The Tradeoff Between Firm Size and Diversity in the Pursuit of Technological Progress

Wesley M. Cohen
Steven Klepper

1. Introduction

Since the writings of Schumpeter, the role of firm size in promoting technical advance has preoccupied scholars of technological change. As international competitive pressures have intensified in recent years, the role of firm size has emerged as a central concern of policymakers and industrialists as well. In the United States, some argue that the small entrepreneurial firm is the primary vehicle through which new ideas are introduced into the marketplace and that the diversity of ideas and approaches flowing from small firms represents American manufacturing's key competitive advantage (e.g., Gilder, 1988; Rodgers, 1990; Shaffer, 1990). On the other side, it is argued that only large firms or consortia can command the resources necessary to field the large research efforts required to keep up with the large, often cooperatively organized research operations of Japan, and, more recently, Europe (e.g., Ferguson, 1988; Norris, 1983; Noyce, 1990). Over the course of his career, Schumpeter himself spanned the poles of this debate. In his earlier work, Schumpeter (1934) saw the small-scale entrepreneur as the key to capitalism's vitality. Later Schumpeter (1942) argued that the large scale enterprise was the principal engine of technological progress, although he feared that the bureaucratic character of the large modern corporation would undermine entrepreneurial initiative and eventually sap capitalism of its technological vitality.

Final version accepted on June 13, 1991

Carnegie-Mellon University
Department of Social & Decision Sciences
Pittsburgh, PA 15213-3890, U.S.A.

In recent years, this controversy has been dominated, with some exceptions (e.g., Jorde and Teece, 1990), by industrialists and policy analysts. Most of the arguments advanced by both sides are based upon specific industries, most notably the semiconductor and computer industries. These arguments have largely failed to galvanize economists. Not only have the arguments been casually developed, but there are strong counterarguments to each, leaving us with little a priori basis for favoring either position. For example, proponents of large scale enterprise often allude to scale economies and capital market imperfections to rationalize the need for large firms and consortia. Scale economies, however, bear only on the optimal size of the R&D effort and relate only indirectly to the optimal size of the firm. With regard to capital market imperfections, there is no consensus among economists as to the import of such imperfections for investment in general, much less R&D in particular. Those arguing in behalf of small size have claimed that small firms, unencumbered by bureaucracy, provide both the freedom and economic incentives that stimulate creativity and agility in response to economic opportunity. On the other hand, while large firms may be more bureaucratic, they provide a superior human and capital infrastructure to support innovative activity.

To the extent that economists have probed the relationship between firm size and innovative activity, the evidence for either side of the debate has also not been compelling. Indeed, the predominant finding from a voluminous literature on the subject is that in most industries, above a modest threshold firm size large firms are no more research intensive than smaller firms. This implies that consolidation of firms above the threshold would have no effect on total industry R&D

Small Business Economics **4**: 1—14, 1992.

expenditures, which has been widely interpreted to indicate that above a modest firm size neither small nor large firm size confers any advantage in R&D (e.g., Kamien and Schwartz, 1982; Scherer, 1980; Baldwin and Scott, 1987).

We propose that even if small and large firms are equally capable at R&D, there may nonetheless be important social advantages associated with *both* large and small firm size. We suggest that there is merit to both sides of the debate about the optimal firm size for R&D, but not for the reasons typically advanced. Our demonstration of the social advantages of small and large firm size proceeds in two stages. First, drawing heavily on prior work (Cohen and Klepper, 1990, 1991), we suggest that both firm size and technological diversity influence firm R&D expenditures within industries and we highlight evidence that supports this claim. In the second stage, we go beyond our earlier analysis, and, in a more speculative spirit, consider not only how firm size and diversity influence firm R&D expenditures, but also how they affect the technical advance generated by the expenditures. We argue there are virtues both to having a large number of small firms in an industry and to consolidating output in a few large firms.

The social advantages in our framework of an industry composed of numerous small firms rests on two notions suggested by Nelson's (1981) argument about the importance of competition and diversity for technological change. The first notion is that in the typical industry undergoing technological change, there are many productive ways of innovating. The other notion is that firms have different capabilities and perceptions which lead them to pursue different sets of approaches to innovation. In such a world, dividing up industry output over a greater number of small firms increases the chances that any given approach to innovation will be pursued, thereby increasing the diversity of technological efforts in the industry. While increasing the number of firms does not necessarily benefit individual firms in the industry, it promotes technical advance and, hence, benefits society by increasing the number of productive approaches to innovation that are collectively pursued in the industry. From this perspective, the source of the social advantage associated with small firm size is not smallness per se but the greater number of firms that small size implies given some industry demand.

Our argument about the social advantages of large firm size stems from the idea that large firms possess an advantage in appropriating the returns from innovation due, in part, to imperfections in the market for information (Arrow, 1962). This causes the returns to large firms from any given R&D effort to be greater than the returns to small firms, leading large firms to conduct more socially desirable R&D than smaller firms. Consequently, consolidating output over a smaller number of firms will result in the performance of more socially desirable R&D for each approach to innovation that is pursued. This benefit is realized independent of whether there are any economies of scale in the conduct of R&D or whether large firms have superior ability to finance R&D.

Given the social advantages associated with both large firm size and more numerous small firms, there will always be a tradeoff associated with changing the number of firms within an industry. Reducing the number of firms will increase the average firm size, which will increase the level of innovative effort applied to each approach to innovation that is pursued. This comes, however, at the cost of reducing the number of productive approaches to innovation that are collectively pursued in the industry. On the other hand, increasing the number of firms will increase the diversity of approaches that are pursued in the industry but reduce the level of effort for each approach. The optimal number of firms in each industry will depend on the relative importance of these two effects, which we argue will vary across industries.

We speculate that market forces alone are not likely to lead to the socially optimal average firm size and number of firms in each industry. For example, the eventual displacement of larger incumbent firms by smaller new firms will compromise the appropriability advantages of size. Alternatively, even when entry of new firms can enhance social welfare by increasing the diversity of approaches to innovation pursued within an industry, we argue that such entry will not necessarily be forthcoming. Although these opposing forces could conceivably balance out to yield some optimal number of firms, we have no reason to expect this to occur in the typical industry.

