Diversification, Industry Structure, and Firm Strategy: An Organizational Economics Perspective

Peter G. Klein
Division of Applied Social Sciences
University of Missouri
and
Department of Strategy and Management
Norwegian School of Economics and Business
pklein@missouri.edu

Lasse B. Lien
Department of Strategy and Management
Norwegian School of Economics and Business
Lasse.Lien@nhh.no

This version: 5 March 2009

Prepared for Jackson A. Nickerson and Brian S. Silverman, eds., Economic Institutions of Strategy, vol 26 of Advances in Strategic Management, forthcoming 2009. We thank Nicolai Foss, Marc Saidenberg, Kathrin Zoeller, and the editors for conversations on this material and Mario Mondelli for research assistance.
1. Introduction

Ronald Coase’s landmark 1937 article, “The Nature of the Firm,” framed the study of organizational economics for decades. Coase asked three fundamental questions: Why do firms exist? What determines their boundaries? How should firms be organized internally? To answer the first question, Coase famously appealed to “the costs of using the price mechanism,” what we now call transaction costs or contracting costs, a concept that blossomed in the 1970s and 1980s into an elaborate theory of why firms exist (Alchian and Demsetz, 1972; Williamson, 1975, 1979, 1985; Klein, Crawford, and Alchian, 1978; Grossman and Hart, 1986). The second question has generated a huge literature in industrial economics, strategy, corporate finance, and organization theory. “Why,” as Coase (1937: 393-94) put it, “does the entrepreneur not organize one less transaction or one more?” In Williamson’s (1996: 150) words, “Why can’t a large firm do everything that a collection of small firms can do and more?” As Coase recognized in 1937, the transaction-cost advantages of internal organization are not unlimited, and firms have a finite “optimum” size and shape. Describing these limits in detail has proved challenging, however.¹

The chapter by Ramos and Shaver in this volume focuses on the distribution of a firm’s activities across geographic space. We focus here on the firm’s activities in the product space, reviewing, critiquing, and extending the literature in organizational economics, strategy, and corporate finance on diversification, asking what determines the optimal boundary of the firm across industries and how these boundary decisions influence industry structure. Below, we first examine the implications of transaction cost economics (TCE) or diversification decisions. TCE is essentially a theory about the costs of contracting, and TCE focuses on the firm’s choice to

¹ For a sample of approaches to understanding the limits to the firm, see Arrow (1974), Williamson (1985: chapter 6), and Klein (1996). On the internal organization of the firm – Coase’s third question – see Argyres’s chapter in this volume. The economics literature on internal organization has tended to draw primarily on agency theory, not transaction cost economics.
diversify into a new industry rather than contract out any assets that are valuable in that industry. While TCE does not predict much about the specific industries into which a firm will diversify, it can be combined with other approaches, such as the resource-based and capabilities views, that describe which assets are useful where.

As we note below, the transaction-cost rationale for unrelated diversification is different from the argument for related diversification. The essence of this argument is that unrelated diversification can be efficient when internal markets can allocate resources—financial capital in particular—better than external markets. We review this argument as it emerged in the transaction cost literature in the 1970s and 1980s and, more recently, theoretical and empirical literature in industrial organization and corporate finance. We move on to discuss how diversification decisions, both related and unrelated, affect industry structure and industry evolution. Here, the stylized facts suggest that diversifying firms have a crucial impact on industry evolution because they are larger than average at entry, grow faster than average, and exit less often than the average firm. We conclude with thoughts on unresolved issues problems and suggest what kinds of research we think are most likely to be fruitful.

2. Why do firms diversify?

The typical firm of an undergraduate economics text is a specialized production process that converts particular combinations of inputs into output. The existence of the firm is given, it produces a single product, and the manager’s task is to maximize profits. Naturally, this entity—Alfred Marshall’s “representative firm”—bears little resemblance to real business firms, which typically produce more than one product, are often vertically integrated, and have complex ownership and governance structures. The transaction cost approach to the firm has focused primarily on the question of vertical integration, or the make-or-buy decision, but transaction
costs also play an important role in determining the distribution of the firm’s activities over industries.

Related diversification\(^2\)

The relatedness hypothesis loosely claims that multi-business firms holding portfolios of similar (related) businesses might obtain efficiency advantages unavailable to non-diversified firms or firms with unrelated portfolios. This immediately raises two questions. What are the relevant kinds of similarity? And under what circumstances do such similarities give related portfolios efficiency advantages? At minimum, it seems reasonable that “relevant similarity” must imply that resources in one industry are substitutes\(^3\) for, or complements to, resources in another industry. If neither is the case—either statically or dynamically—it is hard to make economic sense of relatedness. However, while either substitutability or complementarity is necessary for relatedness to provide efficiency advantages, it is not sufficient.

