10. Diversity: implications for income distribution

David B. Audretsch

As recently as the early 1990s scholars and industry observers predicted, if not the death of Silicon Valley, then its slowdown.[1] For example, in a much-cited article in the *Harvard Business Review* Charles Ferguson (1988, p. 61) argued:

> Fragmentation, instability, and entrepreneurialism are not signs of well-being. In fact, they are symptoms of the larger structural problems that afflict US industry. In semiconductors, a combination of personnel mobility, ineffective intellectual property protection and tax subsidies for the formation of new companies contribute to a fragmented 'chronically entrepreneurial' industry. US semiconductor companies are unable to sustain the large, long-term investments required for continued US competitiveness . . . Personnel turnover in the American merchant semiconductor industry has risen to 20 percent compared with less than 5 percent in IBM and Japanese corporations . . . Fragmentation discouraged badly needed coordinated action – to develop better process technology and also to demand better government support.

A decade later, not only is Silicon Valley thriving but, as *The Economist* pointed out, average pay in Silicon Valley rose between 1995 and 1996 by 5 per cent in real terms, to $43,510, compared to a mere 1 per cent increase to $28,040 for the rest of the country.[2] Despite high production costs, environmental destruction and overall congestion, reports of Silicon Valley's demise were premature.

The purpose of this chapter is to suggest that differences in the distribution of income across regions are likely to grow. As a result of globalization, those regions whose economies are based on routinized economic activity will experience a downward pressure on incomes because the cost of diffusing routinized economic activity across space to lower-cost locations is relatively low. By contrast, those regions whose economies are based on search economic activity will experience growth in incomes because it is costly to diffuse search economic activity across space.

The extent to which firms and individuals are homogeneous or heterogeneous shapes the relative efficiency of routinized and search activities.

Homogeneity is conducive to routinized activity but impedes search activity. Diversity promotes search activity but raises the cost of routinized activity. An implication is that as the comparative advantage of the developed nations of Western Europe and North America shifts away from routinized activities and towards search activities, those organizations able to harness diversity will tend to emerge as the most successful. The income gap will continue to grow between those economic agents and regions engaged in search activity and those engaged in routinized activity.

1. WHY DOES DIVERSITY MATTER?

1.1 New Economic Knowledge

The starting point for most theories of innovation is the firm. In such theories the firm is exogenous and its performance in generating technological change is endogenous. For example, in the most prevalent model found in the literature of technological change, the model of the knowledge-production function formalized by Zvi Griliches (1979), firms exist exogenously and then engage in the pursuit of new economic knowledge as an input into the process of generating innovative activity.

The most decisive input in the knowledge-production function is new economic knowledge. And as Cohen and Klepper (1991, 1992a, 1992b) conclude, the greatest source generating new economic knowledge is generally considered to be R&D. Certainly a large body of empirical work has found a strong and positive relationship between knowledge inputs such as R&D on the one hand and innovative outputs on the other hand.

The knowledge-production function has been found to hold most strongly at broader levels of aggregation. The most innovative countries are those with the greatest investments in R&D. Little innovative output is associated with less developed countries, which are characterized by a paucity of production of new economic knowledge. Similarly, the most innovative industries tend to be characterized by considerable investments in R&D and new economic knowledge. Industries such as computers, pharmaceuticals and instruments are high not only in R&D inputs which generate new economic knowledge, but also in innovative outputs (Audretsch 1995). By contrast, industries with little R&D, such as wood products, textiles and paper, tend to produce only a negligible amount of innovative output. Thus, the knowledge-production model linking knowledge-generating inputs to outputs holds at the more aggregated levels of economic activity.

Where the relationship becomes less compelling is at the desegregated microeconomic level of the enterprise, establishment or even line of business. For example, although Audretsch (1995) found that the simple correlation between R&D inputs and innovative output was 0.84 for four-digit standard industrial classification (SIC) manufacturing industries in the United States, it was only about half (0.40) among the largest US corporations.

The model of the knowledge-production function becomes even less compelling in view of the recent wave of studies revealing that small enterprises serve as the engine of innovative activity in certain industries. These results are startling because, as Scherer (1991) observes, the bulk of industrial R&D is undertaken in the largest corporations; small enterprises account for only a minor share of R&D inputs. Thus the knowledge-production function seemingly implies, as the Schumpeterian hypothesis predicts, that innovative activity favours those organizations with access to knowledge-producing inputs – large incumbent organizations (Schumpeter 1911, 1942). The more recent evidence identifying the strong innovative activity of small firms raises the question: where do new and small firms get innovation-producing inputs, that is, knowledge?

One answer, proposed by Audretsch (1995), is that although the model of the knowledge-production function may be valid, the implicitly assumed unit of observation – the firm – may be less valid. The reason why the knowledge-production function holds more closely for more aggregated degrees of observation may be that investment in R&D and other sources of new knowledge spills over for economic exploitation by third-party firms.

1.2 The Appropriability Problem Revisited

A large literature has emerged focusing on what has become known as the appropriability problem. The underlying issue revolves around how firms which invest in the creation of new economic knowledge can best appropriate the economic returns from that knowledge (Arrow 1962). Audretsch (1995) proposes shifting the unit of observation away from exogenously assumed firms to individuals – agents with endowments of new economic knowledge. As J. de V. Graf (1957) observed:

> When we try to construct a transformation function for society as a whole from those facing the individual firms comprising it, a fundamental difficulty confronts us. There is, from a welfare point of view, nothing special about the firms actually existing in an economy at a given moment of time. The firm is in no sense a 'natural unit'. Only the individual members of the economy can lay claim to that distinction. All are potential entrepreneurs. It seems, therefore, that the natural thing to do is to build up from the transformation function of men,

> rather than the firms, constituting an economy. If we are interested in eventual empirical determination, this is extremely inconvenient. But it has conceptual advantages. The ultimate repositories of technological knowledge in any society are the men comprising it, and it is just this knowledge which is effectively summarised in the form of a transformation function. In itself a firm possesses no knowledge. That which is available to it belongs to the men associated with it. Its production function is really built up in exactly the same way, and from the same basic ingredients, as society's.

