

**COMPETITION TRIBUNAL**

**IN THE MATTER OF** the *Competition Act*, R.S.C. 1985, c. C-34, as amended;

**IN THE MATTER OF** an application by B-Filer Inc., B-Filer Inc. doing business as GPAY GuaranteedPayment and Npay Inc. for an order pursuant to section 103.1 granting leave to make application under sections 75 and 77 of the *Competition Act*;

**AND IN THE MATTER OF** an application by B-Filer Inc., B-Filer Inc. doing business as GPAY GuaranteedPayment and Npay Inc. for an interim order pursuant to section 104 of the *Competition Act*.

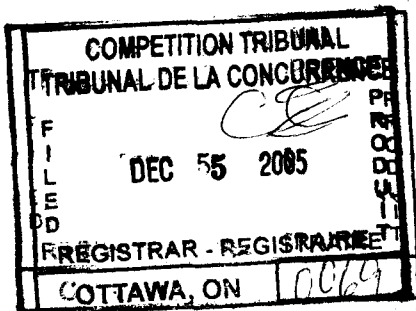
BETWEEN:

**B-FILER INC., B-FILER INC. doing business as  
GPAY GUARANTEEDPAYMENT and NPAY INC.**

Applicants

**THE BANK OF NOVA SCOTIA**

Respondent



**AFFIDAVIT OF AIDAN HOLLIS**  
Affirmed December 2<sup>nd</sup>, 2005

I, **AIDAN HOLLIS**, of the City of Calgary in the Province of Alberta AFFIRM AND SAY AS FOLLOWS:

1. I hold the position of Associate Professor in the Department of Economics at the University of Calgary. My research and teaching focus on industrial organization, and as such relate to the matters discussed in this affidavit. In 2003-4 I was the T.D. MacDonald Chair in Industrial Economics at the Competition Bureau.
2. My curriculum vitae is attached hereto as **Exhibit A**.
3. For the purposes of preparing this Affidavit, I have reviewed the following documents:

a. Reasons for previous Order dated November 4, 2005, granting leave to apply only under section 75 of the Competition Act, of Justice Simpson, in the matter of *B-Filer Inc., B-Filer Inc. doing business as Gpay Guaranteedpayment and Npay Inc. v. The Bank of Nova Scotia*, CT-2005-006, dated November 14, 2005;

b. an Affidavit of Raymond Grace, sworn June 15, 2005; and

c. an Affidavit of Raymond Grace, sworn September 1, 2005.

4. I was asked by the Applicants to address the question: is it possible to quantify the harm to the Applicants caused by being cut off from access to banking services provided to them by Scotiabank. My response is based on the documents reviewed and my professional knowledge and experience.

5. I understand from these documents that the applicants rely on bank accounts with Scotiabank as an essential service for a large component of their business. I assume in what follows that the applicants are unable to replace the accounts with Scotiabank through any other institution.

6. An estimate of the losses that would be incurred by the applicants from not having access to banking services provided by Scotiabank would have two components. First, one would need to be able to extrapolate lost profits for the period of the trial. Second, one would need to be able to extrapolate how lost market share during the period of the trial would likely affect the evolution of market share (and hence profits) following the trial. I examine these in turn.

***Extrapolating past revenues to the period of the trial***

7. In cases where past revenues have been relatively constant, it seems reasonable to extrapolate past revenues over the period of the trial. There are, however, two reasons to be suspect of this approach for the applicants' case. First, economic analysis of growth of small business shows that it is in general very difficult to predict growth rates of small businesses based on past growth rates. Caves (1998), attached as **Exhibit B** hereto, in an article summarizing the economic literature on firm growth rates, notes that "corporate growth rates are not quite random" and that variability in growth rates, as in exit rates, is greater for small firms. Thus, for a small, young firm such as the applicants, the economic literature suggests that it will be very difficult to make reliable predictions about future revenues.

8. Second, the history of the applicants suggests that it will be difficult to extrapolate. I have been informed by Raymond Grace and verily believe that the monthly volume of payments made by the Applicants through Scotiabank increased by over twelve hundred percent (1200%) from January 2004 until September 2005. The volatility and very high growth rate of this business over the course of twenty-one (21) months suggest to me that it would be extremely difficult to use past volumes to predict future volumes of business, even over a relatively short period.

9. Third, this high variance in payment volumes is likely to be in part due to the immaturity of the industry in which it operates. Future demand may very likely depend in part on growth in consumer demand for such payment systems, and will also in part depend on the development of competing technologies, all of which are also relatively new.

10. In these circumstances, it seems to me to be very difficult to estimate the harm to the applicants of not having access to Scotiabank accounts in performing their business, even over the short term.

***Extrapolating lost market share to the period beyond the trial***

11. The second component to estimating possible harm to the applicants involves extrapolating from estimates of lost market share during the period of the trial to permanent impact on market position following the trial. I have been informed by Raymond Grace and verily believe that the business the applicants are involved in has evolved and continues to evolve rapidly, with many new competing entrants arriving over the past two years. In these circumstances, timing is important, as obtaining a strong market position early may allow the firm to become well known and to continue in a strong position.

12. The reasons for such an effect are well documented in the economics literature and are briefly discussed in Arthur (1996), attached as **Exhibit C** hereto. First, there is a “network effect”: the more consumers use a service, the more merchants will offer it, and so the more consumers will use it. In the case of the applicants, they must persuade merchants to offer their payment mechanism. Merchants will only do so if the payment mechanism is effective and can be used by a sufficient number of clients. If clients find they cannot use the service, they will use a different service in the future. Merchants, in turn, may decide not to offer the applicants’ payment mechanism, leading to a downward spiral in sales for the applicants. (Similarly, there can be an upward spiral as merchants, seeking to match service offerings with competitors, offer the applicant’s payment mechanism because of its widespread use.) Second, there is a reputation

effect: the more consumers have used a service, the greater trust customers have in it in the future. This is obviously likely to be important for the applicant's business. Thus, obtaining market share early, in a growing industry in which network effects and reputation are both important, seems likely to be critical in obtaining market share in the future.

13. On this basis, if the applicants are unable to service a substantial part of their customer base for a period of several months, I would expect them to lose market share in the long term, unless their technology and service is so superior to that of all their existing and potential competitors that they are able to re-attract past clients following resolution of the litigation.

14. How large would such a loss in long term market share be? Such a question is inevitably highly speculative in such a dynamic industry.

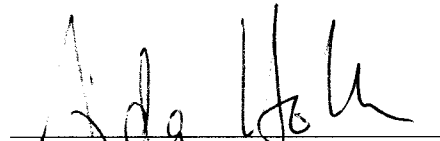
**Summary**

15. In summary, in my opinion the potential harms to the applicant from losing access to banking services from Scotiabank would be unquantifiable, given the growing nature of the industry in which they operate, the volatility and high growth rate of their own business history, and the importance of network and reputation effects in their industry.

**AFFIRMED BEFORE ME**  
at the City of Calgary  
in the Province of Alberta  
on this 2<sup>nd</sup> day of December 2005

  
Commissioner for Taking Affidavits

Name:

  
AIDAN HOLLIS

COMMISSIONER FOR  
OATHS #0694001  
DEBBIE PIAZZA  
EXPIRES FEB 06 / 08

Exhibit A

Resume of Aidan Hollis

THIS IS EXHIBIT A TO  
THE AFFIDAVIT OF \_\_\_\_\_  
AIDAN HOLLIS  
DATED THIS 2 DAY, Dec, 2005

COMMISSIONER FOR  
OATHS #0694001  
DEBBIE PIAZZA  
EXPIRES FEB 06/08  
*Debbie K Piazza*

# Aidan Hollis

Department of Economics  
University of Calgary  
2500 University Drive  
Calgary, T2N 1N4 Canada

Office Tel: (403) 220-5861  
Home Tel: (403) 282-1480  
Facsimile: (403) 220-5861  
e-mail: ahollis@ucalgary.ca

---

## Education

- 1990 - 1996 Ph.D., Economics, University of Toronto  
*Dissertation: The Application of Economic Theory to Competition Policy*
- 1989 - 1990 M.A., Economics, University of Toronto
- 1982 - 1985 M.A., English, Cambridge University

## Professional Experience

- 2003 - 2004 T.D. MacDonald Chair of Industrial Economics,  
Competition Bureau, Industry Canada, Ottawa
- 2003 - Associate Professor (with tenure), University of Calgary
- 1996 - 2003 Assistant Professor, University of Calgary
- 1994 - 1995 Instructor, University of Toronto
- 1987 - 1989 Loans Officer, Midland Bank Seoul Branch
- 1985 - 1987 Trainee, Midland Bank International Division, London

## Awards

### Honours

- 2005- Academic Director, Centre for Regulatory Economics,  
Van Horne Institute
- 2004 "Eminent Scientist of the Year" in the field of Health  
Economics for Canada, as chosen by International  
Research Promotion Council ([www.irpc.org](http://www.irpc.org))

2002 - Research Fellow, Institute of Health Economics

### **Research Grants**

Department of Economics Computer Grant (\$1,300) 2002

University of Calgary SSHRC Development Grant (\$1,000) 2002

SSHRC Standard Research Grant (\$50,450) 1999-2002

University Research Grants Committee, Starter Grant (\$9,376) 1997 - 1999

University of Calgary "Research Excellence Envelope" (\$6,000) 1996 - 1997

### **Scholarships**

Institute of Health Economics Research Grant (\$5,500) 2005

SSHRC Fellowship, 1992 - 1994

Ontario Graduate Scholarship, 1991 - 1992

Connaught Fellowship, 1990 - 1992

Tallow Chandlers' Award, 1984

## **Publications**

### Peer-Reviewed Articles

- Hollis, Aidan, "How do 'authorized generics' affect drug prices?" *Review of Industrial Organization*, forthcoming.
- Hollis, Aidan and Peter Ibbott, "How Parallel Trade Affects Drug Policies and Prices in Canada and the United States." *American Journal of Law & Medicine*, forthcoming.
- Hollis, Aidan, "Comment on 'The Economics of Follow-on Drug Research and Development' by diMasi and Paquette." *Pharmacoeconomics*, forthcoming.
- Boyce John R. and Hollis, Aidan, "Governance of Electricity Transmission Systems" *Energy Economics* 27(2): 237-255, March 2005.
- Hollis, Aidan and Steven Law, "A National Formulary for Canada." *Canadian Public Policy*, 30(4): 445-452, December 2004.
- Hollis, Aidan, and Aslam Anis "Rx for Canada: Close the Internet Pharmacies" C.D.Howe Commentary 205, October 2004.
- Hollis, Aidan, How cheap are Canada's drugs really? *Journal of Pharmacy and Pharmaceutical Sciences*, 7(2): 215-216, June 2004. Reprinted in *Pharma Marketing News* 3(8), September 2004.

- Hollis, Aidan, and Arthur Sweetman. "Microcredit and Famine: The Irish Loan Funds during the Great Famine" *World Development*, 32(9): 1509-1523, September 2004.
- Hollis, Aidan and Lasheng Yuan, "Competition Policy in Open Economies" *International Economic Journal*, 18(2): 179-194, June 2004.
- Hollis, Aidan, "How Brand-Controlled 'Pseudo-Generics' Deter True Generics in the Canadian Pharmaceutical Market," *Canadian Public Policy* 29(1): 21-31, March 2003.
- Hollis, Aidan, "Industrial Concentration, Output, and Trade: An Empirical Exploration," *Review of Industrial Organization*, 23(2): 103-119, March 2003.
- Hollis, Aidan, "The Importance of Being First: Evidence from Canadian Generic Pharmaceuticals," *Health Economics* 11: 723-734, December 2002.
- Hollis, Aidan, "The Link between Publicly Funded Health Care and Compulsory Licensing," *Canadian Medical Association Journal* 167: 765-767, October 2002.
- Hollis, Aidan, "The Strategic Implications of Learning by Doing," *International Journal of the Economics of Business* 9(2): 157-174, July 2002.
- Hollis, Aidan, and Arthur Sweetman, "The Life-Cycle of a Microfinance Institution: An Economic Analysis of the Irish Loan Funds," *Journal of Economic Behavior and Organization* 34(3): 291-311, November 2001.
- Hollis, Aidan, "Coauthorship and the Output of Academic Economists," *Labour Economics* 8(4): 503-530, October 2001.
- Hollis, Aidan, "Adverse Selection, Extended Warranties, and Aftermarkets," *Journal of Risk and Insurance*, 66(3): 321-344, September 1999.
- Gallini, N.T., and Aidan Hollis, "A Contractual Approach to Gray Markets," *International Review of Law and Economics*, 19(1): 1-21, March 1999.
- Hollis, Aidan, and Arthur Sweetman, "Microcredit: What Can We Learn From the Past?" *World Development* 26(10): 1875-1891, October 1998.
- Hollis, Aidan, and Arthur Sweetman, "Microcredit in Pre-Famine Ireland," *Explorations in Economic History* 35: 347-380, October 1998.

#### Magazine Articles and Book Chapters

- Hollis, Aidan, "Neglected Disease Research: Health Needs and New Models for R&D." in P. Illingworth, U. Schlukenk, and G.C. Cohen, *The Power of Pills*, Pluto Press, forthcoming.



- Hollis, Aidan, "Close down our web pharmacies." *National Post* October 29 2004, p. A16.
- Hollis, Aidan, "The economics of compulsory licensing of essential medicines." *Recent Advances and Research Updates*, September 2004, 5(2): 219-225.
- Hollis, Aidan, Review of *An Irish Working Class: Explorations in Political Economy and Hegemony, 1800-1950* by Marilyn Silverman. *Journal of Interdisciplinary History* 33(3): 469-470, Winter 2003.
- Hollis, Aidan, "The use of the 'pseudo-generic' strategy in the Canadian pharmaceutical market." *World Markets Series Business Briefing: Pharmagenics*, September 2002, pp. 99-102.
- Hollis, Aidan, "Women and Microcredit in History: Gender in the Irish Loan Funds," in Gail Campbell, Beverly Lemire and Ruth Pearson, eds., *Women and Credit: Researching the Past, Refiguring the Future*, Oxford: Berg Press, 2002, pp. 73-89.
- Hollis, Aidan, "Closing the FDA's Orange Book," *Regulation* 24(4): 14-17, Winter 2001.

## Presentations in Conferences and Universities

I have made over 60 presentations at conferences and universities.

## Teaching

I have taught graduate and undergraduate courses in microeconomics, industrial organization, and economic regulation at the Universities of Calgary and Toronto.

## Supervision of Graduate Students

### Current students

- E. Sawatzky, MEd project, 2005, "The effect of intellectual property laws on pharmaceutical research location."

### Past students

- M. Tepczynska, MA thesis, 2003, "An Economic Analysis of the Regulating Reserves Market in Alberta."
- W. Brown, MEd project, 2002, "Strategic Issues in the Electrical Power Generation Industry."
- H. Wang, MA thesis, "An Empirical Study of Economies of Scale in the Chinese Railway System."
- R. Williams, MA thesis, 2002, "Williams on Electricity."

- K Aksomitis, MA thesis, 2002, "Strategic Behavior in the Alberta Electricity Market."

I have also acted as examiner for several PhD candidates at the Universities of Calgary and Toronto.

## Other Professional Activities

- I have acted as referee for over twenty journals, for books, institutes, and granting agencies.
- I founded and have organized the Alberta Industrial Organization Conference (now in its third year).
- I have also provided economic commentary from time to time on television news programs and in newspapers.

## Consulting

- Recognized as Expert Witness in Federal Court of Canada.
- Pharmaceutical clients: Apotex, Mylan, Nu-Pharm, Cobalt, Government of Alberta.
- Energy clients: City of Calgary, Balancing Pool (Alberta), ESBI (Alberta), Van Horne Institute.
- Antitrust clients: Industry Canada, Competition Bureau, Dynastream, Prolog Consulting.

Current as of December 2, 2005

**Exhibit B**

This Exhibit contains an article by Caves from 1998.

THIS IS EXHIBIT B TO  
THE AFFIDAVIT OF \_\_\_\_\_  
AIDAN HULLIS  
DATED THIS 2 DAY, Dec, 2005

COMMISSIONER FOR  
OATHS #0694001  
DEBBIE PIAZZA  
EXPIRES FEB 06/08  
*Debra K. Piazza*

# Industrial Organization and New Findings on the Turnover and Mobility of Firms

RICHARD E. CAVES<sup>1</sup>

## 1. Introduction

ALTHOUGH RESEARCH on the turnover of business units has a long tradition, primary data on the full populations of business units (firms, establishments) present in nations' markets were inaccessible until recently. Only in the past decade have economists picked the locks on the doors of numerous national census bureaus and organized the primary records so that the births, deaths, and life trajectories of individual business units can be traced. Commercial data bases that claim similar coverage have also come into use for research on turnover and mobility.

This research has borne as its first fruit a great outpouring of stylized facts where no more than impressions had existed before. Although the importance of these facts for economic behavior and performance is manifest, their development has not been theory-driven. Indeed, identifying the theoretical models on which the stylized facts shed light is itself an exercise in hunting and gather-

ing. The empirical evidence aligns with some obviously salient models of (e.g.) the effect of firms' random growth rates on their industry's concentration and the decisions of potential entrant firms uncertain about their prospective cost levels. But it also spotlights some non-obvious theoretical referents, such as the theory of real options (to explain the varying resource commitments made by entering firms) and the theory of job-matching (to explain the productivity of changes in control of business units).

The newly accessible data on turnover and mobility have attracted economists specializing in several fields. This survey's emphases are driven by the field of industrial organization, although with passing attention to work attuned to labor economics and macroeconomics.<sup>2</sup> Accordingly, the first section summarizes the recently accumulated stylized facts about entry, exit, and the mobility

<sup>2</sup> A sociology-based literature on organizational ecology (e.g. Michael T. Hannan and John Freeman 1989) will be neglected. To an economist's eyes, its treatment of turnover in business populations suffers from eschewing simple priors about business behavior: intended profit-maximization and the need to cover costs to keep a firm's coalition together. But there are compensating strengths, such as the analysis of reasons why a failing coalition frequently cannot imitate the *modus operandi* of a winner (Hannan and Freeman 1989, chap. 4).

<sup>1</sup> Harvard University. Thanks for helpful comments and suggestions to David Audretsch, John R. Baldwin, Tito Boeri, Steve Davies, Paul Geroski, Michael Gort, John Haltiwanger, John E. Jackson, Steve Klepper, Jose Mata, Anita McGahan, Robert McGuckin, Ariel Pakes, Mark J. Roberts, Frank Wolak, and anonymous referees.

of business units. Some of the findings are conditioned on the unit's specific market environment, but most of them deal in overall or average patterns. The second section brings this market context to the fore and seeks to integrate the new findings about turnover and mobility with industrial organization's traditional framework, based on static partial-equilibrium models of markets. The third section relates the new evidence to other lines of empirical research on productivity and productive efficiency, again in the context of firms competing in particular markets.

The data bases that researchers have employed are mostly longitudinal versions of national census data bases. A number of countries are represented, including developing nations. The bulk of the research, and hence this survey's attention, pertains to the United States and Canada; an effort will be made to expose international comparative conclusions, although the market mechanisms at work appear overwhelmingly similar from country to country. Some proprietary data bases have also been employed, and until recently most research on turnover unavoidably depended on data for large, publicly traded companies. Important differences exist among these data bases. Some of them are organized around firms, some around establishments or business units. Many pertain only to manufacturing industries, some to all nonfinancial sectors. They differ importantly in their coverage of small business units: some only sample small units or cut off at some threshold (such as 10 employees), while a few include single-person enterprises. Integrating this research requires many judgment calls on where data-base differences are and are not important. In the first section of this paper, the reader should assume that stated conclusions describe un-

weighted-average patterns drawn from censuses that include quite small firms; because of the sharply skewed distribution of firm sizes, smaller units therefore dominate the conclusions. That fact somewhat mitigates another difference among the data bases: some focus on firms, some on plants that might be either independent firms or dependent units of a multiplant firm. Most small establishments are independent firms, diluting any distortions due to units' heterogeneous independence.

