6

Standards, Modularity, and Innovation: The Case of Medical Practice

Richard N. Langlois
The University of Connecticut

Deborah A. Savage
Southern Connecticut State University
Yale University

The economics of standards and standard setting has grown to considerable prominence in the last few years.[1] Buttressed by influential neoclassical models of network externalities (Farrell & Saloner, 1985; Katz & Shapiro, 1985), this intellectual edifice has as its keystone David's (1985) famous history of the QWERTY keyboard. In David's account, the now-familiar arrangement of keys is a paradigmatic instance of path dependency. The choice of the QWERTY design was essentially a matter of historical accident; and, once that arrangement became dominant, the spiraling benefits of its network of complementary capabilities--notably touch-typing skills--effectively "locked" users into the QWERTY standard.

This chapter takes up the issue of standard setting both in theory and in terms of a historical case study, namely, the setting of standards for the American medical profession in the early 20th century. As a contribution to the literature on standards and path dependency, however, this chapter diverges somewhat from the beaten path. First of all, our case involves behav-

1. For an overview, see David and Greenstein (1990).

ioral standards, not technological standards. Second, our case is one in which the process of path creation did not obviously involve historical accident. Instead, history seems to have been shunted onto its track by underlying economic forces. Indeed, even at the most plastic early stages of the standard-setting process in medicine, seemingly promising proposals for reform and for alternative institutional design were unsuccessful when they did not fit with the underlying logic of professional production. In this respect, our case has more in common with Chandler's (1977) account of the evolution of the modem corporation than it does with QWERTY.

The structure of standards and standard making that emerged from the early century, we argue, was one well adapted both to the provision of medical care in the United States and to the generation of innovation, at least until the modem era of high-cost and high-technology medicine. Even in this modem era, however, reforms are likely to fail that do not take into account the nature of professional production and of behavioral standard setting.

STANDARDS AS INSTITUTIONS

Most analyses of path dependency and lock-in have focused on technological systems in the narrow sense. Favorite examples have included computers, telecommunications systems, and-perhaps especially-various kinds of home entertainment systems like stereos (Langlois & Robertson, 1992), VCRs (Cusumano, Mylonadis, & Rosenbloom, 1992), or high-definition television (Farrell & Shapiro, 1992). In most of these cases, the issue is one of the compatibility of physical components or electronic signals. Occasionally, human behavior is part of what is standardized or coordinated; in the QWERTY case, it was the inertia of human touch-typing skills that generated increasing returns to the dominant standard. But few in the Davidian tradition have focused directly on standards as coordinating human behavior rather than as connecting technology.[2] And this is perhaps surprising. For standards are at base a kind of social institution; and social institutions are recurrent patterns of behavior that help to coordinate human activity (Langlois, 1986b; North, 1990).

As Kindleberger (1983) pointed out, there are basically two classes of standards: those that create economies of scale and those that lower transaction costs. In the former case, economies arise from the increase in the extent of the market that results from reduced variety. For example, in the 1910s the Society of Automotive Engineers set standards for automobile

2. One exception is David (1987) himself, who distinguished between standards of technical design and standards of behavioral performance, even if his own focus seemed to be on the former.

parts that winnowed the kinds of steel tubing in use from 1,600 to 210 and the types of lock washers from 800 to 16 (Epstein, 1928). Independent parts suppliers could then take advantage of longer production runs to reduce costs, which especially helped the smaller car companies that did not have high internal demands for parts.

In the second class of standards, benefits arise because the standards help to reduce the transaction costs of coordination and monitoring. Standards assist in coordination by helping to align expectations. In the classic case, for example, the convention that we all drive on the same side of the road is a standard that reduces the "transaction" costs of ascertaining the intentions of each oncoming driver, not to mention the resource costs of failed coordination. As David (1987) pointed out, behavioral standards of this kind can be thought of as ensuring "interface compatibility" (p. 214) much as do standards of technical design, because such standards help to coordinate the way individuals "connect together" (p. 214). Standards can also reduce measurement and monitoring costs. A single standard of weights and measures, for example, makes the comparison of goods in exchange easier and increases the cost of cheating. More generally, normative standards can reduce costs of monitoring by providing a benchmark against which quality or performance can be judged.[3] This meaning of *standards* will figure prominently in our discussion of medical practice.

The transaction-cost properties of standards are not entirely unrelated to their economies-of-scale function. By regularizing expectations, standards increase the predictability of the extent of the market, which is crucial for large-scale investments in machinery and a more elaborate division of labor. Indeed, the essential tension between flexibility and commitment is perhaps the most intriguing aspect of standard setting. To use the language of Garud and Jain (1996), standards can be at once enabling and constraining. They can be enabling because they create an orderly framework within which economies of scale can develop and technological change can progress effectively. But they can also be constraining in that their necessary rigidity makes costly and thereby inhibits even potentially beneficial change that would require altering the standard. Much of the allure of David's keyboard story comes from the contention that QWERTY is not the best of all possible configurations and that lock in has prevented

3. In a sense, standards are always nonnative in that they take the form: "Do it this way." This is true whether the standard is an injunction to drive on the right of the road or a technical specification constraining design choices. The difference between a coordination standard and a nonnative standard is that the fonner is always self-enforcing, whereas the latter sometimes requires a more complex enforcement mechanism. For example, the standards of cleanliness and efficiency that McDonald's sets for its franchise holders require monitoring of individual locations by company inspectors.

change to a better keyboard.[4] This same logic is true more generally of social institutions. The convention that we all drive on the same side of the road is a standard that brings order out of disorder[5] and increases the efficiency of driving; but to change such a convention can be difficult, as countries like Sweden and Okinawa discovered when they switched sides of the road. We return to the problem of change (innovation) in a system of standards.