But when the lens is shifted away from focusing upon the firm as the relevant unit of observation to individuals, the relevant question becomes: how can economic agents with a given endowment of new knowledge best appropriate the returns from that knowledge?

The appropriability problem confronting the individual may converge with that confronting the firm. Economic agents can and do work for firms, and even if they do not, they can potentially be employed by an incumbent firm. In fact, in a model of perfect information with no agency costs, any positive economies of scale or scope will ensure that the appropriability problems of the firm and individual converge. If an agent has an idea for doing something different than is currently being done by the incumbent enterprises – in terms of a new product or process or organizational structure or management approach – the idea, which can be called an innovation, will be presented to the incumbent enterprise. Because of the assumption of perfect knowledge, both the firm and the agent would agree upon the expected value of the innovation. But to the degree that any economies of scale or scope exist, the expected value of implementing the innovation within the incumbent enterprise will exceed that of taking the innovation outside of the incumbent firm to start a new enterprise. Thus, the incumbent firm and the inventor of the idea would be expected to reach a bargain splitting the value added to the firm by the innovation. The payment to the inventor – in terms of either a higher wage or some other means of remuneration – would be bounded between the expected value of the innovation if it were implemented by the incumbent enterprise on the upper end, and by the return which the agent could expect to earn if she used it to launch a new enterprise on the lower end. Or as Frank Knight (1921, p. 273) observed,

> The labourer asks what he thinks the entrepreneur will be able to pay, and in any case will not accept less than he can get from some other entrepreneur, or by turning entrepreneur himself. In the same way the entrepreneur offers to any labourer what he thinks he must in order to secure his services, and in any case not more than he thinks the labourer will actually be worth to him, keeping in mind what he can get by turning labourer himself.

Thus, each economic agent would choose how to best appropriate the value of his endowment of economic knowledge by comparing the wage he would earn if he remains employed by an incumbent enterprise, w, to the expected net discounted present value of the profits accruing from starting a new firm, π. If these two values are relatively close, the probability that he would choose to appropriate the value of his knowledge through an external mechanism such as starting a new firm, $Pr(e)$, would be relatively low. On the other hand, as the gap between w and π becomes larger, the likelihood of an agent choosing to appropriate the value of her knowledge externally through starting a new enterprise becomes greater:

$$Pr(e) = f(\pi - w) \tag{1}$$

1.3 Asymmetric Knowledge, Transaction Costs and the Principal–Agent Relationship

As Knight (1921) and Arrow (1962) emphasized, new economic knowledge is anything but certain. In addition, substantial asymmetries exist across agents both between and within firms (Milgrom and Roberts 1987). The expected value of a new idea or a potential innovation is likely to be anything but unanimous between the inventor of that idea and the decision-maker or group of decision-makers of the firm confronted with the task of evaluating proposed changes or innovations. In fact, it is because information is not only imperfect but also asymmetric that Knight (1921, p. 268) argued that the primary task of the firm is to process information in order to reach a decision:

> With the introduction of uncertainty – the fact of ignorance and the necessity of acting upon opinion rather than knowledge – into this Eden-like situation [that is, a world of perfect information], its character is entirely changed . . . With uncertainty present doing things, the actual execution of activity, becomes in a real sense a secondary part of life; the primary problem or function is deciding what to do and how to do it.

Alchian (1950) pointed out that the existence of knowledge asymmetries would result in the inevitability of mistaken decisions in an uncertain world. Later, Alchian and Demsetz (1972) attributed the existence of asymmetric information across the employees in a firm as resulting in a problem of monitoring the contribution accruing from each employee and setting the rewards correspondingly. This led them to conclude, 'The problem of economic organization is the economical means of metering productivity and rewards' (Alchian and Demsetz 1972, p. 783).

Combined with the bureaucratic organization of incumbent firms to make a decision, the asymmetry of knowledge leads to a host of agency problems, spanning incentive structures, monitoring and transaction costs. It is the existence of such agency costs, combined with asymmetric information, which not only provides an incentive for agents with new ideas to appropriate the expected value of their knowledge externally by starting new firms, but also with a propensity which varies systematically from industry to industry.

Coase (1937) and Williamson (1975) argued that the size of an (incumbent) enterprise will be determined by answering the question, 'Will it pay to bring an extra exchange transaction under the organizing authority?' (Coase 1937, p. 30). In fact, 'other things being equal, a firm will tend to be larger the less likely the [firm] is to make mistakes and the smaller the increase in mistakes with an increase in the transactions organized' (Coase 1937, p. 24).

Holmstrom pointed out the existence of a bureaucratization dilemma, in which 'to say that increased size brings increased bureaucracy is a safe generalization. To note that bureaucracy is viewed as an organizational disease is equally accurate' (Holmstrom 1989, p. 320).

To minimize agency problems and the cost of monitoring, bureaucratic hierarchies develop objective rules. In addition, Kreps (1991) argues that such bureaucratic rules promote internal uniformity and that a uniform corporate culture in turn promotes the reputation of the firm. These bureaucratic rules, however, make it more difficult to evaluate the efforts and activities of agents involved in activities which do not conform to such bureaucratic rules. As Holmstrom (1989, p. 323) points out:

> Monitoring limitations suggest that the firm seeks out activities which are more easily and objectively evaluated. Assignments will be chosen in a fashion that is conducive to more effective control. Authority and command systems work better in environments which are more predictable and can be directed with less investment information. Routine tasks are the comparative advantage of a bureaucracy and its activities can be expected to reflect that.