We discuss how the failure of private initiatives to lead to a socially optimal state will create a tension from a policy standpoint. Following recent trends, obstacles to cooperative R&D efforts within industries could be removed, which would encourage the development of larger scale innovating units. This will, however, reduce the diversity of technological approaches pursued. Alternatively, diversity could be stimulated by government subsidies to entrants. However, more entry will reduce average firm size, thereby compromising the advantages of size. We suggest that it may well be just this tension that, in part, has given rise to the recent policy debates. We argue that there are policies, however, that can to some degree preserve the benefits of size while at the same time maintain technological diversity.

The paper is organized as follows. In Section 2, we present our framework and briefly review the evidence that supports it. In Section 3, we suggest how larger firm size is tied to technical advance. In Section 4, we suggest how a larger number of firms may yield greater technological diversity and how that diversity, in turn, contributes to technical advance. In Section 5, we discuss the tradeoff implied by the existence of both an appropriability advantage of large firm size and a diversity-inducing advantage of having numerous small firms within an industry. In Section 6, we conclude by highlighting the limitations of our analysis and by suggesting issues for further research.

2. A model of R&D investment

In this section we consider how firms decide how much to spend on R&D. Our conception follows the model of R&D investment developed in Cohen and Klepper (1990, 1991). We discuss the basis for the model and provide a brief description of its implications. We refer the reader to Cohen and Klepper (1990, 1991) for the formal and more extensive development of the model.

The model is based on a distinctive conception of an industry. The industry is assumed to be subject to ongoing technological change. Following Nelson (1990) and others, it is assumed that technological advance can be achieved in many different ways. A distinction is commonly drawn, for example, between process and product innovations. Moreover, there are typically many different ways of improving technology within each of these categories. For example, a production process may be improved through automation and by increasing the scale of production. Similarly, a product may be improved along a range of dimensions. For example, personal computers can be made more user-friendly and they can be made to work faster. These different approaches to enhancing product and process performance are often additive in their effect on technical advance; they do not represent competing or mutually exclusive approaches to satisfying the same performance objective. A firm in principle could pursue all of the approaches or any subset of them.

Firms are assumed to be uncertain about the profitability of pursuing any given approach to innovation, both in terms of the cost of the innovation and its gross returns. For simplicity, we assume that all approaches are profitable to pursue, although the firms do not know that in advance. The uncertainty concerning the profitability of the various approaches to innovation generates differences in firms' expectations about which approaches are worth pursuing. These differences are assumed to reflect the differences in the sorts of expertise possessed by each firm. The differing expectations cause the firms to make different bets on which approaches to pursue.

The firms that make the best bets will experience the greatest rate of innovation and technical advance. It is assumed that eventually all innovations (but not the approaches themselves) will be copied, so that ultimately all benefits from innovation will be passed on to buyers in the form of lower prices. In the interim, however, firms that innovate most will reap economic profits. They will grow at the fastest rates over time, although growth will tend to be incremental, reflecting, among other things, the impact of uncertainty on investment and convex adjustment costs. Those firms that innovate the least will find the price of their product falling below their costs and will eventually exit. Over time the expertise required to evaluate and exploit approaches to innovation will change as the industry's technology evolves. Firms will be limited in their ability to change their expertise,[1] and in the long run new firms with more suitable expertise will replace the industry leaders.

The consequence of this process is that at a given moment in each industry there will be an array of firms of different sizes and capabilities with different views about which approaches to innovation are worth pursuing. The model focuses on how these differences condition firm decisions about which approaches to pursue and about how much to spend on R&D for each approach pursued. The number of firms and the output of each firm are taken as given, having been determined by the evolutionary process leading up to that moment in the industry's history. For simplicity, firms are also assumed to be price takers and to choose their R&D expenditures independent of the R&D expenditures of their rivals.

With regard to the approaches a firm decides to pursue, we assume this choice is determined exclusively by the expertise of the firm. A firm is assumed to pursue a particular approach if it possesses the expertise that would enable it to exploit and recognize the value of the approach. Consequently, in each industry the environment governing the endowment of firms with expertise will ultimately determine the approaches to innovation that are pursued by firms in the industry. We use a very simple model to represent this environment. First, we assume that in each industry the likelihood of a firm possessing the expertise that would lead it to pursue any given approach is independent of the likelihood of the firm possessing the expertise that would lead it to pursue any other approach. This implies that the probability of a firm pursuing any given approach is independent of the probability of it pursuing any other.[2] Second, we assume that in each industry the likelihood of a firm being endowed with the expertise that would lead it to pursue any given approach is independent of the size of the firm. This implies that the probability of a firm pursuing any given approach is independent of its size.

The choice about how much the firm should spend on each approach to innovation it pursues is conceived as follows. Each approach to innovation is assumed to be characterized by a range of projects with different marginal products, where the marginal product of a project is defined in terms of the technical advance (measured in standard units) generated from the project. Ordering the projects by their marginal products defines a diminishing marginal product schedule that relates the marginal product of R&D expenditures on an approach to the level of R&D expenditures on the approach, where it is assumed that the level of R&D expenditures is proportional to the number of projects pursued. All firms that pursue an approach are assumed to face the same marginal product of R&D schedule for the approach.

Although the marginal productivity schedules for all approaches are assumed identical across firms, the marginal revenue schedules associated with the approaches will differ across firms. The marginal revenue a firm earns from a project depends on the marginal product of the project and the level of output over which the innovations from the project are applied. It is assumed that firms will typically not sell their innovations in disembodied form either due to imperfections in the market for information (Arrow, 1962) or because the innovations are more valuable to the innovator than to other firms (Cohen and Klepper, 1990).[3] This implies that firms will appropriate the returns to their innovations primarily through their own output.[4] We also assume that the firm expects its growth due to innovation to be incremental and conditioned by its pre-innovation output level. This implies that the returns a firm earns from any given R&D project will be proportional to its sales prior to the R&D project, which we refer to as the firm's *ex ante* sales.

