Contents

1
Introduction: The Market Process and the Economics of QWERTY

Peter Lewin

Introduction and overview

Some recent claims that high-technology markets fail to operate efficiently have become very influential, not only among economists (theorists and economic historians) but also in the popular press and in public policy. This is a new type of "market failure" mostly associated with the goods (sometimes known as "information goods") that are related to many of the technological advances regarding the generation and use of information in one form or another. In economics, the literature is known generally by association with the technical concepts of "network-effects," "path dependence," and "lock-in." These concepts suggest that the outcomes we observe (for example, in the generation of products to record and play video images, process data, or simply type book chapters) may not be particularly efficient. In this view of the world, random events may "lock us in" to a path and, therefore, an outcome that is socially inferior to an alternative one that is, or was, available. This view of the world has come to be known generally as "the economics of QWERTY." This general field of inquiry in economics is associated with parallel developments in the fields of mathematics and statistics having to do with topics like "chaos" and "complexity."

In the essays reprinted here, Stan Liebowitz and Stephen Margolis have critically examined the various aspects of the economics of QWERTY (to be defined below) and its implications. With eloquence and relevance they call into question the historical accuracy of the standard account of the QWERTY case and of similar "myths" of lock-in. They contend that no plausible case of sub-optimal lock-in has ever been satisfactorily documented. While the conventional wisdom has

been that such inefficiency is widespread and much recent antitrust activity (including the recent Microsoft case) and legislative policy discussion is based on that assumption, Liebowitz and Margolis question the historical evidence, the theoretical basis, and the policy implications drawn from the economics of QWERTY.

Taken by themselves the papers published by Liebowitz and Margolis over the last decade (1990–2000) constitute a remarkably comprehensive and accessible account of this literature As such, they can be read with great profit by anyone wishing to understand what path dependence, lock-in, network effects, etc. are all about and how they relate (if at all) to the recent lawsuit involving Microsoft.

Some brief comments on context

The economics of QWERTY derives from the assertion that the inherited typewriter keyboard, with its layout of keys according to the QWERTY configuration, is an archetypical representation of unfortunate accidents of history leading to inefficient results. The QWERTY keyboard design is claimed to be inferior to more rational designs that should, in a more ideal world, have been adopted. This particular historical example has served as a widely quoted and accepted paradigm case for what is seen as a general phenomenon, namely, the lock-in to inferior standards.

This book's organization of articles reflects the historical progression of Liebowitz and Margolis' research. They began, in 1990, by examining the historical accuracy of the QWERTY story and, stimulated by what they found there, were led to examine other historical cases, like the Beta/VHS videocassette case (see Chapter 5) and the Macintosh/Windows case (see Chapter 6). In each instance they found the evidence for sub-optimal ("inefficient") outcomes wanting. They also re-examined the theoretical basis underlying the conventional wisdom and provided an accessible but rigorous understanding of the concepts involved. Liebowitz and Margolis show also that the concepts of "network externalities" "path dependence", and "lock-in" are ill-defined and inconsistently used and they provide a definitive taxonomic clarification.[1]

In the process of considering both history and theory, the relationship between the two comes into question. Liebowitz and Margolis raise important questions about the proper role of economic modeling. They show that almost all of the discussion in the literature is about alternative economic models and not about the real world. There are an infinite number of possible economic models, many of which

exhibit lock-in. The important question is which models are relevant to economic reality and, even more importantly, to economic policy?

Concerning policy, the literature on network-effects underlies much of the new antitrust policy initiatives that have been manifested in the ongoing case of the Justice Department against Microsoft. Liebowitz and Margolis have analyzed the Microsoft case and the implications for antitrust more generally. They find that these new initiatives make little sense if the objective is to benefit consumers, and are fraught with dangers to the competitive process and the dynamics of innovation (Chapters 9 and 10).

A brief history of the history

Chapter 2 of this volume is a reprint of the article that started it all (Liebowitz and Margolis 1990, provocatively titled "The Fable of the Keys"). It is a critical examination of assertions made by economist Paul David and others that lock-in is historically important. I therefore begin with a brief look at these assertions.

The origin of QWERTYnomics

Paul David (1985) tells us that the QWERTY story is a story of path-dependence. Because, "history matters," sometimes in an irrational way, the historical path that a particular technological development takes can be decisive in locking in an alternative that is, in some meaningful economic sense, inferior to another that is available. This was what happened with the adoption of the QWERTY keyboard. A rival design, by August Dvorak and W. L. Dealey was superior, but lost out to QWERTY. This was because of "*technical interrelatedness, economies of scale*, and *quasi-irreversibility of investment*. They constitute the basic ingredients of what may be called QWERTY-nomics" (David 1985, 334, italics in original).

This unfortunate outcome was not the result "of custom, conspiracy or state control." Rather it reflects the behavior of individuals "held fast in the grip of events long forgotten and shaped by circumstances in which neither they nor their interests figured." They were, to quote Tolstoy, held "in bondage to the whole course of previous history" (David 1985, 332). The whole thing was the result of an attempt to place the keys in such a configuration that would avoid the tendency for them to become jammed. When, with later technology, typewriters no longer used jammable keys, so that this was no longer a relevant consideration, it was "too late" to change. The installed base of soft-

ware in the form of typist-human-capital was too great a barrier in the way of introducing the more rational DSK.[2]

David explicitly links his work with the theoretical contributions of Brian Arthur (1983) and Michael Katz and Carl Shapiro (1983). This literature features situations in which "essentially random transient factors are most likely to exert great leverage" and a "particular system *could* triumph over rivals merely because the purchasers of the software (and or hardware) expected that it would do so" (Katz and Shapiro 1983, 335, italics added).

> From the viewpoint of the formal theory of stochastic processes, what we are looking at now is *equivalent* to a generalized "Polya urn scheme" ... [A]n urn containing balls of various colors is sampled with replacement, and every drawing of a ball of a specified color results in a second ball of the same being returned to the urn; the probabilities that balls of specified colors will be added are therefore increasing (linear) functions of the proportions in which the respective colors are represented within the urn ... [W]hen a generalized form of such a process (characterized by unbounded increasing returns) is extended indefinitely, the proportional share of one of the colors will, with probability one, converge to unity. (David 1985, 335, italics added)

Which of the colors (or rival typewriter keyboards) will gain dominance, however, is "likely to be governed by 'historical accidents,' which is to say, by the particular sequencing choices made close to the beginning of the process" (David 1985, 335).

