Economics has no theory of professions as distinct economic institutions. Economists have, however, written extensively about individual professions such as medicine and law and about professions generally. Almost without exception, economic studies of professions take their existence as given and focus on certain behaviors commonly associated with them, such as licensing or bans on advertising. In taking this approach, economists make the implicit assumption that professions are ordinary, though perhaps objectionable, neoclassical firms whose performance we can assess appropriately using the standards we apply to such firms. Even the most rigorous expositions of economic theory as applied to professions view professional conduct as a simple question of market structure: “The policy maker’s problem . . . reduces to whether professionals should . . . be allowed to retain monopolistic powers” [Shaked and Sutton, 1981, p. 217].

The lack of a theory specific to the professions creates some intriguing problems for the economist’s research agenda. To begin with, although professions are defined as firms, professionals are most often modeled as some kind of non-homogeneous self-employed labor. That is, practitioners are assumed to maximize some sort of personal utility function rather than profit. Of course, no one is entirely happy with this artifice, especially because the comparative-static results are often quite sensitive to these arbitrary specifications of utility functions [Foley, Shaked, and Sutton, 1981]. As there is no theory to help us judge among the alternative formulations, economists tend to model each profession differently. Our dissatisfaction increases as we attempt to apply these models in order to explain the history and organizational
behaviors of professions. How can we explain the long history, continued existence, and general acceptance of self-regulating professional associations, whose constraints on professional behavior would be unacceptable to any standard neoclassical firm and which current economic theory can comprehend only as a cartel?

There is a much larger literature on professions in sociology, which focuses mainly on professions as social groups and on the roles that professionals and professional associations play in society. Each of the competing sociological theories of professions reflects a particular view of the relationship between social groups and economic systems [Abbott, 1988]. However, sociology also tends to view professions simply as some higher-order subset of occupations. As in economics, this approach creates a tension in the literature. First, the approach blurs the distinction between individual practitioners and the profession as a whole. Second, it fails to explain how professional associations differ fundamentally from trade unions.

Nonetheless, a promising research program is emerging among sociologists studying the role of knowledge in occupations. In this view, “when we say a person is ‘skilled,’ ‘semiskilled,’ or ‘professional,’ we are describing what sort of an information processing system he or she is” [Stinchcombe, 1990, p. 32]. As we shall see, this insight is valuable because it directs our attention to the kinds of routines that professionals use in production and to the complexity of their underlying knowledge base.

The major shortcoming of conventional approaches to understanding professions is that, although they purport to study producers, they are in fact analyzing demand-side questions. For instance, when economists ask if production is made more or less efficient by the presence of licensing, they are really asking whether the consumer is better or worse off under licensing. The distinction is important, because under the usual demand-side approach to analyzing professions, “much of the interest hinges on the analysis of consumer choice over quality; it is also an area in which consumers typically have imperfect information regarding rival products” [Foley, Shaked, and Sutton, 1981, p. 1]. The implication is that the existence of professions is prima facie evidence of market failure.
In fact, this kind of analysis most often has nothing to do with the way in which—or even whether—production takes place. In contrast, the premise of this paper is that professions are neither occupations nor firms, but instead represent an example of the network form of organization. Professional networks have evolved and continue to survive because they represent comparatively efficient and adaptable solutions to certain kinds of dynamic production problems. In this paper, I identify some specific kinds of problems that professions solve particularly well. One of the interesting implications of this analysis is that licensing—a central issue for much of the economics debate about professions—is neither a necessary nor a sufficient condition for the existence of professions.

A Theory of Professions

I begin with the following definition of a profession [Savage, 1993]:

A profession is a network of strategic alliances across ownership boundaries among practitioners who share a core competence.

The remainder of this section expands upon this definition.

Competences

In broad terms, dynamic capabilities and competences can explain how firm-specific assets are developed and then adapted as responses to changes in the external and internal competitive environment. The theory of economic capabilities is still in a relatively early stage of development; even the terminology is far from standardized [Teece, Pisano, and Shuen, 1992]. The literature owes a great deal to the work of Penrose [1959] and of Nelson and Winter [1982]. Both of these books focus our attention on the ways that firms adapt in order to survive, on what firms know, and on how they know it.
Neoclassical production theory relies on factors of production as the basic units of analysis. The capabilities framework distinguishes such homogeneous inputs from firm-specific assets because the former lack an organization-specific component. Examples of homogeneous inputs include unskilled labor and information available in the public domain; firm-specific assets include special production facilities and processes and methods of managing innovation and change. These assets cannot easily be either imitated or transferred between organizations, because they embody organizational knowledge that, by its nature, is tacit. Or, the assets may be technically transferable, but only at a transaction cost that is sufficiently high to offset entirely their value to other organizations.

Taken as a group, these firm-specific assets constitute a competence. The value of these competences is that they are not product-specific and may be a source of rent in a variety of applications. Core competences are those that are crucial to an organization’s survival. Good management successfully identifies, nurtures, and develops core competences and constantly samples the market to identify goods and services to produce with them. Some of the literature goes further and identifies distinctive competences as the set of activities that a firm can coordinate better than other firms. Finally, capabilities are the activities that can be undertaken with a set of competences.

This terminology, used increasingly in business strategy literature, appears likely to become the standard for discussing the capabilities and competences of economic institutions. A recent application characterized a competence as “the collective learning in the organization” [Prahalad and Hamel, 1990, p. 82]. However, in terms of its ability to convey the importance of acquiring and managing knowledge, I prefer Nelson and Winter’s approach to understanding how firms exploit specific assets. In their explanation, knowledge is contained in routines. The ability to replicate production tasks is a simple example of a routine—although not necessarily an example of a simple routine; the ability to establish new competences is another. The knowledge contained in routines is not only that of individual employees and managers, but also the larger and largely inseparable knowledge of the organization. Each firm has an underlying strategy that must be consistent with its structure. For example, “a firm whose
strategy calls for being a technological leader that does not have a sizable R&D operation, or whose R&D director has little input into firm decision making, clearly has a structure out of tune with its strategy" [Nelson, 1991, p. 67]. The success of the organization depends on having “practiced organizational routines” [ibid., p.68] that underlie the existence of core competences and capabilities of the firm. The message is that no matter how brilliant the plan, an organization that has the wrong set of complementary capabilities will be unable to implement it.