The assumptions governing the choice of approaches to innovation and the amount spent per approach form the basis for predictions about firm R&D expenditures in an industry. First, assuming no economies of scale in R&D, Cohen and Klepper (1991) demonstrate that the assumption that the firm's expected returns to innovation are scaled by its *ex ante* output implies that in each industry all firms pursuing a particular approach will spend an amount on R&D on the approach that is proportional to their *ex ante* sales. The simple intuition behind this result is that the greater the firm's sales then the larger the returns from any given R&D effort and thus the larger the optimal R&D expenditures of the firm. Furthermore, if the number of approaches to innovation pursued by the firm is uncorrelated with its sales, Cohen and Klepper (1990) show that the overall R&D intensity of the firm (i.e., the ratio of the firm's R&D expenditures on all approaches to its sales) will be independent of its *ex ante* sales. Alternatively expressed, in each industry total firm R&D expenditures will vary across firms proportionally with the *ex ante* sales of the firms.

Second, Cohen and Klepper (1991) demonstrate that if it is assumed that in each industry the marginal product schedule of each approach to innovation is the same and firms are equally likely to adopt each approach to innovation, then the number of approaches to innovation pursued by firms and their overall R&D intensity will be binomially distributed. Intuitively, given that firms will pursue different numbers of approaches, the adoption of each approach by a firm can be thought of as an independent Bernoulli trial. As a consequence, in each industry the number of approaches pursued by firms will be binomially distributed. It was noted above that firms' R&D expenditures on any given approach will be proportional to their sales. Coupled with the assumption that the marginal product schedules of the approaches do not differ, Cohen and Klepper (1991) show that this implies that in each industry firm R&D intensities will also be binomially distributed, with the firms that pursue the greatest number of approaches having the largest R&D intensities and the firms that pursue the smallest number of approaches having the smallest R&D intensities.

Both sets of predictions of the model are supported by evidence on how firm R&D expenditures vary within industries. Cohen and Klepper (1990) note how the predictions of the model concerning the independence of firm R&D intensities and firm sales accords with the large body of evidence about the proportional relationship within industries between R&D expenditures and firm sales noted in the introduction. The predictions of the model concerning the nature of the distribution within industries of firm R&D intensities are supported by the empirical regularities in industry R&D intensity distributions documented by Cohen and Klepper (1991) for a sample of business units[5] belonging largely to the 1,000 leading American manufacturing firms as of 1974—1977. These distributions tend to be unimodal, positively skewed, and to contain a substantial number of nonperformers of R&D, which is consistent with firm R&D intensity being binomially distributed. Moreover, across industries the moments of the distributions are correlated, which is also compatible with the binomial. Indeed, as Cohen and Klepper (1991) show, the model is capable of explaining not only the fact that the moments are correlated but also the signs of the correlations.

Thus, the model can explain a fairly exacting set of regularities that characterize the distribution of R&D intensities within industries. This provides indirect support for the two central tenets of the model: (1) returns to R&D are scaled by the size of the firm; and (2) differences in firm R&D intensities are attributable to differences across firms in the number of approaches to innovation pursued.

3. The advantages of large size

In this section we consider the implications of our framework for the advantages of large firm size, where size refers specifically to the level of output of a particular product. Our discussion is largely based on Cohen and Klepper (1990). Following our earlier work, we assume that the mechanisms within an industry for appropriating rents due to innovation, such as patents, are given. We focus on the welfare implications of the link in our framework between firm size and the appropriability of the returns to innovation.

Since, in our framework, the firm earns returns on its innovations through its own output and growth tends to be incremental, the returns to R&D are scaled by the firm's *ex ante* sales. As we noted, this is the key in our framework to explaining the proportional relationship between R&D effort and firm size. This argument also implies that large firms will have an advantage over smaller firms in R&D competition. To see this, consider two firms that produce the same product and pursue the same number of approaches to innovation. Suppose, for simplicity, that all R&D expenditures result in innovations which lower the average cost of production.[6] If the two firms spend the same amount on R&D then our framework implies that they would achieve the same unit cost reduction. The larger firm, however, would apply the unit cost reduction over a larger level of output, enabling it to earn greater profits from its R&D.[7] Moreover, our framework implies that if two firms pursue the same number of approaches, they will have the same R&D intensity, which implies that the bigger firm will actually perform more R&D than the smaller firm. This reflects the greater return it earns from any given level of R&D due to its greater size, which provides it with an incentive to perform a greater level of R&D than its smaller rival. Since it will earn positive profits from this additional R&D, this reinforces its advantage over its smaller rival.

Our interpretation of the relationship within industries between R&D effort and firm size is quite unconventional. The standard interpretation is that the proportional relationship between R&D effort and firm size suggests no advantage or disadvantage of large firms in R&D, although Fisher and Temin (1973) caution against making inferences about the advantages of firm size in R&D from the relationship between R&D effort and firm size. Of course, all we have shown is that the evidence is compatible with a theory in which there is an appropriability advantage to large firm size in R&D. There is, however, no competing explanation for why R&D effort should be proportional to firm sales.

Our framework implies that the private advantage accruing to large firm size also represents a social advantage. Perhaps the easiest way to see this is to consider the merger of two firms that produce the same product, are the same size, pursue the same approaches to innovation, and spend the same amount on R&D on the same projects within each approach to innovation. The merger would result in two social dividends. First, duplicative R&D would be eliminated, a point suggested by other researchers (e.g., Spence, 1984). Second, and more novel, for each approach to innovation pursued by the two firms, the merger will increase the profitability of R&D *at the margin*, thereby leading the merged firm to undertake socially efficient R&D projects that were not undertaken by either firm before the merger. In effect, the merged firm will be better able to appropriate the returns from its innovations than either firm alone was able to do, causing it to undertake a greater number of socially efficient R&D projects than were collectively undertaken by the two firms before the merger.