The "Fable of the Keys"

Two things are of note in David's account. One concerns the historical accuracy and completeness of the QWERTY story. The other concerns the characterization of it as *equivalent* to a particular stochastic dynamic process. Both of these motivate the essays in this book. I consider the second later.

Concerning the history, Liebowitz and Margolis devote most of their 1990 article (Chapter 2 in this volume) to a careful examination of the historical record. They provide considerable detail, with extensive citations. They point out that the evidence from the many typewriter experiments is ambiguous at best and plausibly tainted by serious conflict of interest and methodological shortcomings. They document the many typewriter competitions that occurred with mixed results

and the rivalrous competition among typewriter producers. In all, the assertion that QWERTY is an inferior standard cannot be sustained.

The case of the typewriter keyboard would appear to be especially suited to an assessment of "efficiency." This is because "what counts" for consumers of typewriter services can be boiled down mainly to two readily measurable dimensions – speed and accuracy in producing text. (Other dimensions, for example, the durability of the typewriter, can be standardized easily for comparison.) So, if tests in these dimensions produced results that clearly contradicted the "market's choice," this would, at the very least, give us pause. One would have to wonder why obvious cost savings had been passed up. Other cases are generally not so readily reducible to clearly measurable dimensions. And when many dimensions are involved, it is hard to know what the relative importance of each is for the consumer. (For example, in videocassettes, consumers care about picture quality, playing time, cassette size, and product durability in ways that are not immediately apparent without resort to observation of their market behavior.) In this respect, the typewriter case is, indeed, a sort of paradigm case.

An early work looking at typing speed was conducted by Dvorak and some coauthors. The detailed examination by Liebowitz and Margolis reveals that, in general, the Dvorak book lacks both sound experimental method and objectivity. A Navy study that held a prominent place in David's telling of the history also fared poorly under scrutiny. Liebowitz and Margolis reveal various deficiencies in the experimental methodology, including a serious truncation of the testing period, but most tantalizing is the objectivity and authorship of the study, which I will not give away here. Perhaps more important was the discovery by Liebowitz and Margolis of a 1956 General Services Administration study by Earle Strong that had been neglected by almost all writers on this subject even though it had considerable national prominence during its undertaking. Liebowitz and Margolis conclude: "Strong conducted what appears to be a carefully controlled experiment designed to examine the costs and benefits of switching to Dvorak. He concluded that retraining typists on Dvorak had no advantages over retraining on QWERTY" and "would never be able to amortize its costs" (see below 37; unless otherwise noted, page numbers will refer to this volume). Liebowitz and Margolis do not consider Strong's study to be without faults, but contend that it should be taken seriously instead of being ignored. Even a current proponent of the Dvorak, like Yamada (1980, 1983), "as much as admits that experimental findings reported by Dvorak and his supporters cannot be assigned much credibility" (38).

Liebowitz and Margolis also consider evidence from the ergonomics literature, which is more current and arguably more "scientific." "The consistent finding in the ergonomic studies is that the results imply no clear advantage for Dvorak" (40). In fact, these studies suggest that there is a strong possibility that "the limitations of typing speed ... [may] have something to do with a mental or, at least, neurological skill and fairly little to do with limitations on the speeds at which fingers can complete their required motions" (41).

Competitions between expert typists provide another type of (limited) evidence. "[T]yping contests and demonstrations of speed were fairly common" at one time involving "many different machines, with various manufacturers claiming to hold the speed record" (43). In the 1880s Remington's champion Frank McGurrin won a number of victories for the QWERTY keyboard. There were other types of machines besides the Dvorak but the evidence is complicated by the lack of standardization of the abilities and training of the various contenders. Touch typing was not common. Suffice it to say that there is absolutely no presumption indicated that QWERTY was an intrinsically inferior design.

The final, and perhaps the most important, type of evidence pertaining to conclusions about the QWERTY standard relates to the details of the market process by which QWERTY emerged. As Liebowitz and Margolis tell us it "was not invented from whole cloth." Quoting Yamada (1983, 177): "Examination of these materials reveals that almost all ideas incorporated into Sholes' [QWERTY] machines, if not all, were at one time or another already used by his predecessors" (41). The emergence of QWERTY was in fact the result of a fairly long and complex rivalrous process between numerous competitors. It is very important to be clear about *exactly* what it is that Liebowitz and Margolis assert in Chapter 2 (the 1990 article) and what they are not asserting. They explicitly state that they are *not* asserting "that QWERTY is proven to be the best imaginable keyboard." Neither are they claiming "that Dvorak is proven to be inferior to QWERTY." Rather their claim is simply "*that there is no scientifically acceptable evidence that Dvorak offers any real advantage over QWERTY*" (Chapter 2, n. 20, my italics).

The theory

In "The Fable of the Keys" Liebowitz and Margolis introduce the subject with a critical survey of the relevant theory of lock-in and

discuss it again in the conclusion, and it forms a large part of the other essays in this book. It is the subject also of a large and growing literature in the journals and advanced texts.

The basics

The relevant theory can be broadly characterized as the theory of *network-effects*. (Liebowitz and Margolis introduced the term "network-effects" to substitute for the formerly more common term network *externalities*, to account for the possibility, indeed the likelihood, that these effects are often internalized.) Network-effects, relating to the consumption of a particular good or service, occur whenever the benefits of consuming that good or service depend positively on the number of individuals who do so. So an additional consumer adds benefits to the consumption of other participants. This phenomenon is not new and is extremely common. Indeed the social institution of "the market" itself is a network. The benefits to all participants often, as Adam Smith realized, depend on its extent. Languages are networks. The value of learning a particular language often depends on how many speakers there already are. In fact, network-effects occur whenever benefits are related positively to the interaction of individuals within the network. Others examples are telephone networks, local area computer networks, clubs, trade associations, and of course the internet.[3] Network-effects are an example of economies of scale (increasing returns to scale) on the demand side as distinct from the more traditional economies of scale in production, with which they sometimes are, but should not be, confused.