Several forms of turnover are addressed in this research, and terminology for identifying them is not fully standardized. We use "turnover" as a general term to embrace three processes: the births and deaths of business units ("entry and exit"), variations in sizes and market shares of continuing units ("mobility"), and shifts between enterprises in the control of continuing business units ("changes in control"). An "establishment" is a plant (in some instances a line of business), and a firm is an independent legal entity (in some instances a set of plants under common control). A "business unit" might be either an establishment or a firm; research in this field commonly employs populations of establishments that include both free-standing single-plant firms and those plants controlled by multiplant firms. We shall not stress the firm/plant distinction except where appropriate to point out differences between the behavior of single-plant firms and of multi-plant firms' dependent plants.

## 2. *Mobility, Entry, and Exit in Populations of Firms and Establishments*

### 2.1 *Background: The Law of Proportionate Effect*

Although the research on turnover is not strongly theory-driven, many earlier

empirical studies of the mobility of firms were shaped by a purely statistical model resting on the hypothesis known as the Law of Proportionate Effect (LPE) or Gibrat's Law: if growth rates of firms in a fixed population are independent of their initial sizes, the variance of growth rates shows no heteroskedasticity with size, and serial correlation of growth rates is absent, then the concentration of the population increases without limit. Many studies sought to test the independence and serial-correlation assumptions themselves, the implication of lognormally distributed firm sizes, or the implication of increasing concentration as the variance of the firm size increases. We can rely on John Sutton's (1997) survey of "Gibrat's legacy" for a review of recent econometric and theoretical advances from Gibrat's beachhead. The original formulation abstracted from the occurrence of entry and exit, which can be integrated into the random-process model in various ways. For organizing the recent profusion of stylized facts, what matters are the empirically established exceptions to and extensions of Gibrat's simple model (basic references include David S. Evans 1987a, 1987b; Bronwyn Hall 1987; Jonathan S. Leonard 1986; M. S. Kumar 1985; and surveys by Tito Boeri 1989 and Joachim Wagner 1992).

1. The variance of firms' proportional growth rates is not independent of their size but diminishes with it. This heteroskedasticity is not always evident among the largest firms but appears in samples that reach below Fortune 500 giants (see Peter E. Hart and Nicholas Oulton 1996; David B. Audretsch 1995a, chap. 4; and Jose Mata 1994). Evans (1987a) reported that the variability over time of a firm's growth rate decreases with size in

85 of 100 U.S. manufacturing industries.<sup>3</sup>

2. Mean growth rates of surviving firms are not independent of their sizes but tend to decline with size and also with the unit's age (given size). Evans (1987a) found the mean growth rate to decline significantly with size in 89 percent of the individual industries that he studied. The pattern is consistent with Galtonian regression and of course removes Gibrat's implication of increasing concentration (for example, Jozef Konings 1995b).

3. Entry and exit are intimately involved in growth-size relations. Entry is more likely to occur into smaller size classes, and the likelihood of a unit's exit declines with its size.

4. Units' growth rates may well be serially correlated, but earlier studies disagreed on the pattern. This is not surprising, as autocorrelation of changes in units' sizes may be regarded either as a statistical error-correction process (negative autocorrelation) or a cumulative economic response to the unit's perceived competitive advantage or disadvantage (positive). For large British firms, Kumar (1985) observed autocorrelation in growth rates measured over blocks of several (usually five) years that is positive but declining over time. Paul Dunne and Alan Hughes (1994), also working with a large sample of British companies and five-year periods,

<sup>3</sup> Sidney S. Alexander (1949) and subsequent researchers investigated whether the decline in the variance of firms' growth rates with size might have a statistical explanation. If a firm of size  $k$  is regarded as a portfolio of  $k$  independent unit-size businesses, we can predict the magnitude of the decline of growth's variance with size. Just as this assumption obviously marks an upper bound on the diversification of larger firms, so does the predicted rate of variance decline exceed those actually estimated (Boeri 1989). In this context S. W. Davies pointed out (personal communication) that it would be attractive to compare this variance-size relation in firms' overall sizes and in their sizes (shares) in individual markets.

detected weak positive autocorrelation for large firms and weak negative autocorrelation for small ones. Bruno Contini and Riccardo Revelli (1989) reported negative autocorrelation for small Italian firms, as did Boeri and Ulrich Cramer (1992) for all German non-agricultural establishments employing 20 or more. In a sample of the larger British firms, Geroski, Stephen Machin, and Chris Walters (1997) estimated significant positive regression coefficients of one year's growth on the growth rate in each of the three preceding years, but they observed that the partial correlation with growth in the preceding year was negative. Wagner's (1992) results for German firms appear consistent with Geroski, Machin, and Walters'. A possible resolution of these conflicting results is suggested in Section 2.2.

### 2.2 *Mobility Patterns: Recent Evidence*

Recent research has turned from a preoccupation with LPE to the direct measurement of turnover itself, recognizing that its significance extends beyond concentration to a broad range of behavioral and normative properties of product (also labor) markets. This new approach addresses both the mobility of continuing firms (this subsection) and turnover due to entry and exit (the next subsection).

Mobility in the shares of a set of continuing firms is usually measured by summing the absolute values of the differences between their activity (output, employment) levels at  $t$  and  $t + 1$  and dividing by the sum of their activity levels at  $t$ . This mobility measure is equivalent to the sum of absolute values of their growth rates weighted by their initial shares, so mobility remains closely linked to the LPE literature (Boeri 1994). These changes are commonly large. Steven J. Davis, John C. Haltiwanger, and Scott Schuh (1996,

Fig. 2.3), working with absolute growth rates of employment, documented that about half of all job changes in U.S. manufacturing are accounted for by units making annual changes of 25 percent or more. John R. Baldwin (1995, Table 4.1) divided Canadian manufacturing firms that survived from 1970 to 1982 into those gaining and those losing employment. While the average continuing firm increased its employment by 1.5 percent annually, the average gainer grew by 7.8 percent while the average loser shrank by 6.3 percent. Findings for other countries are similar: for German nonagricultural sectors during 1977–90, expanding incumbents' employment on average grew 6.2 percent annually, while contracting ones shrank by 5.8 percent (Boeri and Cramer 1992). Although the changes tend to be cyclically sensitive, Baldwin noted that even in the decade's most expansionary year, 34 percent of the continuing firms reduced employment (though not necessarily output). Negative year-to-year autocorrelation appears in the typical firm's employment changes around any long-term trend, shown by comparing average rates of growth for firms that expanded employment in a given year to annualized growth rates for those that expanded cumulatively over longer periods. For example, continuing Canadian firms that expanded over 1970–81 grew by 27.2 percent, equivalent to 2.2 percent annualized, but firms growing in each year within 1970–81 had average annual growth of 8.2 percent. Continuing firms that contracted over 1970–81 (32 percent of all continuing firms) shrank by 11.0 percent, 1.05 percent annualized, but firms shrinking in each year contracted by 5.9 percent on average (Baldwin 1995, Table 4.3).

A second important feature of the mobility pattern is long-run regression to the mean. For example, the initial

average size of Canadian firms that contracted over 1970–76 was 71 percent greater than the initial average size of those that expanded. The difference increases with the length of time over which the gainers and losers are identified. However, when continuing firms are divided into quintiles on the basis of their initial sizes, subsequent gainers predominate at the small end and subsequent losers in the top quintile, but both are well represented throughout; the ratio of the gainers' mean proportional gain to the losers' mean proportional loss rises rapidly and smoothly from the largest to the smallest quintile (Baldwin 1995, Table 4.6). Although many findings in this literature are consistent with pervasive mean regression, other countries' data have not been analyzed to check the short-run (error correction) and long-run (presumably behavior) mean regression found in Canada.

One might suppose that the mobility of firms results from their discrete and unequal adjustments to whatever aggregate expansion or contraction their market is experiencing. However, mobility seems largely independent of the direction and magnitude of the industry-wide change. Timothy Dunne, Mark J. Roberts, and Larry Samuelson (1989a) compared gross rates of increase in employment in expanding plants and of contraction in shrinking plants, after sorting industries into those that (in a five-year period) exhibit growing and shrinking employment overall. Within four-digit industries they found that for every job gained in a growing industry 0.604 are lost, and for every job lost in a contracting industry 0.644 are gained. Davis and Haltiwanger (1992) compared this intraindustry turnover to interindustry shifts. They broke down employment shifts (beyond those needed to accomplish a sector's net overall ex-

TABLE 1  
AVERAGE ANNUAL RATES OF JOB TURNOVER  
(GROSS GAINS PLUS GROSS LOSSES, PERCENTAGE OF  
TOTAL EMPLOYMENT), SELECTED COUNTRIES AND  
TIME PERIODS

Country	Time period	Turnover rate (%)
A. Selected industrial countries		
Canada	1984–90	22.6
Denmark	1984–89	29.8
Finland	1988–91	24.2
France	1985–87	26.3
Germany	1977–89	15.9
Italy	1985–91	23.4
Norway	1976–86	15.6
Sweden	1985–91	22.7
United Kingdom	1982–91	15.2
United States	1976–91	21.5
B. Manufacturing sector, selected countries		
Chile	1979–86	26.8
Colombia	1977–91	24.6
Morocco	1984–89	30.7
Canada	1973–86	20.5
United States	1973–86	19.6

Source: Panel A, Boeri (1994); panel B, Roberts and Tybout (1996, Table 2.1). Sectoral coverage for countries in panel A varies between manufacturing only (United States) and all employment (Germany).

pansion or contraction) into those occurring between and within four-digit U.S. manufacturing industries. Interindustry shifts accounted for only 12 percent, and a more elaborate set of controls (region, plant size, plant age, ownership) still left 61 percent to be explained by random intraindustry mobility.<sup>4</sup>

To explain this high incidence of intraindustry mobility, one can look to

<sup>4</sup> Employment rather than output is widely used as a turnover measure, and it is important to keep in mind that a unit's falling employment might be associated either with falling output or with rising output and productivity. Martin Neil Baily, Eric J. Bartelsman, and Haltiwanger (1996) showed that the distribution of firms by rising and falling employment is largely independent of their distribution by rising and falling productivity.



differences among both countries and industries. The comparisons of countries' average turnover rates in Table 1 are based on employment turnover (jobs created in expanding units plus jobs lost in contracting units, as percentage of total employment). They vary in sectors covered and other details, and they include newly opened and closed business units as well as continuing ones. Variations among industrial countries do not submit to easy explanations, but developing countries seem to experience more turnover.<sup>5</sup> Among the industrial countries there is a weakly evident inverse relationship between average establishment sizes and mobility, consistent with the findings of the LPE literature already mentioned.<sup>6</sup> Table 1's information on selected developing countries supports this interpretation. The work of John E. Jackson et al. (1996) on Poland suggests that turnover in the transition economies is very high.<sup>7</sup>

Other evidence on the correlations of turnover comes from cross-section tests

<sup>5</sup> International comparisons will not be emphasized in this paper because of pervasive differences in national data sets. The most important is the varying truncation of small firms: some include even single-person businesses, while others omit firms smaller than ten employees.

<sup>6</sup> Baldwin and Garnett Picot (1995) compared labor mobility rates between Canada and the United States, broken down by changes in small and large units. The gross flows tend to be larger in Canadian than U.S. small units, and greater for U.S. than Canadian large ones. They conjectured that the Canadian plant-size distribution, more concentrated in small sizes than that of the United States, causes the burden of adjusting to disturbances to be pushed toward the smaller units.

<sup>7</sup> Other noteworthy findings are that privatized and reformed state enterprises have lost less than unreformed ones, and that the annual numbers and sizes of new enterprises have fallen as (perhaps) the richest and most obvious opportunities exposed in the transition were picked off first. The ending of a large disequilibrium has provoked temporary bursts of entry in other settings, such as the United States following World War II (Betty C. Churchill 1954a) and Chile after a massive liberalization of international trade (James R. Tybout 1996a).

within countries. Geroski, Machin, and Walters (1997) measured the dependence of British firms' sales growth on changes in their market values, the growth rates of their industries' outputs and of gross domestic product, and the occurrence of innovations within the firm and industry. Industry-level growth and innovations and aggregate growth exhibit some explanatory power, but instantaneous individual shocks account for most of the variance in growth rates of firms' sales. As they concluded, "corporate growth rates are not quite random." Working with data on plants in 23 four-digit U.S. manufacturing industries, Baily, Charles Hulten, and David Campbell (1992) sought to test competing models of the turnover process on patterns of change in plants' productivity rankings within their industries. They found some evidence of capital vintage effects (a plant's productivity declining with its age), as had T. Y. Shen (1968). These rankings show a lot of continuity, however, and the authors' preferred explanation is that each plant's productivity varies randomly around its own productivity intercept, with regression to the mean evident.

More could be done to pin down industry-specific determinants of mobility, and theoretical models offer help. In "active learning" models, such as Richard Ericson and Ariel Pakes (1995), firms invest in uncertain but expectedly profitable innovations or cost reductions. The firm grows if successful, shrinks or exits if unsuccessful. Capital vintage effects, already mentioned, can account for productivity dispersions that result from a plant's descent through the productivity rankings as it ages, eventually to be refurbished or retired and replaced by a new-vintage plant. Val Eugene Lambson (1991) demonstrated how, even without capital-embodied technical progress, vintage

effects could account for plants' dispersed productivity levels: firms choosing (sunk) technologies must guess about future prices of variable inputs, and the dispersion of firms' average variable costs at any point in time reflects the historic dispersion of those guesses. Inter-industry associations between mobility and industry structure or technology are considered in Section 3.

The heteroskedasticity of firms' growth rates with size noted in the LPE literature suggests that the mobility of an industry's member firms decreases with their market shares. That heteroskedasticity might result from some form of entrenchment of leading firms (suspected although not well explained in the literature of industrial organization), from greater diversification of larger firms even within a well-defined product market, or simply from the way adjustment costs vary with firm size. Baldwin (1995, chap. 5) divided firms in each Canadian manufacturing industry into large and small (around the 50th percentile of output), confirming that the small firms show somewhat more variable shares. The largest four firms, however, do regress toward the mean: over 1970–79 on average they experienced declines in shares. The declines do not vary with rank within the top four, although the variance is somewhat greater for those ranked third and fourth. Nonetheless, thanks to the spread-out distribution of leaders' shares, the predicted duration of tenure in a top rank is quite long: 28, 17, 14, and 12 years for firms ranked 1, 2, 3, and 4 respectively. Geroski and Saadet Toker (1996) performed a similar calculation for 54 British industries, obtaining 17, 10, 11, and 9 years. Is the stickiness of leaders' positions just a reflection of the ubiquitous heteroskedastic variance of growth with size, or is there "something more"? As

Baldwin pointed out, the heteroskedasticity explanation appears sufficient, and an inquiry based on interindustry differences is needed to pin down any structural rigidities or first-mover advantages that might also weigh in. Earlier contributors to the LPE literature concluded that growth rates are almost independent of size among larger firms, but their failure to sort firms into industry groups deprives the result of clear implications.

An important dividend of the studies of mobility is indirect evidence of the character of costs of adjusting actual to desired levels of input or activity. The standard assumption of convex adjustment costs implies that the mass of the distribution of time rates of change will be concentrated around zero. The alternative of fixed costs of adjustment implies a dispersed distribution with modes away from zero. The evidence summarized in Haltiwanger's (forthcoming) survey clearly supports nonlinear adjustment. For example, one inference holds that about 70 percent of a 10 percent disequilibrium between a plant's actual and desired employment will remain three months later, while only 50 percent of a 60 percent disequilibrium will remain. This finding has important implications for macroeconomics as well as supplying a general explanation for high mobility rates. The same literature demonstrates the asymmetry in the adjustment upward and downward of plants' actual to desired capital stocks that is implied by the sunkness of capital. We do not know whether this asymmetry imprints itself on market-share changes and (if so) over what range of industries.

To summarize this section, recent research documents wide variance of business units' rates of growth consistent with fixed costs of adjustment. Large minorities of units contract in

TABLE 2  
AVERAGE RATES OF INDUSTRIES' GROSS ENTRY AND EXIT, EIGHT COUNTRIES (PERCENT)

Country	Time period	Entry rate		Exit rate	
		Number of firms	Market share	Number of firms	Market share
Belgium:	1980-84				
manufacturing		5.8	1.6	6.3	1.9
services		13.0	4.4	12.2	4.1
Canada	1971-79	4.0	3.0	4.8	3.4
Germany	1983-85	3.8	2.8	4.6	2.8
Korea	1976-81	3.3	2.2	5.7	n.a.
Norway	1980-85	8.2	1.1	8.7	1.0
Portugal	1983-86	12.3	5.8	9.5	5.5
UK	1974-79	6.5	2.9	5.1	3.3
US	1963-82	7.7	3.2	7.0	3.3

Source: Cable and Schwalbach (1991, Table 14.1).

an industry or economy that is expanding (and vice versa). The sustained trends in size (share) shown by many units coexist with long-run regression to the mean, although at a rate slow enough for firms to enjoy long tenure in leading positions. Much turbulence in the form of job changes occurs within rather than between industries. The variance of growth rates decreases strongly with the sizes of business units and apparently with the sunkness of capacity.

### 2.3 *Entry Rates and Entrants' Survival*

An important line of research on turnover has tracked entrants to determine their subsequent growth and mortality rates. Studies have documented substantial rates of entry and exit in a number of countries. Data in Table 2 assembled by John Cable and Joachim Schwalbach show average annual entry rates (number of firms) of about 6.5 percent, claiming about 2.8 percent market shares in manufacturing industries; average exit rates are very similar to entry rates.

Entrants suffer from high rates of infant mortality. Churchill (1955) re-

ported that half of all businesses established in the United States during 1946-54 were sold or discontinued within two years; hazard rates declined steadily after the first year. Baldwin's (1995, chap. 2) data show that entrants to Canadian manufacturing experience a first-year hazard rate of about 10 percent; it declines irregularly over a decade to the 5-7 percent range, which still exceeds the 3.5-5 percent range for firms more than one decade old. When the age structure of firms exiting in a given year is analyzed, again about 10 percent entered in the previous year, but about half were more than ten years old. Other studies that confirm the decline of hazard rates over time include Audretsch (1991), who worked with the U.S. Small Business Administration data based on establishment-level records of Dun & Bradstreet.<sup>8</sup>

<sup>8</sup> This data base has been controversial, because the Dun & Bradstreet records are biased toward covering only those establishments and single-plant firms that need to establish credit ratings, and they are updated infrequently. Audretsch and Zoltan J. Acs (1994) summarized the evidence on the quality of this data base, concluding that in practice the patterns that it yields agree with those based on data from official census records.

The hazard rates for entrants reported for various countries seem rather similar, except that Portugal's are distinctly higher: 25 percent in the first year, 16 percent in the second, 13 percent in the third (Mata, Pedro Portugal, and Paulo Guimarzes 1995). Portugal's development status and the coverage of very small firms are both sufficient explanatory factors. Although the pattern of declining hazard rates for an entering cohort is widely reported, Rajshree Agarwal and Michael Gort (1996) showed that it can be overridden by the industry's life cycle associated with the development, maturation, and ultimate displacement of its basic product. Hazard rates, they found, increase over time for early entrants as an industry-wide shakeout eventually sets in. Firms' hazard rates also increase at the end of the cycle. When hazard rates can be measured by month, they increase for the most of the first year; entrant firms likely start with enough resources to sustain themselves that long (Josef Brüderl, Peter Preisendörfer, and Rolf Ziegler 1992). There exists one unexplained exception: hazard rates increase persistently with the ages of small Belgian firms (Konings, F. Roodhooft, and L. Van der Gucht 1996).

Although entrants' rates of infant mortality are high, so are the growth rates of the survivors. Baldwin (1995, pp. 21–27) found that during the 1970s each entrant cohort's share of manufacturing value added increased over time (up to the ten years he could observe), so entry cumulatively contributes a lot to turnover in the enterprise population. Entrants surviving over the years 1970–81 made up 35.5 percent of manufacturing firms in 1981 and accounted for 10.9 percent of employment. If entry and exit are defined not for manufacturing as a whole but for individual (four-digit) industries, plants

that are switched from one industry to another also become exits and entries (respectively). In the average industry, new firms' entries over a decade account for only about one-third as much of employment as existing firms that switched or built new plants.

Similar to Baldwin's findings are those of Timothy Dunne, Roberts, and Samuelson (1988; hereafter DRS) for the United States. They observed firms' entries into individual (four-digit) industries over five-year intervals only, but could extend coverage over a maximum of two decades (1963–82). Their quinquennial hazard rates cannot be compared to Baldwin's annual values, but they appear consistent. In the average industry and five-year period, DRS found that entrants (both new firms and "switchers") account for 39 percent of end-of-period firms (16 percent market share); at the start of a period those destined to exit similarly make up 35 percent of firms (also 16 percent market share). The combined employment share of a given cohort of entrants declines over time, and that decline occurs about equally in the new-firm entrants and the diversifying entrants that have either switched existing capacity or built new plants. In their intensive study of entry into the chemical processing industries, DRS (1989b) established the same conclusion by a different route, using a regression procedure to control for variation in industries' growth rates.