Much of the empirical literature in the Schumpeterian tradition assumes that if R&D effort rises proportionally with firm size, then the size distribution of firms would have no effect on the amount of technological change achieved in the industry. In this view, technical advance is solely a function of the total amount of R&D performed in the industry, notwithstanding who performs it. In contrast, Cohen and Klepper (1990) argue that if innovations cannot be sold and lead to incremental growth, then the total amount of technological change generated in an industry will be a function of not only the total amount of R&D performed by all firms but also the distribution of R&D effort across firms. By redistributing sales and R&D expenditures from smaller to larger firms, total R&D expenditures would not change but a greater number of socially efficient R&D projects would be collectively undertaken, thereby increasing the amount of technological change generated per unit of industry R&D expenditure.

If large firms have an advantage in conducting R&D, as our framework implies, then the appropriability advantage of size would be fully realized under monopoly. We do not, however, commonly observe the emergence of monopoly in R&D intensive industries despite the appropriability advantage of size. An explanation for this is that even successful firms grow only slowly, and that over time technology typically evolves in ways that favor either new firms or firms that were formerly less successful. Indeed, in his development of the notion of creative destruction, Schumpeter (1942) suggested that even monopolists will be displaced in the long run for such reasons.

In the absence of monopoly, firms will have an incentive to cooperate in R&D if they can appropriate the returns to their innovations only through their own output. For any given approach to innovation, all firms pursuing the approach could benefit by pooling their efforts. They would avoid the costly duplication of effort that would otherwise occur. Moreover, because they could apply innovations to their joint output, the firms would collectively profit from R&D expenditures that none of the firms could profit from individually. This suggests that it may be efficient for government to allow cooperative R&D ventures among competing firms pursuing the same approaches to innovation. Indeed, since the 1984 passage of the National Cooperative Research Act, prohibitions against cooperative R&D ventures have been relaxed. A large fraction of the cooperative ventures formed since that time, however, have brought together firms with complementary capabilities rather than firms pursuing the same approaches to innovation (Scott, 1988). This suggests that firms have a range of incentives to cooperate in their R&D efforts, and the one we have identified may not, in practice, be paramount.[8] Moreover, cooperation among firms pursuing the same approaches to innovation may be impeded by difficulties in communication and concerns about competitive advantage. If so, then a more active role for government may be indicated.

It is important to stress that the advantage of size in our framework is based solely on the assumption that firms' returns to innovation are scaled by their *ex ante* output. To the extent that innovations can be sold in disembodied form and/or that firms can reap the returns to their innovations through rapid growth, the advantage of *ex ante* large firm size will be less significant. It is likely that both of these factors will vary across industries and perhaps types of innovations (e.g., product versus process innovations).[9] The greater the extent to which the returns to innovation do not depend on *ex ante* output, then the lower will be the social benefits from cooperation.

4. The advantages of diversity

In this section we consider the role of technological diversity in technical advance and the link between diversity and the number of competitors.

Diversity across firms is incorporated in our framework through the assumption that firms pursue different sets of approaches to innovation. This in turn gives rise to differences in firm R&D intensities that reflect differences in the total number of approaches to innovation pursued by each firm. As we discussed, this form of technological diversity allows us to explain regularities characterizing the nature of the distribution within industries of firm R&D intensities. Of course, this does not mean that diversity is actually an important determinant of R&D intensities within industries, but it does provide indirect support for this view.

In our framework, diversity across firms' innovative activities stems from the existence of different noncompeting approaches to innovation within an industry. We conceive of these different approaches not as substitute or parallel ways of achieving the same technological objective but as independent ways of achieving technological change.[10] The total number of approaches to innovation in an industry can be thought of as being determined by the science and technology underlying R&D in the industry. Which of these approaches are actually pursued by firms in the industry at any given moment will depend upon the expectations and capabilities of firms in the industry. It is certainly possible that some subset of the approaches will not be pursued by any firm in the industry. Indeed, if there are X firms in the industry and the probability of any approach being pursued by any given firm is p, then the likelihood of any approach not being pursued by any firm in the industry is $(1 - p)^X$.

A simple measure of technological diversity in our framework is the number of distinct approaches to innovation that are actually pursued by the firms in the industry. This measure corresponds well to the notion of diversity implicit in the arguments of proponents of small scale enterprise who stress the greater breadth of ideas and types of innovations pursued in industries composed of numerous small firms (e.g., Rodgers, 1990). It follows directly from this measure of technological diversity that the extent of diversity will be directly related to the number of firms in the industry. Since the likelihood of a potential approach to innovation in an industry not being pursued by any firm is $(1 - p)^X$, the expected number of different approaches to innovation pursued in an industry will be an increasing function of the number of firms, X. This implies that on average industries composed of a greater number of firms will be characterized by a greater amount of technological diversity. Note that this greater diversity does not emerge from any superior creativity on the part of smaller firms. It is simply the result of having a greater number of firms in the industry choosing which approaches to innovation to pursue.

The predicted relationship between the number of firms and technological diversity depends upon a key assumption of our framework, namely that the likelihood of a firm pursuing any given approach is independent of the size of the firm. If, alternatively, the probability of a firm pursuing any given approach increased with the size of the firm, it would no longer follow that increasing the number of firms would necessarily increase the expected degree of diversity.[11] While there are undoubtedly industries where the probability of pursuing an approach is correlated with firm size, the evidence concerning the regularities in the industry R&D intensity distributions suggests that the number of such industries is limited. Otherwise, our framework implies that there would be a large number of industries for which the largest firms are characterized by above average R&D intensities,[12] which does not conform with the evidence. Moreover, if the probability of pursuing an approach were an increasing function of the size of the firm, then based on our framework the distribution of firm R&D intensities within an industry would depend on the size distribution of firms and would, in all likelihood, not conform to the regularities in the industry R&D intensity distributions documented by Cohen and Klepper (1991).