Williamson (1975, p. 201) also emphasizes the inherent tension between hierarchical bureaucratic organizations and the ability of incumbent organizations to appropriate the value of new knowledge for innovative activity outside of the technological trajectories associated with the core competence of that organization:

> Were it that large firms could compensate internal entrepreneurial activity in ways approximating that of the market, the large firm need experience no disadvantage in entrepreneurial respects. Violating the congruency between hierarchical position and compensation appears to generate bureaucratic strains,

> however, and is greatly complicated by the problem of accurately imputing causality.

This leads him to conclude:

> I am inclined to regard the early stage innovative disabilities of large size as serious and propose the following hypothesis: An efficient procedure by which to introduce new products is for the initial development and market testing to be performed by independent investors and small firms (perhaps new entrants) in an industry, the successful developments then to be acquired, possibly through licensing or merger, for subsequent marketing by a large multidivision enterprise . . . Put differently, a division of effort between the new product innovation process on the one hand, and the management of proven resources on the other may well be efficient. (Williamson 1975, pp. 205–6)

This model analysing the decision of how best to appropriate the value of new economic knowledge confronting an individual economic agent seems useful when considering the actual decision to launch a new firm taken by entrepreneurs. For example, Chester Carlsson started Xerox after his proposal to produce a (new) copy machine was rejected by Kodak. Kodak based its decision on the premise that the new copy machine would not earn very much money, and in any case Kodak was in a different line of business – photography. It is perhaps no small irony that this same entrepreneurial start-up, Xerox, decades later turned down a proposal from Steven Jobs to produce and market a personal computer because it did not think that a personal computer would sell, and in any case Xerox was in a different line of business – copy machines (Audretsch 1995). After 17 other companies, including IBM and Hewlett-Packard, rebuffed him for virtually identical reasons, Jobs resorted to starting his own company, Apple Computer.

Similarly, IBM turned down an offer from Bill Gates, 'the chance to buy ten percent of Microsoft for a song in 1986, a missed opportunity that would cost $3 billion today'.[3] IBM reached its decision on the grounds that 'neither Gates nor any of his band of thirty-some employees had anything approaching the credentials or personal characteristics required to work at IBM'.[4]

Divergences in beliefs with respect to the value of a new idea need not be restricted to what is formally known as a product or even a process innovation. That economic agents choose to start a new firm because of divergences in the expected value of an idea applies to the sphere of managerial style and organization as well. One of the most vivid examples involves Bob Noyce, who founded Intel. Noyce had been employed by Shockley Semiconductor, which is credited with being the pioneering semiconductor firm. In 1957 Noyce and seven other engineers quit Shockley Semiconductor en masse to form Fairchild Semiconductor, which is considered the start of what is today known as Silicon Valley. Although Fairchild Semiconductor had 'possibly the most potent management and technical team ever assembled' (Gilder 1989, p. 89):

> Noyce couldn't get Fairchild's eastern owners to accept the idea that stock options should be part of compensation for all employees, not just for management. He wanted to tie everyone, from janitors to bosses, into the overall success of the company . . . This management style still sets the standard for every computer, software, and semiconductor company in the Valley today . . . Every CEO still wants to think that the place is run the way Bob Noyce would have run it. (Cringley 1993, p. 39)

Noyce's vision of a firm excluded the dress codes, reserved parking places, closed offices and executive dining rooms, along with the other trappings of status which were standard in virtually every hierarchical and bureaucratic US corporation. But when he tried to impress this vision upon the owners of Fairchild Semiconductor, he was rebuffed. The formation of Intel in 1968 was the ultimate result of the divergence in beliefs about how to organize and manage the firm.

The key development at Intel was the microprocessor. When longtime IBM employee Ted Hoff approached IBM and later DEC with his new microprocessor in the late 1960s, 'IBM and DEC decided there was no market. They could not imagine why anyone would need or want a small computer; if people wanted to use computers, they could hook into timesharing systems' (Palfreman and Swade 1991, p. 108).

2. THE GEOGRAPHY OF SOURCES AND INCENTIVES

2.1 The Selection Mechanism

At the heart of the evolutionary theory proposed by Richard Nelson and Sidney Winter (1982) is the selection mechanism in the economy across diverse alternatives. It is the existence of alternative and competing ways of doing things, ideas and proposed solutions – that is, diversity – which confronts economic agents and institutions with a choice. Most generally considered, diversity represents (1) the simultaneous existence of different possible actions, and (2) a differential in the valuation of potential actions by economic decision-makers. The selection mechanism in the economic process serves to select some of the proposed actions while rejecting others.

The selection mechanism occurs at many different levels. For example, an individual with several different ideas about what action to take must choose one to the exclusion of the others. That is, the most micro level of what Nelson and Winter (1982) term the 'selection environment' actually takes place within the context of the decision-making process by individual economic agents. A second locus of selection occurs within organization units of firms and a third across those organization units. Notice that all of these selection environments are inherently non-market in that they rely upon an internal mechanism for selecting across alternative actions. That is, given a set of diverse alternatives, first the individual economic agent and then the organization selects across those alternatives to eliminate some and choose others. An internal selection environment is based on subjective valuation in that the decision-making unit of observation – an individual or hierarchical bureaucracy – places an implicit value on one action.