**Networks**

Although core competences play an important role in defining professions as knowledge-reliant production organizations, the concept is insufficient to differentiate professions from firms. Professional reliance on the existence of networks serves this purpose. A network is an economic organization that accomplishes the exchange of capital, products, and/or knowledge without explicit equity investment or “ownership.” Applied to the organization of professions, a network implies a community of practitioners operating separately for many purposes, but dependent on the network for the maintenance and development of core competences that earn them rents. This interdependency identifies professional networks as exchange organizations that rely on interaction with and feedback from a network of individual practitioners within the institution. This strategic relationship among professionals creates a competitive advantage for network members in their interactions with other institutions within and outside the network. Network coordination includes a variety of sub-institutions that help the professional network and its members maintain, develop, and validate competences in a variety of formal and informal ways.

The transaction-costs literature has provided us with the insight that, under a variety of very common conditions, internal organization might well be superior to market exchange [Williamson, 1985]. It is, however, a mistake to conclude from this observation that all economic organizations lie on a continuum between markets and firms [Powell, 1990]. Networks are a third form of organizing production and
exchange. The overriding strategy of professional networks is to internalize knowledge and coordinate its transfer without integrating ownership. This allows professionals to remain legally independent while making a credible 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. Without the exchange of cash payments, members exchange information and technology, and they collaborate in production—that is, they share routines—without authoritarian supervision and without integrating external management functions into their day-to-day operations. Strikingly, network members remain competitors across many dimensions.

Note the ways in which this differentiates professions from firms. First, firms in the same industry usually will not expend resources in constructing formal alliances with each other because they are not dependent on other firms for competency-based rent-earning strategies. In contrast, the knowledge, routines, and capabilities that give economic meaning to professional competences reside in organizations and institutions, and not in individuals. This is one reason why successful companies do not fail when one person leaves. Yet in professions, the actual delivery of the product is accomplished by individuals who are members of the network. By definition, the network is a feedback mechanism made up of mortal practitioners and a longer-lived institution. (A useful analogy can be made to the importance of Mozart as an individual. Other musicians can play his music, but they can not write new pieces. The knowledge of his music and methods is (imperfectly) stored in written materials and collective memory.) So the most important reason for the interdependence, and therefore the success, of the network form of institution is the development and maintenance of routines. No single professional knows all of the routines or can manage all of the competences that make up the required capabilities.

Professions are particularly well adapted to dealing with uncertainty because of the system of practitioner-level autonomy. This form of organization allows decision-making processes to be unstructured and varied, while the open communication among professionals allows successful outcomes—innovations, for
example-to be adapted and diffused rapidly without destabilizing related markets. Practitioner-level autonomy succeeds because the contents of certain routines have been developed through the network process—they are the result of past and ongoing, explicit and tacit, negotiations.

Notice that it is not the nature of the product, but the nature of production, that is crucial to the choice of organizational form. Professional production contains many standardized routines, to be sure. However, production itself is far from routinized, especially in the sense that very few of the decisions that professionals make can be completely prespecified. This is in sharp contrast to most kinds of production. For example, an employee of a burger joint, given an order, has only one decision to make: how fast, within a small subset of choices, do I hustle? The decisions about what a hamburger is and how it is to be prepared have already been made: two all-beef patties, special sauce, lettuce, cheese, pickles, onions on a sesame-seed bun. Professionals, on the other hand, are expected to translate requests for uncertain outcomes, loosely called products or services, into production by combining standardized routines and their own experience (a kind of capability) into decisions. This explains why so much innovation—often in the form of improvements to routines—originates at the practitioner level, which is unlikely to have a formal research and development structure.

More often than not, a practitioner’s performance has a direct, if not necessarily immediate, effect on the ability of other practitioners to use their own capabilities. For example, standard terminology and techniques are important in law, medicine, and engineering because independent practitioners need to be able to reconstruct both the process and the outcome of another practitioner’s production decisions. That is, practitioners often play complementary roles, even while competing among themselves.

Professions embody shared competences. The locus of professional production is a well-defined community of practitioners possessing an esoteric knowledge core [Constant, 1984]. Often this means that the decisions of professionals, bringing to bear skills and knowledge in solving a production problem, follow what in the engineering profession is called “next bench design.” In a variety of
ways, an individual practitioner’s decisions are constrained by the capabilities of the network as a whole, because their productive activities must be implemented within the system. Each individual’s routines have to “interface” or be compatible with the routines of others in the network. Each individual practitioner represents embodied or human capital, often in the form of tacit knowledge; but the products and services that each produces require integration of this esoteric knowledge base across practitioners.

There has been a significant amount of economic research into the role of reputation on market outcomes, primarily focusing on its characteristics as a constraint on undesirable professional behavior toward consumers [Klein and Leffler, 1981]. For example, reputation might serve as a quality signal to consumers, allowing some practitioners to raise prices above (theoretically) competitive levels. As a result, professionals have an incentive to over-invest in reputation-enhancing activities [Shapiro, 1986]. These consumer-welfare oriented studies fail to take into account the importance of reputation for producers operating in a network setting, especially where individual practitioners differ in ability, training, and experience. In a network, reputation is a Marshallian external economy; that is, it is a “public” asset that has value to individual members of the network. For this reason, professionals place a high value on their relationships with peers. Network subinstitutions like self-regulation and professional associations use reputations as a basis for solving complex coordination problems. Over time, professionals develop reputations that reflect on the way other members do their jobs. This “stock” reputation effect is exacerbated by the “flow” effect of the entry of new members and the exit of others. Reputations have all the advantages of brand loyalty and most of the problems of public goods.

The strategy of professional networks is to maximize the value of the diversity in individual competences among its members. These differences are the most important sources of innovation and dynamic capabilities for networks, which must continue to produce even when membership is changing, and even though no one practitioner can know all of the production routines.