If the likelihood of pursuing an approach to innovation is independent of the size of the firm then two firms of equal size would together be expected to pursue a greater number of distinct approaches to innovation than one firm twice their size. Having more firms provides more "independent minds" to consider the alternatives, which results in greater technological diversity. There are a number of organizational factors that might explain why the likelihood of pursuing any approach is independent of firm size and, hence, why diversity is promoted by a greater number of firms. We suspect it has something to do with the way R&D proposals are processed within firms. Following the argument suggested by Sah and Stiglitz (1986, 1988), suppose that proposals are initiated by technical staff and then subsequently considered by higher level decision makers. If large firms have more hierarchical levels than small firms and a proposed approach can be rejected at any level, an approach will have a greater chance of being approved by a smaller firm. As long as the number of proposed approaches in the two firms combined and the one large firm are equal, the two smaller firms would then be expected to pursue a greater number of distinct approaches than the large firm. Alternatively, suppose the probability of an approach being rejected is independent of the size of the firm. Even in this case, if all approaches proposed in the large firm are proposed in at least one of the two smaller firms and some are proposed in both of the smaller firms,[13] then the two smaller firms combined will approve a greater number of distinct approaches than the larger firm.[14]

Suppose we are correct and the number of firms is an important determinant of the diversity of approaches to innovation pursued in an industry. In order for there to be an advantage associated with small firm size, there must be some benefit associated with the greater diversity brought about by a greater number of firms. Our framework highlights an obvious benefit from the greater diversity — the exploitation of beneficial approaches to innovation that otherwise would not have been pursued. Indeed, this advantage of diversity was suggested long ago by Alfred Marshall when he stated that, "The tendency to variation is a chief source of progress" (Marshall, 1920, p. 355). Even if diversity does promote progress, as Marshall suggests, this does not mean that social welfare is necessarily promoted by the greater diversity associated specifically with an increase in the number of firms. Holding the output of the industry constant, if the number of firms is increased then the average size of each firm will decline. In the prior section, we showed how in our framework the consolidation of firms would yield social benefits. It follows that reducing the average firm size would impose a social cost. Assuming that the returns to innovation are scaled by *ex ante* output, each firm will spend less on R&D on the approaches it pursues, and for some amount of time, the innovations developed by each firm will be applied over a smaller level of output. The net effect on social welfare of increasing the number of firms depends on the magnitude of these costs relative to the benefits of having additional approaches to innovation pursued. We consider this tradeoff further in the next section.

While it is not straightforward to establish the social advantage of diversity in our framework, it is possible to imagine circumstances in which the increase in diversity brought about by a greater number of firms unequivocally promotes social welfare. For instance, suppose that firms are fully able to sell their innovations and there is no duplication among firms in their R&D efforts. In such a world, all innovations will be applied immediately to the entire output of the industry, and a reduction in the size of each firm will not diminish its incentives to conduct R&D. Then the decrease in firm size associated with the increase in the number of firms will have no effect on the R&D expenditures of the incumbent firms nor on the output over which any innovations resulting from these expenditures are applied. Consequently, any increase in the number of approaches pursued resulting from an increase in the number of firms will provide a net benefit to society. Intuitively, having a greater number of different minds (i.e., firms) evaluate the possible approaches to innovation will diminish the chance that a beneficial approach to innovation will be overlooked.

Our claim that diversity within an industry will increase with the number of firms implicitly assumes that at any given moment there may be firms outside of an industry that would pursue an otherwise unpursued approach if they entered the industry. Stated more baldly, it assumes that there may be potential entrants who recognize an opportunity but who do not enter the industry to exploit it. One reason this could occur is if there is some cost to entering the industry that exceeds the gross profit the entrant could make from pursuing the unexploited opportunity. From a social welfare perspective, however, this cost should be counted against the technical advance generated by the increase in the number of firms. If the social benefit of this technical advance was solely limited to the producer surplus resulting from the entrant's pursuit of a new approach, the net change in social welfare from the increase in the number of firms would be negative since the cost of entry would exceed the gross profits the producer could make from entering. Thus, for the increase in diversity to bring about an increase in social welfare, there must be some kind of wedge between the private and social returns from the pursuit of a new approach.

Our analysis of innovative activity, like many others (e.g., Arrow, 1962; Scherer, 1979; Spence, 1984), provides the basis for such a wedge. In laying out the framework, we noted that eventually all innovations are imitated. To the extent that the costs of imitation are less than the benefits, the social returns to innovation will exceed the private returns.[15] Consequently, it is possible that the pursuit of new approaches to innovation by entrants could be socially beneficial but not profitable to the entrant. In that case, the social benefits from greater diversity would not be realized through private initiative.

Our arguments suggest that if policies to increase the number of firms in an industry were undertaken, diversity would be expected to increase, which, in turn, could increase social welfare. Such policies could include prohibitions on horizontal mergers, subsidies of failing firms, and/or subsidies of entrants. Note that none of these policies requires the government to have any special insights about the sort of innovations an industry should pursue. Indeed, there is no reason to expect the government to be any more enlightened than firms about the best innovations to pursue. The rationale for these policies is simply that by getting more firms in the industry, there will be less chance that beneficial opportunities for technical advance will be unexploited.

It is important to note that the diversity-inducing effect of the presence of many small firms clearly depends upon the vitality of the science and technology underlying technical advance in any given industry, which we take as given. If this science and technology base is moribund, then the number of approaches to innovation that could be pursued will be few. In this case, the relationship between the number of firms and technological diversity, and, in turn, technical advance, will be weaker.

5. The tradeoff

In the last section we argued that it is not possible to bring about an increase in diversity through an increase in the number of firms without compromising some of the benefits associated with large firm size. Not surprisingly, it is also not possible to reap the benefits of large firm size without compromising the advantages associated with greater diversity. In this section we consider further the nature of this tradeoff.