Though common, network-effects are more important for some types of goods than others. (I shall use the word "good" to refer generically to a good or service, any "economic good.") They have been given prominence recently because of the proliferation of so called "knowledge-goods," though, as Liebowitz and Margolis point out knowledge is a dimension of every good. The connection between knowledge-goods and network-effects, however, has been related to the fact that the usefulness of any knowledge possessed, often depends on how many others have similar knowledge (demand-side economies of scale), *and* (not always correctly) to the fact that knowledge consumption is non-rivalrous and that it can often be duplicated without cost (or almost without cost). One person's use does not preclude another's (implying supply-side economies).[4] I discuss this further in a moment.

There is, therefore, a strong connection between networks and *standards*. A standard is a "shared way of doing things, of interacting."

Standards serve to coordinate individual activity by reducing costs of interacting. A common language is a prime example. Common software would be another. Obviously the relative benefits of a particular common standard are related to the presence or absence of devices for cheaply converting from one standard into another – analogous to the presence of a competent language interpreter. Standards may be fixed or flexible to some greater or lesser degree. Many standards, like languages, legal systems, operating systems, etc. evolve over time. Their benefits are a complex function of the degree of stability and flexibility that they exhibit.[5]

Liebowitz and Margolis point out that networks likewise come in many shapes and sizes and vary along a few dimensions. First, networks may be literal or notional. An example of a literal network is a telephone exchange. An example of a notional network is the network of Yale Law School graduates (55). Second, networks may be owned or unowned. This may be crucial in assessing the economic properties of the network. For example, an owned network does not exhibit any "externality problem," even though some of the benefits of consumption of the good involved are "external" to the individual. Though each individual fails to take account of the benefit that he/she confers on others by being in the network, the owner of the network has an incentive to do so and will charge for "membership" accordingly. In contrast an unowned network presents properties that are more interesting and challenging from the standpoint of static allocational efficiency.

As mentioned above, network-effects are economies of scale in demand. As such their existence is an "empirical" matter. Actual networks may exhibit economies or diseconomies of scale. The same is true of supply-side (or production) economies. "The currently popular association of new technology with increasing returns may well be faulty, at least for some technologies" (Liebowitz and Margolis, 1999, 81). In particular, as suggested above, network-effects should not be confused with the decreasing costs of production that occur with time as a result of product-specific learning and general technological advance. Such cost declines have characterized many new industries in the past (like refrigeration, automobiles, etc.) and have been attributed to many of the new "information age" industries of today. Decreasing returns to scale for the latter should also not be simply presumed. For example, it is sometimes argued that software production exhibits a high degree of increasing returns in production. Once the product has been developed, and once production is in place, the marginal cost of

producing an extra copy is negligible so that it is possible to "spread" the fixed setup and development costs over a larger and larger volume of production forever reducing average costs. This, however, is only part of the story. Typically, increases in software consumption eventually imply increases in costs from other sources. As Liebowitz and Margolis illustrate:

> Assume … that there is one technical-support specialist for each 25,000 users [of] … Windows 95. If the hiring of additional technical-support personnel tended to bid up their wages, this diseconomy alone could overwhelm the decreasing average fixed cost. Suppose, for example, that hiring an additional technical-support specialist (for 25,000 additional users) increased the wages of technical-support specialists by $22 a year, or an hourly wage increase of a penny. This small change in wages would be sufficient to make overall average costs increase, not decrease with output. (1999, 81–82, endnote omitted)

Generally, knowledge-goods, like software, are produced and consumed together with a complex of other goods (sales, marketing, public relations, management, distribution, etc.) that may not be subject to increasing returns, and increases in software production may thus be associated in increasing costs from a variety of sources. Bottom line, "without investigation, it is unreasonable to accept that the law of diminishing returns somehow takes a vacation in new-technology industries" (90). Some of what is observed as economies of scale is no doubt explained instead by phenomenal improvements in the technology of production as a result of successful innovative activity, something that is much more difficult to characterize and analyze.

It is undeniably true that the production of software and similar contemporary goods exhibits a particular structure that is worthy of note, namely, instant scalability (a term coined by Liebowitz and Margolis) – the ability to expand production with little or no time lag (213). Replication of these goods is relatively easy. This may be important in considering firm and industry structure and the nature and types of competition one is likely to observe.

The question of efficiency

All this is interesting and relevant to an understanding of the modern economic landscape with its bewildering variety of new "information age" products. But its relevance has been substantially enhanced by

recent discussions about economic efficiency and related policy implications. These discussions take the form of abstract theoretical speculations about the efficiency properties of various processes usually (although not necessarily always) associated with network-effects – processes that exhibit path-dependent lock-in. Crucial to an assessment of these discussions is clarity on the concepts of efficiency, equilibrium, path-dependence, and lock-in. Some brief critical remarks follow, however this is not intended as a comprehensive treatment of these terms.

Efficiency

Economists have searched long and hard for a concept of efficiency that is "objective" or value free. Economic outcomes consist of an array of goods and services and states-of-being of different individuals that are in themselves incommensurate. In order to pronounce one outcome more or less efficient than another one has to have a way of overcoming this "apples and oranges" problem. Commonly one resorts to attempting to appeal to the valuations placed on the outcomes by the affected individuals themselves. One appeals, that is, to individual preferences in deriving efficiency criteria. This obviously involves the decision that individual preferences ought to be what counts when deciding efficiency issues. In itself, however, this is merely a *definition* of efficiency. No value judgment is involved (beyond that of supporting a definition) unless one says something like, "efficiency is good" or "a more efficient outcome is a *preferred* outcome." When we do take this step, as we often do in policy discussions, what we are saying is that we believe that individual preferences ought to count in deciding what economic outcomes are preferable. This will perhaps strike readers as eminently reasonable. If what is "efficient" is defined as what the "people prefer," how could we not be for it? Is it not the quintessence of "economic democracy"?