DRS (1989b) investigated the degree to which the enlarged sizes attained by surviving entrants are due to individual firms' fast growth rather than higher mortality rates of the initially smaller entrants. The initial-size factor is statistically significant for all entrants in their first five-year period and for multi-plant entrants through their observed lives, so survivors' growth and gains in

size relative to incumbent firms stem partly from a selection process.<sup>9</sup>

One is curious about the contrasting findings on whether an entrant cohort's combined output share rises or falls. This outcome is the net result of an initially high but declining hazard rate and high but declining average growth rates of survivors. For Canada, Baldwin (1995, chap. 2) found a net increase; for the United States, DRS (1988, 1989b) reported a net fall. Cohort shares in Portugal decline at rates similar to those in the United States (Mata, Portugal, and Guimaraes 1995). In both the U.S. and Canadian data a fringe of the smallest firms was excluded. Baldwin dealt with annual observations over a shorter period and with the whole manufacturing sector rather than individual industries. Quite possibly entrant cohorts in some settings enjoy short-run gains in aggregate share: for German nonagricultural firms, Boeri and Cramer (1992) found that the level (not share) of employment of an entrant cohort increases for a year or two but then drops below its initial value; for Michigan firms (all industries) Jackson (1996) reported cohort shares of employment to rise for eight years. Given that some early entrants to an industry typically grow large and live a long time, it is almost necessary that the typical entrant cohort's combined share ultimately declines. There is obviously room for further research on the short-run trajectories of entrant cohorts and their variation from industry to industry.

The evidence on entrants' growth and failure rates clearly suggests a stochas-

tic process in which firms make their entry investments unsure of their success and do not initially position themselves at a unique optimal size. By general agreement, a fruitful explanation lies in Boyan Jovanovic's (1982) model of "passive learning" (also Hugo A. Hopenhayn 1992 and Luis M. B. Cabral 1993). The potential entrant is assumed to know the mean and standard deviation of all firms' costs but not its own mean expectation. Upon paying a (non-recoverable) entry fee, it starts to receive noisy information on its true cost level, which in any one period might induce it to expand, contract, or even exit. The consistency of the preceding evidence with Jovanovic's model is clear.<sup>10</sup> Researchers have tested specific implications for the dependence of a firm's growth on its size and age. The link is the proposition that the younger the firm, the more does each observation contribute to its knowledge of its costs. Evans (1987b) analyzed the growth of young firms as a function of their age and initial size, finding that growth diminishes with size (at a decreasing rate) and decreases with age when firm size is held constant. The decrease with age holds both overall and within most individual industries (Evans 1987a), so the passive-learning model is well supported. DRS (1989c) obtained similar conclusions about U.S. manufacturing plants from the Census Bureau's longitudinal data base. As a firm ages and grows more confident about its costs, the mean and variance of its growth rate should decline. Indeed, DRS found that mean rates of (employment) growth decline with age for every

<sup>9</sup> The chemicals sector is not necessarily a representative one for analyzing the fate of entrants. DRS (1989b) showed that in the average manufacturing industry firms operating in 1963 retained in 1982 a 58 percent market share, while in the chemicals industries they retained 70 percent.

<sup>10</sup> Pakes and Ericson (forthcoming) pointed out that the passive-learning model does not necessarily predict hazard rates falling from the outset. They could rise at first, if ill-fated firms need some experience to be sure of their unfitness. This is confirmed in a few studies cited previously.

plant size group except the largest, and failure rates (indicating variance) also decline with age given size.<sup>11</sup> Growth rates and failure rates also decline with size given age. The patterns differ in small ways between single-plant firms and plants belonging to multiplant enterprises, but the basic conclusions seem insensitive to the plant/firm distinction. Intercountry differences are not evident: Roberts (1996a) reported for Colombia the same conclusions about survival's relation to age and size.

Pakes and Ericson (forthcoming) sought to test the passive-learning model in competition with their own active-learning model. The test (on Wisconsin state data) suggests that the passive-learning model fits the retailing sector well, while manufacturing shows patterns that suggest active learning. The test is adroit and suggestive, but suffers in that passive and active learning are not mutually exclusive: opportunities for both could be abundant in one industry, scarce in another.

A consequence of entrants' high rates of early mortality is that, as many cross-section studies have reported, industries with high entry rates will also show high exit rates. That is, the data reject the model that pervades our textbooks: optimal-size firms enter an industry when its equilibrium output expands, exit when it contracts, but never do both at the same time. Cable and Schwalbach (1991, Table 14.2) provided evidence for eight countries. Positive correlations between contemporaneous rates of entry and exit were also reported by Michael E. Beesley and R. T.

<sup>11</sup> Kenneth R. Troske (1996) confirmed from Wisconsin unemployment insurance data that the mean and variance of growth rates fall off as entrants age, with services firms reaching a steady state sooner than manufacturers. Contini and Revelli (1989) found growth to decline with age for Italian firms, with age variations beyond 6 or 7 years having no effect.

Hamilton (1984) for manufacturing industries in Scotland, Geroski (1991a, 1991b) for British industries, and Timothy Dunne and Roberts (1991) for the United States. David I. Rosenbaum and Fabian Lamort (1992) confirmed the positive relationship after controlling for other determinants of entry and exit; for U.S. manufacturing industries they found exits much more responsive to entries over a five-year period than are entries to exits. Agarwal and Gort (1996) observed that these positive entry/exit correlations make most sense for samples of industries in steady states of maturity but varying in structural entry barriers and the sunkness of resource commitments. In early and late phases of a product's life cycle these correlations indeed reverse to negative.

Further evidence of entrants' uncertain fates lies in the association between rates of entry and rates of exit from a given industry over time. DRS (1988, Table 7) observed the usual positive correlations between rates of entry and exit among four-digit U.S. manufacturing industries over a common five-year period. The correlations turn negative when fixed effects confine the variation to temporal changes, but even then the entry rate in a given five-year period is positively correlated with the exit rate in the following five years. For Germany, Boeri and Lutz Bellmann (1995) found a positive entry shock to be followed by an increase in the next year's hazard rate (10 percent significance), although the hazard rate is unrelated to the current year's number of entrants. Leo Sleuwaegen and Wim Dehandschutter (1991) found the same lagged pattern in annual data on Belgian manufacturing industries. Baldwin and Joanne Johnson (1996) employed a hazard-rate analysis of individual entrants' fates to establish that

members of a larger entrant cohort are more likely to exit, and exit is also higher in an industry with a highly concentrated core and a numerous fringe. These studies of intertemporal entry-exit linkages control for macroeconomic conditions in various ways and degrees, but they leave the impression that recent entrants' hazard rates are rather insensitive to the observed variation in the macro environment.

Research on exit has mainly addressed infant mortality, and the prevalent decline of hazard rates with age suggests that geriatric problems are not serious for firms. Frøystein Wedervang (1965, pp. 168–75) early observed that age (i.e., youthfulness) is the more important factor explaining small firms' high exit rates, although size per se may be influential. Troske (1996) found that manufacturing firms five years before their exit are only half the sector's average size, while finance, insurance, and real estate firms destined to exit are then still at the average; both groups' mean growth rates go negative starting three years before exit, with manufacturers plunging more rapidly. Audretsch (1995a, chap. 7) observed that exits by older firms are less sensitive to industry growth disturbances than are exits by younger firms, and the elderly are preserved by the sunkness of resources committed to industries with large minimum efficient scales (MES) of production (also Beesley and Hamilton 1984). Mark Doms, Timothy Dunne, and Roberts (1995) found that firms in machinery industries using various advanced technologies are less likely to exit; because these users do not grow significantly faster than their competitors, sunkness is again suggested. For businesses small enough to be tied to their proprietors' life cycles, the manager's age positively predicts the business's survival in the prime working

years, but eventually age leads to discontinuance or sale of the business (Wedervang 1965, pp. 183–85; Timothy Bates 1990; Thomas J. Holmes and James A. Schmitz, Jr. 1996). Turnover in managers of small businesses tends to predict sales or closures shortly afterward, which suggests that job-matching for managers plays a role in the turnover of small businesses (Holmes and Schmitz 1995).<sup>12</sup>

Casual observation suggests that the oldest firms owe their longevity to trademarks (newspapers, simple consumer goods) that demand little organizational continuity. Organizational geriatrics has received little attention, although firms' exit rates have been studied in the context of industry or product life-cycles. Agarwal and Gort formulated the survival of mature firms as a trade-off between depreciation or obsolescence of their original endowments and the benefits of cumulative learning. Analyzing ten-year survival rates of firms entering during various (of the five) stages of their industry life-cycle, they found hazard rates increasing through the third stage, sufficiently explained by the fading growth opportunities offered by the product market. Hazard rates fall for late entrants, however, consistent with reduced rates of obsolescence of their initial endowments or increased efficacy of cumulative experience. Hazard rates for incumbents are lower than for entrants through all stages of the cycle in "non-technical" products (where experience advantages might be great), higher for "technical" products, where entrants

<sup>12</sup> These patterns point to the ambiguities of defining exit and entry in small businesses that are often bought and sold, or their facilities transformed from one activity to another. Jack's Bar becomes Jill's Bar, a business exit and an entry have occurred, an establishment continues unchanged, and Jack's business did not necessarily fail. See Bruce A. Kirchoff (1994, chaps. 6, 8).

bring the continuing flow of innovations (see also Audretsch 1991). Although their data base gives only limited leverage, Agarwal and Gort were able to calculate "senility points": ages at which hazard rates for incumbents entering in given stages of the life cycle stop falling and turn up again.

Steven Klepper and John H. Miller (1995) focused on the shake-out phase of the industry life cycle, in which the number of firms offering a product declines from its maximum to its long-run "mature" level. The data do not well support a simple model of overshoot entry. The shake-out is prolonged and continues at a steady (rather than declining) pace, suggesting instead a process of continuing competition among incumbents to reduce costs, modeled by Klepper and Elizabeth Graddy (1990) and Klepper (1996). Industry life cycles are considered further in Section 3.

In summary, entrants experience high hazard rates (infant mortality) that decline over time. Successful entrants also achieve high average rates of growth (that also decline as they age). The combined market share of an entering cohort, the net effect of these forces, eventually declines, but may increase for a time after the cohort's entry. The pattern is highly consistent with theoretical models of both passive and active learning. Industries that experience high rates of gross entry also tend to show high rates of exit. Over time, variations of the exit rate lag behind those of the entry rate; there is no strong evidence that exogenous disturbances to the exit rate induce subsequent entry (perhaps indicating just the infrequency of exogenous exit-causing disturbances). Units that exit have experienced declining growth rates for several years but (whether young or old) are still of substantial size relative to their competitors.

#### 2.4 *Interpreting Entry Patterns: Initial Commitments and Subsequent Options*

Firms enter an industry at different initial sizes, and the entrants' size distribution varies from industry to industry. That pattern invites interpretation in terms of entrants' diverse expectations and real options: entrants holding more positive expectations about their untested capabilities—their costs, or the qualities of their assets—make larger initial commitments. Even if the industry's technology supports a large optimal scale, the less confident entrant might rationally start out small, incurring a unit-cost penalty but limiting its sunk commitment while it gathers evidence on its unknown capability. Initially smaller entrants would then be expected to show higher exit rates.<sup>13</sup> Entrants' hazard rates should decline with the size of the minimum sunk resource commitment required to enter (roughly, the amount of costs that must be precommitted), but hazard rates could increase with the size of the irretrievable outlay needed to move from minimal or fringe entry to optimal-scale operation.

The empirical evidence neatly fits this framework of dispersed expectations and real options. First, Churchill (1954b) long ago showed that the size distribution of entrants to the typical industry is stable over time, suggesting some behavioral foundation. Furthermore, the distribution is aligned with the structure (entry barriers) of the entered industry (Mata 1991). Audretsch and Talat Mahmood (1995) as well as

<sup>13</sup> There is an evident problem of distinguishing between the entrant that limits its own initial commitment from self-knowledge of its long-shot status and the small entrant that fails because input-market imperfections denied it access to the optimal entry strategy. For a theoretical model of self-limiting entrants see Murray Z. Frank (1988).



Wagner (1994a) found that entrants' hazard rates decrease with their initial sizes, and Audretsch (1995b) established an inverse relation among industries between entrants' survival rates and the rates at which the lucky survivors grow. Mata (1996) found that better qualified Portuguese entrepreneurs (more schooling, older up to a point) start initially larger firms. Brüderl, Preisendörfer, and Ziegler (1992) found the hazard rates of German entrant firms to be lowered by the entrepreneur's general and industry-specific work experience, after controlling for the business's initial size and access to a broad (national) market; with these factors controlled, access to start-up capital was not a significant additional factor—lenders themselves apparently take these same success factors into account. Baldwin and Mohammed Rafiquzzaman (1995a) observed among Canadian industries a trade-off between the initial productivity levels of entrants who would subsequently survive (for at least ten years) and the rate at which their labor productivity subsequently grew. In each industry they compared the productivity in the first three years of survivors destined to last for a decade and ill-fated entrants who would not; the less the survivors' initial advantage over the quitters, the greater is the productivity growth subsequently achieved by the survivors. There is also an inverse relation between a survivor's initial labor productivity (relative to the productivity of incumbent firms) and the rate at which that productivity subsequently grows. The Baldwin-Rafiquzzaman findings invite the interpretation that successful entrants know or quickly learn their potential for rapid productivity growth. The invitation may be too seductive, because the Canadian data base lacks capital stocks, and Wedervang's (1965, pp. 194–96) study of Nor-

way showed that entrants have low labor productivity but high capital productivity. The winning entrants' labor-productivity gains might derive either from increasing their residual efficiency or exercising the option to make a major capital commitment.<sup>14</sup>

Further evidence of self-selected variation in initial commitments appears in the findings of DRS (1989b, 1989c) about differences between single-plant and multi-plant entrants, and between entries by new firms and firms established elsewhere. Single-plant entrants suffer much higher attrition than multi-plant entrants, consistent with the latter's much larger initial size commitments, reflecting greater confidence about prospects. Correspondingly, new-firm multi-plant entrants do much less well than diversifying multi-plant entrants, who likely have more accurate information on their ability to operate profitably in the entered industry. Capricious capital constraints or the hazards of undiversified life might explain this pattern, but differences in rational self-assessments are strongly suggested.

This real-option perspective is strengthened by evidence of industry-structure influences that both reduce survival and speed the survivors' growth. Audretsch (1995a, pp. 57–62; 1995b) confirmed several: the importance of MES in production<sup>15</sup> and the importance of innovation, both in the

<sup>14</sup> Laura Power (forthcoming) showed that post-entry investment bursts by successful entrants are far from the whole story. She concluded that in general plant-level labor productivity responds only weakly and slowly to "spikes" of machinery investments, even those undertaken after the plant's initial "birth" investment. Substantial and immediate effects turn up only in chemical-process industries. Elsewhere, investment may chiefly expand capacity rather than raise productivity.

<sup>15</sup> It also increases the mobility of small firms relative to large ones (Acs and Audretsch 1990, chap. 7).

industry as a whole and for its smaller firms. Audretsch (1995a, chap. 4) and Audretsch and Mahmood (1995) found that entrants' hazard rates increase with the industry's capital intensity (while older firms' rates decrease).<sup>16</sup> Two considerations let us make sense of this odd-sounding result. First, there is abundant evidence that manufacturing production functions are not homothetic, and capital intensity increases strongly with scale in the typical manufacturing industry (e.g., Caves and Thomas A. Pugel 1980), so small-scale entry need not entail a large sunk capital investment.<sup>17</sup> Second, if factors such as scale economies, capital intensity, and successful innovation can sustain rents to capable incumbents, they induce firms to enter for at least a "look" at their chances of ultimate success (see the model of S. A. Lippman and R. P. Rumelt, 1982).<sup>18</sup> Similar to Audretsch's finding on capital intensity is Marvin B. Lieberman's (1989) conclusion that the learning-curve advantages of leading incumbents of U.S. chemical processing industries do not deter the entry of new competitors but do lower their survival rates. A final evi-

dence of entry as an options purchase lies in Audretsch and Mahmood's (1995) comparison of hazard-rate models fitted to entrant single-plant firms and to new establishments of established firms: the dependent establishment is more likely to exit (with all other factors controlled), consistent with lower sunk costs associated with the entrepreneurial unit; but its survival is not predictable by any of the factors that cogently determine the start-up firm's decision to stick or exit.<sup>19</sup>

With this evidence of entrants' options noted, it still holds (for Portuguese manufacturing, Mata and Jose A.F. Machado 1995) that the initial sizes chosen by entrants increase with the industry's MES, and that (among entrants) the sizes of the larger ones are more strongly influenced by MES in its relation to the market's size and growth.

In summary, the new evidence imputes more rationality to entrants' decisions than has generally been assigned to them. To put the point provocatively, we have thought many entrants fail because they start out small, whereas they may start with small commitments when they expect their chances of success to be small. At the same time, small-scale entry commonly provides a real option to invest heavily if early returns are promising. Consistent with this, structural factors long thought to limit entry to an industry now seem more to limit *successful* entry: if incumbents earn rents, it pays the potential entrant to invest for a "close look" at its chances.

<sup>16</sup> Wagner (1994a, 1994b), however, found for Germany the more conventional conclusion that capital-intensity of the industry raises entrants' survival rate while lowering the entry rate. Boeri and Bellmann (1995) also investigated the joint determinants of entrants' hazard rates and survivors' growth in German industry.

<sup>17</sup> Interestingly, Audretsch and Mahmood (1995) did find that entrants' hazard rates decrease with the industry's human capital intensity measured by its average wage rate. Either human capital is a component of the initially sunk costs, which seems unlikely, or it measures the entrant's opportunity to raise productivity through learning by doing. Nonhomotheticity was confirmed by Bartelsman and Phoebus J. Dhrymes (1992), who found that labor productivity residuals increase with plants' sizes but those from total factor productivity do not.

<sup>18</sup> It would be most desirable to know how much money unsuccessful entrants lose, but dead firms post no losses.

<sup>19</sup> The testing of this model of entrants' options is aided by what appear to be major differences between the typical manufacturing and the typical services industry in the abundance of post-entry growth options. Pakes and Ericson (forthcoming) and Troske (1996) suggested that services entrants quickly settle at their steady-state sizes, while manufacturers take time to grow into them.

### 2.5 *Entry and Exit through Control Changes*

Turnover also occurs in the control of business units through acquisitions, mergers, and sell-offs of plants. From the perspective of static equilibrium models, these changes are commonly regarded as without consequence for industries' behavior and performance. However, control changes can have substantial effects. Large U.S. firms' diversified acquisitions in the 1960s and 1970s drew a justifiably bad press for the subnormal subsequent performance of the acquired business units (e.g., David J. Ravenscraft and F.M. Scherer 1987). When comprehensive census records of control changes came available for analysis, however, a positive relation emerged between control changes and productivity changes. The productivity levels of plants fated for changes in control tend to fall before the change and to recover afterward (Frank R. Lichtenberg and Donald Siegel 1987; also Lichtenberg 1992). Baldwin (1995, chap. 3) analyzed the consequences of control changes for both the market shares and normalized productivity levels of transferred Canadian plants, distinguishing between mergers and spinoffs and between combinations with diversified activities unrelated and related to the unit's industry base. When a firm enters an industry by acquiring a plant, that plant's market share grows somewhat for six years but then drops sharply. The productivity of continuing plants that underwent control changes in the 1970s was typically above average in 1970 but nonetheless higher still at the decade's end. This improvement occurred for all types of control changes except unrelated mergers, and even these showed improvement when the aggregate effect was calculated using plant-size weights rather than unweighted.

