Supplement: 41-58.

Liebowitz, S. J., and Margolis, S. E. (1994) "Network Externalities: An Uncommon Tragedy," *Journal of Economic Perspectives* 8:2, Spring, pp. 133-150.

Loasby, B. J. (1989) "Knowledge and Organization: Marshall's Theory of Economic Progress and Coordination," in Loasby, *The Mind and Method of the Economist* (Aldershot: Edward Elgar).

Loasby, B. J. (1990) "Firms, Markets, and the Principle of Continuity," in J. K. Whitaker, ed., *Centenary Essays on Alfred Marshall* (Cambridge: Cambridge University Press).

Lundvall, B.-Å., ed. (1992) *National Systems of Innovation: Toward a Theory of Innovation and Interactive Learning* (New York: St. Martin's Press).

McHugh, K. S. (1949) "Bell System Patents and Patent Licensing," *Bell Telephone Magazine*, January, pp. 1-4.

Mansfield, E., Rapoport, J., Romeo, A., Villani, E., Wagner, S., and Husic, F. (1977) *The Production and Application of New Industrial Technology* (New York: Norton).

Nelson, R. R., ed. (1993) *National Innovation Systems: A Comparative Analysis* (New York: Oxford University Press).

Nelson, R. R., and Pack, H. (1996) "Firm Competences, Technological Catchup, and the Asian Miracle," paper presented at the International Joseph A. Schumpeter Society meeting, June, Stockholm.

Nelson, R. R., and Winter, S. G. (1982) *An Evolutionary Theory of Economic Change* (Cambridge, Ma.: The Belknap Press of Harvard University Press).

North, D. C. (1981) *Structure and Change in Economic History* (New York: Norton).

North, D. C. (1990) *Institutions, Institutional Change and Economic Performance* (New York: Cambridge University Press).

Pack, H. (1994) "Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings," *Journal of Economic Perspectives* 8:1, Winter, pp. 55-72.

Penrose, E. T. (1959) *The Theory of the Growth of the Firm* (Oxford: Basil Blackwell).

Pigou, A. C. (1932) *The Economics of Welfare* (London: Macmillan), fourth edition.

Polanyi, M. (1958) *Personal Knowledge* (Chicago: University of Chicago Press).

Richardson, G. B. (1960) *Information and Investment.* (Oxford: Oxford University Press).

Richardson, G. B. (1972) "The Organisation of Industry," *Economic Journal* 82, 883-96.

Romer, P. M. (1986) "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94:5, pp. 1002-1037.

Romer, P. M. (1990) "Endogenous Technological Change," *Journal of Political Economy* 98:5, pt. 2, October, pp. S71-S102.

Romer, P. M. (1993a) "Two Strategies for Economic Development:  Using Ideas and Producing Ideas," *Proceedings of the World Bank Annual Conference on Development Economics 1992* (Washington: International Bank for Reconstruction and Development), pp. 63-91.

Romer, P. M. (1993b) "Implementing a National Technology Strategy with Self-Organizing Industry Investment Boards," *Brookings Papers on Economic Activity*, pp. 345-399.

Romer, P. M. (1994) "The Origins of Endogenous Growth," *Journal of Economic Perspectives* 8:1, Winter, pp. 3-22.

Rosenberg, N., and Birdzell, L. E., Jr. (1986) *How the West Grew Rich: the Economic Transformation of the Industrial World* (New York: Basic Books).

Ryle, G. (1949) "Knowing How and Knowing That," in *idem*, *The Concept of Mind*  (London: Hutchison's Universal Library).

Schumpeter, J. A. (1934)  *The Theory of Economic Development* (Cambridge, Ma.: Harvard University Press).

Silver, M. (1984) *Enterprise and the Scope of the Firm* (London: Martin Robertson).

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

Stigler, G., and Becker, G. (1977)  "De Gustibus Non Est Disputandum," *American Economic Review* 67, pp. 76-90.

Stoneman, P. and Diederen, P. (1994) "Technology Diffusion and Public Policy," *Economic Journal* 104: pp. 918-930.

Swan, T. W. (1956) "Economic Growth and Capital Accumulation," *Economic Record* 32: pp. 334-61.

Teece, D. J., and Pisano, G. (1994) "The Dynamic Capabilities of Firms: an Introduction," *Industrial and Corporate Change* 3:3, pp. 537-556.

Tushman, M. and Anderson, P. (1986) "Technological Discontinuities and Organizational Environments," *Administrative Science Quarterly* 31, pp. 439-465.

Weder, R. and Grubel, H. G.(1993) "The New Growth Theory and Coasean Economics:  Institutions to Capture Externalities," *Weltwirtschaftliches Archiv*. 129:3, pp. 488-513.