STANDARDS AND ROUTINES

The idea that predictable, inflexible, standardized behavior generates economies is quite a general one. Influenced by the work of Herbert Simon on rule-following behavior, Nelson and Winter (1982) proposed that much of economic behavior can be discussed in terms of *routines*. Routines are habitual patterns of behavior that embody useful knowledge. Much of this knowledge is in the form that Polanyi (1958) described as tacit; it is skill-like knowledge that cannot be articulated or transmitted explicitly but that must be acquired over time through a process of apprenticeship and trial-and-error learning. And, as Polanyi also suggests, the development of such tacit knowledge economizes on the scarce resource of conscious attention, thus generating economies.

Although routines and standards are clearly related, they are not identical. As Kindleberger (1983) pointed out, standards are public goods; they reflect interpersonally shared knowledge. We might even say that a standard is a certain kind of "public" routine that helps to coordinate private (individual or intraorganizational) routines.

But routines are not only about coordination. As we saw, routines embody potentially useful-we might even say productive-knowledge. In the terminology of Ryle (1949), they reflect "knowledge how." In some cases, such useful knowledge can be knowledge about how to transact, the possession of which thus reduces transaction costs. My internalized knowledge that I always ought to keep to the right (not the left) as another car approaches might be an example, at least if we construe the interaction between oncoming drivers metaphorically as a transaction. But the skillful exercise of a particular technique for suturing an incision would also be a routine, and not one obviously involving the reduction of transaction costs. Useful knowledge applied to problems of transacting is a special case

4. Liebowitz and Margolis (1990, 1994) however, challenged both David's specific contention about QWERTY and his implied contention that problems of lock-in to suboptimal paths are important and ubiquitous phenomena.

5. There is actually a technical sense in which conventions bring order out of disorder; they reduce the entropy of the behavioral environment (Langlois, 1986a; Schotter, 1981).

of a more general phenomenon. As Winter (1988) maintained that one needs to have economic capabilities (an effective repertoire of routines) in order to be able to transact as well as to be able to produce.[6]

Like standards, routines can be both enabling and constraining. The possession of an effective repertoire of routines would be crucial to the successful production of product A; but possessing that repertoire might also inhibit a transition to the production of product B. Routines are generally as hard to unlearn as to learn, which may give the advantage in situations of radical innovation to those who have never learned the routines in the first place.[7] This is no doubt what Schumpeter (1934) had in mind when he wrote that "new combinations are, as a rule, embodied, as it were, in new firms which generally do not arise out of the old ones but start producing beside them; . . . in general it is not the owner of stage-coaches who builds railways" (p. 66).

But perhaps the interesting aspects of enablement and constraint are those involving the interaction of standards (public routines) with the private routines of individuals and organizations. This is what Garud and Jain (1996) seemed to mean when they talked about the degree to which the technological environment is "embedded" in the institutional environment. When there are no standards, there is complete flexibility, but very little enablement, as "customers and vendors might be prone to wait for the emergence of a dominant design before they are induced to make significant investments" (Garud & Jain, 1996, p. 393). But when standards are too tight, they can suffocate progress, leading to a "stuck" technology with little innovation of any kind. Only when the institutional environment (the standards) "just embeds" the technological matrix do those standards most fully enable, and not constrain, technological development. In such a "just embedded" world, technology and standards coevolve, "each of these reciprocally and continually shaping the other" (Garud & Jain, 1996, p. 393).

Now we describe a system in which technology is just embedded in the institutional structure, although we broaden the idea of technological environment to include (primarily in this case) the routines of human behavior. At the level of these routines, what we tell is a story in which new paths are constantly being created and in which lock-in to particular sets of standards (that is, to standardized behavioral routines) is a "wolf" who never quite gets

6. Indeed, there is a developing literature on economic capabilities (Langlois & Robertson, 1995; Teece & Pisano, 1994) that starts from the difficulties of acquiring production knowledge (embodied in routines) rather than from the kinds of informational problems that create transaction costs. Langlois and Robertson (1995) pointed out that an approach to economic organization that begins with the idea of routines is more consonant with the broader new institutional economics--which is concerned with institutions as recurrent patterns of behavior--than is an approach that takes the transaction as the unit of analysis (Williamson, 1985).

7. Langlois and Robertson (1995, chap. 6) provide a more careful analysis of the phenomenon of economic inertia.

to the door. At the higher institutional level, of course, the just enabled system we describe is itself a "path," and we attempt to shed some light on how that path was staked out and on the ways in which it is now appearing to diverge in the woods. Before turning to the case, however, we need to explore the underlying economic considerations that, we argue, helped nudge the institutional structure in the direction it went.

PROFESSIONAL PRODUCTION

Mintzberg (1979) defined as a "bureaucracy" any organization in which behavior is standardized. The hierarchical organization familiar from Weber (1946, 1947) is what he calls the *machine bureaucracy*. But the professions are also bureaucracies in Mintzberg's sense, in that professions coordinate the economic activities of their members, using standardized routines inculcated through what are normally lengthy processes of training and apprenticeship. And, although we might quibble with the term bureaucracy in this case, it is certainly true that standards--publicly shared routines--are crucial to professional production.

In a profession, routines are largely shared in the sense that the abilities and choices of an individual practitioner are shaped by the abilities of those with similar or complementary skills. Although each practitioner produces independently, all practitioners execute their routines in an environment created by other professionals. For example, a lawyer is constrained by the cumulative precedents of previous cases, most of which were decided long before the current generation entered the profession. At the same time, the creative application of existing law generates new opportunities for future practitioners. This is true, too, for physicians, whose day-to-day decisions are affected by the previous treatments administered to patients by other physicians. In all these cases, the shared routines imply public "interfaces" among practitioners, what engineers call "next-bench design": One can rely on the fact that other engineers, or lawyers, or surgeons have made decisions in ways that one can reconstruct by virtue of one's own training and experience (Feynman & Leighton, 1988).