To see this, imagine first the classic situation involving resources that are substitutes across industries (i.e. economies of scope). Suppose a resource in industry A is a perfect substitute for a resource in industry B, meaning that such a resource if developed in industry A can be used in industry B with no loss of productivity. Under the standard economic assumption of perfectly divisible resources, this perfect substitutability provides no advantage to a related diversifier active in both A and B. In contrast, if resources are not perfectly divisible (aka: “lumpy”), then single-business firms or unrelated diversifiers will be left with costly excess capacity, which related diversification can eliminate (Willig, 1979). Penrose (1959) was one of the first writers to relax the assumption of perfect divisibility. She noted that excess capacity arises not only

---

\(^2\) This section draws on Lien and Klein (2006, 2009).

\(^3\) By resources being substitutes across industries we mean that a resource developed in industry A can be deployed in industry B with little or no loss in productivity.
because some resources are inherently indivisible (e.g., half a truck is not half as valuable as a truck), but also because of learning – with accumulated production, new resources are generated, and excess capacity in existing resources is created. These learning effects, combined with resource indivisibilities, suggest that related diversification can improve performance.

However, as Teece (1980, 1982) points out, while the existence of such indivisibilities explains joint production, it does not explain why joint production must be organized within a single firm. If the excess capacity created by indivisibilities can be traded in well functioning markets, single-business firms and unrelated diversifiers can simply sell or rent out their excess capacity, or buy the capacity they need from others. In other words, absent transactional difficulties, two separate firms could simply contract to share the inputs, facilities, or whatever accounts for the relevant scope economies. If they do not, it must be because the costs of writing or enforcing such a contract—due to information and monitoring costs, à la Alchian and Demsetz (1972), or some other appropriability hazard—are greater than the benefits from joint production. Whether the firms will integrate thus depends on the comparative costs and benefits of contracting, not on the underlying production technology. Indeed, if contracting costs are low, the related diversifier may actually compete at a disadvantage relative to the single-business firm, because the diversified firm faces the additional bureaucratic costs of low-powered incentives, increased complexity, and so on (Williamson, 1985).

Another potential source of efficiency gains is not resource substitutability, but resource complementarity (Teece et. al, 1994; Christensen and Foss, 1997; Foss and Christensen, 2001). Complementarities exist when the value of resources in one industry increases due to investment in another industry, or when decisions about resource use in one industry affect similar decisions in another. These positive spillovers create a quantitative and qualitative coordination problem
which may be best managed within a diversified firm (Richardson, 1972; Milgrom and Roberts 1992). Of course, the existence of complementarities does not itself dictate integration. All supply chains involve some form of complementarities, but only some are integrated. Firm diversification is needed to exploit complementarities only if transaction costs prevent specialized firms from realizing these benefits through contract. For resource complementarities, the key source of contractual hazards is not indivisibilities, but the inability to specify contingencies, difficulties in verifying actions to third parties (such as courts), bounded rationality, or some other source of contractual incompleteness. The literature on complementarities as a motive for related diversification is regrettably thin on compared to the literature on substitutability and indivisibility; as we note below, the analysis of complementarities is a fruitful area for future research.

Note that while other approaches to the firm, such as the resource-based and capabilities perspectives, also emphasize the potential uses of resources across industries, and the value of combining specialized resources in particular combinations, they tend to abstract away from contractual hazards. For this reason, the resource-based and capabilities perspectives are not theories of the firm per se, in the sense described above; they deal with the allocation of specific factors to specific activities, not with the boundary of the firm as a legal entity. The resource-based view, for example, focuses on the returns to factors, individually or jointly, but does not explain how joint gains are shared across factor owners, or how the sharing mechanism is implemented and governed. For this we need a contractual explanation.

*Unrelated diversification*

The arguments presented so far explain the decision to exploit resources that are valuable across industries. A central prediction of these literatures is that related diversification should
outperform unrelated, or conglomerate diversification. And yet, the US conglomerates that arose in the 1960s did not, despite the restructuring of the 1980s, disappear from the corporate scene. Rumelt (1982: 361) reports that the percentage of Fortune 500 firms classified as “single business” fell from 42.0 in 1949 to 22.8 in 1959, and again to 14.4 in 1974, while the percentage of “unrelated business” firms rose from 4.1 in 1949, to 7.3 in 1959, to 20.7 by 1974. Servaes (1996), using SIC codes to measure diversification, finds a similar pattern throughout this period. Among firms making acquisitions, the trend is even stronger: pure conglomerate or unrelated-business mergers, as defined by the FTC, jumped from 3.2 percent of all mergers in 1948-53 to 15.9 percent in 1956-63, to 33.2 percent in 1963-72, and then to 49.2 percent in 1973-77 (Federal Trade Commission, 1981).