Internal reputation is intended to convey precisely this information. Some members will contribute more than others to
production, innovation, and strategic decision-making. A variety of network mechanisms use reputations to communicate information about actual and potential contributions by each individual to others in the network. To professionals, reputation signals the ownership of assets that are valuable as inputs into production by other professionals and for which market transfer mechanisms are unlikely to work. Reputation is such an important input into professional production that each member, no doubt, would like to contract with every other member in order to guarantee certain standards of effort and qualification. Such contracts are, however, not feasible. Without the benefit of hierarchical supervision, monitoring, and management of jointly owned rent-earning assets, it would be virtually impossible to construct or enforce countless bilateral or relational contracts between each pair of professionals. The main items that would have to be specified in the contract involve tacit knowledge, autonomous implementation of routines, and adherence to norms. But we know that such contracts would be infeasible: tacit knowledge is not appropriable by contract or ownership; the “essence” of professions depends on autonomy, and the future is not knowable. These problems, along with unenforceability, are presumably what led to the creation of the professional network as an organizational form in the first place.

In place of hierarchical supervision, whether through ownership or contracts, networks rely on reputation as the basis for peer monitoring. Peer monitoring, unlike contracts, relies on self-interest rather than the availability of third-party enforcement. Individuals want to maximize the value of their membership in the network in order to gain access to its shareable assets. They do so by undertaking behaviors that maintain and develop their reputation among their peers. The effect of this investment on the quality of practitioners is a spillover benefit enjoyed by consumers, but it is not designed for them alone. In fact, professionals in a community are often ranked within the profession in a strikingly different order than they are by consumers.

There are many formal and informal sub-institutions that facilitate information flows about reputations. Professionals evaluate one another’s competences directly, as, for example, during consultations. They also do so indirectly through books and journals,
at conferences, or simply in discussion with other members of the network.

For professional networks, peer monitoring works well, compared to other monitoring schemes, across a number of dimensions: “When agents interact to produce output, they acquire low-cost information about colleagues. Mutual monitoring systems derive their energy from the interests of agents to use internal agent markets of organizations to enhance the value of capital” [Fama and Jensen, 1983, p. 310]. Peers are good monitors not only because of what they know, but because they are the ones to whom good monitoring matters most. In an organization without a hierarchy or ownership, residual risk is borne by all members. The members need to know if shared resources are being used efficiently, if individuals are keeping up with changes in the field, and if they are behaving in ways that are generally in line with the network’s strategy.

Ownership Boundaries

To most consumers, the boundaries between professions probably seem as real as the Berlin Wall of the Cold War era: clearly illuminated, carefully guarded, and, if not always convenient, at least unambiguous in intent. Consumers know that (legally, at least) they can purchase certain categories of goods and services only from specified professionals. Suppliers are trained to know the parameters that describe what their final products and services are supposed to look like and, to some extent, how they are to produce them. Although the customary divisions of labor—for example, those between doctors and nurses, doctors and pharmacists, and, in Great Britain, barristers and solicitors—are sometimes awkward and expensive, they are at least reasonably consistent across professions and not altogether counter-intuitive.

On the other hand, professionals, like Cold War politicians, most certainly do not take these boundaries as set in concrete. Professions and paraprofessions constantly jostle for position, seeking to extend their jurisdictions. Professionals also test the boundaries in less formal ways, by building competences and, over time, introducing new products and services that reward their individual efforts.
By focusing on the demand side of the market, economists do tend to take the customary divisions as givens. It follows that, to the extent that they consider the origins and continued existence of the borders at all, economists assume them to be the result of licensing alone. For example, by interpreting the arrangements between professionals and clients solely as a principal-agent problem, we assume that the clients’ problems are solved by finding a contract that aligns the interests of professionals with those of clients. In fact, both demand and supply must be considered. The evidence surely shows that, from the professional’s perspective, the real problem is getting the incentives of the client and other professionals in line with her own so that any production at all will take place.

One of the strengths of this production-based approach is its ability to offer an economic explanation for the observed recent escalation in the number and intensity of border skirmishes [Felsenthal, 1992]. Before I provide an example of this phenomenon, however, the role of ownership in professions must be explained and contrasted with the ownership problem of the firm.

By boundaries in my definition of a profession, I mean the location of production, and especially of potential overlap between practitioners and the profession, professions and other professions, and professions and other economic organizations. What can the nature and location of boundaries tell us about relationships among practitioners in the same profession? Where does a profession end and the market begin?

Recall that the existence of a network allows a community of practitioners to operate independently for some purposes but not for others. That is, professional networks preserve autonomy and authority for professions, while sharing some and building many distinctive competences available across ownership boundaries. This concept clearly differentiates professions from occupations, where each person owns his or her own labor and must seek only an employment relationship in order to use that labor to earn income.

By autonomy I mean that no one except another professional is able to challenge the day-to-day decisions of a professional. The essential routines that constitute the production of professionals are chosen and adapted by the individual practitioner. Along with
autonomy, professional institutions must also inculcate self-restraint if the authority of the professions is to be legitimate. The origins of the capabilities that operationalize autonomy are complex. As with many craft jobs, mechanisms exist through which basic knowledge of routines is conveyed gradually as the apprenticed member passes specified selection points. Each selection point emphasizes a different competence. As a full-fledged member of a network, however, a professional is free to choose the situations in which particular routines will be used; that is, qualified professionals choose their own jurisdictions. They are also free, indeed expected, to use accumulated experience to apply routines in new ways, to stop using routines, and to invent new routines.

By *authority* I mean to emphasize that professionals possess command capabilities not available to non-professional economic agents in a market economy. Examples of this include the attorney’s command over the legal system’s resources and a physician’s command over the health care system’s assets. The source of authority is expert knowledge—that is, the professional not only knows difficult, abstract principles, but also knows routines that allow the application of these principles. Professionals have the ability to solve routine problems easily, and non-routine problems routinely. By authority I do *not* mean the ability to force clients into actions: patients cannot be forced to undergo procedures or even to comply with a physician’s advice; students can be sanctioned by bad grades, but this, as we certainly know, is not a proxy for compliance; and lawyers, too, can only hope that their advice will be heeded.