But what makes a profession different from a machine bureaucracy is the extent to which, along some dimensions, conduct is not standardized. As Stinchcombe (1990, chap. 2) so nicely puts it, professionals are information-processing systems who must wield and apply a wide repertoire of routines to fit widely varying concrete circumstances. The reasons for this reflect the landscapes of both supply and demand. If the price is low enough, a single kind of car--any color Model T you want so long as it's black--will attract buyers, because, in the end, that single kind of car can adequately accommodate a wide range of concrete circumstances. Moreover, there are

large economies of scale to centralized fabrication of cars. By contrast, the product-the service-that professionals offer typically requires finer tuning to a wide variety of concrete circumstances. This is so for reasons that have to do both with preferences and with the technology of provision. (One is willing to tolerate more standardization in a suit of clothes than in a law suit.)

As a result, professionals do not standardize the application of their routines (as does a machine bureaucracy) but only the "toolkit" of routines from which they draw. The particular concrete application of the routines requires on-the-spot professional judgment, a capability Knight thought essential in any situation of true uncertainty (see Langlois & Cosgel, 1993). Like more specific professional routines, judgment is a tacit skill that is cultivated in professional training. Moreover, as we have hinted and as we will discuss in more detail, professionals must also employ judgment in the creation of new routines. As we will see, it is the standardization on toolkits rather than on concrete behavior that provides the necessary contextual flexibility for innovation, making the system of professional practice "just imbedded" in the standards that guide it.

Thus, although professions are not exempt from the Smithian division of labor, and although the professional bureaucracy attempts to create economies of scale through the promulgation of fairly standard toolkits of routines, both the division of labor and standardization take place in a decentralized context. As Savage (1993, 1994) argued, these considerations lead naturally to a theory of the professions as a "production organization."

The conventional neoclassical treatment of the professions focuses almost entirely on the demand side-on the interaction between the professional and the consumer. And it frames this interaction exclusively in terms of the problems of principals and agents (Shaked & Sutton, 1981). In fact, however, one cannot understand the organization of the professions without also considering the supply side (e.g., how professions are organized to employ knowledge usefully and to coordinate among specialized producers). As Jensen and Meckling (1992) point out, economic organization must solve two different kinds of problems: "the rights assignment problem (determining who should exercise a decision right), and the control or agency problem (how to ensure that self-interested decision agents exercise their rights in a way that contributes to the organizational objective)" (p. 251).

Efficiency demands that the appropriate knowledge find its way into the hands of those making decisions. There are basically two ways to ensure such a "collocation" of knowledge and decision making: "One is by moving the knowledge to those with the decision rights; the other is by moving the decision rights to those with the knowledge" (Jensen & Meckling, 1992, p. 253). Markets (in the widest sense of the term) take the latter approach.

The Coase theorem suggests that, so long as decision rights are well defined and alienable, those rights will tend to end up in the possession of those whose specialized knowledge can make the most of them. This also solves the agency problem, because the alienability of the right means that market prices can track the value of the right, which, in turn, creates an incentive for the owner to maximize value by using the right appropriately. But there are also potential costs to such extreme decentralization. These might include the familiar sorts of transaction costs arising from moral hazard and asset specificity. More interestingly, however, they may arise from "dynamic" transaction costs (Langlois, 1992), the costs of bringing otherwise decentralized knowledge together and coordinating it, especially in circumstances involving learning and the generation of new productive knowledge.

The form of organization called the firm is one way to surmount such dynamic transaction costs in some circumstances (Langlois & Robertson, 1995). By concentrating decision rights at the top, the firm can in principle overcome both the narrowness of knowledge of the individual participants and the vestedness of decentralized decision rights. But such centralization of authority comes at the cost of misaligned incentives to the extent that it removes decision rights from the hands of those who must actually execute the routines of production. In a mature firm of the sort Chandler (1977) described, such problems of agency are tolerable because operations are typically characterized by repeated, consistent replication of known routines. Such routines tend to be measurable at various stages of production, and so lend themselves relatively well to formal monitoring schemes, including documentation, accounting trails, and supervision of employees (Barzel, 1982). Clearly, organizations of this sort are not obviously well adapted to the problems of professional production.

But firms and markets are not the only alternatives. There is a growing literature on hybrid forms of organization, forms that offer distinctive solutions to the problems of rights assignment and agency. Principal among these forms are networks (Powell, 1990); the professional network is a particularly important example.

Through formal and informal arrangements, professionals share rent-earning competences without ceding autonomy to a central hierarchy. When professionals locate together in a network, they do not take a joint equity position or even sign a contract. Although remaining legally independent, they make a long-term commitment of their substantial human capital to a "hubless" network organization. Because networks do not integrate ownership, they have a horizontal rather than a hierarchical coordinating structure. In fact, network members remain competitors across many dimensions.

Unlike the ideal of a price-mediated market, however, a professional network is able, with the help of standards, to provide the function of knowl-

edge integration and coordination. Without the exchange of cash payments, members exchange information and technology, and collaborate in production (that is, share routines) without authoritarian supervision, and without integrating external management functions into their day-to-day operations.

Networks are able to transform tacit knowledge into capabilities much more valuable than any individual practitioner could have acquired alone. This is so because the network provides incentives to share skills. Practitioners recognize that they are dependent on the distinctive competences of other practitioners, and that it will be in their individual best interests to share these competences. Indeed, professionals often borrow the routines of others for both exchange and production. For example, von Hippel (1989) showed that engineers routinely share technical information with rivals. Such sharing is most likely to occur when a professional is attempting to solve a new or difficult problem. To put the matter differently, the knowledge, routines, and capabilities that give economic value to professional production lie in the interface between individual practitioners and the system. In the terminology of sociologists, professionals have complex relational roles (Barley, 1990).