Robert H. McGuckin and Sang V. Nguyen (1995) analyzed all plants in the U.S. food and beverage sector that experienced control changes during 1977–82, following them until 1987 and comparing their labor productivity to plants in the same four-digit industry not undergoing control changes. Their main finding resolves an apparent contradiction between Ravenscraft and Scherer (1987) and Lichtenberg and Siegel (1987) concerning the initial (1977) productivity of plants fated for changes in control. Lichtenberg and Siegel, whose data pertain largely to big plants employing more than 250, observed deterioration prior to the control change, while Ravenscraft and Scherer found that control changed for many highly productive small plants. McGuckin and Nguyen concluded that both were right. In their comprehensive data set, the unweighted average initial relative productivity of plants that would undergo control change exceeds unity, and the higher the relative productivity, the more likely the control change. When the average is weighted, however, the ratio lies below unity. And when analysis is confined to large plants, the relation reverses: the likelihood of control change decreases with initial relative productivity. It appears that control changes can either lift the performance of an unproductive large unit or supply resources needed to leverage the strengths of a highly productive small one.

To explain the productivity of control changes, Lichtenberg and Siegel (1987) invoked the theory of job-matching: a continuous stream of disturbances renders some of the ownership links between enterprises and plants (or other major fixed assets) nonoptimal and induces reshuffling through the market for corporate control. Such a model is not only consistent with their overall result, but also can be extended to predict

differences among industries in the incidence of control changes. For the re-shuffling of plants (or lines of business) among firms to be productive, there must be sources of heterogeneity that allow the mismatches to crop up. Assets tied to plants or business units must have these traits: (1) Their qualities or attributes differ from unit to unit, in the sense of either vertical or horizontal differentiation. (2) These heterogeneities cause assets' productivities to vary substantially depending on the other business assets with which they collaborate within the firm. (3) These business assets must be "important"—lumpy or discrete, so that variations in their productivity can warrant incurring the transaction cost of a control change. (4) Transactions in these assets must be subject to market failures that prevent them from being rented or sold directly rather than as components bundled with plants or lines of business. (5) Disturbances continually affect an asset's productivity, not just overall but conditional on the other assets with which it is combined. Job-matching models such as Dale T. Mortensen and Christopher A. Pissarides (1994) seem close in spirit to these conditions. The conditions can be plausibly linked to observable traits of industries (importance of product differentiation, selling costs, or advertising; importance of science base and innovation; importance of scale economies in production). For Canadian manufacturing industries, Baldwin (1995, chap. 11) showed that the volume and productivity of control changes are greater in industries where such inputs are more important.<sup>20</sup>

<sup>20</sup> This result agrees with many findings in the literature on corporate diversification, which attributes the linking within the firm of businesses that serve different markets to just such assets that have the additional property that they can be used productively in more than one market.

### 3. Turnover, Mobility, and Static Elements of Market Structure

These findings on turnover and mobility pose an important challenge for empirical industrial organization, for both the traditional taxonomic (structure-conduct-performance, or SCP) approach and modern market-econometrics approach rest on comparative statics applied to models of market equilibrium. The emphasis here falls on the SCP approach. Its strategy, to inventory regular empirical relationships among elements of market structure, patterns of behavior, and levels of performance, is close in spirit to the turnover literature's quest for empirical regularities (see Richard Schmalensee 1989).<sup>21</sup> Are the stylized facts of the old SCP and the new turnover regularities mutually consistent? Can they shed light on and explain one another? The core of the SCP taxonomy is the causal relationships starting from the number and size distribution (concentration) of participants in a market and the factors limiting their number or access (entry barriers). Concentration itself is regarded as a consequence of factors limiting the equilibrium number of incumbent firms and/or supplying incumbents with first-mover advantages over subsequent entrants. It proves convenient first to align concentration and turnover, including concentration's dependence on entry barriers, then to consider how the structural entry barriers themselves relate to turnover.

#### 3.1 Concentration and Turnover: Contemporaneous Relationships

In the SCP paradigm, sellers' concentration is thought to affect their

<sup>21</sup> Turnover does have important implications for the "new IO" that are illustrated by G. Steven Olley and Pakes (1996).

behavior patterns but also to depend on their past conduct. Concentration is thus both a potential influence on mobility and turnover and (as the LPE shows) a potential consequence of mobility. And it depends on entry and exit in the near-definitional sense that the number of incumbent firms equals cumulative entries minus cumulative exits.

An inverse relation between an industry's concentration and its average rate of turnover due to entry and exit seems well established. Baldwin (1995, Table 8.5) related four-firm concentration ratios in Canadian manufacturing industries in 1979 to both turnover from entry and exit and turnover from incumbents' mobility, each measured over the preceding decade. His regression model also includes the standard measures of structural entry barriers long established as significant determinants of concentration. Turnover due to entry and exit exerts a powerful negative effect. Baldwin (1995, chap. 8) also used principal components summarizing various turbulence measures to supplement the standard cross-section model of concentration's determinants. Again, a principal component weighted positively on turnover due to entry and exit wields a great deal of incremental explanatory power.

Concentration is also commonly included as a regressor in cross-section models seeking to explain entry rates or turnover from entry and exit. Given that such models usually include the structural forces (entry barriers) that limit the equilibrium number of incumbents, it is not clear what behavioral mechanism is being tested. Because of causal links among the regressors (and resulting high multicollinearity), perverse signs commonly appear even when concentration itself takes a significant negative coefficient. The appropriate conclusion about concentration's influ-

ence is moot until we can replace it with measures of the mobility-detering conditions or policies (such as vertical restraints on distributors?) employed by concentrated producers.

The relationship between incumbents' mobility and concentration is more elusive. Contrary to LPE, changes in concentration have no simple empirical relation to the mobility of an industry's firms. In Canadian manufacturing, while concentration showed no trend on average, Baldwin (1995, chap. 7) could regard only about 15 percent of the industries as low in turnover. Although some forms of turnover seem positively correlated with absolute changes in concentration (Baldwin 1995, Fig. 7.2), high mobility and stable concentration are evidently compatible in industries that range widely in concentration levels. In Baldwin's cross-section regression analysis, mobility among continuing firms has a negative but insignificant effect on 1979 concentration levels. When principal components are used, entry/exit turnover remains the dominant influence, and the only significant component strongly weighted on a mobility measure seems to imply that when entry/exit turnover is raised, concentration declines more where incumbents' share changes are similar.

Evidence of relations running from concentration to mobility is similarly thin. Baldwin (1995, chap. 5) did find a negative relation between concentration and mobility of the *leading* firms. However, it turns up as greater mobility for leaders only in the least concentrated quintile of industries, which hardly suggests that collaboration among oligopolists fostered by concentration is what deters mobility. Acs and Audretsch (1990, chap. 7) reported a significant positive influence of concentration on mobility in U.S.

manufacturing industries, but the presence of collinear regressors makes it difficult to interpret. Whatever structural elements determine the differences among industries in incumbents' mobility, concentration is not the dominant one. Mobility's determinants are easily found, however, in more basic elements of market structure. Baldwin and Rafiquzzaman (1995b) employed a classification (originated in the Organization for Economic Cooperation and Development) of manufacturing industries into five broad groups: natural resource-based, labor intensive, scale economies-based, product differentiated, and science-based; partitioning industries into these classes captures a surprisingly large amount of variance in the fundamental conditions of technology and demand among narrowly defined manufacturing industries.<sup>22</sup> The mobility of incumbents tends to be high in labor intensive and product differentiated industries, low in scale-based ones. The pattern is what one would expect if mobility increases with competitors' scope for making uncertain investments that affect their market share, and decreases with the sunkness of costs (which leads to narrowly confined short-run capacities). Baldwin (1995, pp. 344–58) related measures of long-run profitability of Canadian manufacturing industries to various measures of turnover and appropriate control variables. Mobility among incumbents has a marginally significant negative effect, and *net* entry as expected is a highly significant negative influence. Turnover due to entry and exit surprisingly exerts a significant positive influence, perhaps due to causation running from the rents of

successful incumbents to the number of entrants willing to test their luck.<sup>23</sup>

Researchers were led on their quest for a causal influence of concentration on mobility and turnover by the hunch that competition in some sense is a source of turbulence that may be dampened by oligopolistic cooperation. One specific form of competition was tested by Baldwin and Richard E. Caves (1998): international rivalry, measured by imports' penetration of the national market and/or the share of domestic output exported. As to the mechanisms at work, the variance of disturbances might be greater for transactions crossing national borders, or (more subtly) international competition might increase the closeness of substitutes for varieties of a differentiated product, so that any given exogenous disturbance generates larger quantity responses and hence more mobility. They found that international competition does indeed increase mobility and (entry-exit) turnover in Canadian manufacturing industries. The relation is not a particularly strong one, however. Baldwin (1995, pp. 139–47) noted a similar effect of import competition on mobility in both Canada and the United States; although Davis, Haltiwanger, and Schuh (1996, pp. 47–49) reported negative results from a less closely targeted test. Roberts and Tybout (1996, p. 7) also mentioned obtaining negative results for several developing countries. In studying the effects of a large-scale trade liberalization on New Zealand manufacturing, John K. Gibson and Richard I. D. Harris (1996) concluded that "the plants likely to survive trade liberalization were larger, lower cost, older, used specialized capital and were owned by specialized firms

<sup>22</sup> Further evidence appears in the positive rank correlations reported by Roberts (1996b, Table 2.6) between turnover rates in industries matched between countries.

<sup>23</sup> Baldwin suggested industry life cycles as an explanation: at early stages both entry/exit turnover and profitability are high; in maturity both decline.

with few plants." In short, import competition shook out the less efficient and the less "sunk" (also see Joseph E. Flynn 1991, on trade liberalization in the United States).

### 3.2 *Concentration and Turnover: Cumulative Effects*

This completes the descriptive relations found in cross-section between concentration and entry-exit turnover and incumbents' mobility. Much remains to be said, however, about concentration as a *cumulative* result of random processes operating through turnover and mobility, in the manner of Gibrat and the LPE. Two separate theoretical and empirical literatures have evolved, one updating Gibrat, the other resting on the concept of a product life cycle. They lead to broadly similar conclusions but benefit from separate treatment.

That concentration might reflect both structural forces (notably minimum efficient scale [MES] interacting with the size of the market) and random disturbances was recognized theoretically by Herbert A. Simon and Charles P. Bonini (1958) and empirically by Edwin F. Mansfield (1962). Davies and Bruce Lyons (1982) developed this insight into an empirical interindustry test in which the  $n$ -firm concentration ratio is allowed to depend on a lower bound set by MES and market size plus an additional component increasing with the variance of random disturbances that lift leading firms above the MES threshold. Sutton (1997, forthcoming) developed an important theoretical synthesis, building on the approach of Yuji Ijiri and Simon (1977), that rests on two assumptions: (1) New market opportunities arise continually, and the likelihood that the next one will be seized by any incumbent firm is a nondecreasing function of that firm's size (this relaxes

Gibrat's proportionality between size and growth). (2) New firms enter at a constant rate. These assumptions predict a lower bound for concentration as a function of the number of firms in the market that is independent of the rate of entry (proportion of opportunities captured by new firms) but does depend on the assumption that the market grows over time. Sutton (forthcoming, chaps. 10–13) showed that this lower bound to the level of concentration conditional on the number of firms seems highly consistent with data on manufacturing sectors in several countries. In a related contribution Davies and Geroski (1997) linked the random factors determining each leading firm's market share to the random processes influencing an industry's concentration level overall. They devised a way to integrate the determinants of mobility—changes in the market shares of individual leading firms—with the determinants of the change in industry concentration. Thus, the change in each leading firm's share (in U.K. manufacturing industries) is treated as a function of its initial market share and the firm's own rate of spending on advertising and R&D relative to its competitors, so that both discretionary sunk outlays and Galtonian regression can play their roles (also Geroski and Toker 1996).

Sutton's two massive investigations of market structure's determinants (1991, forthcoming) cannot receive just treatment here, but we at least note their links to the literature on turnover and concentration. His investigation of the food-processing industries in various nations focused on the role of "endogenous sunk costs"—outlays establishing trademark goodwill or buyers' perceptions of superior product quality. For reasons indicated by the theory of vertical differentiation, such successful outlays greatly enlarge the innovator's

market share and put pressure on rivals either to imitate the strategy or to exit. Either way, concentration tends to increase, and its minimum bound stays strictly positive as the market's size increases without limit, because a larger market increases the firm's expected returns from these discretionary fixed (and sunk) outlays. The outlays themselves, however, can be regarded as randomly arriving opportunities, and in that sense the concentration arising from endogenous sunk costs depends on the variance of random opportunities and the persistence of their consequences.

In shifting his attention to research-intensive industries, Sutton (forthcoming) retained from his previous study the central role of endogenous fixed costs—now, to develop a vertically differentiated product. The new study deals explicitly with horizontal differentiation and the substitution between innovative and established products. Suppose that an innovative industry offers a line of horizontally differentiated products. The payout (profit, market-share gain) to an endogenous sunk outlay on improving one of these products can still be high if (1) the innovation productivity of R&D outlays is high, (2) this product is an important one for the industry, and/or (3) substitutability in demand between this product and others in the industry's line is high. Implications for concentration then follow as in Sutton (1991), except that the lower bound for concentration now decreases with the heterogeneity of the industry's product line. Once more, the randomness of innovation opportunities and successes links this mechanism to the random-process model. As in Sutton (1991), the theoretical model is supported by an impressive array of statistical tests and case studies.

Sutton's work clearly revitalizes the

random-process approach and gives it potential applicability to a wide range of markets (see William T. Robinson 1993). In ways suggested by the "active learning" models of Richard R. Nelson and Sidney G. Winter (1978) and Ericson and Pakes (1995), parameters of the distribution of random outcomes can be related to observable data to test hypotheses about the richness and variance of opportunities, the persistence of disturbances' effects (mean reversion), and the appropriability of favorable opportunities (alternatively, the contagion of crippling afflictions).

The other empirical and theoretical way to link random processes to concentration lies in the research starting from Gort and Klepper (1982) on the life cycles of industries that arose from important product innovations. Although the theoretical components of this literature grew out of its empirical observations, it is convenient to begin with theory (Klepper and Graddy 1990; Klepper 1996) in order to facilitate comparison to Sutton's approach. A market begins with some major and profitable but not fully appropriable innovation that attracts a queue of potential entrants. Firms that have entered can carry out product and/or process R&D. Product R&D yields modifications that are soon imitated (no enduring vertical differentiation). Process R&D lowers cost toward an (exogenous) attainable minimum and is more profitable for an incumbent, the larger its size. The competitively determined price falls with incumbents' costs. Over time fewer and fewer potential entrants enjoy positive expected profits. Incumbents less successful in lowering costs drop out. The cycle relies on random processes to determine the capabilities of potential entrants in the queue and the successes of incumbents in product and (especially) process innovation.



This model was devised to explain facts emerging from studies of a number of innovative products (Gort and Klepper 1982; Klepper and Graddy 1990; Agarwal and Gort 1996) and more intensive studies of five industries (Klepper 1995; Klepper and Kenneth L. Simons forthcoming). The large-sample studies show an impressive regularity in the gross flows of entrants and quitters as the product passes from an innovation to a mature good. The number of incumbents rises to a peak reached at a time when industry output is still growing. The number then falls off to a plateau level likely to persist until industry output actually declines. The rate of product innovation peaks early in the cycle, the rate of process innovation later. Within a few years after the industry's origin, the survival rates of the earliest entrants come to exceed those of all subsequent entering cohorts. Correspondingly, the leading firms in the mature industry were usually among the early entrants; the basis for this first-mover advantage (skill and luck at process innovation, in the model) does not come particularly clear in the empirical studies (Klepper 1995). Contrary to other authors, Klepper (1997; Klepper and Simons forthcoming) argued that the shakeout of firms in the latter part of the cycle arises not from exogenous developments (major product innovations, emergence of a standard product configuration) but is implied by the basic conditions that drive the whole cycle.

To conclude, Sutton's and the life-cycle approaches to turnover and evolving concentration differ in many ways, but are clearly complementary in the opportunities that they open for empirical research. In general, incorporating turnover into traditional industrial organization clarifies how underlying structure shapes the environment in

which market outcomes are determined. However, the evidence on turnover also deepens our anxiety as to what measurable features of markets are truly exogenous. While making the problem harder, the turnover literature does tell us where to look: the random drawings come from distributions that differ from market to market. They vary in where the disturbances strike, how large are their means and variances, and how durable are their consequences. The relevant parameters can be related coherently to "bedrock" characteristics of technology and tastes, although pinning down the connections will be a major challenge for empirical researchers. Turnover also hurls down a considerable challenge to those who seek their empirical research agendas from modern game theory. The outcomes that reflect adroit play of strategic advantages may be few relative to the outcomes in which the winner of a commitment game is the one who guessed most accurately the magnitude of Nature's next draw.

### 3.3 *Structural Entry Barriers and Turnover*

Standard structural barriers to entry first identified by Joe S. Bain (1956) have a well-established ability to predict industries' concentration levels (even if the normative interpretation of those barriers remains a festering issue). The theory of contestable markets flagged the need to establish the basis in sunk or committed costs for anything labeled a barrier to entry, and Ioannis N. Kessides (1990a, 1990b) found that both entry and concentration depend on the sunkness of incumbents' commitments.<sup>24</sup> The coefficient of variation among industries of entrants' survival

<sup>24</sup> Mata (1995) found entry into Portuguese manufacturing to decline with the sunkness of incumbents' capital, but sunkness deters exit only through its effect on gross entry.

rates exceeds that of entry rates (Audretsch 1995b), and barriers must clearly be regarded as affecting survival as well as entry.

To ensure the coherence of their mechanisms, these entry barriers need to be related to the uncertainty of entrants' investments and its consequences for the numbers of entrants and commitments they choose. Entry barriers based on scale economies we usually suppose to have no randomness in their effects (but cf. David E. Mills and Lawrence Schumann 1985). Those based on proprietary intangible assets (advertising and buyer goodwill, proprietary innovations) and even learning by doing are another story, as we saw in Section 2.4. Entry-detering assets based on intangibles are related to random processes and turnover in two ways. First, incumbents' advantages acquired through luck or skill are potentially wasting assets, at risk of losing their quasi-rents and deterrence potential due to taste shifts, other firms' innovations, and the like. Second, before entrants commit to a market, their qualifications will differ in unknown ways. Occasionally the random entrant will turn up with assets that excel those of some incumbents, supplying them with what George S. Yip (1982) called "gateways to entry." The larger the rents earned by successful incumbents and the smaller the entrant's sunk admission charge, the more entrants will make the attempt, and the higher will be their infant mortality rate (compare Lippman and Rumelt 1982). Thus, the entry barrier more essentially deters *successful* entry than it does *gross* entry.

A good deal of empirical evidence confirms this integration of structural entry barriers with random processes. Acs and Audretsch (1987, 1988) addressed the issue of innovation as a barrier or a gateway to entry using data on

the number of innovations introduced in 1982 and classified by four-digit U.S. industry and size of the innovating firm. Small firms on average have higher rates of innovation (per employee) than large firms. Small firms' innovation rates are lower, relative to their large competitors' rates, in the less innovative industries and those with "heavy industry" characteristics (high concentration, capital intensity, unionization).<sup>25</sup> When the determinants of the innovation rates themselves are analyzed (Acs and Audretsch 1988), the small firms' rate surprisingly is not less sensitive to the industry's level of company-financed R&D spending. The small firms' rate is reduced by the industry's advertising outlays but increased by the human capital of its labor force, consistent with small firms' rate being high early in a basic innovation's life cycle and declining (absolutely and relative to large competitors) as the cycle proceeds. Correspondingly, the opportunity to make and appropriate innovations, a gateway to entry early in the process, becomes a barrier to entry in the mature stage (this model was sketched by Dennis C. Mueller and John E. Tilton 1969; William J. Abernathy and James M. Utterback 1978 discussed its implications for the sunkness of incumbents' resource commitments). With other factors controlled, the small firms' innovation rate is a significant positive influence on the rate of entry into industries (Audretsch 1995a, chap. 3). Klepper and Simons (forthcoming), studying a group of industries that ultimately became highly concentrated, confirmed innovation's

<sup>25</sup> Gort and Akira Konakayama (1982), although working with net-entry data, modeled entry and exit as gross flows determined in the setting of innovative, growing markets. They inferred that both entry and exit increase with the industry's rate of patenting and the growth rate of output per firm, while entry (but not exit) also increases with the industry's incidence of major innovations.

decline over the life cycle and conversion to an entry barrier; in these industries, process and (largely) product innovations were dominated by larger incumbents throughout the cycle.