The market provides another selection environment. As a result of the selection mechanism of the market, only a subset of the diverse alternatives tends to survive. A rich body of literature has shown that as the duration of any period increases, fewer of the original diverse alternatives tend to survive. At the same time, the conditional likelihood of surviving into the next time period increases as the number of time periods which have already been survived increases (Audretsch 1995).

For example, Audretsch (1995) shows both theoretically and empirically that the act of creating a new firm is the result of diversity with respect to the valuation of new ideas or potential innovations. My model suggests that (uncertain) knowledge asymmetries combined with high transaction costs result in individual economic agents deciding to start a new firm. Divergences in the expected value regarding new knowledge lead some economic agents to value any given idea (potential innovation) more than other agents, including those involved in the decision-making process of incumbent firms. When such divergences occur and an agent chooses to exercise what Albert O. Hirschman (1970) has termed exit – rather than voice or loyalty – and the agent departs from an incumbent organization to launch a new enterprise, then who is right, the departing agent or those agents in the organizational hierarchy who, by assigning the new idea a relatively low expected value, have effectively driven the agent with the potential innovation away? *Ex post* the answer may not be too difficult, but given the uncertainty inherent in new knowledge, the answer is anything but trivial a priori.

Thus, when a new firm is launched its prospects are shrouded in uncertainty. If the new firm is built around a new idea – that is, a potential innovation – it is uncertain whether there is sufficient demand for the new idea or whether some competitor will have the same or even a superior idea. An additional layer of uncertainty pervades a new enterprise: it is not known how competent the new firm really is in terms of management, organization and workforce. At least incumbent enterprises know something about their underlying competencies from past experience.

The initial condition of not just uncertainty but greater degree of uncertainty *vis-à-vis* incumbent enterprises in the industry is captured by Boyan Jovanovic (1982). Jovanovic presents a model in which the new entrants, which he refers to as entrepreneurs, face costs which not only are random but also differ across firms. A central feature of the model is that a new firm does not know what its cost function is (that is, its relative efficiency) but rather discovers this through the process of learning from its actual post-entry performance. In particular, Jovanovic (1982) assumed that entrepreneurs are unsure about their ability to manage a new firm start-up and therefore their prospects for success. Although entrepreneurs may launch a new firm based on a vague sense of expected post-entry performance, they only discover their true ability – in terms of managerial competence and of having based the firm on an idea which is viable in the market – once their business is established. Those entrepreneurs who discover that their ability exceeds their expectations expand the scale of their business, whereas those discovering that their post-entry performance is less than commensurate with their expectations will contract the scale of output and possibly exit the industry. Thus Jovanovic's (1982) model is a theory of noisy selection, in which efficient firms grow and survive and inefficient firms decline and fail.

The role of learning in the selection process has been the subject of considerable debate. On the one hand is the Lamarckian assumption that learning refers to adaptations made by the new enterprise. Those new firms which are the most flexible and adaptable will be the most successful in adjusting to whatever the demands of the market are. As Nelson and Winter (1982, p. 11) point out, 'Many kinds of organizations commit resources to learning; organizations seek to copy the forms of their most successful competitors.'

On the other hand is the interpretation that the role of learning is restricted to discovering if the new firm is viable in terms of the product as well as the production process. Under this interpretation the new enterprise is not necessarily able to adapt or adjust to market conditions, but receives information based on its market performance with respect to its fitness in terms of meeting demand most efficiently *vis-à-vis* rivals. The theory of organizational ecology proposed by Michael T. Hannan and John Freeman (1989, p. 132) most pointedly adheres to the notion that 'individual organizations are characterized by relative inertia in structure'. That is, firms learn not in the sense that they adjust their actions as reflected by their core identity and purpose, but in the sense that they adjust their perception. What is learned, then, is whether the firm has 'the right stuff', but not how to change that stuff.

Audretsch (1995) shows that the process of firm selection in markets apparently revolves around two driving selection mechanisms. The first is the gap between the size of the firm and the minimum efficient scale (MES) of output. The greater this gap is, the greater the growth rates of surviving firms tend to be but the smaller the likelihood of firm survival is. Since the variance of new firm start-up sizes is low relative to the variance in the MES levels of output, it is essentially the degree of scale economies which determines the extent of this gap and therefore the severity of this market selection mechanism.

The second selection mechanism in markets is the degree of uncertainty inherent in the nature of the product being sold and in how to produce it. In highly innovative industries this selection mechanism plays a more important role. In environments where innovation is relatively less important this selection mechanism plays less of a role.

2.2 Search versus Routine

As Knight (1921, p. 199) pointed out, uncertainty is the result of possessing only partial or bounded knowledge: 'The essence of the situation is action according to *opinion*, of greater or less foundation and value, neither entire ignorance nor complete and perfect information, but partial knowledge.' In fact, it is the fundamental condition of incomplete knowledge which leads Arrow (1983) to focus on the firm as an organization whose main distinction is processing information. As March and Simon (1993, p. 299) argue, 'Organizations process and channel information.' But as Arrow (1985, p. 303) emphasizes, 'The elements of a firm are *agents* among whom both decision making and knowledge are dispersed . . . Each agent observes a random variable, sometimes termed a signal . . . Each agent has a set of actions from which *choice is to be made. We may call the assignment of signals to agents the* information structure and the choice of decision rules the decision structure.'

How will economic agents and ultimately hierarchical organizations respond when confronted with incomplete knowledge? Knight's answer is 'differently' because agents differ in 'their capacity by perception and inference to form correct judgements as to the future course of events in the environment' (Knight 1921, p. 241). In addition, there are differences in 'men's capacities to judge means and discern and plan the steps and adjustments necessary to meet the anticipated future situation'. This is to say that different economic agents confronted with the same signal, in Arrow's

(1985) terms, or simply with incomplete information, in Knight's terms, will respond differently because they have different sets of experiences from which to evaluate that incomplete information.