As is well known, however, there are numerous practical difficulties in deciding what "people prefer" and, indeed, what this means. When changes are contemplated in which all of the individuals who are involved clearly gain from the change (that is, can be confidently said to prefer that the change be made), then there is little ambiguity and we have the well known Pareto improvement. (So, for example, in the typewriter case discussed above, one may be able to argue that, other things constant, a keyboard layout that is able to unambiguously deliver a faster typing speed would definitely be preferred by *everyone* concerned.) The most common difficulty comes from situations in

which some individuals gain and others lose. In such "mixed" situations, we have to resort to so-called compensation tests – that is, somehow judging whether the gains outweigh the losses. If we take this leap, we are, in effect, saying that, in deciding matters of efficiency, the distribution of gains between individuals is not relevant. Of course, a standard defense is that it might be relevant, but that it is a separate issue – we ought to make the pie as large as possible before we consider how it ought to be divided up. This involves a new additional value judgment and clearly is a much less plausible and easily defensible position than one that simply says "people prefer the change." To be sure, it is still a kind of economic democracy – it says something like "more people prefer it" or "the intensity of the preferences for outweigh the intensity of the preferences against." (Again, in the keyboard case, we would not consider it reasonable for holdouts of "inferior" keyboard layouts to be able to block the adoption of a "superior" layout. This judgment may be couched in Pareto efficiency terms by noting that the truly "superior" keyboard would be able to deliver cost savings in excess of the losses suffered by those wedded to an "inferior" one.)

This efficiency standard is widely accepted in economic discussions and has penetrated deeply into the policy and legal environments. The situation is complicated because the word "efficiency" has a very strong colloquial connotation and in economic policy discussions is often confused with what is meant by efficiency in the natural sciences where inputs and outputs are so much more easily identified and evaluated and no compensation criteria are necessary. It lends to economic policy discussions a spurious aura of being "scientific."

Economists encourage this impression in spite of the fact that they are well aware of the insurmountable obstacles to arriving at unambiguous decisions about which changes are efficient and which not. These involve the well-known impossibility of discerning individuals' preferences, having to use *hypothetical* market valuations instead, and of having to posit unknowable counterfactuals, often having to do with unknowable (even unimaginable) futures. The real drawback, however, of this traditional efficiency standard, I suggest, has not so much to do with its theoretical conception *per se*, as it does with the way in which it is traditionally used in economics, that is, in *the context of the static model of resource allocation*. It is in this context that it has encouraged the kind of attacks, in the name of efficiency, on the unfettered emergence and development of products and standards that we are witnessing in this literature.

In a static context, in which the value of all potential resource uses are known (either with certainty or "probabilistically"), in which technology is unchanging, the set of products is fixed, and there are no external effects of any kind or any elements of monopoly, it is well known that a "competitive solution" is also Pareto optimal and, therefore, efficient in the sense discussed above. This idealized situation of neoclassical "perfect competition" has unrealistically and unreasonably served as a standard of comparison for actual real-world situations. In particular, in the context of network-effects and standards, economists have thought it relevant and meaningful to argue that the presence of such effects suggests that private markets might provide an "inefficient" result. The works reprinted in this book suggest that these attacks are ill-informed and based on an unwarranted presumption of knowledge as well as an irrational concept of efficiency.

Equilibrium

Equilibrium is closely connected to the concept of efficiency. It is, however, even more widely and less self-consciously used. I have elsewhere dealt at some length with the different possible meanings of equilibrium and their implications (Lewin 1997, 1999, Chapter 2). Of particular interest to us here is the tendency for writers to make connections between theoretical processes that end in some sort of equilibrium with actually existing and observed processes and outcomes in the real world.

As with the concept of efficiency, the concept of equilibrium has migrated from the natural sciences, where it connotes some type of stasis, a stable configuration of variables. In economics one cannot understand equilibrium without reference to human behavior and, therefore, to human cognition. Following Hayek many theorists define equilibrium in terms of the plans and expectations that individuals have. A situation is said to be in equilibrium when people's plans are mutually consistent and realistic; that is to say, when people have plans that are mutually compatible and can be successfully carried out. In such a situation there are no surprises, no one has any reason to "change his/her mind." There is no change in an economically meaningful sense.

Such an equilibrium never exists as such, though aspects of individual plans must be consistent if we are to have life as we know it. In the realm of economic activity, however, and particularly in the area of the "new technology" industries there is no equilibrium to speak of. The whole process is driven by *differences* in opinion and perception

between rival producers and entrepreneurs. Where people have different expectations about the same situation, at most one of them can be right. The values they place on the resources at their disposal or which they trade, are not, in any meaningful sense, equilibrium values. They reflect only a "balance" of expectations about the possible uses of the resources. One cannot use such values meaningfully in any assessment of efficiency (in the sense discussed above).

Path-dependence

The concept of path-dependence is also not new and it also has links to the natural sciences. In economics it gives expression to the common sense idea that equilibrium values might depend on the path taken to get to equilibrium. This is most obvious already in the realization that "false trading" may imply income effects that affect the final prices and quantities of a set of products in a general equilibrium system. But it is much more general and, one suspects, ubiquitous. One should surely not be surprised to find that the equilibrium values of most economic systems are likely to be affected by events that lead up to the attainment of equilibrium, that is, the equilibrium values are not insensitive to the path taken to get to equilibrium. The fact that, as explained, we may never get to equilibrium speaks to the relevance of the whole discussion. May we assume that it is relevant to an assessment of which path is chosen even if equilibrium is never attained?

Lock-in

In the current discussion path-dependence gains added relevance because it is seen to attach to systems that exhibit network-effects. In particular, the fact that the benefits of being in the network depend, in part, on how many individuals already belong, suggests that, among competing networks, whichever gets started first may foreclose the development of the others simply by virtue of being there first, and not from any economic merit. This could then be seen as a "market failure," a failure of the market to unfailingly deliver the "best" standard. To be sure, lock-in may or may not be a problem. It is a problem only if one becomes locked-in to an inferior standard.

We are now in a position to consider some of the theoretical contributions in this field and the role of Liebowitz and Margolis in all of this.