MODULARITY AND INNOVATION

Sanchez and Mahoney (1996) argued for a kind of duality between the structure of products and the structure of organizations. Modular organizations are conducive to modular products, that is, to products with standardized interfaces; at the same time, modularity in product designs is conducive to modular organizational design. Professional networks are very much the kind of loosely coupled system they describe as modular. In this case, however, the "product" is in the nature of behavioral routines, and these are modular in the sense that standards have created a widely shared toolkit from which professionals can draw their repertoires. At the same time, the organization (the network) is modular in the sense that, at least in principle, there are sharply defined boundaries between the subspecialties that make up the profession. Mintzberg (1979) talks of the process of "pigeonholing" (p. 352), in which clients are sorted into categories according to which subset of standardized tools best fit their needs. Professional subspecialties are one kind of pigeonhole. And all professionals know more or less the kind of skills each subspecialty represents and how and when to interact with it.

In practice, of course, the boundaries between professions are not always sharp. Boundary disputes occur frequently, and boundaries can overlap for long periods of time (Halpern, 1992; Savage, 1993, 1994). But this only means that, as in technological systems, professions reflect continual inter-

action among standards, boundaries, and exogenous conditions. The result is the kind of ongoing process of modularization and remodularization that Garud and Kumaraswamy (1995) liken (in the technological context) to rebuilding a ship plank by plank even as it sails.

The degree and character of the modularity of a system determines the pace and direction of technological change in that system. From the perspective of product design, a system in which the parts are standardized but the connections among the parts are not lends itself to architectural innovation (Henderson & Clark, 1990), that is, to innovative recombinations of standardized parts. By contrast, a system in which the interfaces among the parts are standardized lends itself to modular innovation (Langlois & Robertson, 1992), that is, to innovation in which the parts improve in performance without changing the way in which they are hooked together. A more significant distinction in this context would be that between *systemic innovation* and *autonomous innovation* (Langlois & Robertson, 1992; Teece, 1986). A systemic innovation is one that requires simultaneous change in several stages of production. In the case of professional production, this would mean change that spills across professional boundaries. By contrast, autonomous innovation is change that can take place safely within existing boundaries.

Within the boundaries of a professional subspecialty, practitioners engage in architectural innovation. As we saw, in day-to-day practice, the professional uses judgment to select and apply more or less standard routines in new combinations in response to unique circumstances. In the large, however, professions tend to lend themselves more naturally to autonomous innovation. Individual practitioners can improve their repertories of routines, and those improvements can diffuse quickly to others with similar training, so long as the boundaries between standard subspecialties change relatively slowly. As Mintzberg (1979) noted, it is "the pigeonholing process that enables the Professional Bureaucracy to decouple its various operating tasks and assign them to individual, relatively autonomous professionals. Each can, instead of giving a great deal of attention to coordinating his work with his peers, focus on perfecting his skills" (p. 353).

Another way to look at the issue of modularity and innovation is to think of the structure of an organization as defining a kind of cognitive or perceptual structure (Langlois, 1997). In a sense, this is the Sanchez and Mahoney argument in a different guise. The cognitive structure of an organization is not just a matter of information flows along an organization chart; rather, it is strongly influenced by the modular character of its products and of its structure. For example, when IBM switched from a regime of architectural innovation to one of modular innovation based around the 360/370 series of computers, the organization's perceptual system changed dramatically (Langlois, 1997). In the case of professional networks, as we see, the decen-

tralized structure creates a perceptual system that is open to a range of new ideas and that, like a market, is capable of rapid trial and error learning[8] (Nelson & Winter, 1977). But it is also a system whose perceptual ability is circumscribed by the standards that constitute and modularize it.

In order to make these arguments more concrete, we turn to our case study: the history of standard setting and innovation in medicine (here defined broadly to include surgery as well as medicine proper). This is a story of the process by which, and the reasons for which, medical "toolkits" came to be standardized. And it illustrates one possible way in which paths can be created.

MEDICAL STANDARDS

As Savage (1993, 1994) argued, each profession possesses a core competence that defines what we might call its natural boundaries, that is, the boundaries one would observe in the absence of important institutional overlays, such as legal restrictions.[9] Surgeons and physicians, for example, were distinct professions since at least the Middle Ages. Once members of the guild of barbers, surgeons have always possessed a core competence in wound management (Wangensteen & Wangensteen, 1978), in contrast to the physician's core competence in diagnosis. By the late 19th century, a convergence of changes in technology and practice emerged to greatly amplify the competences of physicians and, especially, of surgeons.[10] The adoption of anesthetics by the late 1840s, coupled later in that century with the techniques of antiseptic and aseptic practice (which preceded a very slowly emerging consensus around the germ theory of disease) turned surgery from a painful and often deadly route of last resort into a powerful and widely applicable approach to therapy.[11] Improved diagnostic techniques,

8. Garud, Kumaraswamy, and Prabhu (1996) discussed another kind of network-what they call a generative network-in similar terms. Although professional networks are a somewhat different kind of learning system (largely because of their modular structure), they share with generative and other networks the ability to tap into a wider variety of information sources than can a more centralized organization.

9. This is in contrast to many traditional accounts of professions, which see boundaries solely as the product of legal or other institutional restrictions.

10. In the terminology of Tushman and Anderson (1986), these innovations were competence enhancing for physicians and surgeons. Such need not always be the case. As Savage (1993, 1994) argued, technological change has been competence destroying for pharmacists, whose traditional core competence had been the certification of the strength and purity ofmedicinal drugs. The increased complexity of pharmaceuticals and the higher capital intensity of testing equipment has shifted that competence upstream to the drug manufacturers.