Moreover, despite evidence of de-diversification or refocus during the 1980s (Lichtenberg, 1992; Liebeskind and Opler, 1995; Comment and Jarrell, 1995), major U.S. corporations continue to be diversified. Montgomery (1994) reports that for each of the years 1985, 1989, and 1992, over two-thirds of the Fortune 500 companies were active in at least five distinct lines of business (defined by 4-digit SIC codes). As she reminds us, “While the popular press and some researchers have highlighted recent divestiture activity among [the largest U.S.] firms, claiming a ‘return to the core,’ some changes at the margin must not obscure the fact that these firms remain remarkably diversified” (Montgomery, 1994: 163). Baldwin et al. (2000) estimate that 71 percent of corporate diversification among Canadian companies occurs across two-digit SIC codes. In the developing world, conglomerates are even more important, accounting for a large share of economic activities in countries like India and Korea (Khanna and Palepu, 1999, 2000). On the whole, the evidence suggests that appropriately organized conglomerates can be efficient (Klein, 2001; Stein, 2003).
Arguments about resource substitutability and complementarity do not apply to unrelated diversification. Can unrelated diversification be efficient, or is it simply a manifestation of agency costs, a form of empire building, or a response to antitrust restrictions on horizontal expansion? Williamson (1975: 155-75) offers one efficiency explanation for the multi-industry firm, an explanation that focuses on intra-firm capital allocation. In his theory, the diversified firm is best understood as an alternative resource-allocation mechanism. Capital markets act to allocate resources between single-product firms. In the diversified, multidivisional firm, by contrast, resources are allocated via an internal capital market: funds are distributed among profit-center divisions by the central headquarters of the firm (HQ). This miniature capital market replicates the allocative and disciplinary roles of the financial markets, ideally shifting resources toward more profitable activities.4

According to the internal capital markets hypothesis, diversified institutions arise when imperfections in the external capital market permit internal management to allocate and manage funds more efficiently than the external capital market. These efficiencies may come from several sources. First, HQ typically has access to information unavailable to external parties, which it extracts through its own internal auditing and reporting procedures (Williamson, 1975: 145-47). Second, managers inside the firm may also be more willing to reveal information to HQ than to outsiders, since revealing the same information to the capital market would also reveal it to rivals, potentially hurting the firm's competitive position. Third, HQ can intervene selectively, making marginal changes to divisional operating procedures, whereas the external market can

---

4 A related literature looks at internal labor markets, focusing primarily on firms’ choices to rotate individuals among divisions and departments and to promote from within, rather than hire for top positions from outside (see Lazear and Oyer, 2004, and Waldman, 2007, for overviews). This literature has tended to focus on promotion paths and human-resource practices, rather than the internal structure and diversification level of the firm, however. For a related approach relating human-resource issues to firm scope see Nickerson and Zenger, 2008).
discipline a division only by raising or lowering the share price of the entire firm. Fourth, HQ has residual rights of control that providers of outside finance do not have, making it easier to redeploy the assets of poorly performing divisions (Gertner, Scharfstein, and Stein, 1994). More generally, these control rights allow HQ to add value by engaging in “winner picking” among competing projects when credit to the firm as a whole is constrained (Stein, 1997). Fifth, the internal capital market may react more “rationally” to new information: those who dispense the funds need only take into account their own expectations about the returns to a particular investment, and not their expectations about other investors’ expectations. Hence there would be no speculative bubbles or waves.

Bhide (1990) uses the internal-capital-markets framework to explain both the 1960s and 1980s merger waves, regarding these developments as responses to changes in the relative efficiencies of internal and external finance. For instance, the re-specialization or refocus of the 1980s can be explained as a consequence of the rise of takeovers by tender offer rather than by proxy contest, the emergence of new financial techniques and instruments like leveraged buyouts and high-yield bonds, and the appearance of takeover and breakup specialists like Kohlberg Kravis Roberts which themselves performed many functions of the conglomerate HQ (Williamson, 1992). Furthermore, the emergence of the conglomerate in the 1960s can itself be traced to the emergence of the multidivisional corporation. Because the multidivisional structure treats business units as semi-independent profit centers, it is much easier for a multidivisional corporation to expand via acquisition than it is for the older unitary structure. New acquisitions can be integrated smoothly when they can preserve much of their internal structure and retain control over day-to-day operations. In this sense, the conglomerate could emerge only after the

---

5 Of course, large blockholders, such as institutional investors in the Anglo-American system, or banks under universal banking, can also influence intra-firm decision-making.
innovation of the multidivisional firm had diffused widely throughout the corporate sector. Likewise, internal capital markets will tend to add value where the external capital markets are hampered by regulation, inefficient legal structures, and other institutional impediments, explaining the prevalence of diversified business groups in emerging markets (Khanna and Palepu, 1999, 2000).6

If unrelated diversification is primarily a response to internal-capital-market advantages, rather than a manifestation of agency problems, then unrelated diversifiers should perform better than specialized firms, particularly when external capital markets are weak. And yet, the evidence on the value of unrelated diversification is mixed. Consider, for example, the “diversification-discount” literature in empirical corporate finance. Early studies by Lang and Stulz (1994), Berger and Ofek (1995), Servaes (1996), and Rajan, Servaes, and Zingales (2000) found that diversified firms were valued at a discount relative to more specialized firms in the 1980s and early 1990s. Lang and Stulz (1994), for example, find an average industry-adjusted discount—the difference between a diversified firm’s $q$ and its pure-play $q$—ranging from $-0.35$ for two-segment firms to $-0.49$ for five-or-more-segment firms. Bhagat, Shleifer, and Vishny (1990) and Comment and Jarrell (1995) document positive stock-price reactions to refocusing announcements.7 The apparent poor relative performance of internal capital markets has been explained in terms of rent seeking by divisional managers (Scharfstein and Stein, 2000), bargaining problems within the firm (Rajan, Servaes, and Zingales, 2000) or bureaucratic rigidity (Shin and Stulz, 1998). For these reasons, it is argued, corporate managers fail to allocate investment resources to their highest-valued uses, both in the short and long term.