Kessides (1986) concluded that the sunk (capitalized) value of industry advertising outlays serves as a goodwill entry barrier, but the current flow (including entrants' outlays) operates as a gateway. We lack direct tests of the barrier-vs-gateway duality comparable to Acs–Audretsch in product-differentiation entry barriers, but the parallel seems clear. Robinson and Claes Fornell (1985) identified high barriers with pioneering trademarks immune to obsolescence (also Mata 1995). The increase in the productivity of large-scale advertising associated with the rise of television increased concentration (Willard F. Mueller and Richard T. Rogers 1980) but it also increased mobility (Mark Hirschey 1981). Robert J. Stonebraker (1976) in a neglected paper showed that the uncertainty of profits of small (fringe) firms in an industry increases with the product differentiation entry barriers protecting large incumbents.

To complete the analysis of turnover and entry barriers, we refer briefly to the large literature on the determinants of *net* entry (surveyed by Geroski 1991a, 1995). It concludes that net entry should be represented as an adjustment process that depresses industry profits to the "limit" level that renders further entry unprofitable, or that (in the "free entry" case) increases the number of incumbents to their zero-profit equilibrium number.<sup>26</sup> When the

<sup>26</sup> Another element recently added to this model is a supply-side component to depict the individual's trade-off between selling labor services and undertaking the entrepreneurial role. The implication that (*cet. par.*) an increase in the unemployment rate should raise the number of entering firms was confirmed by Evans and Linda S. Leighton (1990), among others.

dependent variable becomes the *gross* number of entrants, another block of determinants is needed to model the replacement of incumbents, including expected failures among the entrants (Baldwin 1995, chap. 14). Most studies have modeled this replacement component rather crudely by including the number of incumbent firms, when the gross number of entrants is the dependent variable. Enrico Santarelli and Alessandro Starlacchini (1994) employed the prevalence of small firms (whose mortality accounts disproportionately for turnover). Boeri and Bellmann (1995) added the lagged number of entrants as an explanatory variable, though with only modest statistical success, and Sleuwaegen and Dehandshutter (1991) found that entry increases with lagged exit. Baldwin (1995, chap. 14) observed a positive effect of the variability of industry sales on the number of entrants, suggesting that it reflects the decreased survival rate and increased entry needed to sustain expected zero profits in equilibrium.<sup>27</sup>

A result regularly reported in the entry-determinants literature is that, while structural barriers deter entry by newly created firms, they have little significant effect (sometimes even a positive effect) on entry by initially large newcomers, by firms established in other industries, or by multinational

<sup>27</sup> The positive effect of sales variability on gross entry is complemented by its negative effect on new-plant construction by incumbent firms (Baldwin 1995, Table 3.7); the incumbents are likely to build larger plants that are low-cost but better suited to stable environments (Mills and Schumann 1985). Regarding gross entry studies in general, the consideration of entrants' turnover calls into question researchers' practice of measuring entry by the market share achieved by entrants arriving over some time interval. That variable is the outcome of the gross number of entrants, their early hazard rates, and the growth rates achieved by the survivors. These three components do not have identical determinants, and it seems desirable to address them one by one.

firms. Baldwin and Paul K. Gorecki (1987; also Geroski 1991b and Mata 1993) explicitly treated the different types of entrants as responding to different replacement incentives (e.g. foreign-controlled entrants expect to replace foreign-controlled incumbents). Short-run entry inducements (profits, market growth) typically have little statistical effect on these well-endowed entrants. The pattern is consistent with their entries being driven largely by a replacement mechanism: apparently they expect to pass successfully through an entry “gateway” and compel incumbents to contract or exit.

### 3.4 Market Structure, Mergers, and New Entry

Actual entry and structural entry barriers must also, in light of the new evidence on turnover, be related to the incidence of changes in corporate control. Control changes (see Section 2.5) are not neutral with respect to the productivity and market shares of the acquired business units. They are “entry-like” in their potential competitive effect on other firms, which makes their relation to structural entry barriers important for an understanding of those barriers’ overall effect. Baldwin (1995, p. 48) found that during 1970–79 rates of entry by acquisition and “green field” among Canadian manufacturing industries were *negatively* correlated,  $-0.18$ , statistically significant. Furthermore (Table 3.6), while the green-field entry rate decreases with four-firm concentration (regarded as a proxy for structural entry barriers), entry by acquisition significantly increases with it. See Nils-Henrik Mørch von der Fehr (1991) for related results.

Behind the inverse relation between green-field and acquisition entry rates lies the fact that acquisition rates are high in just those industries surrounded

by structural barriers to entry. Baldwin (1995, chap. 11) found that both the incidence and the productivity of control changes are high in industries that are surrounded by structural entry barriers—production scale economies, product differentiation, control of proprietary technologies. Structural entry barriers thus exert their effect not simply by limiting an industry’s equilibrium number of firms but also by altering the gross number and failure rates of entrants trying their luck, the mixture of types of entrant, and the relative occurrence of entry via new plants (or plants switched from other industries) and acquisition of incumbent capacity.

## 4. Turnover, Productivity, and Efficiency

### 4.1 Contribution of Turnover to Productivity Growth

The simple Darwinian interpretation that we reflexively assign to business units’ turnover implies that the more efficient units displace the less efficient, so that average productivity rises. Researchers on turnover have seized the opportunity to measure this contribution and impute industry-level productivity gains to increases in the productivity of the average unit and the displacement of the less by the more efficient. Bartelsman and Dhrymes (1992) dramatized the role of turnover. In U.S. manufacturing, the unweighted mean total factor productivity (TFP) of large plants (over 250 employees) showed a sustained decline through 1972–84, though with a definite upturn in 1984–86. Over the same period, aggregate (i.e., weighted) TFP at the two-digit industry level dipped in the early 1970s but subsequently achieved a sustained increase. Individual plants’ rankings in the productivity distribution they found to be quite stable, so much of the

discrepancy depends on the process of reallocating shares toward the more efficient.

Baldwin (1995, Table 9.1) considered the responsiveness of turnover to productivity differences among units. He divided all plants in each industry around median labor productivity and found that units exiting in 1970–79 were significantly concentrated below the 1970 median, while the previous decade's (surviving) entrants were more prevalent above the 1979 median. The pattern holds for other classes of entrants and exits but with exceptions: plants closed by continuing firms, and entrants who switch plants previously classified to other industries. With their sizes controlled, entrant plants of new firms became 16 percent more productive than continuing plants by 1979, and new plants of continuing firms 31 percent more productive. Continuing plants that were less productive in 1970 had a somewhat higher chance of gaining market share during the following decade (regression to the mean, once again), but in 1979 the percentage above the median that had gained share greatly exceeded the percentage below the median that had gained (Baldwin 1995, Table 9.2). Continuing plants that gained share were in 1970 insignificantly more productive than those destined to lose share, but they wound up 34 percent more productive in 1979.

Another contribution to the relation between turnover and productivity is Byong-Hyong Bahk and Gort (1993), who investigated related issues concerning the maturation of new plants in fifteen manufacturing industries. There is potential inconsistency between the recent findings on the growth and maturation of new business units and the traditional presumption that a plant or firm has a unique or at least a minimum optimal scale of operation. One way to

reconcile the dynamic evidence with the static presumption is to establish what role vintage effects and learning-by-doing play in the plant's productivity level. Bahk and Gort followed to 1986 plants that had begun operation between 1973 and 1982, establishing that (overall, and within most industries) both average capital vintage and cumulative output have significant effects; the evidence does not distinguish clearly between cumulative plant output and the passage of time as factors governing the rate at which productivity improves.<sup>28</sup> They also found that the explanatory power of cross-section regressions of plants' input-output relations increases with plants' ages, consistent with Jovanovic's process of time-related learning.

Imputing industrywide productivity growth to components of unit-level productivity growth, turnover, and mobility presents a problem of statistical decomposition. It can be done in various ways resting on different economic assumptions. For example, do the shares gained by entrants come at the expense of exiting firms, and those of expanding incumbents from the shrinking incumbents? Or do all the expanders push indiscriminately against all the contractors? Baily, Hulten, and Campbell (1992) decomposed five-year productivity growth for 23 U.S. manufacturing industries in a way implying that the mobility of continuing plants adds roughly 50 percent (their Table 1) to those plants' own productivity growth. Low-productivity plants are more likely to exit (although initial productivity does not predict a plant's switch to

<sup>28</sup> Their findings on learning must be treated with caution. An entrant discovering immediately that it is blessed with low costs will elect to produce a large output. A young firm's cumulative large output might therefore reflect either learning-by-doing or early confirmation of its innate efficiency.

another industry), but net entry-exit is inferred to add little because of the entrants' low initial levels of productivity. Haltiwanger (forthcoming) reported a somewhat different decomposition applied to all U.S. manufacturing industries over 1977-87. It attributes 54 percent of industrywide productivity growth to within-plant increases but divides the share-change contribution into two components: share changes weighted by the difference between the unit's initial productivity and industry productivity; and the covariance between share changes and productivity changes. The latter term isolates the productivity winners' share gains and accounts for 38 percent of industrywide productivity growth. The former picks up mean regression and accounts for -10 percent. Net entry accounts for 18 percent.

Results for other countries suggest roughly the same relative importance for turnover in industrywide productivity growth. Baldwin (1995, chap. 9) used both statistical inference and a series of alternative assumptions to pin down these replacement patterns, concluding that on any reasonable procedure, 40 to 50 percent of Canadian industries' productivity growth in the 1970s could be laid to turnover. Baldwin (1996) also analyzed turnover and productivity growth in various subperiods of the years 1973-90, finding little or no evidence of interrelation in their movements. Analyzing all manufacturing and mining establishments in Israel, Zvi Griliches and Haim Regev (1992) imputed a larger proportion of productivity growth to within-plant improvements and less to mobility, possibly because of shorter (three-year) periods of observation. Tybout (1996b) reported complex patterns in several developing countries.

The observation period's duration

conspicuously affects the estimated contribution of entry and exit on industry productivity. All studies agree that in the short run, turnover from entry and exit appears to make a minimal contribution to an industry's productivity growth, because the quitters' productivity and the initial productivity levels of the entrants are similar and both below those of continuing firms, and because entrants account for a small share of activity. The productivity of surviving entrants grows rapidly, however, as they exploit their revealed competence (including the small-firm innovations stressed by Acs and Audretsch) and exercise of their investment options (see section 2.4). Exiting firms have experienced declining productivity and would presumably have deteriorated further if not pushed over the edge. In the long run, therefore, entry-exit turnover is important for industry-level productivity gains. Roberts and Tybout (1996, chap. 1) stressed this conclusion for the developing countries that they studied. Besides the evidence on developed countries reviewed in section 2, Geroski (1989) found that overall productivity growth in 79 British manufacturing industries (1976-79) increased significantly with the lagged rate of gross entry of new firms.

#### 4.2 *Turnover and Productive Efficiency*

Although comparing productivity gains through turnover and through improved performance in the individual unit puts turnover's importance in useful perspective, the underlying mechanisms are not really commensurate. Given random shocks to firms' or plants' productivity levels, mobility and turnover should continuously enlarge the winners and shrink the losers whether the industry's overall rate of productivity growth is high or low. More fundamental is the relation between turnover and

the dispersion at a point in time of the efficiency levels of competing business units, for that dispersion provides the opening for turnover to raise productivity. If we regard the production function as defining an industry's efficient frontier, then the reshuffling of the units' shares should directly register the penalties that the market imposes on units whose performance is slipping relative to the frontier and the rewards to those climbing toward it. Mobility should reduce the weighted average shortfall from the frontier, just as disturbances that widen the efficiency distribution should increase turnover.

Research on productive efficiency has recently drawn upon the same primary industrial census records as the research on turnover. Studies in a number of countries have used one or another technique to infer a frontier of technical efficiency for each manufacturing industry and test in cross-section hypotheses about factors that might cause or perpetuate inefficiency (Caves and David R. Barton 1990; Caves and Associates 1992; David G. Mayes 1996). These studies assume that the gap between average and best-practice productivity for an industry's representative business unit measures an equilibrium outcome with determinants that are stable over time. On that assumption they perform cross-section (inter-industry) tests of hypotheses about structural and behavioral forces that could create a dispersion of units' efficiency levels beneath the frontier and either speed or delay its shrinkage through the mobility process. The results support the general framework and confirm a number of specific hypotheses (Caves and Associates 1992, chap. 1; also Mayes, Christopher Harris, and Melanie Lansbury 1994, chap. 6). The confirmed determinants include such factors as industries' R&D intensi-

ties and rates of productivity growth—sources of turbulence and obsolescence that continually drag some units beneath the productivity frontier. They also include others (competitiveness, unionization of its workforce, exposure to international competition) that should either speed or delay mobility processes. Thus, the evidence on productive inefficiency reveals the sources of opportunities for productivity-raising turnover, just as the evidence on turnover shows the strength of the forces keeping productive inefficiency in check.<sup>29</sup>

Baldwin's (1995, Table 12.6) data dramatize the interplay of turnover and individual units' productivity changes for Canadian manufacturing in the 1970s: while each class of share gainers displaced a less productive set of losers, average productive efficiency fell 5 percent because continuing plants that lost market share experienced large declines in their relative productivity, while continuing plants that gained share raised theirs only a little. He found some evidence on what causes units to slip farther beneath the frontier. The increase in an industry's overall labor-productivity level is negatively correlated with the change in its productive efficiency (10 percent significance), which in turn is negatively correlated with its productivity gain through the turnover of

<sup>29</sup> Studies of industries' productive efficiency estimated in successive years further underline the importance of the relation between efficiency and turnover. For both Britain in the 1960s and 1970s (Sheryl Bailey 1992) and Korea during 1978–88 (Seong Min Yoo 1992) annual estimates of productive inefficiency seem to vibrate around stable means, although for Britain in the 1980s Lansbury and Mayes (1996) found upward trends prevailing. In Korea Yoo observed that the rank correlations of industries' efficiency levels are low even between adjacent years. Bailey analyzed what determines the amount of intertemporal variation in an industry's efficiency level, concluding that variability increases with the incidence of disturbances and decreases with the industry's speed of adjustment and flexibility of resource use.

market shares. That is, faster technical advance pushes the laggards and nonadopters farther back and prompts more turnover to displace them. The new research on turnover has an important potential for extending our knowledge of what determines productivity dispersions. Timothy Dunne (1994), for example, concluded that the adoption of advanced production technologies is not biased away from old plants in the way vintage models suggest.

In a regression analysis, Baldwin (1995, pp. 318–26) related market shares gained by an industry's entrants and expanding incumbents to the productivity differentials that propel this turnover and the various types of disturbances that disperse plants' productivity levels. The exogenous variables were filtered through principal components, which makes the influences of the underlying regressors difficult to summarize. Nonetheless, it is clear that forces in both groups exert significant influences. Investigating the determinants of productive efficiency in U.S. manufacturing industries, Caves and Barton (1990, chap. 6) concluded that it decreases with the importance of product innovations to the industry and also with process innovations underlying capital-vintage effects (also Shen 1968). Lansbury and Mayes (1996) observed a negative influence of productivity growth on productive efficiency in Britain, but Akio Torii (1992) found a positive effect in Japan.

Another apparently important influence on mobility and entry–exit turnover, and thereby productivity, is the stage of the business cycle. Davis and Haltiwanger (1990, 1992) observed that gross turnover moves contracyclically, with the effects most pronounced in larger and older plants and plants belonging to multiplant firms (also see Jeffrey R. Campbell 1997). Apparently “creative destruction” is involved when

adverse macroeconomic conditions force the process of adjustment (and the incurring of adjustment costs). The same result was found for Canada (Baldwin 1995, chap. 6) and the United Kingdom (Konings 1995a). In Britain, turnover tends to occur within industries during booms, between sectors during recessions. In Belgium no contracyclical job destruction was found, but the authors noted that their data covered a period lacking the major recessions that occurred in Britain and the United States (Konings, Roodhooft, and Van de Gucht 1996).<sup>30</sup> Roberts (1996b) also found no evidence of it in the developing countries that he studied, which might well lack a core of larger and older units vulnerable to being dislodged in a recession. Indeed, in developing countries entry–exit turnover is substantially more important relative to incumbents' mobility than in developed countries.<sup>31</sup> Christopher L. Foote (1997) showed that these disparate results on gross turnover components might be an artifact of countries' different net growth rates.

## 5. Conclusions

*Plus ça change, plus c'est la même chose.* Turnover processes are ubiquitous among plants and firms classified to an industry. They are also stable, explicable, and can be embraced within the traditional thinking based on market-equilibrium models that underlies the bulk of empirical research in industrial organization. The abundant new evidence on turnover processes reviewed in this paper provides a deeper

<sup>30</sup> Also, the Belgian study, unlike the others, covered very small firms, and contracyclicality would be expected more in the larger units.

<sup>31</sup> However, job turnover is concentrated within (rather than between) industries and regions no less in developing than in developed countries, contrary to what one might expect.



understanding of why concentration levels are what they are, why they change when they do, and how structural entry barriers affect the behavior of actual entrants. We know much more about why industries contain firms of diverse sizes and not alike-as-peas incumbents reflecting some iron law of optimal scale. We understand that the diverse fates and frequent failures of new firms reflect hidden information and option-value considerations, and need not be written off to "cannon fodder" and "animal spirits."

It is difficult to summarize a summary. However, if in a biblical mood one seeks ten substantive conclusions, they would be these:

1. In the typical industry and overall, mean growth rates of firms decrease with their initial sizes among small firms; for initially large firms growth rates and size are unrelated. The variance of growth rates decreases with firm size. Entry into and exit from the distribution occur mainly in the smaller size classes. These processes are typically consistent with the size distribution of firms (concentration) being stable over time.

2. The mobility of incumbent firms is substantial and consistent with the assumption that adjustment costs are not smoothly convex but have a fixed component. Long-run regression to the mean is clearly evident, although at a rate consistent with slow turnover in the ranks of the typical industry's largest firms.

3. Gross entry is substantial in most industries. It is much larger than net entry, due to high rates of infant mortality. Successful entrants grow rapidly, so that an entrant cohort's initial market share falls slowly. The pattern is consistent with Jovanovic's model of entrants learning their specific capabilities.

4. Entrants select different initial

sizes reflecting both the structure of the entered market and their own perceived capabilities. The process works as if entrants obtain options to make larger further investments after learning or confirming their capabilities.

5. The turnover processes described so far show very little qualitative difference among countries. Less developed countries appear to exhibit more turnover associated with their concentration on activities with smaller sunk costs. Large macroeconomic disturbances (wars, major trade liberalizations, major privatizations) explicably affect the turnover process.

6. Turnover through changes in the control of business units operates in the manner of a job-matching process driven by continual disturbances to the optimal match.

7. Concentration in manufacturing industries is negatively related to turnover from entry and exit but largely independent of incumbents' current mobility, except that mobility is greater in the *least* concentrated industries. Changes in concentration depend on mobility-related factors. Mobility depends strongly on basic features of an industry's technology and demand conditions. Concentration strongly depends, however, on cumulative effects of past mobility rates, in ways shown by modern random-process models and analyses of industry life-cycles.

8. The traditional structural entry barriers affect both the number of entries and the entrants' survival rate; for some barriers the latter effect seems more important. Due to random qualities of new firms' endowments, entry barriers can become entry gateways for lucky entrants.

9. Productivity growth for an industry as a whole depends to an important degree on the redistribution of shares toward the more productive units and not

just on growth of the units' individual productivity.

10. Research on productive efficiency (gap between an industry's average and best-practice productivity) shows what factors determine the opportunities for productivity-raising turnover, thus complementing the evidence on turnover that shows the strength of the forces keeping productive inefficiency in check.

Much has been accomplished in a short time to set the stylized facts about turnover, but the research opportunities remain rich. Hazard-rate analyses are just coming into use to isolate the causes and correlates of individual units' fates. In particular, little is known about what factors systematically trigger the expansion or contraction of incumbent business units, and what temporal relations exist between changes in their unit profitability and scales of operation. Substantial work has been done on the factors associated with the persistence of successful business units' high profits (e.g., Geoffrey F. Waring 1996; Anita M. McGahan and Michael E. Porter 1996), but it needs to be extended and related to the general properties of turnover processes. We have far to go in relating mobility to its determinants in basic conditions of technology and demand. Because reallocations of activity from the less efficient to the more efficient are so important for the optimal use of resources, more evidence is needed on how competitive conditions within an industry affect the speed with which the more efficient displace the less efficient.