When diversity is promoted through a greater number of firms, average firm size is reduced, which compromises the advantages associated with large size firms. Consider, alternatively, the costs associated with promoting the advantages of large firm size through the consolidation of firms. By reducing the number of firms in the industry, the expected number of approaches to innovation collectively pursued in the industry will decline. Then, on average, a smaller fraction of the beneficial opportunities for innovation will be pursued, which will reduce social welfare. Thus, just as increasing diversity through an increase in the number of firms has offsetting costs, so does increasing firm size through the consolidation of firms.[16]

Thus, in our framework, there is a tradeoff between the advantages associated with small firms and those associated with large firms. In order to have more approaches to innovation pursued, it is necessary to sacrifice some intensity of effort for each approach, and *vice versa*. The nature of this tradeoff is complicated and will depend on a number of factors. Analyzing the optimal average firm size for any industry in our framework requires a formal model of this tradeoff, which is beyond the scope of the paper. We can, however, use our framework to speculate about the role of factors that would be expected to shape the nature of the tradeoff: the extent to which firms can either sell their innovations or grow rapidly due to innovation, and the vitality of an industry's technology, as represented by the number of different approaches to innovation potentially available in the industry. To the extent that firms can realize rents due to innovation via rapid growth, licensing, and other mechanisms that do not exploit the firm's *ex ante* output, the appropriability advantage of size will be limited. In that case, an industry composed of a greater number of firms will be desirable. To the extent that the science and technology base of an industry is such that the potential number of approaches to innovation available to the industry is small, the diversity advantage associated with numerous small firms will be limited. In that case, an industry composed of a smaller number of firms will be desirable.

Consider the case of an industry in which there are a large number of approaches to innovation and firms can appropriate the returns to innovation only through their own *ex ante* output, which is the case where the tradeoff between the advantages associated with large and small firm size is the most acute. We suspect a substantial number of technologically progressive industries will be characterized by just such an acute tradeoff. There is no reason to believe that in these types of industries market forces alone will result in an optimal number of firms. In any given industry, there likely will be either too little entry to promote sufficient diversity or entry may be so unencumbered that the appropriability advantage of larger firm size is excessively compromised.

If market forces cannot be expected to lead to the optimal number of firms, policy makers will be faced with the challenge of deciding whether the number of firms in each industry is too large or too small. This will require making extremely difficult judgments. It is possible, however, that we have overblown the challenge posed by the tradeoff and policies can be designed to achieve the benefits of both diversity and large size. Consider, for example, current proposals that government condone firm cooperation on "generic" research (e.g., Nelson, 1990). Such proposals can be interpreted as suggesting that firms pursue separately R&D on approaches which they do not pursue in common and pursue together R&D on approaches which they do pursue in common. In this fashion, cooperative agreements could be crafted that would permit the industry and society to benefit from the appropriability advantage of large firm size without compromising the benefits of diversity. Indeed, cooperation on something resembling what we call generic research is apparently one of the models for cooperative research employed in the Japanese manufacturing sector.

6. Conclusion

Our analysis lends support to both sides of the debate concerning the optimal firm size for achieving technical advance. It provides a basis for why industries composed of many small firms will tend to exhibit greater diversity in the approaches to innovation pursued, and why greater diversity will contribute to more rapid technological change. It also provides a basis for why industries populated by larger firms will achieve a more rapid rate of technical advance on the approaches to innovation that are pursued. These arguments together suggest that a tradeoff exists between the appropriability advantage of large size and the advantages of diversity that accrue from numerous small firms.[17]

Our analysis has been more appreciative than rigorous and, indeed, often explicitly speculative. While we attempted to raise important questions, our framework requires more structuring before we can be confident about any of our conclusions. Even in its inchoate form, however, our analysis demonstrates that much needs to be done before the current debate about firm size can seriously inform policy. If we accept the plausibility of our basic framework, it focuses attention on a range of issues and questions. The fundamental premise of our analysis is that firm capabilities and perceptions differ within industries. This premise is not, however, widely reflected in analyses of industry behavior and performance, which typically take some representative firm as their starting point. Indeed, the analytic utility of our particular premise deserves scrutiny. Are differences in firm capabilities and perceptions as critical to explaining the industry patterns in innovative activity and performance as we suggest? Do these differences persist? Is our abstract characterization of these differences and their effects on innovative activity up to the task of providing a basis for policy?

These intraindustry differences in capabilities and perceptions underpin the hypothesized relationship in our framework between the number of firms within an industry and the number of distinct technological activities pursued by the industry as a whole. Surely this hypothesis should be tested. To establish the relationship between numbers of firms and technological diversity, we also made two important assumptions, which themselves should be examined. First, we assumed that firms independently decide upon which approaches to innovation to pursue. This assumption precludes the clustering of firms around innovative activities due to imitation, a phenomenon highlighted by Nelson (1981) and Scott (1991). To the degree that innovative activities yield relatively fast, public results, the assumption may be suspect. While our evidence indirectly suggests that such clustering may not be critical for explaining innovative activity in a wide range of industries, more research would be helpful. Second, we assumed that the number of approaches to innovation pursued by firms is independent of their size, implying large and small firms will tend to pursue the same number of approaches. This assumption probably does not apply to the smallest firms within an industry, particularly to the extent that such firms are often not full line manufacturing firms. Does it apply, however, to the medium to large firms that account for the preponderance of R&D and economic activity in the manufacturing sector? While our evidence again provides indirect support for this claim, more empirical and theoretical research is indicated.

We also made other claims and assumptions that deserve further attention. For example, we argued that greater technological diversity stimulates technical advance and provides gross increments to social welfare. Assuming it exists, the mechanism linking diversity and technical advance has never been examined empirically and is not obvious. Our assumption that expected firm growth due to innovation is incremental played an important role in permitting us to hypothesize an appropriability advantage of large size. Again, both the assumption and its alleged effect on innovative activity are worth examining. Finally, we also need to test whether the relationship between R&D and firm size within industries depends upon appropriability conditions, particularly upon the extent to which firms can sell their innovations or grow rapidly due to innovation.[18] In conclusion, this litany of reasonable but unsubstantiated assumptions and arguments should make clear that this paper is only a modest beginning of a daunting research agenda.

Acknowledgements

We are indebted to Mark Kamlet, Jonathan Leland and F. M. Scherer for their suggestions.

Notes

[1] An important theme of the work on industry evolution is that it is difficult for a firm to change deliberately its core expertise (e.g., Nelson and Winter, 1982; Klepper and Graddy, 1990). The corporate strategy literature also implicitly makes this point when arguing that firms should exploit their "core competences" (e.g., Andrews, 1971; Porter, 1980), suggesting that it is difficult for firms to transform their capabilities.

[2] This is a strong assumption which we plan to relax in subsequent development of our model. It can be rationalized, however, by simply defining two approaches as distinct if the likelihood of the approaches being pursued by any given firm are independent.