Like Nelson and Winter (1982), March and Simon (1993, p. 309) emphasize the role of established routines in the functioning of organizations:

> The process of gaining individual expertise by coding experience into recognition/action pairs is paralleled by organizational processes for developing pairings between rules and situations . . . Organizations are collections of roles and identities, assemblages of rules by which appropriate behavior is paired with recognized situations . . . These are developed in an organization through collective experience and stored in the organizational memory as standard procedures . . . Organizations turn their own experience as well as the experience and knowledge of others into rules that are maintained and implemented despite turnover in personnel and without necessary comprehension of their bases. As a result, the processes for generating, changing, evoking, and forgetting rules become essential in analysing and understanding organizations.

As long as new information is consistent with the routines established in an organization, it will be processed by economic agents and a decision-making hierarchy in a manner which is familiar. New information under the routinized regime is familiar turf for organizations. A more fundamental problem arises, however, when the nature of that new information is such that it can no longer be processed by the familiar routines. Under these circumstances the organizational routines for searching out new relevant information and making (correct) decisions on the basis of that information break down. And it is under such information conditions that divergences tend to arise not only among economic agents in evaluating that information, but between agents and organizational hierarchies.

If each economic agent were identical, such divergences in beliefs would not arise. The greater the degree of homogeneity among agents, the greater the tendency will be for beliefs in evaluating uncertain information to converge. But individuals are not homogeneous. Rather, agents have varied personal characteristics and different experiences which shape the lens through which each agent evaluates where to get new information and how to assess it. That is, reasonable people confronted with the same information may evaluate it very differently, not just because they have different abilities but because each has had a different set of life experiences which affects the decision-making process. Perhaps this helps to explain why IBM, for all its collective knowledge, not to mention resources, was proven wrong about its early rejection of the minicomputer. Steve Jobs, a college dropout, was able to see something that the decision-making hierarchy at IBM did not. After all, Jobs emerged from the milieu of computer 'hackers' and 'freaks' in Northern California, which provided him with experience and knowledge unavailable to the IBM decision-makers, who generally lived in upper-middle-class East Coast residential areas such as White Plains, north of New York City.

Thus, to some extent the phenomenon of the establishment of a new firm represents not just imperfect information but a diverse population of economic agents. That is, diversity in the population of economic agents may ultimately lead to diversity in the types of firms populating the enterprise structure. And to some extent these diverse firms represent experiments based on differing visions about the product and how to produce it.

Diversity also is the source of the high degree of turbulence which is experienced in the United States and, increasingly, in other developed nations. That is, industrial markets are characterized by a high degree of churning. It took the two decades of the 1950s and 1960s for one-third of the Fortune 500 companies to be replaced by new additions. In the 1970s it took the entire decade to replace one-third of the Fortune 500. By contrast, in the 1980s it took just five years for one-third of the Fortune 500 to be replaced (Audretsch 1995).

3. THE SPATIAL INCOME DISTRIBUTION

3.1 Innovation

The emergence of a recent literature (re)discovering the importance of economic geography might seem paradoxical in a world increasingly dominated by e-mail, faxes and electronic communications superhighways. Why should geographic proximity matter when technology has advanced in a manner which has drastically reduced the cost of transmitting information across geographic space? The answer posited by Audretsch and Feldman (1996), Audretsch and Stephan (1996) and Feldman (1994a, 1994b) is based on a key distinction between information on the one hand and tacit knowledge on the other. Although the costs of transmitting information may be invariant to distance, the cost of transmitting knowledge and especially tacit knowledge rises with distance. Geographic location and proximity to the source matter in the transmission of tacit knowledge because face-to-face contact is the most effective and economical mode of transfer. Thus, Glaeser et al. (1992, p. 1127) characterize the Marshall–Arrow–Romer model as suggesting that 'intellectual breakthroughs must cross hallways more easily than oceans and continents'.

This model is consistent with anecdotal evidence. For example, a survey of nearly 1000 executives located in America's 60 largest metropolitan areas ranked Raleigh/Durham as the best city for knowledge workers and for innovative activity. *Fortune* magazine reports:

> A lot of brainy types who made their way to Raleigh/Durham were drawn by three top research universities. . . . US businesses, especially those whose success depends on staying atop new technologies and processes, increasingly want to be where hot new ideas are percolating. A presence in brain-power centers like Raleigh/Durham pays off in new products and new ways of doing business . . . Dozens of small biotechnology and software operations are starting up each year and growing like kudzu in the fertile business climate.

Considerable evidence has been found suggesting that location and proximity clearly matter in exploiting knowledge spillovers. Not only have Jaffe et al. (1993) found that patent citations tend to occur more frequently within the state in which they were patented than outside of that state, but Audretsch and Feldman (1996) found that the propensity of innovative activity to cluster geographically tends to be greater in industries where new economic knowledge plays a more important role.

3.2 Diversity versus Specialization

Despite the general consensus which has now emerged in the literature that knowledge spillovers within a given location stimulate technological advance, there is little consensus as to exactly how this occurs. The contribution of the knowledge-production function approach was simply to shift the unit of observation away from firms to a geographic region. But does it make a difference how economic activity is organized within the black box of geographic space? Political scientists and sociologists have long argued that differences in the cultures of regions may contribute to differences in innovative performance across regions, even holding knowledge inputs such as R&D and human capital constant. For example, Saxenian (1990) argues that a culture of greater interdependence and exchange among individuals in the Silicon Valley region has contributed to a superior innovative performance than is found around Boston's Route 128, where firms and individuals tend to be more isolated and less interdependent.