Exogenous shocks to an industry provoke diverse responses in individual units, and the relation between the overall disturbance and the pattern of responses should be pursued. This is especially important for the foundations of macroeconomics. There, analysis has

begun on the relation between changes in aggregate demand and plant-level investment responses (Ricardo J. Caballero, Eduardo M. R. A. Engel, and Haltiwanger 1995) and on the relative roles for determining fluctuations in U.S. aggregate manufacturing employment's growth of aggregate shocks and of changes in the distribution of idiosyncratic shocks (Caballero, Engel, and Haltiwanger 1997). The same questions arise for shocks originating in major innovations, international comparative advantage, trade policy (especially the formation of free-trade areas), and others.<sup>32</sup>

Turnover in the control of business units, noted only briefly in section 2.5 of this survey, deserves much more attention than it has received. This turnover has been regarded chiefly through the lens of corporate governance and contract theory, not as a type of job-matching problem triggered by changes in the optimal combination of heterogeneous business assets under particular managerial roofs.

National differences noted in this survey suggest further leads. Major reforms in national economic systems generate enormously heterogeneous disturbances to the nation's plants and firms, as in the transition economies and developing economies that have undertaken broad-based privatizations. Understanding their consequences requires following the turnover of individual units. Indeed, economic development seems to involve raising the capability to coordinate and manage larger business units with more complex teams of inputs, and that process itself

<sup>32</sup> An important limitation of this survey from the viewpoint of macroeconomics is its emphasis on unweighted-average behavior patterns. Size-weighted patterns matter for aggregate activity, an important point in the dispute over firm size and job creation (Davis, Haltiwanger, and Schuh 1996, chap. 4)

appears primarily in the differential fates of diverse business units.

## REFERENCES

- Abernathy, William J. and James M. Utterback. 1978. "Patterns of Industrial Innovation," *Tech. Rev.* 80:7, pp. 40-47.
- Acs, Zoltan J. and David B. Audretsch. 1987. "Innovation, Market Structure, and Firm Size," *Rev. Econ. Statist.*, 69:4, pp. 567-74.
- . 1988. "Innovation in Large and Small Firms: An Empirical Analysis," *Amer. Econ. Rev.* 78:4, pp. 678-90.
- . 1990. *Innovation and Small Firms*. Cambridge, MA: MIT Press.
- Agarwal, Rajshree and Michael Gort. 1996. "The Evolution of Markets and Entry, Exit and the Survival of Firms," *Rev. Econ. Statist.*, 78:3, pp. 489-98.
- Alexander, Sidney S. 1949. "The Effect of Size of Manufacturing Corporation on the Distribution of the Rate of Return," *Rev. Econ. Statist.*, 31:3, pp. 229-35.
- Audretsch, David B. 1991. "New-Firm Survival and the Technological Regime," *Rev. Econ. Statist.*, 73:3, pp. 441-50.
- . 1995a. *Innovation and Industry Evolution*. Cambridge, MA: M.I.T. Press.
- . 1995b. "Innovation, Growth, and Survival," *Int. J. Ind. Organ.*, 13:4, pp. 441-57.
- and Zoltan J. Acs. 1994. "New-Firm Startups, Technology, and Macroeconomic Fluctuations," *Small Bus. Econ.*, 6:6, pp. 439-49.
- Audretsch, David B. and Talat Mahmood. 1995. "New Firm Survival: New Results Using a Hazard Function," *Rev. Econ. Statist.*, 77:1, pp. 97-103.
- Bahk, Byong-Hyong and Michael Gort. 1993. "Decomposing Learning by Doing in New Plants," *J. Polit. Econ.*, 101:4, pp. 561-83.
- Bailey, Sheryl D. 1992. "The Intraindustry Dispersion of Plant Productivity in the British Manufacturing Sector, 1963-79," in Caves and Associates 1992, pp. 329-84.
- Baily, Martin Neil; Eric J. Bartelsman and John Haltiwanger. 1996. "Downsizing and Productivity Growth: Myth or Reality?" *Small Bus. Econ.*, 8:4, pp. 259-78.
- Baily, Martin Neil; Charles Hulten, and David Campbell. 1992. "Productivity Dynamics in Manufacturing Plants," *Brookings Pap. Econ. Activity: Microeconomics*, pp. 187-249.
- Bain, Joe S. 1956. *Barriers to New Competition*. Cambridge, MA: Harvard U. Press.
- Baldwin, John R. 1995. *The Dynamics of Industrial Competition*. Cambridge: Cambridge U. Press.
- . 1996. "Productivity Growth, Plant Turnover and Restructuring in the Canadian Manufacturing Sector," in Mayes 1996, pp. 245-62.
- and Richard E. Caves. 1998. "International Competition and Industrial Performance: Allocative Efficiency, Productive Efficiency, and Turbulence," in *The Economics and Politics of International Trade*. Gary Cook, ed. London: Routledge, Volume II, pp. 57-84.
- Baldwin, John; William Chandler, Can Le, and Tom Papailiadis. 1994. *Strategies for Success*. Catal. No. 61-523RE. Ottawa: Statistics Canada.
- Baldwin, John and Paul K. Gorecki. 1987. "Plant Creation versus Plant Acquisition: The Entry Process in Canadian Manufacturing," *Int. J. Ind. Organ.*, 5:1, pp. 27-41.
- Baldwin, John and Joanne Johnson. 1996. "Survival of New Canadian Manufacturing Firms: The Importance of Financial Structure," working paper, Micro-Economic Studies and Analysis Division, Statistics Canada.
- Baldwin, John and Garnett Picot. 1995. "Employment Generation by Small Producers in the Canadian Manufacturing Sector," *Small Bus. Econ.*, 7:4, pp. 317-31.
- Baldwin, John and Mohammed Rafiqzaman. 1995a. "Selection versus Evolutionary Adaptation: Learning and Post-Entry Performance," *Int. J. Ind. Organ.*, 13:4, pp. 501-22.
- . 1995b. "Restructuring in the Canadian Manufacturing Sector from 1970 to 1990: Industry and Regional Dimensions of Job Turnover," Research Paper Series No. 78, Analytical Studies Branch, Statistics Canada.
- Bartelsman, Eric J. and Phoebus J. Dhrymes. 1992. "Productivity Dynamics: U.S. Manufacturing Plants, 1972-1986," Discussion Paper No. 92-1, Center for Economic Studies, U.S. Bureau of the Census.
- Bates, Timothy. 1990. "Entrepreneur Human Capital Inputs and Small Business Longevity," *Rev. Econ. Statist.*, 72:4, pp. 551-59.
- Beesley, Michael E. and R. T. Hamilton. 1984. "Small Firms' Seedbed Role and the Concept of Turbulence," *J. Ind. Econ.*, 33:2, 217-31.
- Boeri, Tito. 1989. "Does Firm Size Matter?" *Gior. degli Econ. e Annali di Econ.*, 48:9-10, pp. 477-95.
- . 1994. "Why Are Establishments So Heterogeneous?" *Small Bus. Econ.*, 6:6, pp. 409-20.
- and Lutz Bellmann. 1995. "Post-Entry Behavior and the Cycle: Evidence from Germany," *Int. J. Ind. Organ.*, 13:4, pp. 483-500.
- Boeri, Tito and Ulrich Cramer. 1992. "Employment Growth, Incumbents and Entrants: Evidence from Germany," *Int. J. Ind. Organ.*, 10:4, pp. 545-65.
- Brüderl, Josef; Peter Preisendörfer, and Rolf Ziegler. 1992. "Survival Chances of Newly Founded Business Organizations," *Amer. Soc. Rev.*, 57:2, pp. 227-42.
- Caballero, Ricardo J.; Eduardo M. R. A. Engel, and John C. Haltiwanger. 1995. "Plant-Level Adjustment and Aggregate Investment Dynamics," *Brookings Pap. Econ. Activity*, 2, pp. 1-54.
- . 1997. "Aggregate Employment Dynamics: Building from Microeconomic Evidence," *Amer. Econ. Rev.*, 87:1, pp. 115-37.
- Cable, John and Joachim Schwalbach. 1991. "In-

- ternational Comparisons of Entry and Exit," in Geroski and Schwalbach 1991, pp. 257-81.
- Cabral, Luis M. B. 1993. "Experience Advantages and Entry Dynamics," *J. Econ. Theory*, 59:2, pp. 403-16.
- Campbell, Jeffrey R. 1997. "Entry, Exit, Embodied Technology, and Business Cycles," Working Paper No. 5955, National Bureau of Economic Research.
- Caves, Richard E. and Associates. 1992. *Industrial Efficiency in Six Nations*. Cambridge, MA: M.I.T. Press.
- Caves, Richard and David R. Barton. 1990. *Efficiency in U.S. Manufacturing Industries*. Cambridge, MA: M.I.T. Press.
- Caves, Richard and Thomas A. Pugel. 1980. *Intraindustry Differences in Conduct and Performance*. Monograph Series 1980-2. NY: Graduate School Bus. Admin., NYU.
- Churchill, Betty C. 1954a. "Recent Business Population Movements," *Surv. Curr. Bus.*, 34:1, pp. 11-16.
- . 1954b. "Size Characteristics of the Business Population," *Surv. Curr. Bus.*, 34:5, pp. 15-24.
- . 1955. "Age and Life Expectancy of Business Firms," *Surv. Curr. Bus.*, 35:12, pp. 15-19.
- Contini, Bruno and Riccardo Revelli. 1989. "The Relationship between Firm Growth and Labor Demand," *Small Bus. Econ.*, 1:4, pp. 309-14.
- Davies, Stephen and Paul Geroski. 1997. "Changes in Concentration, Turbulence, and the Dynamics of Market Shares," *Rev. Econ. Statist.*, 79:3, pp. 383-91.
- Davies, Stephen and Bruce R. Lyons. 1982. "Seller Concentration: The Technological Explanations and Demand Uncertainty," *Econ. J.*, 92:4, pp. 903-14.
- Davis, Steven J. and John C. Haltiwanger. 1990. "Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications." *NBER Macroeconomics Annual*, pp. 123-68.
- . 1992. "Gross Job Creation, Gross Job Destruction, and Employment Reallocation," *Quart. J. Econ.*, 107:3, 819-63.
- Davis, Steven; John Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: M.I.T. Press.
- Doms, Mark; Timothy Dunne, and Mark J. Roberts. 1995. "The Role of Technology Use in the Survival and Growth of Manufacturing Plants," *Int. J. Ind. Organ.*, 13:4, pp. 523-45.
- Dunne, Paul and Alan Hughes. 1994. "Age, Size, Growth, and Survival: UK Companies in the 1980s," *J. Ind. Econ.*, 42:2, pp. 115-40.
- Dunne, Timothy. 1994. "Plant Age and Technology Use in U.S. Manufacturing Industries." *RAND J. Econ.*, 25:3, pp. 488-99.
- and Mark J. Roberts. 1991. "Variation in Producer Turnover across US Manufacturing Industries," in Geroski and Schwalbach 1991, pp. 187-203.
- Dunne, Timothy; Mark Roberts, and Larry Samuelson. 1988. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries," *RAND J. Econ.*, 19:4, pp. 495-515.
- . 1989a. "Plant Turnover and Gross Employment Flows in the U.S. Manufacturing Sector," *J. Labor Econ.*, 7:1, pp. 48-71.
- . 1989b. "Firm Entry and Post-entry Performance in the U.S. Chemical Industries," *J. Law Econ.*, 32:2, Part 2, pp. S233-71.
- . 1989c. "The Growth and Failure of U.S. Manufacturing Plants," *Quart. J. Econ.*, 104:4, pp. 671-98.
- Ericson, Richard and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Rev. Econ. Stud.*, 62:1, pp. 53-82.
- Evans, David S. 1987a. "The Relationship between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries," *J. Ind. Econ.*, 35:4, pp. 567-81.
- . 1987b. "Tests of Alternative Theories of Firm Growth," *J. Polit. Econ.*, 95:4, pp. 657-74.
- and Linda S. Leighton. 1990. "Small Business Formation by Unemployed and Employed Workers," *Small Bus. Econ.*, No. 4, 2:4, pp. 319-30.
- Flynn, Joseph E. 1991. "The Determinants of Exit in an Open Economy," *Small Bus. Econ.*, 3:3, pp. 225-32.
- Foote, Christopher L. 1997. "The Surprising Symmetry of Gross Job Flows," working paper, Harvard U.
- Frank, Murray Z. 1988. "An Intertemporal Model of Industrial Exit," *Quart. J. Econ.*, 103:2, pp. 333-44.
- Geroski, Paul A. 1989. "Entry, Innovation, and Productivity Growth," *Rev. Econ. Statist.*, 71:4, pp. 572-8.
- . 1991a. *Market Dynamics and Entry*. Oxford: Basil Blackwell.
- . 1991b. "Domestic and Foreign Entry in the United Kingdom: 1983-1984" in Geroski and Schwalbach 1991, pp. 63-88.
- . 1995. "What Do We Know about Entry?" *Int. J. Ind. Organ.*, 13:4, pp. 421-40.
- Geroski, Paul; Stephen Machin, and Chris Walters. 1997. "Corporate Growth and Profitability," *J. Ind. Econ.*, 45:2, pp. 171-89.
- Geroski, Paul A. and Joachim Schwalbach, eds. 1991. *Entry and Market Contestability*. Oxford: Blackwell.
- Geroski, Paul A. and Saadet Toker. 1996. "The Turnover of Market Leaders in UK Manufacturing Industry, 1979-86," *Int. J. Ind. Organ.*, 14:2, pp. 141-58.
- Gibson, John K. and Richard I. D. Harris. 1996. "Trade Liberalisation and Plant Exit in New Zealand Manufacturing," *Rev. Econ. Statist.*, 78:3, pp. 521-29.
- Gort, Michael and Steven Klepper. 1982. "Time Paths in the Diffusion of Product Innovations," *Econ. J.*, 92(3), pp. 630-53.
- Gort, Michael and Akira Konakayama. 1982. "A

- Model of Diffusion in the Production of an Innovation," *Amer. Econ. Rev.*, 72:5, pp. 1111-20.
- Griliches, Zvi and Haim Regev. 1992. "Productivity and Firm Turnover in Israeli Industry: 1979-1988," Working Paper No. 4059, National Bureau of Economic Research.
- Hall, Bronwyn H. 1987. "The Relationship between Firm Size and Firm Growth in the US Manufacturing Sector," *J. Ind. Econ.*, 35:4, pp. 583-600.
- Haltiwanger, John. Forthcoming. "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence," *St. Louis Fed. Reserve Bank Econ. Rev.*
- Hannan, Michael T. and John Freeman. 1989. *Organizational Ecology*. Cambridge, MA: Harvard U. Press.
- Hart, Peter E. and Nicholas Oulton. 1996. "Growth and Size of Firms," *Econ. J.*, 106:3, pp. 1242-52.
- Hirschey, Mark. 1981. "The Effect of Advertising on Industrial Mobility, 1947-72," *J. Bus.*, 54:2, pp. 329-39.
- Holmes, Thomas J. and James A. Schmitz, Jr. 1995. "On the Turnover of Business Firms and Business Managers," *J. Polit. Econ.*, 103:5, pp. 1005-38.
- . 1996. "Managerial Tenure, Business Age, and Small Business Turnover," *J. Labor Econ.*, 14:1, pp. 79-99.
- Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60:3, pp. 1127-50.
- Ijiri, Yuji and Herbert A. Simon. 1977. *Skew Distributions and the Sizes of Business Firms*. Amsterdam: North-Holland.
- Jackson, John E. 1996. "Firm Size and the Dynamics of a Market Economy," Working paper, Dept. Polit. Sci., U. Michigan.
- ; Jacek Klich, Krystyna Poznanska, Leszek Zienkowski, and Józef Chmiel. 1996. "Economic Change in Poland: 1990-1994," Working paper, Dept. Polit. Sci., U. Michigan.
- Johnson, Joanne; John Baldwin, and Christine Hinchley. 1997. *Successful Entrants: Creating the Capacity for Survival and Growth*. Catalogue No. 61-524-XPE. Ottawa: Statistics Canada.
- Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry," *Econometrica*, 50:3, pp. 649-70.
- Kessides, Ioannis N. 1986. "Advertising, Sunk Costs, and Barriers to Entry," *Rev. Econ. Statist.*, 68:3, pp. 84-95.
- . 1990a. "Toward a Testable Model of Entry: A Study of the US Manufacturing Industries," *Economica*, 57:2, pp. 219-38.
- . 1990b. "Market Concentration, Contestability, and Sunk Costs," *Rev. Econ. Statist.*, 72:4, pp. 614-22.
- Kirchhoff, Bruce A. 1994. *Entrepreneurship and Dynamic Capitalism: the Economics of Business Firm Formation and Growth*. Westport, CT: Praeger.
- Klepper, Steven. 1995. "Evolution, Market Concentration, and Firm Survival," working paper, Carnegie Mellon University.
- . 1996. "Entry, Exit, Growth, and Innovation over the Product Life Cycle," *Amer. Econ. Rev.*, 86:3, pp. 562-83.
- . 1997. "Industry Life Cycles," *Ind. and Corp. Change*, 6:1, pp. 1-37.
- and Elizabeth Graddy. 1990. "The Evolution of New Industries and the Determinants of Market Structure," *RAND J. Econ.*, 21:1, pp. 27-44.
- Klepper, Steven and John H. Miller. 1995. "Entry, Exit, and Shakeouts in the United States in New Manufactured Products," *Int. J. Ind. Organ.*, 13:4, pp. 567-91.
- Klepper, Steven and Kenneth L. Simons. Forthcoming. "Technological Extinctions of Industrial Firms: An Inquiry into their Nature and Causes," *Ind. and Corp. Change*.
- Konings, Jozef. 1995a. "Job Creation and Job Destruction in the UK Manufacturing Sector," *Oxford Bull. Econ. Statist.*, 57:1, pp. 5-24.
- . 1995b. "Gross Job Flows and the Evolution of Size in U.K. Establishments," *Small Bus. Econ.*, 7:3, pp. 213-20.
- ; F. Roodhooft, and L. Van de Gucht. 1996. "The Life Cycle of New Firms and Its Impact on Job Creation and Job Destruction," Working Paper No. 9669, Faculty Econ. and Applied Econ., Katholieke Universiteit Leuven.
- Kumar, M. S. 1985. "Growth, Acquisition Activity and Firm Size: Evidence from the United Kingdom," *J. Ind. Econ.*, 33:3, pp. 327-38.
- Lambson, Val Eugene. 1991. "Industry Evolution with Sunk Costs and Uncertain Market Conditions," *Int. J. Ind. Organ.*, 9:2, pp. 171-96.
- Lansbury, Melanie and David Mayes. 1996. "Shifts in the Production Frontier and the Distribution of Efficiency," in Mayes 1996, pp. 66-88.
- Leonard, Jonathan S. 1986. "On the Size Distribution of Employment and Establishments," Working Paper No. 1951, National Bureau of Economic Research.
- Lichtenberg, Frank R. 1992. *Corporate Takeovers and Productivity*. Cambridge, MA: M.I.T. Press.
- and Donald Siegel. 1987. "Productivity and Changes in Ownership of Manufacturing Plants," *Brookings Pap. Econ. Activity*, 3, pp. 643-73.
- Lieberman, Marvin B. 1989. "The Learning Curve, Technology Barriers to Entry, and Competitive Survival in the Chemical Processing Industries," *Strategic Manag. J.*, 10:5, pp. 431-47.
- Lippman, S. A. and R. P. Rumelt. 1982. "Uncertain Imitability: An Analysis of Interfirm Differences in Efficiency under Competition," *Bell J. Econ.*, 13:2, pp. 418-38.
- Mansfield, Edwin F. 1962. "Entry, Innovation, and the Growth of Firms," *Amer. Econ. Rev.*, 52:5, pp. 1023-51.