[3] In recent years, it has been recognized that many innovations are most valued by the innovators themselves because they are either idiosyncratic to the innovator (Nelson, 1989) or require idiosyncratic expertise for their exploitation that is only generated in the course of the innovation process itself (Cohen and Levinthal, 1989, 1990).

[4] The survey results of Levin *et al.* (1987) provide a strong empirical basis for the claim. They find that the most effective mechanism for appropriating rents due to innovation are first-mover advantages, secrecy, and complementary sales and service, all of which are implemented via the firm's own sales.

[5] A business unit represents a firm's activity in a given industry.

[6] The following argument applies equally to product innovations (Cohen and Klepper, 1990).
[7] Cf. Nelson, Peck and Kalachek (1967) and Scherer (1970), who also note this potential advantage of size.
[8] See Scott (1988), Link and Bauer (1989), and Katz and Ordover (1990) for a discussion of the range of factors that may affect the benefits and costs of R&D cooperation.
[9] For example, the ability of firms to sell their innovations in disembodied form will vary across industries according to the ease of defining and enforcing property rights over innovations.
[10] To make the idea of different, noncompeting approaches to technological change within an industry more concrete, consider the semiconductor industry. Historically, research in semiconductors has been conducted in many areas, including silicon manufacture, low-cost methods for semiconductor manufacture, and the design of various devices such as MOS, bipolar, integrated circuit, and memory devices. Each of these can be thought of as different but complementary approaches to innovation in semiconductors. Also, consistent with our general framework, semiconductor firms have differed considerably in the sets of these approaches that they pursued. Texas Instruments, for example, possessed a broad range of expertise in solid state physics, materials properties and preparation, production process chemistry, advanced photolithographic processes, semiconductor device and design technologies, and pursued a broad range of approaches to innovation. In contrast, companies such as Mostek, American Micro-Systems and SEMI possessed more circumscribed expertise in device design technologies, and pursued a narrower range of approaches to innovation that focused on the design of MOS, bipolar, or memory devices.

Our characterization of approaches as noncompeting contrasts with others' conceptualization of different approaches to innovation. For example, in the analyses of Evenson and Kislev (1976) and Nelson (1982), different approaches represent distinct, competing ways of achieving the same technological end. Differentiating approaches in this way is relevant when there is uncertainty about how best to achieve a particular technological objective. In our framework, however, we abstracted from this kind of uncertainty. Our firms differ, in contrast, in the objectives (i.e., approaches) they pursue, reflecting differences in their expertise. While there is often uncertainty about how best to achieve a particular technological objective, we have abstracted from this kind of uncertainty because it would not readily explain the regularities in industry R&D intensity distributions. Indeed, it does not readily explain why R&D intensities of firms in an industry should differ at all in any systematic way (Cohen and Klepper, 1991).
[11] Holding industry output constant, an increase in the number of firms will reduce the average firm size. If the probability of a firm pursuing an approach were positively related to its size, then the increase in the number of firms would decrease the probability of any firm pursuing any given approach. Then the probability of any given approach not being pursued by any firm in the industry could fall with an increase in the number of firms, which would cause the expected amount of diversity to fall.
[12] In our framework, if the probability of pursuing an approach were positively related to the size of the firm, then larger firms would be expected to pursue a greater number of approaches to innovation, causing them to have above average R&D intensities.
[13] This is plausible if some of the approaches proposed in the large firm are proposed by multiple individuals. In that case, if the larger firm were broken into two firms, there would be some probability that the technical personnel in each of the smaller firms would propose the same approach.
[14] Another reason firms of equal size might pursue a greater number of distinct approaches than one firm twice their size is that most firms tend to funnel control over their R&D operations for a given industry through one individual, and that individual will naturally reflect his/her own expertise and perceptions in the decisions that are made. Thus, regardless of the firm's size, the approaches that are pursued will tend to reflect one individual's judgment. Indeed, a recent article in the *The Wall Street Journal* (Rigdon, 1991) about Kodak's research on filmless cameras highlighted the importance — and difficulty — of decentralizing high-level R&D decision-making authority when a firm becomes aware of a distinct and viable new approach to innovation within an industry. It noted that while the research team working on new filmless camera technology has been given its own budget and autonomy over hiring decisions, the same individual who oversees research on more conventional film technologies also oversees this research, which may ultimately inhibit Kodak's willingness or ability to pursue the new technology.
[15] In a static setting, imitation reinforces the gross social benefits accruing from greater diversity because the benefits from the pursuit of new approaches will be reaped by more than just the firms pursuing the new approaches. Imitation may also yield dynamic dividends. As noted by Marshall (1920, p. 271), "... if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas ..." Thus, pursuing a given approach may stimulate other firms "to improve it, variegate it, more generally contribute to its further advance," (Nelson, 1989, p. 6) even in ways that cannot be anticipated when the approach is first pursued. Von Weizsacker (1980) also recognized this point when he explored the implications of "sequential innovations" for the welfare effects of entry barriers.
[16] We managed to sidestep this cost in the example we used earlier to illustrate the benefits of large firm size by focusing on the merger of two firms that pursued the same approaches to innovation. This insured that the merger of the firms did not reduce the diversity of approaches pursued in the industry. However, firms generally will not pursue the identical set of approaches to innovation, so that consolidations will be expected to compromise the collective number of approaches to innovation pursued in the industry.
[17] Others, such as Nelson (1981), have also recognized a tradeoff between the diversity-inducing advantage of more competitive industry structures and advantages of large firm size, but not the particular tradeoff we have identified.
[18] Cohen and Klepper (1990) demonstrate that if firms can sell some fraction of their innovations in disembodied form or

if growth due to innovation is unconditioned by existing output levels, then large firm size will confer less of an advantage and R&D effort should rise less than proportionally with firm size.

References

Andrews, K. R., 1971, *The Concept of Corporate Strategy*, Homewood, Illinois: Dow Jones-Irwin.

Arrow, K., 1962, 'Economic Welfare and the Allocation of Resources for Inventions', in R. R. Nelson (ed.), *The Rate and Direction of Inventive Activity*, Princeton, NJ: Princeton University Press for the National Bureau of Economic Research.