11. The full emergence of modem surgery had to await the development of antibiotics drugs (such as sulfa) in the 1930s. These "three As" -- anesthesia, antisepsis, and antibiotics -- are generally seen as the transforming forces of modem surgery. As we argued, however, they amplified rather than created the surgeon's core competence in wound management, greatly extending the boundaries of that specialty into realms of therapy that had been the province of the physician as well as into new realms of possibility.

including laboratory tests and Roentgen's X-ray device, had a similar, if less dramatic, effect for physicians (Gelijns & Rosenberg, 1995).

As the 20th century began, the medical professions started to consolidate the dramatic changes of the 19th century. In a sense, these professions experienced the kind of maturation process one often finds in the life cycle of technological innovations (Abernathy & Utterback, 1978). One part of the consolidation process was the setting of standards (notably in the areas of medical education) and the keeping of medical records.

Traditionally, the education of physicians and surgeons was a matter of apprenticeship. By the 19th century in the United States, medical school training (often in conjunction with internship) had become the norm, with some 85% of physicians trained between 1811 and 1820 having taken medical degrees (Rothstein, 1987). But even late in the 19th century, many of these medical schools were small proprietary affairs that granted degrees easily and quickly. The conventional account of the transformation of medical education that occurred in the early 20th century lays great stress on the normative character of the standards that were adopted. Many point to the influential report in 1910 by Abraham Flexner, who was working for the Carnegie Foundation for the Advancement of Teaching in close collaboration with the Council on Medical Education of the American Medical Association. Flexner deplored the state of most American medical schools and argued for standards based around the example of Johns Hopkins, America's premier medical school at the time. The Flexner report, and the press coverage it generated, is often credited with a precipitous decline in proprietary medical schools and the rapid tightening of requirements by those (largely now affiliated with major universities) that remained (Kaufman, 1976).

On closer examination, however, it is clear that the standards involved were as much conventional as normative. For one thing, Flexner himself emphasized facilities, physical plant, and such measurable variables as enrollments, entrance requirements, and number of teachers, not the quality of teaching in the medical schools (Rothstein 1987). In essence, Flexner was implicitly arguing for the comparison of medical schools with the outward characteristics of a (presumed high-quality) exemplar, not for the direct assessment of quality.

Moreover, it is clear that the changes toward standardization-and away from proprietary medical schools, which had always had a high mortality rate-were essentially in place by the time Flexner wrote (Ludmerer, 1985; Rothstein, 1987). In part, American medical education in the 19th century had reflected the diverse and still largely rural character of the country. By century's end, however, may of the same forces that Chandler (1977) described in manufacturing were also overtaking the medical professions, namely, the increased scale and integration of the American economy in an

era of population growth and reduced transportation and communications costs.[12] One historian of medical education puts it this way:

> Medical education could not remain immune from the processes that were so radically transforming the traditional way of living and working in America. The country had become much more closely integrated than at any previous time in history, and as a result of new technological breakthroughs in transportation and communication, Americans had become much more mobile as well. Medical education felt the effect of these changes. If a college student in Indiana wished to attend a medical school in California, or if a medical student at Oklahoma, a two-year school, desired to do his clinical training in Minnesota, or if a graduate of the University of Michigan wanted to practice in the state of Wisconsin, a certain amount of uniformity in medical education was mandatory. Absolute equivalency was not necessary, only a high degree of similarity. However, this was the direction that medical education in the early 1900s was already taking. The innumerable discussions occurring among representatives of medical schools from all parts of the country served to reconcile differences and guarantee that courses and requirements would mean the same thing everywhere. (Ludmerer, 1985, p. 89.)

As we saw, standardizing professional education is one important way to produce standardized professional "toolkits" and to help ensure "next-bench design" among cooperating but decentralized professionals. We can read the revolution in medical education of the early century as generating on a national scale a process that is typical of professional production.

One might expect that, as medical schools became more closely integrated with hospitals, many of the same forces that helped standardize medical education would also tend to standardize hospitals. To the extent that this happened, however, it did so in an indirect way. If one imagines hospitals to be hierarchical firm-like organizations, then it should clearly be the managers ofhospitals at the forefront of standardizing hospital structure and practice. And, in 1913, the American Hospital Association (AHA), a group dominated by physicians who had become hospital superintendents, decided in principle on standards for hospitals. But the organization lacked the political clout and the resources to implement its recommendations (Stevens, 1989). In part, this impotence reflected the odd position of the superintendent, who was caught between donors and medical staff (especially admitting physicians and surgeons) without the authority typical of managers in corporations. As Stevens (1989) noted, the superintendents tended as a result to organize themselves "around the hospital's hotel-management and financial concerns rather than in relation to its medical function-the entire purpose for which, in theory, the hospital existed" (p. 73).

From the point of view of the theory of professions just outlined, this result may seem less surprising. As Savage and Robertson (1998) argued, a

12. Indeed, it is a recognized shortcoming of Chandler's (1977) account that he ignores the role in economic growth of noncorporate sectors like the professions (Landes, 1991; Supple, 1991).

hospital is not an organization like a firm. Rather, it is an example of what they call the *coancillary* institution of a profession. For reasons that we have already suggested, decision rights in medical practice reside largely with the decentralized professionals. Those professionals also possess significant decision rights over institutions like hospitals that are complementary to professional practice but that the professionals do not own. In this light, it is not surprising that hospital superintendents should have been relegated principally those (complementary) tasks that fall outside the competences of the medical personnel.