Theory and efficiency

It is fair to say that the theoretical contributions in this field of economics are almost exclusively in the form of a series of exercises

designed to show how various types of sub-optimality can occur. A typical example is the set of articles collected in the *Journal of Industrial Economics* (March 1992). In the introductory overview, Richard Gilbert (the editor of the symposium articles) provides a sampling of the findings:

> The need for standardization is a constraint on product variety … The five papers that appear in this symposium address how successfully markets make this tradeoff[6] … Unfortunately, coincidence between the compatibility choice that is best for producers and the choice that is *best for economic performance* is not likely to hold in many real situations. (Gilbert 1992, 1, italics added)

For example (it is worth quoting at length to get the flavor of the assertions made):

> Katz and Shapiro (1992) showed that in a market with network externalities, the sponsors of technologies that differ in the size of the installed base may have different preferences for compatibility. For example, a dominant firm might prefer a technological design that is incompatible with other alternatives, thereby denying a new entrant the benefits of the installed base … [We may require] firms to produce products that conform to set standards … [but this] is a potentially costly requirement. Standards limit flexibility … [and] may constrain technological progress … [an alternative is] the development of products that allow consumers to use different technologies. Farrell and Saloner (1992) study the economics of (two way) "converters" …
>
> *Markets fail* to give consumers the right incentive for investment in imperfect converters …
>
> *Markets may fail* to give consumers the correct incentives to join one network instead of another. As a result, the success of a network is likely to be determined by consumers' expectations about which network will prevail and by choices made in the past …
>
> [C]*onverters can exacerbate the problem* [italics in original] of incorrect market incentives. Converters encourage consumers to choose the technology that best suites their private preferences. But consumers fail to take account of the benefits that their patronage would confer on an alternative network [and this] … does not generate as much total benefit for society as a whole. In Farrell and Saloner (1992) standardization provides greater total surplus unless

> consumers have sharply differentiated preference for different technologies. (Gilbert 1992, 1–3, italics added, except where noted)

And so on.

> [There is] a common theme. When production and consumption decisions are interrelated, either through network-effects or through complementary products, a competitive market does not necessarily send the right signals to firms and consumers for the delivery and purchase of goods and services. The *market fails* to reward consumers for the benefits they bring to a network or for the costs they impose by leaving the network. As a result, consumers who follow their own private interests may support more (incompatible) competing products than *would be desirable for total economic surplus*. The market would make the wrong tradeoff between product variety and network economies of scale, sometimes leading to too much variety and not enough technological compatibility. (Gilbert 1992, 7, italics added)

Note that either too much or too little variety my emerge from these models – they are models in which sub-optimal "lock-in" *and* "lock-out" may occur. Sub-optimal standards may be adopted too early (as with QWERTY) or the adoption of an optimal standard may be sub-optimally delayed – there may be too much or too little variety as opposed to uniformity. So uncertainty about the emergence of optimality applies on both sides of the issue and would necessarily also be attached to any remedial policy. It is noteworthy that nowhere in Gilbert's introduction, nor in the papers are we told how we could identify such "market failures" or what we could or should do about them. Almost as an afterthought Gilbert adds two caveats:

> Market forces might produce new institutions, *not addressed in these models* to deal with these inefficiencies … In addition there are a myriad of unknowns concerning the performance of new technologies, the ability of firms to deliver desired services, and consumer behavior, *all of which could influence the efficient structure* of supply in markets with network-effects and complementary products. (Gilbert 1992, 7, italics added)

Models and methods

The above is offered as typical of the kind of work that is being done in this area. It is this type of work that Liebowitz and Margolis criticize in

much of their work. In doing so, they raise crucial questions not only about this work, but also about the relationship between models, methods and reality in general.

Their attack proceeds on two broad fronts:

1. They point out that the building of models is not a substitute for empirical, historical research – the investigation of real case studies in order to decide which of the infinite types of models that can be constructed is likely to be relevant, and they provide a number of such studies.
2. They subject the prevailing theory to in-depth examination and demonstrate that a much wider range of results than those typically derived (and illustrated above) is not only possible but is likely. Applying what we know about the historical functioning of markets there are theoretical (in addition to empirical reasons) for doubting the existence of the so-called inefficiencies purported to be characteristic of these situations.

I consider these in turn, the first point more briefly.

"Economists often like to make their work appear to offer the same certainty as mathematics, presenting much of their work as 'proofs'" but "proofs in economic models are not proofs about the world. However much a model may prove, it can never prove that the assumptions it starts off with are the right ones" (Liebowitz and Margolis, 1999, 50). An examination of the historical case record suggests that "QWERTY worlds are awfully hard to find" (1999, 50). They look at a series of cases including the VHS versus Beta case for videocassettes, the WindowsDOS versus Macintosh case, and a whole series of cases in the software industry (to which we shall return). In each case they find no evidence for any kind of inefficient lock-in such as the dominance of early starters, lack of rivalrous activity, absence of technological innovation, or even the diminution of competitive activity.

Of course this raises the basic methodological issue, alluded to above, of how such investigations should proceed in the first place. This is an issue we shall have to examine at some length below.

Concerning the second point, Liebowitz and Margolis provide a series of crucial theoretical insights. Let us begin with the concept of path-dependence. Liebowitz and Margolis link this concept with its potential policy relevance. It is not hard to agree that "history matters." In this context, and in a number of their other comments, Liebowitz and Margolis concede the serious limitations of the model of

perfect competition and related constructs. Outcomes in the world depend in a variety of ways on history. So much is clear from observing that the capital stock consists of durable items that are often highly specific in form and function. "Mistakes" in investment endure over time. That is a form of simple path-dependence. Persistence in human affairs can be called *first-degree path dependence* (97). It has no obvious policy implications. This does not imply it should be ignored, quite the contrary an understanding of the present (without any necessary efficiency judgment) demands an examination of the past.

A slightly "stronger" form of path-dependence follows from the observation that many outcomes are (in whole or in part) the subject of regret. That is to say, a retrospective evaluation of an outcome may evoke the opinion that it is not the most preferred of the alternatives that *were* available. This is called *second-degree path-dependence* (97). It is also likely to be quite common, although not so common as first-degree path dependence. We should note that its identification relies on the (necessarily speculative) identification of counterfactual historical alternatives. Except insofar as an assessment of the past is informative for the future, it too has no obvious policy implications. History is history, what's done is done.