The limits of autonomy and authority define the boundaries of individual professional practice and, in this restricted sense, the limits of professional ownership within a horizontal governance structure as well. For professions, the benefits of adopting a firm-like institutional structure would be more than offset by the high costs of attempting to write contracts between individual practitioners for the exchange of tacit knowledge. This is especially true because individual effort is a large part of earning income even from shared routines, so that individual practitioners would be particularly sensitive to any appropriation of their individual rents. Network “ownership” emerges as a solution to these problems.
This section confronts the theory of professions adumbrated above with the history of pharmacy as a profession separate from medicine and chemistry. The story of this separation involved both different professional tasks or skills and different professional goals. The central problem that I will address in this paper is the identification of the appropriate location of production. I solve this problem by identifying a unique, constant, and identifiable core task that has successfully differentiated pharmacy from all other related professions: the certification of the strength and purity of medicinal drugs. All of the other tasks that have been identified with pharmacy at one time or another, or in one place or another, have been peripheral to this core; as such, they have been episodically jettisoned by pharmacists or raided by other professions. Pharmacists have had to respond to changes in science and technology as well as to developments in the economic organization of trade and production in order to be able to continue to fulfill their core responsibility. Their predominant response has been to institutionalize their expertise, which has resulted in increasingly inflexible boundaries between the professions. In spite of this rigidity, controversies (which I call border skirmishes) continue to occur, as the realities of change overwhelm the existing institutions. Especially over the last century, changes in technology, coupled with an increasing understanding of disease, have altered the location of both drug research and drug production. In the twentieth century, therapeutic drugs have become sufficiently complex that both pharmacists and physicians have yielded to pharmaceutical chemists in both of these functions.

Boundary disputes in the provision of medicinal substances can be traced back as far as recorded history [Savage, 1993]. Much of the secondary-source background material for this section draws on History of Pharmacy [Kremers and Urdang, 1976], the most comprehensive of the few available texts on pharmacy history. It is also the source of the following passage, which is an excellent representation of the official rhetoric of the debate: “These advances in pharmaceutical knowledge and technic on the one side and medical knowledge on the other, in connection with a growing recognition of governmental responsibility for the health of the people..., fostered a division of labor between
pharmacy and medicine, and finally the creation of the public health system in which the profession of pharmacy was given a definite place of its own” [p. 27]. Amazingly, the authors are describing conditions in eighth-century Damascus! Leaving aside the question of historical accuracy, the conclusion is, at best, optimistic. The tasks that are most closely identified with medicine and pharmacy have been and continue to be performed in a variety of professional and physical locations by a variety of practitioners. In this paper, I will focus on the development of the boundaries of the pharmacy profession in the United States.

Not surprisingly, there are more differences than similarities between the economic and institutional forms of pharmacy in the United States and in Europe. Even across Europe there are many dissimilarities; but throughout the Continent pharmacists have in common their historical roots in guilds [Gilmour, 1932]. Apothecaries’ guilds were customarily granted monopolies over the sale of specific services, drugs, and other goods resulting in more clearly defined and legally protected boundaries between the professions than in America. In Colonial America, by contrast, consumers were happy to obtain familiar drugs from whatever source they could. Generally, this meant waiting for shipments from European cities. Although there was some interest on the part of colonial physicians in identifying indigenous plants and producing medicines, they had neither the time nor the financial support for it. The result of these beginnings is that in America medicines were sold and compounded by physicians who had dispensaries in their offices; by professional pharmacists, to the extent that they emigrated here in the early years of expansion (pharmacy schools were formed much later than medical schools); by general stores, which engaged in barter for all kinds of goods; and by wholesalers and importers, who capitalized on the demands of various ethnic groups for familiar kinds of drugs from the “old country.”

The contrast with European pharmacies is clearly quite sharp. Though individual American pharmacists exhibited the same recognition of the task of certifying the quality of drugs as their counterparts in the old country, they could not depend on having a prescription-compounding practice sufficient to support the customary level of innovation and testing. In addition, they were competing directly with physicians, who were at least as educated and who were
firmly attached to including compounding and dispensing among their tasks. Physicians’ interest in this aspect of their practice was no doubt threefold. First, dispensing produced additional income for the practice; second, even if they were willing to write a prescription, there were only a few professional pharmacies capable of filling it correctly; and, finally, in the virtual absence of legislative standards of any kind, the safest drugs were those secured and prepared by the physician himself. As a result, dispensing became a traditional task of practicing medicine and, as late as the 1940s, the majority of doctors’ offices included facilities for preparing drugs.

A watershed event for the institutionalization of pharmacy boundaries in America occurred in the 1840s, when a pharmacist, Ewen McIntyre, “discovered that a portion of supposed calcium carbonate, imported from England, was in fact calcium sulfate” and further investigation proved that many of the drugs coming into the country were “substituted, adulterated, or deficient in strength” [Kremers and Urdang, p. 198]. The traditional institutions that had existed in Europe to deal with these problems did not exist in the United States, so pharmacists banded together to push for government standards. This action emphasizes pharmacy’s recognition of its historical core concern with the safety of medicinal drugs. The episode led to the creation of legal standards for imported drugs, a solution well-suited to the realities of politics and economics at that time. The federal law went into effect in 1848. However, although it specified standards for and required inspection of imported drugs, the law proved to be useless. Political appointees with little training and insufficient technical skills served as inspectors. Furthermore, the legal specifications for strength and ingredients were, by *Pharmacopoeia* standards, so imprecise as to be unenforceable [Nitardy, 1934].

This chaotic situation provided the impetus for the creation of a national professional association of pharmacists. At the urging of local pharmacy and medical associations and colleges, a letter was sent to call a convention “for the purpose of considering the propriety and practicability of fixing a set of standard strengths and qualities of drugs and chemicals for the government of the United States Drug Inspectors” [Anonymous, 1852]. This afforded pharmacists an opportunity to organize for broader purposes, and resulted in the drafting of a
constitution in 1852. The 1856 revision of the Constitution of the American Pharmaceutical Association addresses various aspects of pharmacy practice, including business relations with physicians, education, and entry. Significantly, however, the first section of Article 1 defines the aim of the organization as “to improve and regulate the drug market, by preventing the importation of inferior, adulterated or deteriorated drugs, and by detecting and exposing home adulteration.”

Consistent with their roots in European pharmacy practice, pharmacists (and physicians) in the early republic bundled the task of assuring quality and purity with other tasks, including compounding and research and development. In the process of preparing drugs for patients and physicians, individual pharmacists continually added to the accumulated knowledge of interactions between chemical substances, and some speculated on the way in which these drugs worked in the body. They developed dozens of ways to administer drugs, and in the process learned how to purify, make extracts, distill, and infuse. To their minds, all of these were a necessary part of the compounding function. Throughout the seventeenth and eighteenth centuries, most of the new remedies and preparation methods that appeared in formularies were the creations of community pharmacies. However, because medical treatments in America originated in Europe and especially England, early pharmacists found that much of their business depended on wholesale importing of medicinal substances to be provided to pharmacies and physicians throughout America [Mason, 1901].