In studying the networks in California's Silicon Valley, Saxenian (1990, pp. 96–7) emphasizes that it is the communication between individuals which facilitates the transmission of knowledge across agents, firms and even industries, and not just a high endowment of human capital and knowledge in the region:

> It is not simply the concentration of skilled labor, suppliers and information that distinguish the region. A variety of regional institutions – including Stanford University, several trade associations and local business organizations, and a myriad of specialized consulting, market research, public relations and venture capital firms – provide technical, financial, and networking services which the region's enterprises often cannot afford individually. These networks defy sectoral

> barriers: individuals move easily from semiconductor to disk drive firms or from computer to network makers. They move from established firms to start-ups (or vice versa) and even to market research or consulting firms, and from consulting firms back into start-ups. And they continue to meet at trade shows, industry conferences, and the scores of seminars, talks and social activities organized by local business organizations and trade associations. In these forums, relationships are easily formed and maintained, technical and market information is exchanged, business contacts are established, and new enterprises are conceived . . . This decentralized and fluid environment also promotes the diffusion of intangible technological capabilities and understandings.[5]

Though economists tend to avoid attributing differences in economic performance to cultural differences, there has been a series of theoretical arguments suggesting that differences in the underlying structure between regions may account for differences in rates of growth and technological change. In fact, a heated debate has emerged in the literature about the manner in which the underlying economic structure within a geographic unit of observation might affect economic performance. One view, which Glaeser et al. (1992) attribute to the Marshall–Arrow–Romer eternality, suggests that an increased concentration of a particular industry within a specific geographic region facilitates knowledge spillovers across firms. This model formalizes the insight that the concentration of an industry within a city promotes knowledge spillovers between firms and therefore facilitates innovative activity. An important assumption of the model is that knowledge externalities with respect to firms exist, but only for firms within the same industry. Thus, the relevant unit of observation is extended from the firm to the region in the tradition of the Marshall–Arrow–Romer model and in subsequent empirical studies, but spillovers are limited to the relevant industry.

By contrast, restricting knowledge externalities to occur only within the industry may ignore an important source of new economic knowledge – interindustry knowledge spillovers. Jacobs (1969) argues that the most important source of knowledge spillovers are external to the industry in which the firm operates and that cities are the source of considerable innovation because the diversity of these knowledge sources is greatest in cities. According to Jacobs, it is the exchange of complementary knowledge across diverse firms and economic agents which yields a greater return on new economic knowledge. She develops a theory which emphasizes that the variety of industries within a geographic region promotes knowledge externalities and ultimately innovative activity and economic growth.

The extent of regional specialization versus regional diversity in promoting knowledge spillovers is not the only dimension over which there has been a theoretical debate. A second controversy involves the degree of competition prevalent in the region, or the extent of local monopoly. The

Marshall–Arrow–Romer model predicts that local monopoly is superior to local competition because it maximizes the ability of firms to appropriate the economic value accruing from their innovative activity. By contrast, Jacobs (1969) and Porter (1990) argue that competition is more conducive to knowledge externalities than is local monopoly.[6] It should be emphasized that by local competition Jacobs does not mean competition within product markets as has traditionally been envisioned within the industrial organization literature. Rather, Jacobs is referring to the competition for the new ideas embodied in economic agents. Not only does an increased number of firms provide greater competition for new ideas, but greater competition across firms also facilitates the entry of a new firm specializing in some particular and new product niche. This is because the necessary complementary inputs and services are likely to be available from small specialist niche firms but not necessarily from large, vertically integrated producers.

The first important test of the specialization versus diversity theories to date has focused not on gains or innovative activity but on employment growth. Glaeser et al. (1992) employ a data set on the growth of large industries in 170 cities between 1956 and 1987 in order to identify the relative importance of the degree of regional specialization, diversity and local competition in influencing industry growth rates. The authors find evidence which contradicts the Marshall–Arrow–Romer model but is consistent with Jacobs's theory. However, their study provided no direct evidence on whether diversity is more important than specialization in generating innovation.

Feldman and Audretsch (1999) identify the extent to which the organization of economic activity either is concentrated or, alternatively, consists of diverse but complementary economic activities, and how this composition influences innovative output. We ask the question: does the specific type of economic activity undertaken within any particular geographic concentration matter? To consider this question we link the innovative output of product categories within a specific city to the extent to which the economic activity of that city is concentrated in that industry or, conversely, diversified in complementary industries sharing a common science base.

To systematically identify the degree to which specific industries share a common underlying science and technology base, Feldman and Audretsch (1999) rely upon a deductive approach which links products estimated from their closeness in technological space. They use the responses of industrial R&D managers to a survey by Levin et al. (1987). To measure the significance of a scientific discipline to an industry, the survey asked: 'How relevant were the basic sciences to technical progress in this line of business over the past 10–15 years?' The survey uses a Likert scale of 1 to 7, from least important to most important, to assess the relevance of basic scientific research in biology, chemistry, computer science, physics, mathematics, medicine, geology, mechanical engineering and electrical engineering. Any academic discipline with a rating greater than 5 is assumed to be relevant to a product category. For example, basic scientific research in medicine, chemistry and chemical engineering is found to be relevant to product innovation in drugs (SIC 2834).

Feldman and Audretsch (1999) then use cluster analysis to identify six groups of industries which rely on similar rankings for the importance of different academic disciplines. These six groups reflect distinct underlying common scientific bases.