A much stronger form of path-dependence refers to outcomes that are judged to be inferior *and were known to be inferior when the past decisions that led up to them were taken*. This is referred to as *third-degree path dependence* (98). Liebowitz and Margolis point out that this type of path-dependence implies the notion of *remediability*, that is, because of some remediable impediment, like the costs of coordinating decision-makers, an outcome that is less preferred by everyone concerned nevertheless emerges. It is only third-degree path dependence that has any possible policy relevance. As they argue: "for an inefficiency to be economically relevant, there must be some better alternative that is *feasible* in light of the *information* the we have at the time that we are making a decision" (1999, 54, italics added).

This simple taxonomy has at least two very important implications:

1. It focuses attention on the key ingredient of any discussion of policy and inefficiency, namely, *the importance of knowledge and who has it*. Asserting that one path is economically inferior to another must presume some knowledge on behalf of the economic-theorist-cum-policy-maker. And if that knowledge is available to the policy-maker, it is presumably also available to the economic agents concerned. This suggests the second implication.

2. If a path-dependent inefficiency is known *ex ante* to exist, then, by definition of efficiency as discussed above, this implies the potential that a Pareto improvement can profitably be made. That is, there is scope for someone to profitably remedy the inefficiency, since the gains available outweigh the losses that would be produced by such a remedy.

What this discussion clearly does is to place these issues firmly within the realm of familiar Coasian transaction costs economics (Coase 1960). The Coase theorem suggests that, absent transaction costs and transaction-impeding wealth effects, apparent externalities and other inefficiencies would spontaneously be removed by the market process. The identification of any such inefficiencies thus must be seen to rely on these broadly construed "transaction costs." This is relevant to third-degree path-dependent inefficiencies. If such inefficiencies exist, that is, if everyone would prefer, for example, the adoption of a particular standard, but because they expect everyone else to adopt an inferior standard, themselves all choose the inferior standard, so that we become locked-in to a standard that is Pareto dominated then such an inefficiency may be said to exist because of the high costs of coordinating the activities of the numerous agents around the adoption of the "correct" standard, that is, because of high transaction costs. More generally, if such a lock-in exists because the agents are ignorant of the advantages of the alternative standard, this too may be characterized as a transaction cost problem, since, if it were possible to cost-effectively inform such agents of their errors and to facilitate a coordinated alternative, it would be done. In fact, from one perspective, all transactions costs are *information costs* (Dahlman 1979). In sum, "it must be possible for someone to have credible information that a better allocation [of resources] is available" (Liebowitz and Margolis, 1999, 56) for path-dependence to be policy relevant.

Liebowitz and Margolis are clearly skeptical of claims regarding the existence of policy remediable inefficiencies, or "market failures." This skepticism reflects their conviction that "mechanisms exist to obviate such failures" (1999, 68). Where networks can be owned, benefits will tend to be internalized. Where this is not possible, other mechanisms exist to internalize (in whole or in part) the benefits available, for example, through the provision of complementary goods. In addition, the literature on network-effects is misleading to the extent that it tends to emphasize the possibilities for the emergence of inefficiencies. Liebowitz and Margolis provide an extensive examination of the rele-

vant theory, one that reveals a much wider range of possibilities, even while staying within the static allocative framework.

Concerning the benefits of particular standards, Liebowitz and Margolis point out that often these benefits are tied less closely to the total number of other users of the standard and more closely to the number of users who actually interact. What concerns them are ways to achieve greater coordination and synchronization with this smaller subset. These "synchronization effects" (network-effects) may coexist with increasing, decreasing or constant returns to scale. Increasing the number of interactors in a network may add to the value of being in the network, though these additions may diminish. To obtain the total value to the consumer of any good, subject to network-effects, one must add this total "synchronization value" to the value that the consumer would place on the good as a sole user, the "autarky value." If the good has a positive supply price then its net value will be the difference between its total value to the consumer and its supply price. "[I]t is only when the net-value function slopes upward that choices between standards are fundamentally different in character from choices of other goods" (133–134). And if the supply curve slopes upward it is possible that the net-value function will slope downward.

Liebowitz and Margolis use this analysis to show that network-effects do not necessarily, or even probably, imply increasing returns to one standard that can be expected to dominate. Multiple formats are not uncommon and theoretical considerations are quite compatible with this. Niche formats are examples (Macintosh, linex, Betamax, all exist in smaller specialized markets). This result is reinforced if we assume that consumers have different tastes. Thus, "the mere existence of synchronization (network) effects is insufficient to establish the winner-take-all choice with respect to standards" (Liebowitz and Margolis, 1999, 99) and in fact the case for single dominance is quite weak.

This suggests that one may expect to see competitive strategies in which entrant firms try to specialize in their products that appeal strongly to particular groups of consumers, while incumbents, on the other hand, might try to create products that appeal to the widest possible audience attempting to foreclose opportunities for competitors (Liebowitz and Margolis, 1999, 106).

They show further that even when there are increasing returns and/or network-effects more than one standard may emerge.

Economic models are like recipes. You tend to get what you put into to them, but the way they taste depends crucially on how you mix the ingredients. Even models that stay within the static alloca-

tive equilibrium framework can vary a great deal in their implications. Liebowitz and Margolis show that winner-take-all dominance is in no way a necessary property of models that incorporate network-effects. The bias observed in the literature in that direction is a result of the presumptions of the modelers. There is no way to assess their relevance for economic policy without some way of deciding how closely they correspond to reality.

Policy

The static welfare framework is difficult to apply in the modern market environment. Market processes are truly dynamic in the sense that they take place in "real time," they are evolutionary processes that are driven by the diversity of perceptions and expectations that individuals have of the value of resources and of the process itself. They are processes that are "open ended" and are never in equilibrium. They are processes that are characterized by "radical uncertainty" and *novelty*, their outcomes are inherently unpredictable. For such processes the traditional types of efficiency assessments based on static models of resource allocation are completely meaningless. There is no way, for example, of applying the traditional utility calculus to choices among technologies that are not yet available, but that might emerge as a result of becoming "locked-in" to a particular dominant standard. Such a standard may appear to be inferior to others that are available, but it also may lead in the future to the discovery and application of complementary technologies that are vastly superior in a number of dimensions. How are we to assess the likelihood that policy action *itself* may be an "accident" that locks us into an inferior path? Insofar as new technologies are based on future knowledge and insofar as future knowledge cannot be available in the present, we cannot consider future technologies as part of today's choice-set. Neither can we include the new products, new methods of production, and new modes or organization that they bring with them. Technological change emerges from the complex interaction of individual visions (including those of the policy-makers) that are partial and incomplete and that are subject to the trial-and-error processes of the market.