In European countries, guilds had long-established traditions of compounding quantities of drugs for their communities, and this served as the basis for the eventual development of larger-scale drug and chemical manufacturing enterprises. In America, the tasks of preparing basic ingredients and compounding medicines remained shop-level activities far into the nineteenth century. Eventually research breakthroughs in pharmaceutical chemistry led to methods of synthesizing the active ingredients of medicinal plants [Wardell, 1979]. As a result of cumulative changes in economic institutions, technical processes, and scientific knowledge, drug and chemical manufacturing shifted from shops to small laboratories (some of which were in pharmacy and medical schools) and finally into larger, specialized firms. Throughout this transition, pharmacists continued their obsession
with standards and their strategy of defining pharmacy as the keeper of those standards. The institutions were changing in fundamental ways, but pharmacists were determined to play a key role in defining what the new institutions would look like. Individually and collectively, they adapted and innovated their own institutions to assure that pharmacy would be influential in setting and enforcing standards affecting the provision of medicinal drugs. The first Pharmacopoeia of the United States of America (USP), published on December 15, 1820, and the revisions that followed provide an illuminating record of the impact of economic and technological changes on the production and distribution of drugs in the United States in general and on pharmacy practice in particular. The first edition primarily reflects the concerns of physicians regarding the potential problems of using drugs prepared by someone other than the attending physician. It differentiates among drugs compounded by physicians, by local pharmacists, and by “manufactories.” The concerns of physicians were well-founded, since there were as yet no pharmacy schools or major pharmacy associations in America. Without an official pharmacopoeia, each physician and pharmacist relied on local formularies or, more often, on their European training and literature. As a result, standards of drug preparation and usage differed greatly across communities.

The Philadelphia College of Pharmacy and Science was organized in 1821, and its faculty and laboratories were instrumental in facilitating the Second Revision (1842) of the USP. As a result of this collaboration, the pharmacopoeia included detailed descriptions of brand-new techniques like extraction by percolation and of the latest methods for testing the purity of many of the drugs listed in the text.

When time came for the Third Revision (1851 edition), the convention explicitly recognized pharmacists as parts of a separately represented and trained profession with special expertise in preparing medicines. Pharmacy’s preeminent position was all but assured by 1863, when the Fourth Revision instituted the practice of including findings recently published in the Proceedings of the American Pharmaceutical Association. This edition also contains some clues about the ways that technological changes were beginning to alter pharmacy practice. Among these are explicit recognition of the need to coordinate future publications with the British Pharmacopoeia, which
indicates the importance of widening the professional network in response to the increasing formalization of applied scientific knowledge; and a movement away from using the old system of apothecaries’ weights and measures, no doubt a response to the standardization of manufacturing technology.

The Sixth Revision (1882 edition) reveals the combined effects of the overwhelming changes in technology and science and the strategic choices of the pharmacy profession. This passage from Kremers and Urdang [p. 267] conveys how fundamental these changes were to the location of pharmacy practice:

The new Pharmacopoeia turned more sharply away from the outworn concept of community pharmacy as the place of manufacture of most of the pharmaceutical preparations. Instead it tried to establish in the pharmacy another kind of responsibility: examination of medicinal substances by the pharmacist as a check on quality. Casual mention of a few tests was replaced with detailed tests for identifying and determining the purity of many of the drugs. Detailed processes for assaying the alkaloids appeared for the first time....Symbolic formulas and molecular weights were introduced.

The revision does not explicitly discuss the redirection of pharmacy’s network strategy. We can, however, interpret its description of good pharmacy practice as a call for pharmacists to assume the role of agent for the consumer of medicinal drugs, whether physician or patient. The pharmacist is to stand between manufacturers and consumers, who will rely on the expertise of individual pharmacists and network institutions like the USP to evaluate the quality and purity of drugs.

The Seventh Revision (1894 edition) provides more evidence of the institutional response required to deal with the rapid pace of change. By this time, some states have adopted the USP as a legal standard. Patent medicines and proprietary drugs, a problem of growing proportions and increasing severity, are excluded from the USP altogether. The convention used metric measurements to specify
standards of purity that were known to be feasible given the current capabilities of large-scale manufacturers.

Further adaptations to economic realities are seen in the Eighth Revision (1905 edition), which acknowledged the importance of including information about patent medicines in reference form. The compromise reached by the convention allowed consideration of drugs whose ingredients and strengths were ascertainable, while continuing to exclude drugs with secret ingredients. This revision was notable, too, for its inclusion of anti-diphtheria serum, the first “biological” drug to be described in a USP.

In the 1905 edition, we begin to sense that the boundary between pharmacists and physicians is becoming more rigid. For example, this edition introduced directions for average doses, rather than maximums and minimums. These guidelines encouraged pharmacists to double-check prescriptions written by physicians and de-emphasized the role of pharmacists in prescribing directly. This is, perhaps, tacit recognition that the core task of certifying quality was, for the time being, no longer strategically linked to diagnosing and treating disease, and, at the same time, that physicians had their hands full as well and no longer were a strategic threat in the pharmaceutical sphere. In a sense, the USP had come full circle: once written as a technical aid for dispensing physicians, it had now become a general reference book and catalogue for prescribing physicians and a technical reference manual for pharmacists. In keeping with this trend, the USP included even more tests and assays intended for pharmacists’ use.

Notice that the boundaries were not becoming more rigid because of laws forbidding physicians to compound or pharmacists to prescribe, but because the relevant components of the institutional networks are strategically concentrating their energies on those tasks most crucial to protecting and extending their core competences. The relative competences of the combined medical/pharmaceutical convention that produced the USP was recognized by the federal government, which granted it full legal recognition under the Food and Drug Act of 1906.

The preface to the Eleventh Revision (1936 edition) notes that the task of listing and testing all of the available drugs, plus preparing articles on new types of drugs and chemical methods, was now so great
that not only pharmacists and physicians, but manufacturing companies, university laboratories, and government scientists from the United States and abroad were necessary to accomplish the huge task of assembling the edition.