Nor, thus, is it surprising that the effective move to standardization in hospitals was taken by the American College of Surgeons (ACS), not by the AHA. In 1913, the ACS set up a committee on the standardization of hospitals, installing as chair a Boston surgeon named Ernest Amory Codman (Howell, 1995; Stevens, 1989). Codman proposed what he called an "end-result system," that is, a system of monitoring hospital care (especially surgeries) directly by outcomes. Part of the system would require that hospitals set up detailed systems of record keeping. The policy the ACS eventually settled on in 1917 jettisoned the end-result system but retained, and, indeed, focused on, the standardization of patient records. Coupled with the original mission of the ACS-to recognize and thereby certify the competence of surgeons-the resulting system of standards fit well with the professional nature of surgical production.

Stevens (1989) saw these developments as a matter of surgeons jealously guarding their prerogatives:

> Standardizing the surgeon emphasized professional authority. Standardizing the work suggested the surgeon was a mere craftworker, or even a mechanic working in an organization. Scientific management in industry ranged from time and motion studies to sweeping organizational analyses--activities that surgeons did not want. Codman's methods would put trustees and administrators as policemen over medical work.. . . Surgeons, as a group, were willing to upgrade standards but not to lose their professional prerogatives to individuals they did not trust. (p. 77)

But, as Savage and Robertson (1998) argued, the issue was not one of trusting hospital administrators. The issue was one of the optimal collocation of knowledge and decision rights; and an arrangement in which professionals retain authority and autonomy is arguably a system of monitoring with characteristics superior to one in which the hospital is recast as a hierarchical organization in the model of a firm.

Moreover, it is in the network's interest to discourage the kind of monitoring Codman was proposing; this is so because end-result monitoring (or outcomes monitoring, as it is now called) changes the focus from the network to the individual practitioner. The potential punitive uses of external monitoring would cause individual practitioners to be less likely to share informa-

tion about their practices and would therefore diminish the value of shared competences, with destructive repercussions for the network and, withal, for patient choice. It is characteristic of professional networks to provide practitioners with up-to-date information about the latest developments in their profession, as occurs at professional meetings. But external monitoring presupposes some proprietary use of information, again lessening the incentive to share.

In a sense, the model of self-governance that developed out of the ACS standardization program may be thought of as a way of building monitoring into the production process rather than imposing it from outside. With standardized and open records, the ACS hoped to disseminate information about and to evaluate the procedures and methods of their fellow surgeons. They instituted a policy of regular local meetings at which clinical experiences were reviewed and interpreted. The process enabled surgeons to educate themselves about the safety and efficacy of competing procedures, and to decide which to include in surgical training and continued education. It was also the College of Surgeons-the specialty most dependent on emerging hospital capabilities-that recommended that hospitals adopt an open, but defined, medical staff model.[13] The ACS recognized that the future lay not in keeping fellow surgeons out, but in monitoring the quality and ensuring the cooperation of those who shared hospital resources. Over time, other specialties adopted similar formats, and in 1952 the various accrediting organizations merged to form the Joint Commission of Medical Accreditation, which is now called the Joint Commission on Accreditation of Healthcare Organizations.

MEDICAL INNOVATION

We interpret the processes of standardization that took place in the early 20th century not as a fundamental break with the existing system of medical production in the United States but rather as an accommodation to that existing system of technological and demographic changes that the 19th century had left behind. By standardizing the professional (medical education) and by standardizing the interfaces among professionals (medical records), the medical profession, like a Chandlerian firm, had seized the scale economies of a national market. And, by linking these standards to the open-staff

13. In an open-staff model, the physicians and surgeons affiliated with a hospital evaluate the credentials of other professionals who wish hospital privileges. But the incumbent staffconsider only the technical credentials of the applicants, not the commercial feasibility or desirability of admitting them to staff. (Thus the staff is "open" to anyone who is deemed competent.) "Economic credentialling" is becoming more common today, however, as hospitals in the eraofmanaged care move closer to aclosed-staff model and to a more hierarchical organizational form (Savage & Robertson, 1998).

model of hospital governance, the profession had brought the new technologies under professional control (like operating theaters and X-ray machines) whose scale ofoperation was larger than the individual practitioner. In this way, the medical profession arguably retained the porous network structure it had long possessed.

It is our contention that these standard-setting events of the early 20th century provided the institutional framework that governed medical production at least until the modem era of managed care. These events created the path down which American medicine traveled. Perhaps things might have happened differently. If the AHA or ACS had been able to impose end-result standards of the sort Codman had proposed, perhaps hospitals would have become something more like Chandlerian firms, which would have affected the rate and direction of technological change (among many other things). On the other hand, our reading of the history does not suggest that the path taken resulted in an obviously inferior or undesirable system; quite the opposite. It is not inexplicable that the ACS, not the AHA, set the standards, and that Codman's ideas fell so easily by the wayside. As we suggested, there is an important logic to the location of medical decision rights in the hands of the decentralized practitioners rather than in any central location. It would have required some large countervailing consideration-as may be present today in the form of more or less exogenous changes in the system of health-care financing-to overcome the organizational economics of professional production.

Our argument is that, by permitting a continuation of the earlier professional form of organization, the standards of the early 20th century turned out to be enabling in the sense of Garud and Jain (1996). As we saw, innovation in a professional network is likely to be both architectural (involving the individual practitioner's toolkit of routines) and autonomous (focused, at least initially, within existing professional or subprofessional boundaries). In addition, innovation in such a network is likely to benefit from a wide variety of information sources. Because professionals possess only a narrow range of capabilities (in at least some contrast, perhaps, to the research laboratories of the large corporation), they must of necessity rely on knowledge developed elsewhere, often in distant fields of knowledge, and must adapt those outside discoveries to their own local needs. Moreover, because professionals possess localized expertise that is hard to transmit to others, those professional are more likely to see opportunities for innovation than are the holders of the complementary outside knowledge. As a result, we would expect practitioners to be important sources of changes in practice and in technology. This was certainly the case in the 19th century. And, despite the increasing complexity of technology, it remains the case to a large extent even today.