---

6 Another possibility is that internal capital markets add value when top managers have particular political skills, skills that can be leveraged widely across industries—particularly likely in developing economies.

7 Matsusaka (1993), Hubbard and Palia (1999), and Klein (2001) argue, by contrast, that diversification may have created value during the 1960s and early 1970s by creating efficient internal capital markets.
On the other hand, as pointed out by Campa and Kedia (2002), Graham, Lemmon, and Wolf (2002), Chevalier (2004), and Villalonga (2004), diversified firms may trade at a discount not because diversification destroys value, but because undervalued firms tend to diversify. Diversification is endogenous and the same factors that cause firms to be undervalued may also cause them to diversify. Campa and Kedia (2002), for example, show that correcting for selection bias using panel data and fixed effects and two-stage selection models substantially reduces the observed discount (and can even turn it into a premium).8

Seemingly absent from this literature, however, is the role of organizational structure. All diversified firms are not alike. Some are tightly integrated, with strong central management; others are loosely structured, highly decentralized holding companies. Sanzhar (2004) and Klein and Saidenberg (2009) show that many of the effects of diversification described in the literature are also visible in samples of multi-unit firms from the same industry, suggesting that studies of diversification tend to conflate the effects of diversification and organizational complexity. For this reason, organizational scholars may have much to contribute to the debate about the value of unrelated diversification.

A modest literature emerged in the late 1970s attempting to classify firms according to their organizational structure and see how organizational structure affects profitability and market value. The inspiration was Chandler’s work on the emergence of the multidivisional, or “M-form,” corporation. The M-form corporation is organized into divisions by geographic area or product line. These divisions are typically profit centers with their own functional subunits. The firm’s day-to-day operations are decentralized to the divisional level, while long-term strategic

---

8 There are also important data and measurement problems. Most studies use Tobin’s q to measure divisional investment opportunities, but it is marginal q—which may not be closely correlated with observable q—that drives investment (Whited, 2001). SIC codes are also typically used to measure diversification and to identify industries, but the SIC system contains significant errors (Kahle and Walkling, 1996) and cannot reliably distinguish between related and unrelated activities (Teece, Dosi, Rumelt, and Winter, 1994; Klein and Lien, 2008).
planning is centralized at the corporate office. Williamson’s (1975: 150) “M-form” hypothesis stated that the M-form structure is generally superior to the older, unitary (U-form) structure, as well as overly decentralized firms organized as mere holding companies (H-form). Inspired by the M-form hypothesis, attempts were made to classify firms into these categories and see if one type outperforms the others. Argyres’s chapter in this volume provides details on this literature. In short, the empirical evidence on the performance effects of internal organization is mixed, though qualitative fieldwork offers the potential for further insight.

A newer (and more successful, by modern standards) literature relates indirect, but observable, measures of internal organization to firm performance and behavior, using large samples of firms. Examples include the empirical literatures on related and unrelated diversification discussed earlier in the chapter. One approach is to assume that organizational form is correlated with observable characteristics such as the number of industry segments, the distribution of activities across industries, or some measure of relatedness. Another approach is to infer organizational form from past performance, such as prior acquisitions (Hubbard and Palia, 1999; Klein, 2001). Other papers look directly at resource transfers between a firm’s divisions to see how such transfers are organized and governed (Shin and Stulz, 1998; Rajan, Servaes, and Zingales, 2000).

The main advantage of this approach, over the older approach used in the M-form literature, is that it is more straightforward to implement. The results do not rely on the researcher’s discretion in assigning firms to categories. The tradeoff is that the newer literature uses much cruder measures of organizational form, and hence ignores important differences among firms that appear superficially similar (e.g., firms with the same number of divisions). Of course, it is not clear to what extent variations in diversification are correlated with variations in
organizational structure. Indeed, the literatures in strategy and empirical corporate finance probably conflate the two (Klein and Saidenberg, 2009). But it is not clear how organizational structure can be measured more cleanly in a large sample of firms. Other proxies include segment or subsidiary counts within a single industry (Sanzhar, 2004; Klein and Saidenberg, 2009), the ratio of administrative staff to total employees (Zhang, 2005), the number of positions reporting directly to the CEO (Rajan and Wulf, 2006), and the average number of management levels between the CEO and division managers (Rajan and Wulf, 2006). All have advantages and disadvantages.