Can one then say anything about efficiency? If one can it will have to be at another level and it will have to be analytically less precise than (though not entirely unrelated to) static Pareto criteria. It is at the level of the institutional framework that efficiency judgments will have to be made. If we can learn anything from history perhaps it is that

certain kinds of social, legal, moral, and economic institutions are generally more conducive to the generation of innovation and prosperity that are others. Policy regimes rather than policies should perhaps be the context for this discussion.

An example is the issue of monopoly policy in high-technology industries (for an in-depth survey see Teece and Coleman 1998). Teece and Coleman argue that the nature of these industries has made the current antitrust environment obsolete. "[A]ntitrust policy cannot realistically aspire to produce 'optimal' outcomes, where 'optimality' is measured against some theoretically defined efficiency or consumer welfare criteria" (Teece and Coleman 1998, 815). What is required is a new way of thinking attuned to the truly dynamic processes of rapid and unpredictable innovation and change, in which competition and monopoly do not necessarily mean the same things as they used to.

If Liebowitz and Margolis are correct in arguing that network-effects are seldom if ever an effective mechanism for isolating firms from competition, then the kind of criteria by which we should judge the presence or absence of monopoly are not the usual market share criteria. As they point out "in this world, the firm competes not to take its share of today's market; it competes to take the market for its share of days. Legislators and courts may choose to make some ways of competing legal and some not, but the old structuralist approaches to antitrust will only be misleading" (1999, 13). The old notion of competition as a state of affairs needs to be replaced by competition as a process over time in which some firms displace others, sometimes as the dominant or only firm in the industry. Competition may indeed show up as "serial monopoly." "Competition here takes a very different form from that found in text book models of perfect competition … There really is no 'competing' by firms in models of perfect competition, except to keep costs down" (1999, 63–64).[7] We should not base policy "on a worldview that is taken a bit too literally from an intermediate macroeconomics textbook. In the real world, information is not perfect; the future is not known with certainty; products are not perfectly homogeneous; and sometimes what we like is influenced by what others like … From the perspective offered by … the textbook model of the ideal economy, we might well be confused by strategic alliances, by technology-sharing agreements, by long-term guarantees, or by the continuous addition of functionality to a product." And they reject the claims "that winners might more appropriately be chosen by policymakers than by people making choices in a free market" (1999, 233).

Implicit in Liebowitz and Margolis' work is the role of knowledge – how it changes over time and how it is related to information. Much more can be said about this. In particular, innovative environments are ones where people are free to make mistakes, that is, where expectations at any time are wrong. Put another way, when we realize that at any point of time different entrepreneurs have different and inconsistent expectations about the same business environment, we must realize that at most only one of them can turn out to be right. Knowledge, unlike information, cannot be shared, and what one perceives as knowledge another will see as (an unjustified) expectation. This is fundamentally a disequilibrium situation (meaning that different people have different expectations). *In such an environment why should one assume that policy-makers are ever in possession of privileged knowledge?* Why are they likely to be right more often than any particular entrepreneur about which technology will turn out to be the best standard for consumers? More to the point, is it not likely that they are in a worse position than private businesses in this regard? In this way their work is in sympathy with some much more fundamental and far-reaching critiques of mainstream economics.

"Testing" the market process

How then does one evaluate the market process? Liebowitz and Margolis have provided a variety of in-depth historical examinations. They uncover the "problematic" nature of such "evidence" for traditional antitrust economics.[8] For example, if we "define standard A to be superior if, for all consumers and any given market share the net value of A is higher than the net value of B when B has the same market share" then it is likely that standards that create greater "social wealth" would tend to dominate and market share winning behavior may appear to be "predatory." Furthermore, "if a clearly superior technology were offered [to the market] first, we would be unlikely to see a sustained attempt to dislodge the leader by owners of inferior technologies, unless they expect that they can achieve their ends through political means, inasmuch as their expenditures in the market are likely to be futile" (Liebowitz and Margolis, 1999, 109–110). In fact, there is no presumption that policy attempts to remove a particular firm's dominance may not *itself* be responsible for us getting locked-in to an inferior standard.

This is relevant to the case of Microsoft. In order to investigate whether claims that Microsoft has achieved an inefficient monopoly are credible, Liebowitz and Margolis investigated a variety of different software products by looking at market shares in relation to reviews in consumer reports (Chapter 10). Computer magazines frequently and

extensively evaluate rival products for quality and functionality. They are obviously not "objective" measures of quality, but they probably come as close as one can to ascertaining whether the products consumers use most fulfill their requirements better or worse than the alternatives. It is possible to argue that consumer reports actually provide a self-fulfilling confirmation of the tests, because consumers buy those products that the reports recommend, but this would be an extreme stretch. It is probably true that most computer users don't read the reports but become aware of the characteristics of the products by use or word of mouth over time.

Liebowitz and Margolis looked at spreadsheets, word processors, (both stand-alone and in suites), financial packages, desktop publishers, online services, and browsers. They found that (see Chapter 10 below):

- "[T]he products that have prevailed have always been the products that were regarded as best by software consumers and by the reviewers in specialist magazines" (1999, 237).
- "[A] large market share, even an extraordinarily large market share, is not a good predictor of a price increase" (1999, 238).
- Changes in market share are smooth and continuous. There is no evidence of "tipping" (the reaching of a threshold beyond which lock-in sets in).
- In many products Microsoft achieved dominance in the Macintosh market considerably earlier than in the PC market.
- Microsoft tended to charge lower prices in the market where it was dominant than in the market where it was competing.
- Rivalrous competition continues to be vigorous.