Most of the important surgical advances in the 19th century were the result of professions. These included anesthesia, (suggested in one of its forms by a dentist from Hartford, Connecticut), the various tentative approaches to antiseptic and aseptic technique, and even the development of rubber gloves, which required the collaboration of a rubber company (Wangensteen & Wangensteen, 1978). Many diagnostic tools, including the electrocardiogram and the electroencephalogram, were also developed by physicians.

What is more striking is that medical practitioners remain important in technological change into the modem era. The principal change is that today it is work at academic medical centers rather than by isolated practitioners that seems to be important. Gelijns and Rosenberg (1999) noted that, although academic medical centers conduct almost no basic research in medical devices, they are significant players in the development of new devices-a pattern in stark contrast to that in other academic fields. In their study of innovation in endoscopy, Gelijns and Rosenberg (1995) discovered that innovation was a network process in which the network of medical researchers allied itself with other networks, especially technologists in industrial firms who possessed complementary knowledge. In one case, however, it was the medical researchers who actually solved important technical and manufacturing problems that had eluded the industrial scientists with whom they collaborated. In some cases, more than collaboration was required. In a study of the development of medical lasers, Spetz (1995) found that advances often came from academic physicians who had significant training in other fields, notably physics.

CONCLUSIONS

We argued that standard setting is a phenomenon that goes well beyond the well-trod ground of technical standards. Behavioral and other knowledge standards are a fertile area of study. We also argued that the creation of paths in the process of standardization does not always involve the arbitrary force of historical accident. Sometimes the underlying economics of production provide the contours in which standardization must ultimately rest. In the case of medical standards, the path that was created in the early 20th century in the United States was strongly influenced by the economics of professional production, by technological change in medicine and surgery, and by the dynamic of American economic growth in the period. Once established, however, those standards proved "enabling" in that they created a decentralized network that was open to ideas from outside and was able to collaborate easily in the interdisciplinary fashion that proved crucial for the development of new devices and techniques.

ACKNOWLEDGMENT

Deborah Savage gratefully acknowledges the support of the Robert Wood Johnson Foundation. Opinions expressed in this chapter are those of the authors and not necessarily those of the Foundation.

REFERENCES

Abernathy, W., & Utterback, J. M. (1978, June, July). Patterns of industrial innovation. *Technology Review*, 41-47.

Barley, S. R. (1990). The alignment of technology and structure through roles and networks. *Administrative Science Quarterly*, 35, 61-103.

Barzel, Y. (1982). Measurement costs and the organization of markets. *Journal of Law and Economics*, 25, 27-48.

Chandler, A. D., Jr. (1977). The visible hand: The managerial revolution in American business. Cambridge, MA: The Belknap Press of Harvard University Press.

Coase, R. (1960). The problem of social cost. *Journal of Law and Economics*, 2, 3.

David, P. A. (1985). Clio and the economics of QWERTY. *American Economic Review*, 75(2), 332-337.

David, P. A. (1987). Some new standards for the economics of standardization in the information age. In P. Dasgupta and P. Stoneman (Eds.), Economic policy and technological perfonmance (pp. 206-239). Cambridge, England: Cambridge University Press.

David, P. A., & Greenstein, S. (1990). The economics of compatibility standards: An introduction to recent research. *Economics of Innovation and New Technology*, 1 (1-2), 3-41.

Epstein, R. C. (1928). The automobile industry: Its economic and commercial deveIopment. Chicago: A. W. Shaw.

Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *Rand Journal of Economics*, 16(l), 70-83.

Farrell, J., & Shapiro, C. (1992). Standard setting in high-definition television. *Brookings Papers on Economic Activity* (Microeconomics), l-77.

Feynman, R. P. [as told to Leighton, R.] (1988). What do you care what other people think? New York: Norton.

Garud, R., & Jain, S. (1996). The embeddedness of technological systems. *Advances in Strategic Management*, 13, 389-408.

Garud, R., & Kumaraswamy, A. (1995). Coupling the technical and institutional faces of Janus in network industries. In W. R. Scott & S. Christensen (Eds.), The institutional constuction of organizations: International and longitudinal studies. Thousand Oaks, CA: Sage.

Garud, R., Kumaraswamy, A., & Prabhu, A. (1996). Networking for success in cyberspace. Unpublished manuscript, New York University.

Gelijns, A. C., & Rosenberg, N. (1995). From the scalpel to the scope: Endoscopic innovations in gastroenterology, gynecology, and surgery. In N. Rosenberg, A. C. Gelijns, & H. Dawkins (Eds.), Medical Innovation at the Crossroads. Volume V: Sources of medical technology: universities and industries (pp. 67-96). Washington, DC: National Academy Press.

Gelijns, A. C., & Rosenberg, N. (1999). Diagnostic devices: An analysis of comparative advantages. In D. C. Mowery & R. R. Nelson (Eds.) , The sources of industrial leadership (pp. 3 12-358). New York: Cambridge University Press.

Grossman, S., & Hart, 0. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94, 691-719.

Halpern, S. A. (1992). Dynamics of professional control: Internal coalitions and crossprofessional boundaries. *American Journal of Sociology*, 97(4), 994-1021.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 359-30.

Howell, J. D. (1995). Technology in the hospital: Transforming patient care in the early twentieth century. Baltimore. MD: Johns Hopkins University Press.

Jensen, M. & Meckling, W. (1992). Specific and general knowledge, and organizational structure. In L. Werin &a H. Wijkander (Eds.), Contract Economics, pp. 251-274. Oxford, England: Basil Blackwell.

Katz, M., &a Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review*, 75(3), 424-440.

Kaufman, M. (1976). American medical education: The formative years, 1765-1910. Westport, CT: Greenwood Press.