To test the hypothesis that the degree of specialization or, alternatively, diversity as well as the extent of local competition within a city shapes the innovative output of an industry, Feldman and Audretsch (1999) estimate a model where the dependent variable is the number of innovations attributed to a specific four-digit SIC industry in a particular city. To reflect the extent to which economic activity within a city is specialized, we include as an explanatory variable a measure of industry specialization which was used by Glaeser et al. (1992) and is defined as the 1982 share of total employment in the city accounted for by industry employment in the city, divided by the share of US employment accounted for by that particular industry. This variable reflects the degree to which a city is specialized in a particular industry relative to the degree of economic activity in that industry which would occur if employment in the industry were randomly distributed across the United States. A higher value of this measure indicates a greater degree of specialization of the industry in that particular city. Thus, a positive coefficient would indicate that increased specialization within a city is conducive to greater innovative output and would support the Marshall–Arrow–Romer thesis. A negative coefficient would indicate that greater specialization within a city impedes innovative output and would support Jacobs's theory that diversity of economic activity is more conducive to innovation than is specialization of economic activity.

To identify the impact of an increased presence of economic activity in complementary industries sharing a common science base on the innovative activity of a particular industry within a specific city, a measure of the presence of science-based related industries is included. This measure is constructed analogously to the index of industry specialization, and is defined as the share of total city employment accounted for by employment in the city in industries sharing the science base, divided by the share of total US employment accounted for by employment in that same science base. This variable measures the presence of complementary industries relative to what the presence would be if those related industries were distributed randomly across the United States. A positive coefficient of the presence of science-based related industries would indicate that a greater presence of complementary industries is conducive to greater innovative output and would lend support to the diversity thesis. By contrast, a negative coefficient would suggest that a greater presence of related industries sharing the same science base impedes innovation and would argue against Jacobs's diversity thesis.

The usual concept of product market competition in the industrial organization literature is typically measured in terms of the size-distribution of firms. By contrast, Jacobs's concept of localized competition emphasizes instead the extent of competition for the ideas embodied in individuals. The greater the degree of competition among firms, the greater will be the extent of specialization among those firms and the easier it will be for individuals to pursue and implement new ideas. Thus the metric relevant to reflect the degree of localized competition is not the size of the firms in the region relative to their number (because, after all, many if not most manufacturing product markets are national or at least interregional in nature) but rather the number of firms relative to the number of workers. In measuring the extent of localized competition we again adopt a measure used by Glaeser et al. (1992), which is defined as the number of firms per worker in the industry in the city relative to the number of firms per worker in the same industry in the United States. A higher value of this index of localized competition suggests that the industry has a greater number of firms per worker relative to its size in the particular city than it does elsewhere in the United States. Thus, if the index of localized competition exceeds 1, then the city is locally more competitive than other American cities.

In Feldman and Audretsch (1999) the regression model is estimated based on the 5946 city-industry observations for which data could be collected. The Poisson regression estimation method is used because the dependent variable is a limited dependent variable with a highly skewed distribution. By focusing on innovative activity for particular industries at specific locations, Feldman and Audretsch (1999) find compelling evidence that specialization of economic activity does not promote innovative output. Rather, the results indicate that diversity across complementary economic activities sharing a common science base is more conducive to innovation than is specialization.

CONCLUSIONS

An important impact of globalization has been to shift the comparative advantage of the leading developed nations in Western Europe and North

America away from routinized economic activity towards search activity. An important implication of globalization is that in a world where diffusion costs are relatively low and large wage differentials exist across geographic space, routine economic activity tends to be transferred out of the high-cost *Standort* (location) to lower-cost locations. The telecommunications revolution has rendered this just as true for information-based economic activity as for manufacturing activities.

Income differentials across geographic space can only be maintained by engaging in economic activity in which the cost of diffusing that activity across space is high. Whereas the cost of diffusing routine economic activity across geographic space is relatively low, the cost of diffusing search activity is high. Thus, the comparative advantage of high-wage locations is shifting to knowledge-based search activity and away from routine activity.

Homogeneity, in both the underlying population and the enterprise structure, is more conducive to routinized economic activity. Homogeneity across economic agents reduces the cost of transactions, resulting in efficiency gains for routinized economic activity.

By contrast, heterogeneity, in both the underlying population and the enterprise structure, is more conducive to knowledge-based innovative activity. Such diversity is the driving force behind knowledge spillovers. Thus, those geographic regions which comprise diverse economic agents engaged in knowledge-based economic activity, which does not costlessly diffuse across space, are likely to experience rapid increases in income, while those regions based on homogeneous economic agents engaged in routinized economic activity are likely to experience a relative decline in income.

NOTES

1. I would like to thank Professor Erik Reinert for his helpful suggestions along with those made by the other participants in the Oslo conference.
2. 'The valley of money's delight', *The Economist*, 29 March 1997, special section, p. 1.
3. 'System error', *The Economist*, 18 September 1993, p. 99.
4. Paul Carrol, 'Die offene Schlacht', *Die Zeit*, **39**, 24 September 1993, p. 18.
5. Saxenian (1990, pp. 97–8) claims that even the language and vocabulary used by technical specialists is specific to a region: 'A distinct language has evolved in the region and certain technical terms used by semiconductor production engineers in Silicon Valley would not even be understood by their counterparts in Boston's Route 128.'
6. Porter (1990) provides examples of Italian ceramics and gold jewellery industries in which numerous firms are located within a bounded geographic region and compete intensively in terms of product innovation rather than focusing on simple price competition.

REFERENCES

Alchian, Almerin (1950), 'Uncertainty, evolution, and economic theory', *Journal of Political Economy*, **58**, 211–21.

Alchian, Almerin and H. Demsetz (1972), 'Production, information costs, and economic organization', *American Economic Review*, **62**, 777–95.