- Mata, Jose. 1991. "Sunk Costs and Entry by Small and Large Plants," in Geroski and Schwalbach 1991, pp. 49-62.
- . 1993. "Entry and Type of Entrant: Evidence from Portugal," *Int. J. Ind. Organ.*, 11:1, pp. 101-22.
- . 1994. "Firm Growth during Infancy," *Small Bus. Econ.*, 6:1, pp. 27-39.
- . 1995. "Sunk Costs and the Dynamics of Entry in Portuguese Manufacturing," in *Market Evolution: Competition and Cooperation across Markets and Time*. A. van Witteloostuijn, ed. Dordrecht: Kluwer Academic, pp. 267-84.
- . 1996. "Markets, Entrepreneurs, and the Size of New Firms," *Econ. Letters*, 52:1, pp. 89-94.
- and Jose A. F. Machado. 1996. "Firm Start-up Size: A Conditional Quantile Approach," *European Econ. Rev.*, 40:6, pp. 1305-23.
- Mata, Jose and Pedro Portugal. 1994. "Life Duration of New Firms," *J. Ind. Econ.*, 42:3, pp. 227-45.
- Mata, Jose; Pedro Portugal, and Paulo Guimaraes. 1995. "The Survival of New Plants: Start-up Conditions and Post-Entry Evolution," *Int. J. Ind. Organ.*, 13:4, pp. 459-81.
- Mayes, David G. 1996. *Sources of Productivity Growth*. Cambridge: Cambridge U. Press.
- , Christopher Harris, and Melanie Lansbury. 1994. *Inefficiency in Industry*. Hemel Hempstead: Harvester Wheatsheaf.
- McGahan, Anita M. and Michael E. Porter. 1996. "The Emergence and Sustainability of Abnormal Profits," working paper, Harvard Business School.
- McGuckin, Robert H. and Sang V. Nguyen. 1995. "On Productivity and Plant Ownership Change: New Evidence from the Longitudinal Research Database," *RAND J. Econ.*, 26:2, pp. 257-76.
- Mills, David E. and Lawrence Schumann. 1985. "Industry Structure with Fluctuating Demand," *Amer. Econ. Rev.*, 75:4, pp. 758-67.
- Mortensen, Dale T. and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment," *Rev. Econ. Stud.*, 61:3, pp. 397-415.
- Mueller, Dennis C. and John E. Tilton. 1969. "Research and Development Costs as a Barrier to Entry," *Canad. J. Econ.*, 2:4, pp. 570-79.
- Mueller, Willard F. and Richard T. Rogers. 1980. "The Role of Advertising in Changing Concentration of Manufacturing Industries," *Rev. Econ. Statist.*, 62:1, pp. 89-96.
- Nelson, Richard R. and Sidney G. Winter. 1978. "Forces Generating and Limiting Concentration under Schumpeterian Competition," *Bell J. Econ.*, 9:2, pp. 524-48.
- Olley, G. Steven and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64:6, pp. 1263-97.
- Pakes, Ariel and Richard Ericson. Forthcoming. "Empirical Implications of Alternative Models of Firm Dynamics," *J. Econ. Theory*.
- Power, Laura. Forthcoming. "The Missing Link: Technology, Investment, and Productivity," *Rev. Econ. Statist.*
- Ravenscraft, David J. and F. M. Scherer. 1987. *Mergers, Sell-offs, and Economic Efficiency*. Washington, DC: Brookings Institution.
- Roberts, Mark J. 1996a. "Colombia, 1977-85: Producer Turnover, Margins, and Trade Exposure," in Roberts and Tybout 1996, pp. 227-59.
- . 1996b. "Employment Flows and Producer Turnover," in Roberts and Tybout 1996, pp. 18-42.
- and James R. Tybout, eds. 1996. *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*. New York: Oxford U. Press for World Bank.
- Robinson, William T. 1993. "Are Sutton's Predictions Robust? Empirical Evidence on Price Competition, Advertising, and the Evolution of Concentration," working paper, School of Business Administration, U. Michigan.
- and Claes Fornell. 1985. "Sources of Market Pioneer Advantages in Consumer Goods Industries," *J. Marketing Res.*, 22:3, pp. 305-17.
- Rosenbaum, David I. and Fabian Lamort. 1992. "Entry Barriers, Exit, and Sunk Costs: An Analysis," *Appl. Econ.*, 24:3, pp. 297-304.
- Santarelli, Enrico and Alessandro Starlacchini. 1994. "New Firm Formation in Italian Industry: 1985-89," *Small Bus. Econ.*, 6:2, pp. 95-106.
- Schmalensee, Richard. 1989. "Inter-Industry Studies of Structure and Performance," in *Handbook of Industrial Organization*. Richard Schmalensee and Robert D. Willig, eds. Amsterdam: North-Holland Publishing Co., 1989. II, pp. 951-1009.
- Schwalbach, Joachim. 1991. "Entry, Exit, Concentration, and Market Contestability," in Geroski and Schwalbach 1991, pp. 121-40.
- Shen, T. Y. 1968. "Competition, Technology, and Market Shares," *Rev. Econ. Statist.*, 50:1, pp. 293-310.
- Simon, Herbert A. and Charles P. Bonini. 1958. "The Size Distribution of Business Firms," *Amer. Econ. Rev.*, 48:4, pp. 607-17.
- Sleuwaegen, Leo and Wim Dehandschutter. 1991. "Entry and Exit in Belgian Manufacturing," in Geroski and Schwalbach 1991, pp. 111-20.
- Stonebraker, Robert J. 1976. "Corporate Profits and the Risk of Entry," *Rev. Econ. Statist.*, 58:1, pp. 33-39.
- Sutton, John. 1991. *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*. Cambridge, MA: M.I.T. Press.
- . 1997. "Gibrat's Legacy," *J. Econ. Lit.*, 35:1, pp. 40-59.

- . Forthcoming. *Technology and Market Structure: Theory and History*. Cambridge, MA: M.I.T. Press.
- Torii, Akio. 1992. "Technical Efficiency in Japanese Industries," in Caves and Associates 1992, pp. 31–119.
- Troske, Kenneth R. 1996. "The Dynamic Adjustment Process of Firm Entry and Exit in Manufacturing and Finance, Insurance, and Real Estate," *J. Law Econ.*, 39:2, pp. 705–35.
- Tybout, James R. 1996a. "Chile, 1979–86: Trade Liberalization and Its Aftermath," in Roberts and Tybout 1996, pp. 200–26.
- . 1996b. "Heterogeneity and Productivity Growth: Assessing the Evidence," in Roberts and Tybout 1996, pp. 43–72.
- von der Fehr, Nils-Henrik Mørch. 1991. "Domestic Entry in Norwegian Manufacturing Industries," in Geroski and Schwalbach 1991, pp. 89–110.
- Wagner, Joachim. 1992. "Firm Size, Firm Growth, and Persistence of Chance: Testing Gibrat's Law with Establishment Data from Lower Saxony, 1978–1989," *Small Bus. Econ.*, 4:2, pp. 125–31.
- . 1994a. "The Post-Entry Performance of New Small Firms in German Manufacturing Industries," *J. Ind. Econ.*, 42:2, pp. 141–54.
- . 1994b. "Small Firm Entry in Manufacturing Industries: Lower Saxony, 1979–1989," *Small Bus. Econ.*, 6:3, pp. 211–23.
- Waring, Geoffrey F. 1996. "Industry Differences in the Persistence of Firm-specific Returns," *Amer. Econ. Rev.*, 86:5, pp. 1253–65.
- Wedervang, Frøystein. 1965. *Development of a Population of Industrial Firms: The Structure of Manufacturing Industries in Norway, 1930–1948*. Oslo: Universitetsforlaget.
- Yip, George S. 1982. *Barriers to Entry: A Corporate-strategy Perspective*. Lexington, MA: Lexington Books.
- Yoo, Seong Min. 1992. "Technical Efficiency over Time in Korea, 1978–88: Exploratory Analysis," in Caves and Associates, pp. 313–27.

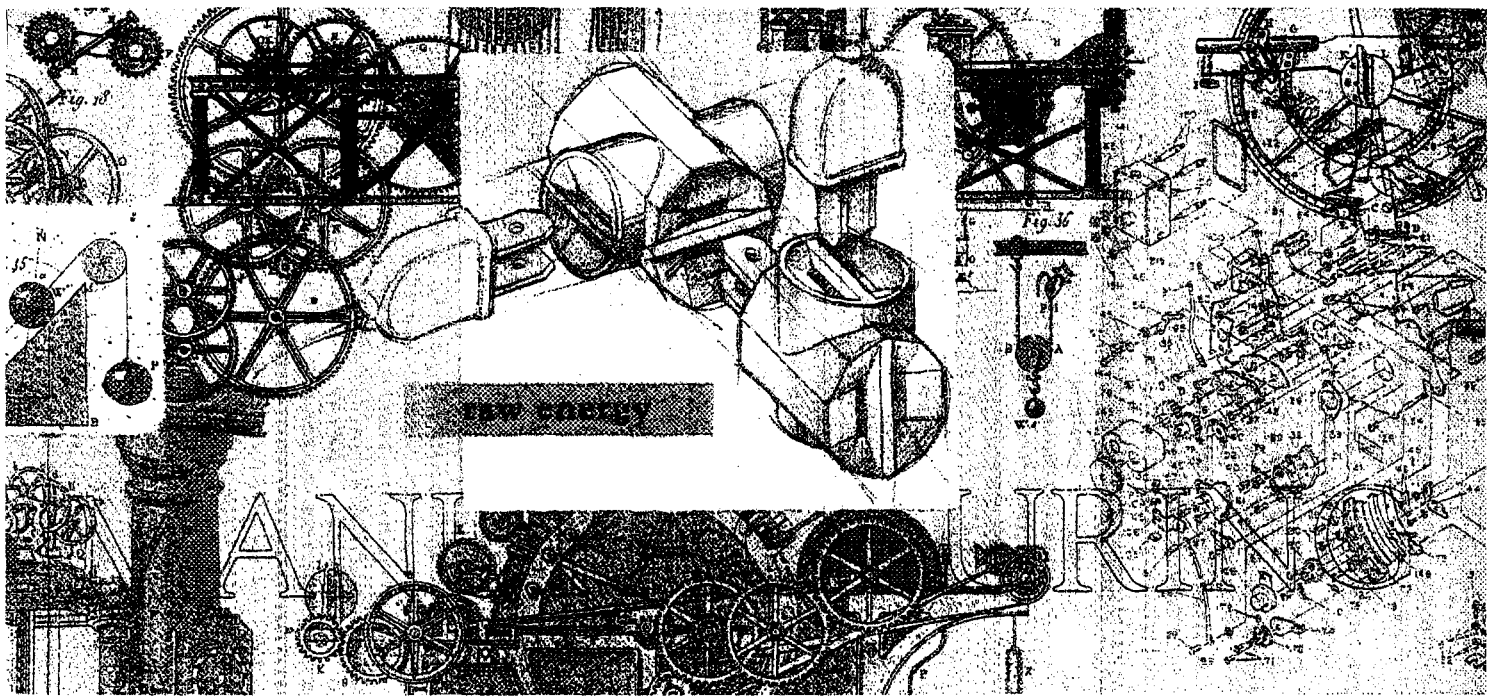
Exhibit C

This Exhibit contains an article by Arthur from 1996.

THIS IS EXHIBIT C TO  
THE AFFIDAVIT OF \_\_\_\_\_  
Arlen Helle  
DATED THIS 2 DAY, Dec, 2005

COMMISSIONER FOR  
OATHS #0694001  
DEBBIE PIAZZA  
EXPIRES FEB 06/08  
*Deborah K. Piazza*





*Increasing returns cause products that are ahead to get further ahead.*

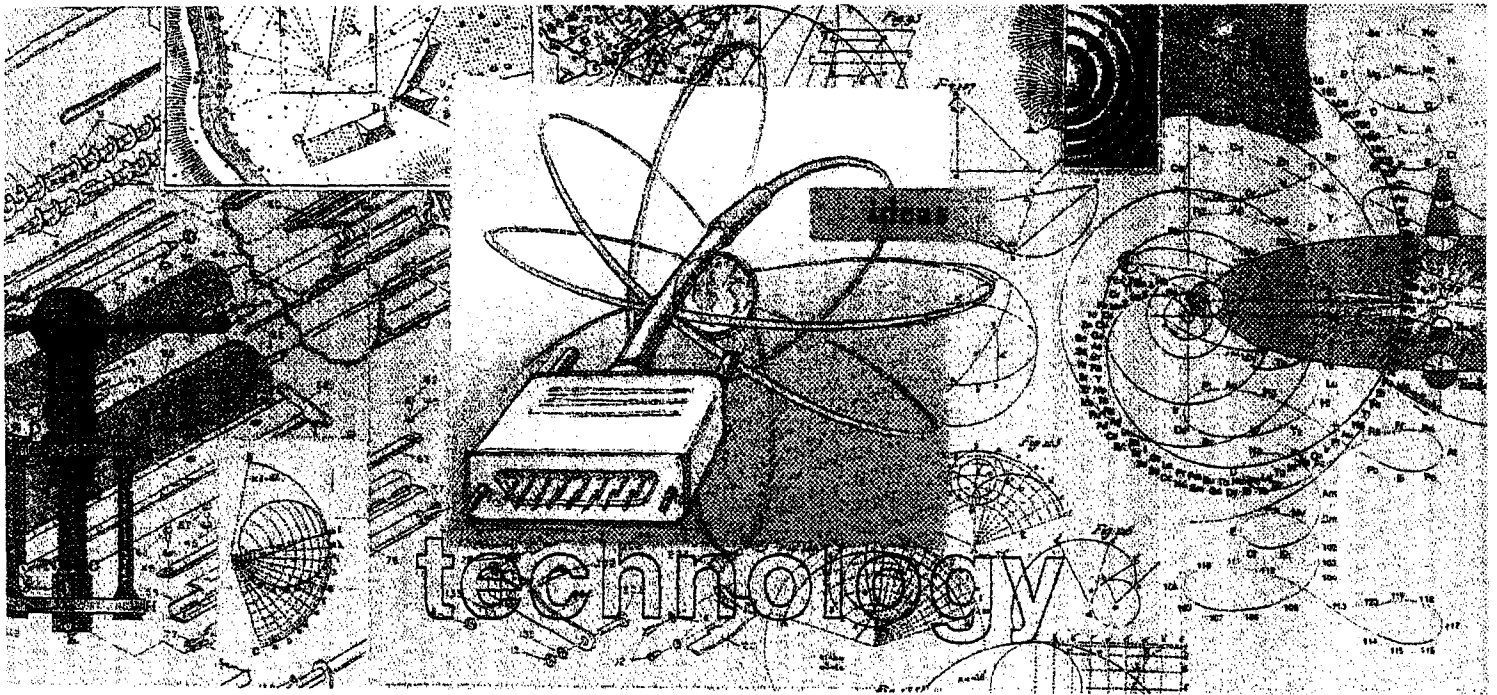
# Increasing Returns and the

by W. Brian Arthur

Our understanding of how markets and businesses operate was passed down to us more than a century ago by a handful of European economists—Alfred Marshall in England and a few of his contemporaries on the continent. It is an understanding based squarely upon the assumption of diminishing returns: products or companies that get ahead in a market eventually run into limitations, so that a predictable equilibrium of prices and market shares is reached. The theory was roughly valid for the bulk-processing, smokestack economy of Marshall's day. And it still thrives in today's economics textbooks. But steadily and continuously in this century, Western economies have undergone a transformation from bulk-material manufacturing to design and use of technology—from processing of resources to processing of information, from application of raw energy to application of ideas. As this shift has occurred, the underlying mechanisms that determine economic behavior have shifted from ones of diminishing to ones of *increasing* returns.

Increasing returns are the tendency for that which is ahead to get further ahead, for that which loses advantage to lose further advantage. They are mechanisms of positive feedback that operate—within markets, businesses, and industries—to reinforce that which gains success or aggravate that which suffers loss. Increasing returns generate not equilibrium but instability: If a product or a company or a technology—one of many competing in a market—gets ahead by chance or clever strategy, increasing returns can magnify this advantage, and the product or company or technology can go on to lock in the market. More than causing products to become standards, increasing returns cause busi-

*W. Brian Arthur is the Dean and Virginia Morrison Professor of Economics and Population Studies at Stanford University in Stanford, California, and Citibank Professor at Santa Fe Institute in Santa Fe, New Mexico. He is the author of *Increasing Returns and Path Dependence in the Economy* (University of Michigan Press, 1994). His Web site is [www.santafe.edu/arthur](http://www.santafe.edu/arthur).*



*The concept has revolutionized economics. Business is next.*

# New World of Business

nesses to work differently, and they stand many of our notions of how business operates on their head.

Mechanisms of increasing returns exist alongside those of diminishing returns in all industries. But roughly speaking, diminishing returns hold sway in the traditional part of the economy – the processing industries. Increasing returns reign in the newer part – the knowledge-based industries. Modern economies have therefore bifurcated into two inter-related worlds of business corresponding to the two types of returns. The two worlds have different economics. They differ in behavior, style, and culture. They call for different management techniques, strategies, and codes of government regulation.

They call for different understandings.

## Alfred Marshall's World

Let's go back to beginnings – to the diminishing-returns view of Alfred Marshall and his contemporaries. Marshall's world of the 1880s and 1890s was

one of bulk production: of metal ores, aniline dyes, pig iron, coal, lumber, heavy chemicals, soybeans, coffee – commodities heavy on resources, light on know-how. In that world it was reasonable to suppose, for example, that if a coffee plantation expanded production it would ultimately be driven to use land less suitable for coffee. In other words, it would run into diminishing returns. So if coffee plantations competed, each one would expand until it ran into limitations in the form of rising costs or diminishing profits. The market would be shared by many plantations, and a market price would be established at a predictable level – depending on tastes for coffee and the availability of suitable farmland. Planters would produce coffee so long as doing so was profitable, but because the price would be squeezed down to the average cost of production, no one would be able to make a killing. Marshall said such a market was in perfect competition, and the economic world he envisaged fitted beautifully with the Victorian values of his time. It

was at equilibrium and therefore orderly, predictable and therefore amenable to scientific analysis, stable and therefore safe, slow to change and therefore continuous. Not too rushed, not too profitable. In a word, mannerly. In a word, genteel.

With a few changes, Marshall's world lives on a century later within that part of the modern economy still devoted to bulk processing: of grains, livestock, heavy chemicals, metals and ores, foodstuffs, retail goods – the part where operations are largely repetitive day to day or week to week. Product differentiation and brand names now mean that a few companies rather than many compete in a given market. But typically, if these companies try to expand, they run into some limitation: in numbers of consumers who prefer their brand, in regional demand, in access to raw materials. So no company can corner the market. And because such products are normally substitutable for one another, something like a standard price emerges. Margins are thin and nobody makes a killing. This isn't exactly Marshall's perfect competition, but it approximates it.

### The Increasing-Returns World

What would happen if Marshall's diminishing returns were reversed so that there were *increasing* returns? If products that got ahead thereby got further ahead, how would markets work?

Let's look at the market for operating systems for personal computers in the early 1980s when CP/M, DOS, and Apple's Macintosh systems were competing. Operating systems show increasing returns: if one system gets ahead, it attracts further software developers and hardware manufacturers to adopt it, which helps it get further ahead. CP/M was first in the market and by 1979 was well established. The Mac arrived later, but it was wonderfully easy to use. DOS was born when Microsoft locked up a deal in 1980 to supply an operating system for the IBM PC. For a year or two, it was by no means clear

which system would prevail. The new IBM PC – DOS's platform – was a kludge. But the growing base of DOS/IBM users encouraged software developers such as Lotus to write for DOS. DOS's prevalence – and the IBM PC's – bred further prevalence, and eventually the DOS/IBM combination came to dominate a considerable portion of the market. That history is now well known. But notice several things: It was not predictable in advance (before the IBM deal) which system would come to dominate. Once DOS/IBM got ahead, it locked in the market because it did not pay for users to switch. The dominant system was not the best: DOS was derided by computer professionals. And once DOS locked in the market, its sponsor, Microsoft, was able to spread its costs over a large base of users. The company enjoyed killer margins.

These properties, then, have become the hallmarks of increasing returns: market instability (the market tilts to favor a product that gets ahead), multiple potential outcomes (under different events in history, different operating systems could have won), unpredictability, the ability to lock in a market, the possible predominance of an inferior product, and fat profits for the winner. They surprised me when I first perceived them in the late 1970s. They were also repulsive to economists brought up on the order, predictability, and optimality of Marshall's world. Glimpsing some of these properties in 1939, English economist John Hicks warned that admitting increasing returns would lead to "the wreckage of the greater part of economic theory." But Hicks had it wrong: the theory of increasing returns does not destroy the standard theory – it complements it. Hicks felt repugnance not just because of unsavory properties but also because in his day no mathematical apparatus existed to analyze increasing-returns markets. That situation has now changed. Using sophisticated techniques from qualitative dynamics and probability theory, I and others have developed methods to analyze increasing-returns markets. The theory

In 1939, English economist  
John Hicks warned that admitting  
increasing returns would lead to  
"the wreckage of the greater part of  
economic theory."