Summary

As the foregoing discussion demonstrates, there is a robust literature on how, where, and when firms will diversify, either into related or unrelated industries. Theory and evidence suggest that firms diversify when they have valuable and difficult-to-imitate resources that are valuable across industries, or are complementary to resources in other industries, and where these gains cannot be realized by contracting among independent firms. Firms also diversify when they have effective internal resource-allocation mechanisms, particularly when background institutions and external capital markets are undeveloped. And yet, many important questions remain.

First, as noted above, the literature on complementarities is thinner than the literature on substitutability. There is growing interest among economists in organizational complementarities (Milgrom and Roberts, 1992, 1995; Ichniowski, Shaw, and Prennushi, 1997; Bresnahan, Brynjolfsson, and Hitt, 2002; James, Klein, and Sykuta, 2008), but these ideas have not been widely applied to questions of firm scope. Just as organizational practices, governance, and
ownership tend to cluster in particular combinations, industry activities may tend to cluster, in
t ways that cannot be managed effectively across independent firms.

Another understudied area is divestment (Mahoney and Pandian, 1992). Are exit decisions
driven by the same factors that drive entry decisions? Lien and Klein (2009) suggest that
relatedness, as measured by observed patterns of industry combinations, drives both entry and
exit. Firms are more likely to enter industries in which their resources can add value, and less
likely to exit these related industries. The evolution of industry structure thus reflect changes in
patterns of relatedness, changes that are driven by prior changes in technology, competition,
regulation, and the like. Klein, Klein, and Lien (2009) show that divestitures of previously
acquired assets are not, in general, predictable ex ante, implying that exits may reflect an
efficient form of experimentation, rather than the reversal of previously inefficient decisions. In
general, however, the literature has focused much more on entry than exit.

Finally, the relationship between the institutional environment and diversification strategy
remains underdeveloped. As Bhide (1990), (Khanna and Palepu, 1999, 2000), and others have
argued, “optimal” diversification depends on the legal, political, regulatory environments as
much as competitive conditions and the state of technology. Comparative work on diversification
across institutional contexts, and the political economy of diversification, is sorely needed.9

---

9 One instance of government policy that may have affected the conglomerate movement of the 1960s, although it
has been mostly ignored in the literature, is the Vietnam War. Several of the larger, highly visible 1960s
conglomerates like ITT, Litton, and Gulf & Western sought growth by expansion from their original businesses into
the most glamorous, rapidly growing areas of the time, namely aerospace, navigation, defense-related electronics,
and the like all industries into which the federal government was pumping billions of dollars. Gulf & Western, LTV,
Litton, and Textron all had significant Defense Department contracts; indeed, several highly visible conglomerates
first diversified in the early 1960s precisely to get into the high-growth industries of the time like aerospace,
navigational electronics, shipping, and other high-technology areas, all defense related.

A US Antitrust Subcommittee’s report (1971, p. 360) on conglomerates, noting that in 1969 Litton was number
21 on the Defense Department’s list of the 100 largest military prime contractors, described the company as follows:
“Sophisticated in the interrelationships between the government and private sectors of commercial activities, Litton
has sought to apply technological advances, novel management techniques, and system concepts developed in
government business to an expanding segment of the commercial economy.” These “system concepts” were the
3. What does diversification imply for industry structure and industry evolution?

How can organizational economics inform our understanding of industry evolution? At the extreme one might say that industry structure is important because organizational economics is relevant. In other words, with no transaction costs, the Coase Theorem implies that industry structure is unrelated to efficiency and profit maximization (Coase, 1960). Firms within an industry—and their customers and suppliers—can realize all possible efficiency gains through contracting and side payments. Exploiting profit opportunities would not require ownership changes or changes in firm size, as they could be realized by contracting between independent parties, regardless of whether the industry in question is a monopoly, duopoly, or fragmented. Industry structure would accordingly be indeterminate (Furubotn, 1991), since no particular industry structure is better than another. Industry structure would also be uninteresting, because it would not impact profit or efficiency. But transaction costs are pervasive, and therefore industry structure is interesting.

Industry structure evolves as a function of three key processes; entry, exit, and market-share dynamics. Organizational economics is relevant for the evolution of industry structure to the degree that it is relevant for understanding these three processes. Each of these three broad processes can further be subdivided into finer categories. Entry for example, occurs either through diversification by established firms or through the formation of new firms. Exit may occur via bankruptcy, closure or divesture. Market share changes happen organically or via the market for corporate control. To understand industry dynamics, then, we must focus on entry,
exit, growth, and decline by existing and new firms. This is a huge topic; we restrict our attention here to the link between diversification and industry structure.

First, the stylized facts. Diversifying entrants enter at a bigger scale and are more likely to survive and grow than de novo entrants (Baldwin, 1995; Dunne, Roberts, & Samuelson, 1989; Klepper & Simons, 2000; Siegfried & Evans, 1994; Geroski, 1995; Sharma & Kesner, 1996). Consequently, diversifying entrants pose a bigger threat, in increasing rivalry and challenging incumbents’ market share, than de novo entrants. An important question is therefore how “vulnerable” a given industry is to diversifying entry.