In 1950, the convention was headed by Lloyd C. Miller, who was neither a physician nor a pharmacist, but a biochemist with a pharmacology background. These qualifications were no coincidence. The Sixteenth Revision (1960 edition) stressed that the “feature that distinguished this revision from all of its recent predecessors is the progress made in adopting new analytic techniques....[U]se has been made of nonaqueous titremetry, complexometry, ultraviolet and infrared spectrophotometry, column and paper chromatography and . . .phase solubility” [p. xvi]. Clearly, the time had come when a community pharmacy could no longer be expected to house the kind of equipment that would allow the pharmacist to perform in-store tests on individual drugs from purchased inventory. Ironically, pharmacy schools had just agreed on a restructured curriculum that required extensive training in theoretical pharmaceutical chemistry. The goal was to ensure that pharmacists would enter practice with the ability to assess all of the new drugs entering the market. As individuals, they could; but as retail community pharmacists, they could not.

The Eighteenth Revision (1970 edition) spells out the problem facing pharmacists quite vividly. The preface directly addresses the adoption of “good manufacturing processes,” which included a variety of sophisticated tests and procedures. One of the effects of the adoption of these criteria, even by willing commercial producers, is that production of drugs by any but large manufacturers becomes virtually impossible. Interestingly, the manufacturing specifications are contained in a section of the revision separated from the information intended for the dispensing pharmacist in a community setting. There can be no doubt, even to the casual follower of the revisions, that pharmaceutical companies are now the real players in the production of drugs for consumers. It is at this point that professional pharmacies in the United States begin a desperate scramble to find a way to salvage their profession by finding strategies to protect the rent-earning power of their core competency in providing safe and effective drugs.
Throughout the eighteenth, nineteenth, and early twentieth centuries, the retail drug industry was characterized by consumer access to a variety of types of sellers. As late as the 1940s, the majority of physician’s offices included facilities for compounding drugs. Prescriptions were not required by law or custom, and consumers were free to purchase whichever drugs they felt would benefit them most. Advertisements for medicines were aimed exclusively at the consumer, not at physicians or pharmacists. As a result, retail pharmacies in this country had never counted on earning the major part of their revenues from their prescription practice. As late as 1931, fewer than one percent of all businesses calling themselves pharmacies received as much as half of their revenues from their prescription departments [Jordaan, 1931].

The changes that began in the late nineteenth century continued. If it was increasingly difficult for local pharmacists to compound drugs in their shops, it was becoming virtually impossible for physicians to do so in their offices. These technological impediments, combined with significant institutional changes detailed below, primarily in laws, added to the likelihood that physicians would write prescriptions for pharmacists to fill. Over the next few decades, pharmacy prescription practice grew rapidly, and by 1962, 25 percent of all pharmacies earned half or more of their income from prescriptions [Olsen, 1973]. In a very real sense, the sale of drugs had never been the main concern of early American pharmacists. Their economic success originated in a core competency in certifying quality, but it was achieved through wholesaling, retailing, employment in physician’s laboratories, and in a variety of other venues.

Kremers and Urdang [p. 315], themselves pharmacists, documented one of the effects of the transition to larger-scale manufacture. In the 1930s, 75 percent of all prescriptions still required compounding skills. That is, the physician’s written prescription was a formula composed of a variety of substances that the dispenser had to combine using technically sophisticated methods. By 1950, the number of prescriptions fitting this description fell to about 25 percent; by 1962, to less than 4 percent; and by 1973, to less than one percent.

Looking at these trends, we can conclude that physicians were writing more prescriptions than ever before, with the intention that they be filled by pharmacists. At the same time, the medicines they specified
were increasingly the finished products of large manufacturing companies. In fact, a growing number of physicians had begun adopting the practice of prescribing drugs by trade name rather than by ingredients. A 1973 survey reported that almost 90 percent of all drugs were prescribed this way. As a result, only 12 percent of prescribed drugs were included in the USP [Kremers and Urdang, p. 315].

These trends raise two issues. First, what accounts for the drastic increase in prescription writing by physicians? Second, what was the effect of the decrease in shop-based compounding on the location of the performance of the core task of certifying quality, strength, and efficacy, and on the boundaries of the pharmacy profession as a whole?

In wrestling with the problem of whether and how to include patent drugs, the Eighth Revision of the USP (1905 edition) was in actuality attempting to adapt this institution to the realities of the U.S. drug industry. Seventy-five million dollars a year were spent on patent medicines, the overwhelming majority of which were either worthless or dangerous. The USP was designed to help pharmacists certify quality by setting standards and specifying tests. However, patent medicines were by nature secret (that is, not patented) formulas. Increasingly, independent pharmacies lacked the technical facilities needed to reveal the contents of snake-oil medicine bottles. In addition, the formulas of individual remedies were changed at the whim of the manufacturer, which prevented even a slow accumulation of information about the many medications on the market. A private study by muckraker Samuel Hopkin Adams [1907] showed that many contained alcohol, opium, or cocaine as their active ingredients. They were habit-forming, and long-term use caused lung disease and other illnesses.

Since the standards approach was proving inadequate, pharmacy associations supported the adoption of specific laws as a way to maintain their core task of certifying quality. In keeping with the training of its membership and the information on testing and safety available to pharmacists through the USP, the American Pharmaceutical Association pushed for standards that would require all drugs to be clearly labeled with the names and amounts of every ingredient. However, the Proprietary Association of America, which was founded in 1881 to represent the interests of manufacturers of home (patent)
medications and cosmetics, had a head start. The products of its members were marketed directly to consumers rather than to physicians, and the advertising fees its members paid were virtually the sole support of the print media. Furthermore, the sale of these products accounted for a large part of the revenues of general stores. As a result, the watered-down law that ultimately emerged was the 1906 Pure Food and Drug Act, which simply allowed for penalties against producers of misbranded drugs. A misbranded drug, as described in Section 5 of the act, was one whose label contained false or misleading statements about its contents: “That drug shall be deemed misbranded...if the package containing it, or its label, shall bear any statement regarding the existence or nonexistence or the amount or purity of any ingredient or substance contained therein, which statement shall be false or misleading.”