All in all, the evidence that Liebowitz and Margolis present strikes this writer as sufficient to cast substantial doubt on the case that the US government has brought against Microsoft. In some products Microsoft does not dominate, in others it has won its dominance by producing (often with extreme effort) better products. Where it does dominate, "one does not need to appeal to other factors such as those the government has focused on in its case against Microsoft … to explain Microsoft's increasing share of the market" (Liebowitz and Margolis 1999, 202–203).

Concluding remarks

In the final analysis, Liebowitz and Margolis' work is noteworthy for its solidity (its groundedness). They do not plumb the depths of method-

ology. Yet at the same time, their work transcends their idiom and is pregnant with profound and far-reaching implications. Common-sense and basic historical research leads quickly to highly significant theoretical results and empirical generalizations, not to mention political implications. The hidden (and no doubt unconscious) message is much more revolutionary than it appears.

Liebowitz and Margolis provide a formidable array of objections to some standard arguments. Anyone interested in the issues surrounding the emergence of new technologies and standards and the related policy questions cannot afford to be ignorant of their work. One may hope that this book will at the very least stimulate discussion and further research.[9]

Notes

1. They also argued that, far from this literature constituting a "new economics" that moves beyond the old obsolete macroeconomics, the latter is fully capable of explaining everything that these situations presented. The economics of natural monopoly, of externalities and public goods, of transactions costs and other related well-established themes serve to establish that these new "information age" industries present nothing that is really new in principle, though some interesting new perspectives do emerge. (For example, in reexamining the alleged market failure due to consumption scale economies – the fact that the value of the product depends in part on how many people use it – they point out that this can be interestingly interpreted as the traditional "tragedy of the commons" in reverse.) This general line of reasoning is capable of being misunderstood. To claim that traditional microeconomics can be made to incorporate almost every imaginable case is probably true, but this does not contradict the criticism of the use of perfect competition as a standard in policy discussions.
2. "The occurrence of this 'lock-in' as early as early as the mid-1890's *does* appear to have owed something also to the high cost of software 'conversion' and the resulting *quasi-irreversibility of investments* in specific touch-typing skills. Thus, as far as keyboard conversion costs were concerned, an important asymmetry had appeared between the software and the hardware components of the evolving system: the costs of typewriter software conversion were going up, whereas the costs of typewriter hardware conversion were coming down" (David 1985, 335–6, first italics added).
3. In a sense, network-effects are complementarities in consumption and production and were anticipated by Marshall in his work on industrial districts, trade alliances, etc.
4. Actually, this discussion conflates "knowledge" with "information" in an illegitimate way. Information is public in a way that knowledge is not (Lewin 1997, 1999 Chapters 3 and 12). This will become more important below when I discuss the role of knowledge in policy and research.
5. Of course a vast literature on this exists. The work of Hayek (1988) on the evolution of social institutions is relevant as is that of North (1990). See also Lewin (1997) for a discussion of the connection between stability and flexibility in evolved institutions.

6. This is a typical misstatement. The papers do not address anything about real-world markets. There is no real "history" in them. Instead they address how it is possible to produce sub-optimality in theoretically conceived simulations of real world markets.
7. Cf. Hayek (1978).
8. "[A]nything that a firm does to compete can be, at some point, viewed as an attempt to monopolize. And anything that a firms does to improve its products, extend its standards, or reach additional markets will look like an attempt to monopolize" (Liebowitz and Margolis, 1999, 11).
9. For more information on the debate between Liebowitz and Margolis and Paul David see Lewin (2000) and Chapter 11 below.

References

Arthur W. B. 1983. "On Competing Technologies and Historical Small Events: The Dynamics of Choice Under Increasing Returns." Technological Innovation Program *Workshop Paper*, Department of Economics, Standford University, (November).

———— 1989. "Competing Technologies, Increasing Returns, and Lock-in by Historical Events," *Economic Journal*, 99: 116–131.

Coase, R. H. 1960. "The Problem of Social Cost," *Journal of Law and Economics*, 3: 1–44.

Dahlman, C. 1979. "The Problem of Externality," *Journal of Law and Economics*, 22: 141–163.

David, P. A. 1985. "Clio and the Economics of QWERTY," *American Economic Review*, 75: 332–337.

Farrell, J. and Saloner, G. 1992. "Converters, Compatibility, and Control of Interfaces," *The Journal of Industrial Economics*, 40: 9–36.

Gilbert R. J. 1992. "Symposium on Compatibility: Incentives and Market Structure," *The Journal of Industrial Economics*, 40: 1–8.

Hayek, F. A. 1978. "Competition as a Discovery Procedure," in Hayek, F. A., *New Studies in Philosophy, Politics and Economics and the History of Ideas*, London: Routledge.

———— 1988. *The Fatal Conceit: The Errors of Socialism*, Chicago: University of Chicago Press.

Katz, M. L. and Shapiro, C. 1983. "Network Externalities, Competition and Compatibility," *Woodrow Wilson School Discussion Paper in Economics*, 54, Princeton University September.

———— 1992. "Product Introduction with Network Externalities," *The Journal of Industrial Economics*, 40: 55–84.

Lewin P. 1997. "Hayekian Equilibrium and Change," *Journal of Economic Methodology*, 4: 245–266.

———— 1999. *Capital in Disequilibrium: The Role of Capital in a Changing World*, New York and London: Routledge.

———— 2000. "The Market Process and the Economics of QWERTY: Two Views," *Review of Austrian Economics*, forthcoming.

Liebowitz, S. J. and Margolis S. E. 1990. "The Fable of the Keys," *Journal of Law and Economics*, 33: 1–25.

———— 1999. *Winners, Losers and Microsoft: Competition and Antitrust in High Technology*, Oakland: The Independent Institute.

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

Teece, D. J. and Coleman, M. 1998. "The Meaning of Monopoly: Antitrust Analysis in High-Technology Industries," *The Antitrust Bulletin*, Fall–Winter: 801–857.

Yamada, H. 1980. "A Historical Study of Typewriters and Typing Methods from the Position of Planning Japanese Parallels," *Journal of Information Processing*, 2: 175–202.

———— 1983. "Certain Problems Associated with the Design of Input Keyboards for Japanese Writing," in Cooper W. E. (ed.), *Cognitive Aspects of Skilled Typewriting*, New York: Springer.

Author Index

Subject Index