Another relevant empirical finding is that diversification patterns are not random. While the performance effect of related diversification is controversial, the broad tendency of firms to diversify in a related manner is not (Lemelin, 1982; Chatterjee and Wernerfelt, 1991; Montgomery and Hariharan, 1991; Teece et al. 1994; Silverman, 1999). Indeed, in terms of predicting which industries a diversifying entrant chooses to enter, relatedness seems by far the most important determinant (Silverman, 1999; Sharma; 1998; Lien and Klein, 2009).

Klepper and Simons’s (2000) account of the US television-manufacturing industry, and how it was shaped and ultimately dominated by diversifying entrants from radio manufacturing, nicely ties these observations together. From an industry perspective we may also note that some industries are not very closely related to any other industries, while still other industries are closely related to several. By means of an analogy; some industries have several close neighbors, and others do not (Santalo & Becerra, 2008; Lien & Foss, 2009). Also, some industries are related to fragmented industries, while other industries are related to concentrated industries (Scott, 1993).
Conditions such as these are key determinants of how likely a given industry is to become the target of diversifying entry. All of them are inextricably linked to the notion of relatedness. But as argued above, the notion of relatedness is itself inextricably linked to transaction costs.

We now move on to focus on the consequences of diversification for industry structure.\(^{10}\) Note that within the received literature it has been more common to study the opposite, namely how industry structure affects diversification.\(^{11}\) For example, consider the debate on diversification and performance (beginning with Rumelt, 1974; Bettis, 1981; Christensen & Montgomery, 1981; Rumelt, 1982). In their critique of Rumelt (1974) Christensen and Montgomery (1981) argued that the relatedness/performance link in (an updated version of) Rumelt’s sample was strongly influenced by industry characteristics: Controlling for such characteristics largely eliminated Rumelt’s (1974) finding of a positive relatedness/performance link. However, this line of argument takes industry characteristics as exogenously given, whereas our point of departure is that industry structure is to a considerable extent an outcome of relatedness and diversification decisions.

One way to determine how related various industries are to each other is to consider how often a pair of industries are combined inside a firm, compared to what one would expect if diversification patterns were random (Teece et al., 1994; Lien and Klein, 2006, 2009). A pair of industries are related to the extent that this difference is positive, and unrelated to the extent that it is negative.\(^{12}\) Note also that such a survivor-based measure of relatedness will incorporate transaction costs considerations to the extent that these considerations are reflected in firms’ actual diversification decisions.

---

\(^{10}\) This section draws on Lien and Foss (2009).

\(^{11}\) For an important exception see Adner and Zemsky (2006). These authors develop a game-theoretic model in which relatedness impact diversification decisions, which in turn impacts industry structure.

\(^{12}\) Following Teece, et al. (1994), the difference between expected and actual combinations is calculated as a hypergeometric situation in order to remove the effect of the size of the industries in question.
This measure allows one to calculate the relatedness between a focal industry and all other industries in the economy, and subsequently the same can be repeated using each industry in the economy as the focal one. The pattern that emerges if this is done is that industries differ significantly in how closely related they are to their closest neighboring industries. For example, for each industry one may calculate the sum of the relatedness scores for the four closest related industries. The mean of this sum will then describe how close the average industry has its four closest neighboring industries. This is done in Table 1 below for a comprehensive dataset from the 1980s (Trinet). A striking observation is how much variation the table reveals. The standard deviation is nearly 50% of the mean in all four data years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Relatedness to the 4 closest industries (mean of sum)</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>82.8</td>
<td>39.4</td>
<td>10.7</td>
<td>273.2</td>
<td>856</td>
</tr>
<tr>
<td>1983</td>
<td>87.4</td>
<td>41.7</td>
<td>7.9</td>
<td>274.4</td>
<td>846</td>
</tr>
<tr>
<td>1985</td>
<td>85.6</td>
<td>40.2</td>
<td>6.8</td>
<td>257.3</td>
<td>848</td>
</tr>
<tr>
<td>1987</td>
<td>82.2</td>
<td>38.9</td>
<td>9.65</td>
<td>236.1</td>
<td>838</td>
</tr>
</tbody>
</table>

N = Number of industries included

Table 1 provides data on how industries differ in terms of how related the 4 closest related industries are. Put differently, it shows how industries differ in terms of how close the 4 closest neighboring industries are. Relatedness is calculated using a survivor-based measure of relatedness as detailed in the text above.

This suggests that some industries are closer to their neighboring industries than others. What we know less about are the consequences of this observation, and how changes in this distance
will affect industry structure. In the following we will suggest this as a potentially fruitful research area..