Pharmacists had also requested explicit labeling about the strength of active ingredients. Instead, though the bill recognized the USP as the official standard, it allowed manufacturers to use drugs of greater or lesser strength in their products. Even worse, from the standpoint of physicians and pharmacists, the federal law effectively superseded the control that some states had granted to state pharmacy boards to apply the USP as a standard for the regulation of drug production and sales within their borders. Not surprisingly, the bill did almost nothing to protect the ability of pharmacists to certify the purity, safety, and efficacy of drugs sold in their shops. Instead, it meant that a patent medicine company’s therapeutic claims for a product were limited only by the imagination of its marketing department. Scanty information on the label meant less liability for the firm.

The next critical juncture for pharmacists was the Massengill Elixir Sulfanilimide disaster in 1937 [Wilson, 1942]. Although the story is actually quite complicated, we can summarize it briefly here. The crisis occurred when the company created a liquid form of the drug sulfanilimide, which had been used successfully and safely for several years in capsule and tablet form. First developed in France in the mid-1930s, sulfanilimide was found to kill streptococci bacteria. Many companies competed in marketing versions of the drug. In order to improve its early market performance, therefore, salesmen for Massengill and other companies requested a liquid version that would
be especially attractive to the parents of young children. It proved difficult, however, for pharmaceutical chemists to find a solvent for the powdered drug. Massengill finally stumbled upon a terrifically successful one, diethylene glycol. Almost as soon as the drug hit the market, reports of painful and torturous deaths were reported to the American Medical Association, the American Pharmaceutical Association, and the Food and Drug Administration. This should come as no surprise to us, since the solvent is well known to us as antifreeze. A successful recall, in which, notably, pharmacy records played an important part, was instituted; but about one hundred people died, many of them children.

This episode highlighted the inadequacies of the 1906 law, which did not prohibit the sale of dangerous or untested substances, but addressed only the problem of mislabeling. Ironically, however, it was misbranding that allowed the early and successful recall to occur and, importantly, sanctions to be applied. The USP defined an elixir as a sweetened, aromatic mixture of alcohol and water, which this mixture was not. If the label had read tincture (meaning an alcohol solution), it is unlikely that the government or the courts could have enforced any penalties against the company.

The Elixir Sulfanilimide tragedy leads us to the answer to the question of why prescription drug dispensing rose so quickly in its importance to pharmacy incomes. Partly in response to the disaster, the federal government passed a revised Food, Drug and Cosmetic Law in 1938. Among its provisions was the requirement that manufacturers provide labels for their drugs that included the identity of ingredients, directions for use, specific dosages, and warnings about potential problems. The extent of any deviation of the drug from the USP standards of strength, quality, and purity also had to be clearly described on the label. The law reestablished the liability of the producer if the specified dosage caused injury and required tests proving that claims on the label were true. Once again, pharmacists encouraged this legislation as a new institution that could aid them in certifying the purity, safety, and efficacy of drugs to the general public. Once again, too, the drug producers reacted by proposing amendments that weakened the bill. This time, the loophole excepted from these stringent labeling requirements those drugs that were to be repackaged for consumer use
(presumably by pharmacists or small manufacturers) and those for which a written prescription was provided to the consumer by a physician, dentist, or veterinarian. The justification for this was that physicians would provide detailed information directly to the patient.

The unintended and, as far as I have been able to uncover from the literature of the period, unforeseen result was that drugs were effectively but arbitrarily divided into over-the-counter (OTC) and prescription-only drugs. The option depended only on the willingness of the manufacturer to assume responsibility for labeling. Pharmacists criticized the amendment for its effect on the public. Clearly, for many drugs it would be advantageous for manufacturers to avail themselves of the opportunity to avoid testing and labeling. Pharmacists would be unable to sell these drugs without prescriptions, and consumers would have to pay for an office visit to procure them. The amendment’s sponsors responded that the “bill is not intended to restrict in any way the availability of drugs for self-medication. On the contrary, it is intended to make self-medication safer and more effective” [H.R. 2139, 1938, p. 8]. In fact, of course, consumer choice was curtailed, with debatable benefits. Ironically, moreover, the provisions inserted at the behest of the drug companies had the unintended effect of creating a guaranteed income for pharmacies and of formalizing the institutional boundaries between physicians and pharmacists.

Not surprisingly, the arbitrariness of the OTC/prescription-only designation became unworkable within a few decades. In the 1950s identical medicines were sold under both categories, differing only because one producer chose to label and another to let the pharmacists dispense the drug. The Durham-Humphrey Act (both Carl T. Durham and Hubert H. Humphrey were pharmacists) amended the Federal Food, Drug and Cosmetic Act (Section 503b, 1952) and empowered the Food and Drug Administration to decide whether labeling or prescriptions would be required, using pharmaceutical guidelines. The bill was endorsed by the National Association of Retail Druggists, who had by this time split with the American Pharmaceutical Association because of the latter’s ignorance of, or at least insensitivity to, the business problems of pharmacists. The American Pharmaceutical Association opposed the legislation, because they believed that consumers had a
right to self-medicate, under the condition that pharmacists could provide safe and effective drugs. They also seemed concerned that the bill further limited the ability of pharmacists to use their skills [Herzog, 1955].

Peter Temin [1979] views the adoption of the 1938 law requiring prescriptions as the first manifestation of the growing official, paternalistic belief that consumers were unable to understand or use drugs without official supervision, regardless of how carefully they were labeled. He concludes by confessing that he finds this attitude inexplicable, especially since he can detect no changes in the technology of drug production, and no significant increase in the number of new and complex drugs. However, as the preceding review of the successive revisions of the USP shows, there was a fundamental change in the technology of drug production and distribution, especially as it affected pharmacists and physicians as producers and distributors. These well-recognized and fundamental changes were more subtle than “simple” cases of rapid technological change because of their cumulative effects on professional institutions, but no less revolutionary.

In this case, the problem was that while pill-stamping machines still looked more or less the same as they had in preceding decades, the medicines rolling out of them were very different. Even before 1938, Temin’s watershed for the creation of new drugs, many important scientific advances had occurred. Organic chemistry had reached the point where pharmaceutical researchers could create drugs that interacted with biological material: synthesis of molecules led to aspirin production in 1880 and to barbiturates in 1903. Five thousand sulfanomides alone were synthesized and studied beginning in the 1930s, when the drug was extracted from a dye called prontosil rubrum. These techniques are clearly advances in the technology of producing and especially developing drugs, and they resulted in an increase in the number of new drugs introduced.