Arrow, Kenneth J. (1962), 'Economic welfare and the allocation of resources for invention', in R.R. Nelson (ed.), *The Rate and Direction of Inventive Activity*, Princeton, NJ: Princeton University Press, pp. 609–26.

Arrow, Kenneth J. (1983), 'Innovation in large and small firms', in J. Ronen (ed.), *Entrepreneurship*, Lexington, MA: Lexington Books, pp. 15–28.

Arrow, Kenneth J. (1985), 'Informational structure of the firm', *American Economic Review*, **75** (2), 303–7.

Audretsch, David B. (1995), *Innovation and Industry Evolution*, Cambridge, MA: MIT Press.

Audretsch, David B. and Maryann P. Feldman (1996), 'R&D spillovers and the geography of innovation and production', *American Economic Review*, **86** (3), 630–40.

Audretsch, David B. and Paula E. Stephan (1996), 'Company–scientist locational links: the case of biotechnology', *American Economic Review*, **86** (3), 641–52.

Coase, R.H. (1937), 'The nature of the firm', *Economica*, **4** (4), 386–405.

Cohen, Wesley M. and Steven Klepper (1991), 'Firm size versus diversity in the achievement of technological advance', in Z. Acs and David B. Audretsch (eds), *Innovation and Technological Change: An International Comparison*, Ann Arbor, MI: University of Michigan Press, pp. 183–203.

Cohen, Wesley M. and Steven Klepper (1992a), 'The anatomy of industry R&D intensity distributions', *American Economic Review*, **82** (4), 773–99.

Cohen, Wesley M. and Steven Klepper (1992b), 'The tradeoff between firm size and diversity in the pursuit of technological progress', *Small Business Economics*, **4** (1), 1–14.

Cringley, Robert X. (1993), *Accidental Empires: How the Boys of Silicon Valley Make Their Millions, Battle Foreign Competition, and Still Can't Get a Date*, New York: Harper Business.

Feldman, Maryann P. (1994a), *The Geography of Innovation*, Boston, MA: Kluwer Academic Publishers.

Feldman, Maryann P. (1994b), 'Knowledge complementarity and innovation', *Small Business Economics*, **6** (5), 363–80.

Feldman, Maryann P. and David B. Audretsch (1999), 'Innovation in cities: science-based diversity, specialization and localized competition', *European Economic Review*, **43**, 409–29.

Ferguson, Charles H. (1988), 'From the people who brought you voodoo economics', *Harvard Business Review*, **87** (May/June), 55–62.

Gilder, George (1989), *Microcosm*, New York: Touchstone.

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman and Andrei Shleifer (1992), 'Growth of cities', *Journal of Political Economy*, **100** (4), 1126–52.

Graf, J. de V. (1957), *Theoretical Welfare Economics*, Cambridge: Cambridge University Press.

Griliches, Zvi (1979), 'Issues in assessing the contribution of R&D to productivity growth', *Bell Journal of Economics*, **10** (1), 92–116.

Hannan, Michael T. and John Freeman (1989), *Organizational Ecology*, Cambridge, MA: Harvard University Press.

Hirschman, Albert O. (1970), *Exit, Voice, and Loyalty*, Cambridge, MA: Harvard University Press.
Holmstrom, Bengt (1989), 'Agency costs and innovation', *Journal of Economic Behavior and Organization*, **12**, 305–27.
Jacobs, Jane (1969), *The Economy of Cities*, New York: Random House.
Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *Quarterly Journal of Economics*, **63** (3), 577–98.
Jovanovic, Boyan (1982), 'Selection and the evolution of industry', *Econometrica*, **50** (2), 649–70.
Knight, Frank H. (1921), *Risk, Uncertainty and Profit*, New York: Houghton Mifflin.
Kreps, David (1991), 'Corporate culture and economic theory', in J. Alt and K. Shepsle (eds), *Positive Perspectives on Political Economy* , Cambridge: Cambridge University Press, pp. 119–33.
Levin, R.C, A.K. Klevorick, R.R. Nelson and S.G. Winter (1987), 'Appropriating the returns from industrial research and development', *Brookings Papers on Economic Activity*, 783–820.
March, James G. and Herbert A. Simon (1993), 'Organizations revisited', *Industrial and Corporate Change*, **2** (3), 299–316.
Milgrom, Paul and John Roberts (1987), 'Information asymmetries, strategic behavior, and industrial organization', *American Economic Review*, **77** (2), 184–93.
Nelson, Richard R. and Sidney G. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.
Palfreman, Jon and Doron Swade (1991), *The Dream Machine: Exploring the Computer Age*, London: BBC Books.
Porter, Michael (1990), *The Competitive Advantage of Nations*, New York: Free Press.
Saxenian, Anna Lee (1990), 'Regional networks and the resurgence of Silicon Valley', *California Management Review*, **33** (1), 89–112.
Scherer, F.M. (1991), 'Changing perspectives on the firm size problem', in Z.J. Acs and D.B. Audretsch (eds), *Innovation and Technological Change: An International Comparison*, Ann Arbor, MI: University of Michigan Press, pp. 108–18.
Schumpeter, Joseph A. (1911), *Theorie der wirtschaftlichen Entwicklung: Eine Untersuchung über Unternehmergewinn, Kapital, Kredit, Zins und den Konjunkturzyklus*, Berlin: Duncker & Humblot.
Schumpeter, Joseph A. (1942), *Capitalism, Socialism and Democracy*, New York: Harper and Row.
Williamson, Oliver E. (1975), *Markets and Hierarchies: Antitrust Analysis and Implications*, New York: Free Press.