One obvious question one may ask is whether having the closest neighboring industries close will work to reduce concentration in the focal industry or increase it? Our reasoning suggests that it will probably depend on the level of concentration within those industries, for the following reasons; First, if the industries close to the focal industry are concentrated, there is a smaller pool of potential diversifying entrants (*ceteris paribus*). In other words, the threat of direct entry from such an industry is smaller, weakening an important mechanism that may otherwise contribute to reduce concentration. Second, concentrated neighboring industries are themselves likely to be difficult to enter, reducing the number of entrants that can enter the focal industry indirectly, that is, using neighboring industries as stepping stones. Third, high levels of economies of scope or positive spillovers between neighboring industries can create an entry barrier that is shared between the industries, facilitating concentration in both the focal industry and its neighboring industries.

In total, the implication is that concentration levels should correlate across related industries. Actually, there is empirical evidence to support this. Table 2 below shows the correlation between concentration (C4) in a focal industry and a summary measure of the concentration level in the four closest neighboring industries.

<table>
<thead>
<tr>
<th>Correlation with concentration level in the 4 closest industries</th>
<th>Significance (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>0.410</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 2 provides data on the correlation of concentration ratios (C4), between a given industry and its 4 closest related neighboring industries.

<table>
<thead>
<tr>
<th>Year</th>
<th>Correlation</th>
<th>p-value</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>0.371</td>
<td>0.000</td>
<td>846</td>
</tr>
<tr>
<td>1985</td>
<td>0.358</td>
<td>0.000</td>
<td>848</td>
</tr>
<tr>
<td>1987</td>
<td>0.435</td>
<td>0.000</td>
<td>838</td>
</tr>
</tbody>
</table>

As Table 2 reveals these correlations are quite strong (although they could admittedly appear for other reasons than those suggested here). It would seem that the data suggest that having your neighbors close is good (for allowing high concentration) if they are concentrated, but bad if they are fragmented. 13 It would also seem to suggest that any technological or other change that serves to fragment (concentrate) a neighboring industry will tend to fragment (concentrate) the focal industry industry.

What about changes in transaction costs? Assume that there is a gain from coordination across a pair of industries, but that achieving this coordination through market contracting for some reason becomes more costly. This could for example because of a weaker appropriability regime, or some other increase in contractual hazards. The analysis supplied here (and in the previous section) implies that this essentially amounts to an increase in relatedness between the two industries. This will increase the likelihood that firms from either of the industries will enter the other. At the same time it will decreases the likelihood that outside firms will enter either, since it creates a need to be active in both to be competitive in either.

13 We are not suggesting here that firms can freely choose the degree relatedness between industries. By and large we think this is exogenously given. However, in adapting it might be advantageous to know how industry structure in one industry is affected by changes in relatedness with others. The findings in Scott (1993) can be read to support this. Scott finds that an industry with high levels of multimarket contact with other industries has higher average ROA. However, the relationship only holds if these other industries are concentrated. If not, the relationship is negative.
Which effect will dominate? We don’t know, but one plausible conjecture is that a fragmented industry that experiences increasing transaction costs with a concentrated industry will experience increased concentration, while a concentrated industry experiencing increased transaction costs with a fragmented industry will experience fragmentation. Our reasoning here is that for in the former situation the threat from direct entry from the other industry is limited, while the opportunity to benefit from increased entry barriers against outside entrants is likely to be important, thus a net increase in the equilibrium concentration level is likely. For the latter type of situation, the opposite seems likely. The effect of decreases in transaction costs might follow the same logic but with all signs reversed.

In summary, we have argued that while many researchers in the strategy field have devoted considerable attention to how industry structure influences diversification decisions, we have much less knowledge about how diversification decisions influence industry structure. Diversification research has traditionally looked at resources and tried to determine in which industries those resource will be useful. But an alternate way to approach this is by looking at the industry as the unit of analysis. Which industries are susceptible to diversifying entry, and are there performance implications for industries that are susceptible to such entry?

Another approach would be to link changes in transaction costs to implications for industry structure and dynamics. If there is an increase in the hazards associated with contracting out excess resources for use in industry X, then firms in related industries will be more likely to diversify into industry X, thus leading to changes in structure, performance, and dynamics as proposed above. For resources that are substitutes, this might result from regulatory or legal changes that weaken appropriability regimes. For complementary resources, the cause might be technological shocks that increase the benefit of being in both industries. Indeed, it would be
particularly interesting to study whether a change in contractual hazards in industry Y affects industry structure and dynamics in industry X via the mechanisms noted above.

We also think there is more work to be done on the competitive interactions among entrants and incumbents. There is a large game-theoretic literature in industrial organization on entry (e.g., Gilbert, 1989), and the relationship between incumbents and potential and actual entrants is complex. The analysis above focuses on the advantages of diversifying entrants over incumbents, but incumbents may also have advantages (for example, the ability of single-industry incumbents to make credible threats of aggressive competition after entry, on the grounds that they cannot “retreat” to another industry). Incorporating these kinds of considerations into a theory of diversification and industry structure presents an exciting research opportunity.