Whatever the intentions and origins of the 1938 law, the result was a change in the ability of both physicians and pharmacists to accomplish their traditional tasks in their traditional way. Physicians became the consumers’ agents in the procurement of drugs. Pharmacists still pursued their goal of certifying the purity, efficacy, and safety of drugs, but in a new way. Compounding became increasingly rare, but
repackaging became more common. Pharmacists assumed primary responsibility for maintaining a large inventory of unadulterated and properly stored drugs, a task that had been theirs for centuries. In addition, their core responsibility took on additional dimensions. They could best accomplish the task of quality certification by standing between physicians and patients. For example, they checked prescriptions for errors, monitored drugs for interactions, and supervised ongoing drug use through careful record keeping and contact with customers through refilling.

The core competency of pharmacy has remained unchanged throughout American history, although the boundaries of the profession, its sub-institutions, and its strategies have evolved considerably. This process continues today. In April 1987, Congressman Ron Wyden introduced legislation to amend the federal Food and Drug Act to prohibit physicians from tilling the prescriptions that they have written for their patients [HR 2168]. This resurgence of physician dispensing is another episode in the long history of boundary disputes between pharmacists and physicians. Pharmacists, too, have been testing the borders. Since May 1986, pharmacists in Florida have been allowed to prescribe certain classes of prescription-only drugs without consultation with or permission from a physician. Complicating the issue further are laws in some states that allow nurses to write prescriptions and, finally, a concerted effort on the part of psychologists to be able to prescribe medications as well.

During the last few decades, pharmacy began to lose important parts of its professional status. The changing technology of drug research, development, and production, as well as changes in prescription drug laws and distribution sites, reduced the scope of activities and the range of authority available to individual pharmacists. At the same time, their training has become more rigorous; they are clearly overqualified for the simple routines of dispensing medicines.

According to surveys, the public continued to rank pharmacy as the most respected profession. Building on this external reputation base and their specialized capabilities, pharmacists devised a new and controversial strategy that appears to be succeeding in re-invigorating their network. Most colleges now award a non-thesis doctorate, called the Pharm.D., as the basic pharmacy degree. This allows pharmacists
to interact more equally with Ph.D. chemists and biologists as well as with physicians in both hospital and community settings. As a result, and as the theory predicts, these networks are welcoming pharmacy into their midst because pharmacists now possess knowledge that these other networks want to internalize in a form that they can use.

This is an excellent example of how capabilities help explain professionalization and deprofessionalization. Pharmacists have been reluctant to give up tasks like compounding and research even when technological and legal imperatives make it clear that they would have little financial incentive to maintain them. The strategy behind this reluctance has been explained in a somewhat different context by Nelson [1991]. He explains that routines should be maintained and practiced for many reasons. One is that sometimes the information acquired in performing routines is more valuable than the any direct payoff from the routine. For example, routines often generate innovations that benefit other capabilities. Another is that abandoned routines may play a part in the maintenance of other capabilities. Finally, routines, once abandoned, may become impossible to recover. Individual practitioners may lose the requisite skills very quickly, and the network may lose access to necessary complementary assets. If, as in drug development, technology radically alters the location of production, network members who did not keep up with the advances would quickly find themselves out of the feedback loop. We are therefore not surprised to discover that pharmacy-level drug compounding is having a small, and very controversial, resurgence; that pharmacy schools are making the curriculum more rigorous; or that pharmacists are advocating for themselves a larger role in front-line health care.

Conclusions

As we have seen, shared routines depend on shared competences and capabilities. Sharing occurs through mechanisms designed to facilitate collecting, evaluating, and disseminating information; many of these sub-institutions are built into the process of producing and exchanging the products and services of the network. In professional networks, each practitioner processes a constant flow of information
while engaging in shared routines implemented in the context of independent practice. Independent practice exposes them to a broad but, because of their own incentives and capabilities, unique set of experiences in the larger market, both with clients and competitors. This information is likely to be idiosyncratic: too abstract to be internalized by a hierarchical organization, too tacit to be bundled and sold in an external market. The minute doses of exposure provide incremental additions to knowledge that, even if appropriable, would be difficult to put a value on ex ante.

The advantage of networks is that they aggregate, transform, and disseminate tacit knowledge without appropriating it or organizing it internally. Sharing knowledge is cheaper than trying to enforce contracts for nonappropriable and noncontractible assets. So trading information through formal and informal means is, at the outset, smart competitive strategy, not anticompetitive collusion.

Professional networks evolve capabilities that allow them to coordinate production and exchange activities that markets and firms are unable to organize. Over time, however, the capabilities of other institutions change as well. In principle, then, markets and firms can begin to replace professional networks. Pharmacists faced precisely this kind of competition from other organizational forms. Changes in technology and drug development allowed pharmaceutical companies to compete for drug-manufacturing routines. Markets, too, improved and became more stable, partly because of changes in law and advancements in information flows.

Professionals remain reluctant to abandon routines traditionally associated with their boundaries, however. This is partly because they have established their professional boundaries on the basis of reputation and shared routines, both of which can change only slowly. Even when it becomes feasible for a market to provide an isolated professional routine, professional networks will be reluctant to allow them to do so. One reason is that tacit, nonappropriable knowledge is often embedded in seemingly straightforward technologies. Allowing the firm and market free use of the technology undermines the ability of the profession to control the quality of the technologies of underlying routines, which together are crucial to maintaining vital network capabilities.
Pharmacy has faced, and continues to face, competition from drug companies, which have a clear advantage over pharmacists in large-scale drug research and production. These firms and the market have found it possible to usurp some, but not all, of the relevant production routines.

One of the reasons that pharmacy has not been absorbed is that it is a network-based rather than a firm-based institution. It is harder for network institutions to disappear from the economy than it is for firms, which can simply dissolve if the bottom line turns red. In contrast, the network exists as long as there are any shared routines with the ability to generate rents. Pharmacy held on, with front-line pharmacists bearing the brunt of the painful adjustment, until their network affiliates—pharmacists in the academy—developed a new strategy and convinced pharmacists in other subspecialties to change their strategies. The Pharm.D. places pharmacists squarely in the academic network, which has long been an excellent base from which to build a knowledge base with marketable capabilities. Placed in the context of the history of pharmacy, the current debates about the role of pharmacists in providing health care is another episode in a long battle over the boundaries between competing institutional forms.

References