Baldwin, W. L. and J. T. Scott, 1987, *Market Structure and Technological Change*, Chur: Harwood Academic Publishers.

Cohen, W. M. and S. Klepper, 1990, 'A Reprise of Size and R&D', mimeo, Carnegie Mellon University.

Cohen, W. M. and S. Klepper, 1991, 'The Anatomy of Industry R&D Intensity Distributions', *American Economic Review*, forthcoming.

Cohen, W. M. and D. Levinthal, 1989, 'Innovation and Learning: The Two Faces of R&D', *Economic Journal* **99**, 569—596.

Cohen, W. M. and D. Levinthal, 1990, 'Absorptive Capacity: A New Perspective on Innovation and Learning', *Administrative Science Quarterly* **35**, 128—152.

Evenson, R. E. and Y. Kislev, 1976, 'A Stochastic Model of Applied Research', *Journal of Political Economy* **84**, 265—281.

Ferguson, C. H., 1988, 'From the People Who Brought You Voodoo Economics', *Harvard Business Review* (May—June), 55—62.

Fisher, F. M. and P. Temin, 1973, 'Returns to Scale in Research and Development: What Does the Schumpeterian Hypothesis Imply?', *Journal of Political Economy* **81**, 56—70.

Gilder, G., 1988, 'The Revitalization of Everything: The Law of the Microcosm', *Harvard Business Review* (March—April), 49—61.

Jorde, T. M. and D. J. Teece, 1990, 'Innovation and Cooperation: Implications for Competition and Antitrust', *The Journal of Economic Perspectives* **4**, 75—96.

Kamien, M. I. and N. L. Schwartz, 1982, *Market Structure and Innovation*, Cambridge: Cambridge University Press.

Katz, M. L. and J. A. Ordover, 1990, 'R&D Cooperation and Competition', *Brookings Papers on Economic Activity, Microeconomics*, 137—203.

Klepper, S. and E. Graddy, 1990, 'The Evolution of New Industries and the Determinants of Market Structure', *The Rand Journal of Economics* **21**, 27—44.

Levin, R. C., A. Klevorick, R. Nelson and S. Winter, 1987, 'Appropriating the Returns from Industrial Research and Development', *Brookings Papers on Economic Activity* **3**, 783—820.

Link, A. N. and L. L. Bauer, 1989, *Cooperative Research in U.S. Manufacturing — Assessing Policy Initiatives and Corporate Strategies*, Lexington, Mass.: Lexington Books.

Marshall, A., 1961, *Principles of Economics*, New York: The Macmillan Company, 9th (variorum) edition, Vol. 1, with annotations by C. W. Guillebaud (based on New York: The Macmillan Company, 8th edition, 1920).

Nelson, R. R., 1981, 'Assessing Private Enterprise: An Exegesis of Tangled Doctrine', *The Bell Journal of Economics* **12**, 93—111.

Nelson, R. R., 1982, 'The Role of Knowledge in R&D Efficiency', *Quarterly Journal of Economics* **97**, 453—470.

Nelson, R. R., 1989, 'What Is 'Commercial' and What Is 'Public' About Technology, and What Should Be?', mimeo, presented at the Conference on Economic Growth and the Commercialization of New Technologies, Stanford University, September 11—12, 1989.

Nelson, R. R., 1990, 'Capitalism as an Engine of Progress', *Research Policy* **19**, 193—214.

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

Nelson, R. R., M. J. Peck and E. D. Kalachek, 1967, *Technology, Economic Growth, and Public Policy*, Washington: Brookings Institution.

Norris, W. C., 1983, 'How to Expand R&D Cooperation', *Business Week*, April 11, p. 21.

Noyce, R. N., 1990, 'Cooperation Is the Best Way to Beat Japan', *New York Times*, July 9, Section 3, p. 2.

Porter, M. E., 1980, *Competitive Strategy: Techniques for Analyzing a Business, Industry and Competitors*, New York: Free Press.

Rigdon, J. E., 1991, 'Kodak Tries to Prepare for Filmless Era Without Inviting Demise of Core Business', *The Wall Street Journal*, April 18, 1991, B1, B5.

Rogers, T. J., 1990, 'Landmark Messages from the Microcosm', *Harvard Business Review* (January—February), 24—30.

Sah, R. J. and J. E. Stiglitz, 1986, 'The Architecture of Economic Systems: Hierarchies and Polyarchies', *American Economic Review* **76**, 716—727.

Sah, R. J. and J. E. Stiglitz, 1988, 'Committees, Hierarchies and Polyarchies', *Economic Journal* **98**, 451—470.

Scherer, F. M., 1970, *Industrial Market Structure and Economic Performance*, 1st edition, Chicago: Rand McNally College Publishing Company.

Scherer, F. M., 1979, 'The Welfare Economics of Product Variety: An Application to the Ready-to-Eat Cereals Industry', *Journal of Industrial Economics* **28**, 113—134.

Scherer, F. M., 1980, *Industrial Market Structure and Economic Performance*, 2nd edition, Chicago: Rand McNally College Publishing Company.

Scherer, F. M. and D. Ross, 1990, *Industrial Market Structure and Economic Performance*, 3rd edition, Boston: Houghton Mifflin.

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

Schumpeter, J. A., 1942, *Capitalism, Socialism, and Democracy*, New York: Harper.

Scott, J. T., 1988, 'Diversification Versus Cooperation in

R&D Investment', *Managerial and Decision Economics* **9**, 173—186.

Scott, J., 1991, 'Research Diversity and Technical Change', in Z. Acs and D. Audretsch (eds.), *Innovation and Technical Change*, Ann Arbor: University of Michigan Press, pp. 132—151.

Shaffer, R. A., 1990, 'Let a Thousand Companies Fight', *New York Times*, July 9, Section 3, p. 2.

Spence, A. M., 1984, 'Cost Reduction, Competition, and Industry Performance', *Econometrica* **52**, 101—121.

Von Weizsacker, C. C., 1980, *Barriers to Entry*, New York: Springer-Verlag.