4. Discussion, conclusions, and outstanding issues and problems

We hope this brief sketch illustrates the depth and variety of the existing literature while illustrating the many challenges that remain in developing a fuller understanding of diversification, industry structure, and firm strategy. Specifically, some parts of this literature—for instance, the relatedness-performance link—are fairly mature. A new study using cross-sectional data to relate some aspect of diversification strategy to firm performance is unlikely to be published in a top journal in strategy (or any field) unless it analyzes a new sample or institutional setting, uses a novel econometric technique or a particularly strong identification strategy, tests a new theoretical model, investigates novel hypotheses, and so on. Moreover, young scholars must recognize that the literature on diversification and industry structure spans several academic disciplines, including strategy, industrial economics, corporate finance, organizational and labor economics, and others. Some of these disciplines, particularly those
based in economics, have very high econometric standards, particularly regarding identification (see Angrist and Krueger, 2001, for discussion).

One potentially fruitful approach for strategy scholars, particularly those informed by TCE, is to provide qualitative, case-study evidence that complements the econometric results in the established literature. Consider, for example, Mayer and Argyres’s work on learning between contractual partners (Mayer and Argyres, 2004; Argyres and Mayer, 2007). TCE and the incomplete-contracting perspective have tended to focus on “optimal” contractual arrangements, and the relationship between these arrangements and the characteristics of the underlying transactions, assuming that the competitive selection process works to eliminate inefficient choices.14 Parties are modeled as far-sighted agents who anticipate potential hazards and contract around them. Mayer and Argyres (2004) examined a particular trading relationship over time and found, surprisingly, that the parties actually experienced many of the hazards TCE argues they should avoid. It was only through experimentation and learning that “optimal” contractual arrangements were discovered. The dynamics of this kind of relationship are difficult to pick up in large datasets (see Costinot, Oldensi, and Rauch, 2009, for one attempt).

Another important issue relates to comparative-statics versus evolutionary explanations for organizational form more generally. Can a single theory or theories explain both the optimal boundary of the firm across industries and the optimal means of changing that boundary? For example, one could take a comparative-statics approach in which appeals to scope economies, transaction costs, internal-capital-market efficiencies, resource substitutability and complementarity, and the like, explain optimal boundaries as a function of some exogenous characteristics, and argue that firm boundaries tend to change only when those characteristics

14 The usual justification is Alchian’s (1950) and Friedman’s (1953) arguments for profit maximization based on an efficient selection mechanism. See Lien and Klein (2009) for further discussion.
change, i.e., in response to exogenous shocks. Alternatively, one could posit one theory of optimal boundaries and a different theory of boundary changes (appealing to experimentation, error, learning, the selection mechanism, and so on). As noted above, the theory and practice of organizational adaptation has received less attention from TCE and strategy scholars than the theory of optimal boundaries (exceptions include Argyres and Liebeskind, 1999, 2002; Mayer and Argyres, 2004; Milgrom and Roberts, 1995; Argyres and Mayer, 2007). The economics literature on the diffusion of technological innovation (Hall and Khan, 2003) provides a useful analytical framework, but has not widely been adopted to the study of organizational innovation.15

A similar question relates to the application of strategy theories to questions about diversification and organization. Does the RBV or the internal-capital-markets approach or neoclassical economics explain the optimal scope of the firm's activities while a different theory, like TCE, explains the choice of contractual form? I.e., do theories of relatedness explain optimal scope independent of organizational form, or do they also explain whether the relevant efficiencies are best exploited via a wholly owned subsidiary, a partially owned subsidiary, an alliance, a joint venture, an informal network, etc.? As noted above, theories of resource substitutability and complementarity implicitly assume sufficient transaction costs in factor markets to prevent scope economies from being exploited fully through contract. However, the RBV literature has not devoted as much attention to the details of contractual form as the TCE (and law-and-economics) literatures on firm and industry structure.

15 Gibbons (2005) points out that Williamson’s writings offer two distinct theories of the firm, one the familiar asset specificity and holdup theory, also associated with Klein, Crawford, and Alchian (1978), the other an “adaptation” theory in which the main advantage of firm over market is the ability to facilitate coordinated adaptation. While concepts of adaptation and coordination feature prominently in Williamson (1991), they have not been picked up widely by strategy scholars, despite an obvious connection to theories of organizational change.
We have also tried to illustrate the variety of empirical approaches that have appeared in the literature. It is critical to identify, and understand, the strengths and weaknesses of large-sample, quantitative research on these questions compared to smaller-sample, more qualitative work. As noted above, research on the M-form has tended to rely on small samples and “deep” classifications of organizational form and diversification levels; even Rumelt’s (1974) quantitative approach, which proved highly influential in strategy work on relatedness, relies on subjective classifications. (The SIC system, in a sense, also relies to some degree on subjective classifications of industries.) While strategy scholars have employed both small- and large-sample empirical techniques, the industrial economics and corporate finance literatures have tended to favor large-sample econometric studies using panel data, instrumental variables, or some other means to address endogeneity, selection bias, and other forms of unobserved heterogeneity.

5. References