Kindleberger, C. P. (1983). Standards as public, collective and private goods. *Kyklos*, 36(3), 377-396.

Landes, D. S. (1991). Introduction: On technology and growth. In P. Higonnet, D. S. Landes, & H. Rosovsky (Eds.), Favorites of fortune: Technology, growth and economic development since the Industrial Revolution (pp. l-29). Cambridge, MA: Harvard University Press.

Langlois, R. N. (1986a). Coherence and flexibility: Social institutions in a world of radical uncertainty. In l. Kirzner (Ed.), Subjectivism, intelligibility, and economic understanding: Essays in honor of the eightieth birthday of Ludwig Lachmann (pp. 171-191). NewYork: New York University Press.

Langlois, R. N. (1986b). The new institutional economics: An introductory essay. In R. N. Langlois (Ed.), Economics as a process: Essays in the new institutional economics (pp. l-25). New York: Cambridge University Press.

Langlois, R. N. (1992). Transaction-cost economics in real time. *Industrial and Corporate Change*, 1(1), 99-127.

Langlois, R. N. (1997). Cognition and capabilities: Opportunities seized and missed in the history of the computer industry. In R. Garud, P. Nayyar, & Z. Shapira (Eds.), Technological learning, oversights and foresights (pp. 7 l-94). New York: Cambridge University Press.

Langlois, R. N., & Cosgel, M. M. (1993). Frank Knight on risk, uncertainty, and the firm: A new interpretation. *Economic Inquiry*, 31, 456-465.

Langlois, R. N., & Robertson, P. L. (1992). Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries. *Research Policy*, 21(4), 297-313.

Langlois, R. N., & Robertson, P. L. (1995). Firms markets, and economic change: A dynamic theory of business institutions. London: Routledge.

Liebowitz, S. J., & Margolis, S. E. (1990). The fable of the keys. *Journal of Law and Economics*, 33(l), l-25.

Liebowitz, S. J., & Margolis, S. E. (1994). Network externalities: An uncommon tragedy. *Journal of Economic Perspectives*, 8(2), 133-150.

Ludmerer, K. M. (1985). Learning to heal: The development of American medical education. New York: Basic Books.

Mintzberg, H. (1979). The structuring of organizations. Englewood Cliffs: Prentice-Hall.

Nelson, R. R., &Winter, S. G.. (1977). In search of more useful theory of innovation. *Research Policy*, 5, 36-76.

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. (1990). Institutions, institutional change, and economic performance. New York: Cambridge University Press.

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

Powell, W. W. (1990). Neither market nor hierarchy: Network forms of organization. *Research in Organizational Behavior*, 12, 295-336.

Rothstein, W. G. (1987). American medical schools and the practice of medicine : A history. New York : Oxford University Press.

Ryle, G. (1949). Knowing how and knowing that. In idem, *The concept of mind.* London: Hutchison's Universal Library.

Sanchez, R., &a Mahoney, J. T. (1996). Modularity, flexibility, and knowledge management in product and organizational design. Strategic Management Journal, 17, 63-76.

Savage, D. A. (1993). Change and response: An economic theory of professions with an application to pharmacy. Unpublished Ph.D. dissertation, the University of Connecticut.

Savage, D. A. (1994). The professions in theory and history: The case of pharmacy. *Business and Economic History*, 23(2), 130-160.

Savage, D. A., & Robertson, P. L. (1998). The maintenance of professional authority: The case of physicians and hospitals in the United States. In P. L. Robertson (Ed.), Authority and control in modern industry (pp. 155-172). London: Routledge.

Schotter, A. (1981). The economic theory of social institutions. New York: Cambridge University Press.

Schumpeter, J. A. (1934). The Theory of economic development. Cambridge, MA: Harvard University Press.

Shaked, A., & Sutton, J. (1981). The self-regulating profession. *Review of Economic Studies*, 48, 217-234.

Smith, A. (1976). An enquiry into the nature and cause of the wealth of nations. Oxford: Clarendon Press, Glasgow edition. [First published in 1776.]

Spetz, J. (1995). Physicians and physicists: The interdisciplinary introduction of the laser to medicine. In N. Rosenberg, A. C. Gelijns, & H. Dawkins (Eds.), Sources of medical technology: Universities and industries (pp. 41-66). Washington: National Academy Press.

Stevens, R. (1989). In Sickness and in wealth: American hospitals in the twentieth century. New York: Basic Books.

Stinchcombe, A. L. (1990). Information and Organizations Berkeley: University of California Press.

Supple, B. (1991). Scale and scope: Alfred Chandler and the dynamics of industrial capitalism. *Economic History Review*, 44, 500-514.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing, and public policy. *Research Policy*, 15, 285-305.

Teece, D. J., & Pisano, G. (1994). The dynamic capabilities of firms: An introduction. *Industrial and Corporate Change*, 3(3), 537-556.

Tushman, M. L., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly* 3 1, 439-465.

von Hippel, E. (1989). Cooperation between rivals: Informal know-how trading. In B. Carlsson (Ed.), Industrial dynamics: Technological, organizational and structural changes in industries and firms. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Wangensteen, O. H., & Wangensteen, S. D. (1978). The rise of surgery: From empiric craft to scientific discipline. Minneapolis: University of Minnesota Press.

Weber, M. (1946). From Max Weber Essays in sociology (H. H. Gerth & C. W. Mills, Eds. & Trans.). New York: Oxford University Press.

Weber, M. (1947). The theory of social and economic organization (A. M. Henderson & T. Parsons, Trans., T. Parsons, Ed.). New York: Oxford University Press.

Williamson, 0. E. (1985). The economic institutions of capitalism. New York: The Free Press.

Winter, S. G. (1988). On Coase, competence, and the corporation. *Journal of Law, Economics, and Organization*, 4(l), 163-180.