But Hicks had it wrong.

of increasing returns is new, but it already is well established. And it renders such markets amenable to economic understanding.

In the early days of my work on increasing returns, I was told they were an anomaly. Like some exotic particle in physics, they might exist in theory but would be rare in practice. And if they did exist, they would last for only a few seconds before being arbitrated away. But by the mid-1980s, I realized increasing returns were neither rare nor ephemeral. In fact, a major part of the economy was subject to increasing returns – high technology.

Why should this be so? There are several reasons:

**Up-front Costs.** High-tech products—pharmaceuticals, computer hardware and software, aircraft and missiles, telecommunications equipment, bio-engineered drugs, and suchlike—are by definition complicated to design and to deliver to the marketplace. They are heavy on know-how and light on resources. Hence they typically have R&D costs that are large relative to their unit production costs. The first disk of Windows to go out the door cost Microsoft \$50 million; the second and subsequent disks cost \$3. Unit costs fall as sales increase.

**Network Effects.** Many high-tech products need to be compatible with a network of users. So if much downloadable software on the Internet will soon appear as programs written in Sun Microsystems' Java language, users will need Java on their computers to run them. Java has competitors. But the more it gains prevalence, the more likely it will emerge as a standard.

**Customer Groove-in.** High-tech products are typically difficult to use. They require training. Once users invest in this training—say, the maintenance and piloting of Airbus passenger aircraft—they merely need to update these skills for subsequent versions of the product. As more market is captured, it becomes easier to capture future markets.

In high-tech markets, such mechanisms ensure that products that gain market advantage stand to gain further advantage, making these markets un-

stable and subject to lock-in. Of course, lock-in is not forever. Technology comes in waves, and a lock-in such as DOS's can last only as long as a particular wave lasts.

So we can usefully think of two economic regimes or worlds: a bulk-production world yield-

ing products that essentially are congealed resources with a little knowledge and operating according to Marshall's principles of diminishing returns, and a knowledge-based part of the economy yielding products that essentially are congealed knowledge with a little resources and operating under increasing returns. The two worlds are not neatly split.

Some products—like the IBM PC—  
start in the increasing-returns  
world but later in their life  
cycle become virtual commodities  
that belong to Marshall's  
processing world.

Hewlett-Packard, for example, designs knowledge-based devices in Palo Alto, California, and manufactures them in bulk in places like Corvallis, Oregon, or Greeley, Colorado. Most high-tech companies have both knowledge-based operations and bulk-processing operations. But because the rules of the game differ for each, companies often separate them—as Hewlett-Packard does. Conversely, manufacturing companies have operations such as logistics, branding, marketing, and distribution, which belong largely to the knowledge world. And some products—like the IBM PC—start in the increasing-returns world but later in their life cycle become virtual commodities that belong to Marshall's processing world.

## The Halls of Production and the Casino of Technology

Because the two worlds of business—processing bulk goods and crafting knowledge into products—differ in their underlying economics, it follows that they differ in their character of competition and their culture of management. It is a mistake to think that what works in one world is appropriate for the other.

There is much talk these days about a new management style that involves flat hierarchies, mission orientation, flexibility in strategy, market

positioning, reinvention, restructuring, reengineering, repositioning, reorganization, and re-everything else. Are these new insights or are they fads? Are they appropriate for all organizations? Why are we seeing this new management style?

Let us look at the two cultures of competition. In bulk processing, a set of standard prices typically emerges. Production tends to be repetitive—much the same from day to day or even from year to year. Competing therefore means keeping product flowing, trying to improve quality, getting costs down. There is an art to this sort of management, one widely discussed in the literature. It favors an environment free of surprises or glitches—an environment characterized by control and planning. Such an environment requires not just people to carry out production but also people to plan and control it. So it favors a hierarchy of bosses and workers. Because bulk processing is repetitive, it allows constant improvement, constant optimization. And so, Marshall's world tends to be one that favors hierarchy, planning, and controls. Above all, it is a world of optimization.

Competition is different in knowledge-based industries because the economics are different. If knowledge-based companies are competing in winner-take-most markets, then managing becomes redefined as a series of quests for the next technological winner—the next cash cow. The goal becomes the search for the Next Big Thing. In this milieu, management becomes not production oriented but mission oriented. Hierarchies flatten not because democracy is suddenly bestowed on the workforce or because computers can cut out much of middle management. They flatten because, to be effective, the deliverers of the next-thing-for-the-company need to be organized like commando units in small teams that report directly to the CEO or to the board. Such people need free rein. The company's future survival depends upon them. So they—and the commando teams that report to them in turn—will be treated not as employees but as equals in the business of the company's success. Hierarchy dissipates and dissolves.

Does this mean that hierarchy should disappear in meatpacking, steel production, or the navy? Contrary to recent management evangelizing, a style that is called for in Silicon Valley will not necessarily be appropriate in the processing world. An aircraft's safe arrival depends on the captain, not on the flight attendants. The cabin crew can usefully be "empowered" and treated as human beings. This approach is wise and proper. But forever there will be a distinction—a hierarchy—between cockpit and cabin crews.

In fact, the style in the diminishing-returns Halls of Production is much like that of a sophisticated modern factory: the goal is to keep high-quality product flowing at low cost. There is little need to watch the market every day, and when things are going smoothly the tempo can be leisurely. By contrast, the style of competition in the increasing-returns arena is more like gambling. Not poker, where the game is static and the players vie for a succession of pots. It is casino gambling, where part of the game is to choose which games to play, as well as playing them with skill. We can imagine the top figures in high tech—the Gateses and Gerstners and Groves of their industries—as milling in a large casino. Over at this table, a game is starting called multimedia. Over at that one, a game called Web services. In the corner is electronic banking. There are many such tables. You sit at one. How much to play? you ask. Three billion, the croupier replies. Who'll be playing? We won't know until they show up. What are the rules? Those'll emerge as the game unfolds. What are my odds of winning? We can't say. Do you still want to play?

High technology, pursued at this level, is not for the timid.

In fact, the art of playing the tables in the Casino of Technology is primarily a psychological one. What counts to some degree—but only to some degree—is technical expertise, deep pockets, will, and courage. Above all, the rewards go to the players who are first to make sense of the new games looming out of the technological fog, to see their shape, to cognize them. Bill Gates is not so much a wizard of technology as a wizard of precognition, of discerning the shape of the next game.

We can now begin to see that the new style of management is not a fad. The knowledge-based part of the economy demands flat hierarchies, mission orientation, above all a sense of direction. Not five-year plans. We can also fathom the mystery of what I've alluded to as *re-everything*. Much of this "re-everything" predilection—in the bulk-processing world—is a fancy label for streamlining, computerizing, downsizing. However, in the increasing-returns world, especially in high tech, re-everything has become necessary because every time the quest changes, the company needs to change. It needs to reinvent its purpose, its goals, its way of doing things. In short, it needs to adapt. And adaptation never stops. In fact, in the increasing-returns environment I've just sketched, standard optimization makes little sense. You cannot optimize in the casino of increasing-returns games. You can be smart. You can be cunning. You can position. You can observe. But when the games them-

selves are not even fully defined, you cannot optimize. What you *can* do is adapt. Adaptation, in the proactive sense, means watching for the next wave that is coming, figuring out what shape it will take, and positioning the company to take advantage of it. Adaptation is what drives increasing-returns businesses, not optimization.

## Playing the High-Tech Tables

Suppose you are a player in the knowledge-industry casino, in this increasing-returns world. What can you do to capitalize on the increasing returns at your disposal? How can you use them to capture markets? What strategic issues do you need to think about? In the processing world, strategy typically hinges upon capitalizing on core competencies, pricing competitively, getting costs down, bringing quality up. These are important also in the knowledge-based world, but so, too, are other strategies that make use of the special economics of positive feedbacks.

Two maxims are widely accepted in knowledge-based markets: it pays to hit the market first, and it pays to have superb technology. These maxims are true but do not guarantee success. Prodigy was first into the on-line services market but was passive in building its subscriber base to take advantage of increasing returns. As a result, it has fallen from its leading position and currently lags the other services. As for technology, Steve Jobs's NeXT workstation was superb. But it was launched into a market already dominated by Sun Microsystems and Hewlett-Packard. It failed. A new product often has to be two or three times better in some dimension – price, speed, convenience – to dislodge a locked-in rival. So in knowledge-based markets, entering first with a fine product can yield advantage. But as strategy, this is still too passive. What is needed is *active* management of increasing returns.

One active strategy is to discount heavily initially to build up an installed base. Netscape handed

out its Internet browser for free and won 70% of its market. Now it can profit from spin-off software and applications. Although such discounting is effective – and widely understood – it is not always implemented. Companies often err by pricing high initially to recoup expensive R&D costs. Yet even smart discounting to seed the market is ineffective unless the resulting installed base is exploited later. America Online built up a lead of more than 4.5 million subscribers by giving away free services. But because of the Internet's dominance, it is not yet clear whether it can transform this huge base into later profits.

Let's get a bit more sophisticated. Technological products do not stand alone. They depend on the existence of other products and other technologies. The Internet's World Wide Web operates within a grouping of businesses that include browsers, on-line news, E-mail, network retailing, and financial services. Pharmaceuticals exist within a network of physicians, testing labs, hospitals, and HMOs. Laser printers are part of a grouping of products that include computers, publishing software, scanners, and photo-input devices. Unlike products of the processing world, such as soybeans or rolled steel, technological products exist within local groupings of products that support and enhance them. They exist in mini-ecologies.

This interdependence has deep implications for strategy. When, in the mid-1980s, Novell introduced its network-operating system, NetWare, as a way of connecting personal computers in local networks, Novell made sure that NetWare was technically superior to its rivals. It also heavily discounted NetWare to build an installed base. But these tactics were not enough. Novell recognized that NetWare's success depended on attracting software

applications to run on NetWare – which was a part of the ecology outside the company's control. So it set up incentives for software developers to write for NetWare rather than for its rivals. The software writers did just that. And by building NetWare's

Adaptation means watching for  
the next wave and positioning the  
company to take advantage of it.

Adaptation is what drives  
increasing-returns businesses,  
not optimization.



## In the Case of Microsoft...

What should be legal in this powerful and as yet unregulated world of increasing returns? What constitutes fair play? Should technology markets be regulated, and if so in what way? These questions have come to a head with the enormous amount of publicity generated by the U.S. Justice Department's current antitrust case against Microsoft.

In Marshall's world, antitrust regulation is well understood. Allowing a single player to control, say, more than 35% of the silver market is tantamount to allowing monopoly pricing, and the government rightly steps in. In the increasing-returns world, things are more complicated. There are arguments in favor of allowing a product or company in the web of technology to dominate a market, as well as arguments against. Consider these pros and cons.

**Convenience.** A locked-in product may provide a single standard of convenience. If a software company such as Microsoft allows us to double-click all the way from our computer screen straight to our bank account (by controlling all the technologies in between), this avoids a tedious balkanizing of standards, where we have to spend useless time getting into a succession of on-line connection products.

**Fairness.** If a product locks in because it is superior, this is fair, and it would be foolish to penalize such success. If it locks in merely because user base was levered over from a neighboring lock-in, this is unfair.

**Technology Development.** A locked-in product may obstruct technological advancement. If a clunker such as DOS locks up the PC market for ten years, there is little incentive for other companies to develop alternatives. The result is impeded technological progress.

**Pricing.** To lock in, a product usually has been discounted, and this established low price is often hard to raise. So monopoly pricing—of great concern in bulk-processing markets—is therefore rarely a major worry

Added to these considerations, high tech is not a commodity industry. Dominance may consist not so much in cornering a single product as in successively taking over more and more threads of the web of technology, thereby preventing other players from getting access to new, breaking markets. It would be difficult to separate out each thread and to regulate it. And of course it may be impracticable to regulate a market before it forms—before it is even fully defined. There are no simple answers to antitrust regulation in the increasing-returns world. On balance, I would favor a high degree of regulatory restraint, with the addition of two key principles.

□ **Do not penalize success.** Short-term monopolization of an increasing-returns market is correctly perceived as a reward or prize for innovation and risk taking. There is a temptation to single out dominant players and hit them with an antitrust suit. This reduces regulation to something like a brawl in an Old West saloon—if you see a head, hit it. Not a policy that preserves an incentive to innovate in the first place.

□ **Don't allow head starts for the privileged.** This means that as a new market opens up—such as electronic consumer banking—companies that already dominate standards, operating systems, and neighboring technologies should not be allowed a ten-mile head start in the land rush that follows. All competitors should have fair and open access to the applicable technologies and standards.

In practice, these principles would mean allowing the possibility of winner-take-all jackpots in each new subindustry, in each new wave of technology. But each contender should have access to whatever degree possible to the same technologies, the same open standards, so that all are lined up behind the same starting line. If industry does not make such provisions voluntarily, government regulation will impose them.

success, they ensured their own. Novell managed these cross-product positive feedbacks actively to lock in its market. It went on to profit hugely from upgrades, spin-offs, and applications of its own.

Another strategy that uses ecologies is linking and leveraging. This means transferring a user base built up upon one node of the ecology (one product) to neighboring nodes, or products. The strategy is very much like that in the game Go: you surround neighboring markets one by one, lever your user base onto them, and take them over—all the time enhancing your position in the industry. Microsoft levered its 60-million-person user base in DOS onto Windows, then onto Windows 95, and then onto

Microsoft Network by offering inexpensive upgrades and by bundling applications. The strategy has been challenged legally. But it recognizes that positive feedbacks apply across markets as well as within markets.

In fact, if technological ecologies are now the basic units for strategy in the knowledge-based world, players compete not by locking in a product on their own but by building *webs*—loose alliances of companies organized around a mini-ecology—that amplify positive feedbacks to the base technology. Apple, in closing its Macintosh system to outsiders in the 1980s, opted not to create such a web. It believed that with its superior technology, it could

hold its increasing-returns market to itself. Apple indeed dominates its Mac-based ecology. But this ecology is now only 8% of the personal computer business. IBM erred in the other direction. By passively allowing other companies to join its PC web as clones, IBM achieved a huge user base and locked in the market. But the company itself wound up with a small share of the spoils. The key in web building is active management of the cross-company mutual feedbacks. This means making a careful choice of partners to build upon. It also means that, rather than attempting to take over all products in the ecology, dominant players in a web should allow dependent players to lock in their dependent products by piggybacking on the web's success. By thus ceding some of the profits, the dominant players ensure that all participants remain committed to the alliance.

Important also to strategy in knowledge-based markets is psychological positioning. Under increasing returns, rivals will back off in a market not only if it is locked in but if they *believe* it will be locked in by someone else. Hence we see psychological jockeying in the form of preannouncements, feints, threatened alliances, technological preening, touted future partnerships, parades of vaporware (announced products that don't yet exist). This posturing and puffing acts much the way similar behavior does in a primate colony: it discourages competitors from taking on a potentially dominant rival. No moves need be made in this strategy of premarket facedown. It is purely a matter of psychology.

What if you hold a losing hand? Sometimes it pays to hold on for residual revenue. Sometimes a fix can be provided by updated technology, fresh alliances, or product changes. But usually under heavy lock-in, these tactics do not work. The alternatives are then slow death or graceful exit—relinquishing the field to concentrate on positioning for the next technology wave. Exit may not mean quitting the business entirely. America Online, CompuServe, Prodigy, and Microsoft Network have all ceded dominance of the on-line computer networking market to the Internet. But instead of exiting, they are steadily becoming adjuncts of the Net, supplying content services such as financial quotations or games and entertainment. They have lost the main game. But they will likely continue in a side game with its own competition for dominance within the Net's ecology.

Above all, strategy in the knowledge world requires CEOs to recognize that a different kind of economics is at work. CEOs need to understand which positive and negative feedback mechanisms

are at play in the market ecologies in which they compete. Often there are several such mechanisms—interbraided, operating over different time frames, each needing to be understood, observed, and actively managed.

## What About Service Industries?

So far, I've talked mainly about high tech. Where do service industries such as insurance, restaurants, and banking fit in? Which world do they belong to? The question is tricky. It would appear that such industries belong to the diminishing-returns, processing part of the economy because often there are regional limits to the demand for a given service, most services do consist of "processing" clients, and services are low-tech.

The truth is that network or user-base effects often operate in services. Certainly, retail franchises exist because of increasing returns. The more McDonald's restaurants or Motel 6 franchises are out there geographically, the better they are known. Such businesses are patronized not just for their quality but also because people want to know exactly what to expect. So the more prevalent they are, the more prevalent they can become. Similarly, the larger a bank's or insurance company's customer base, the more it can spread its fixed costs of headquarters staff, real estate, and computer operations. These industries, too, are subject to mild increasing returns.

So we can say more accurately that service industries are a hybrid. From day to day, they act like bulk-processing industries. But over the long term, increasing returns will dominate—even though their destabilizing effects are not as pronounced as in high tech. The U.S. airline business, for example, processes passengers day to day. So it seemed in 1981 that deregulation should enhance competition, as it normally does under diminishing returns. But over the long term, airlines in fact experience a positive feedback: under the hub-and-spoke system, once an airline gets into trouble, it cannot work the feeder system for its routes properly, its fleet ages, it starts a downward spiral, and it loses further routes. The result of deregulation over the long term has been a steady decline in large carriers, from 15 airlines in 1981 to approximately 6 at present. Some routes have become virtual monopolies, with resulting higher fares. None of this was intended. But it should have been predicted—given increasing returns.

In fact, the increasing-returns character of service industries is steadily strengthening. One of the marks of our time is that in services everything is



going software – everything that is information based. So operations that were once handled by people – designing fancy financial instruments or automobiles or fashion goods, processing insurance claims, supplying and inventorying in retail, conducting paralegal searches for case precedents – are increasingly being handled by software. As this reengineering of services plays out, centralized software facilities come to the fore. Service providers become hitched into software networks, regional limitations weaken, and user-base network effects kick in.

This phenomenon can have two consequences. First, where the local character of service remains important, it can preserve a large number of service companies but clustered round a dominant software provider – like the large numbers of small, independent law firms tied in to the dominant computer-search network, Lexis-Nexis. Or physicians tied in to an HMO. Second, where locality is unimportant, network effects can transform competition toward the winner-take-most character we see in high tech. For example, when Internet-based retail banking arrives, regional demand limitations will vanish. Each virtual bank will gain in advantage as its network increases. Barring regulation, consumer banking will then become a contest among a few large banking networks. It will become an increasing-returns business.

Services belong to both the processing and the increasing-returns world. But their center of gravity is crossing over to the latter.

### Thoughts for Managers

Where does all this leave us? At the beginning of this century, industrial economies were based largely on the bulk processing of resources. At the close of the century, they are based on the processing of resources *and* on the processing of knowledge. Economies have bifurcated into two worlds – intertwined, overlapping, and different. These two worlds operate under different economic principles. Marshall's world is characterized by planning, control, and hierarchy. It is a world of materials, of processing, of optimization. The increasing-returns world is characterized by observation, positioning, flattened organizations, missions, teams, and cunning. It is a world of psychology, of cognition, of adaptation.

Many managers have some intuitive grasp of this new increasing-returns world. Few understand it thoroughly. Here are some questions managers need to ask themselves when they operate in knowledge-based markets:

**Do I understand the feedbacks in my market?** In the processing world, understanding markets means understanding consumers' needs, distribution channels, and rivals' products. In the knowledge world, success requires a thorough understanding of the self-negating and self-reinforcing feedbacks in the market – the diminishing- and increasing-returns mechanisms. These feedbacks are interwoven and operate at different levels in the market and over different time frames.

**Which ecologies am I in?** Technologies exist not alone but in an interlinked web, or ecology. It is important to understand the ecologies a company's products belong to. Success or failure is often decided not just by the company but also by the success or failure of the web it belongs to. Active management of such a web can be an important magnifier of increasing returns.

**Do I have the resources to play?** Playing one of the increasing-returns games in the Casino of Technology requires several things: excellent technology, the ability to hit the market at the right time, deep pockets, strategic pricing, and a willingness to sacrifice current profits for future advantage. All this is a matter not just of resources but also of courage, resolution, will. And part of that resolution, that courage, is also the decisiveness to leave the market when increasing returns are moving against one. Hanging on to a losing position that is being further eroded by positive feedbacks requires throwing reinforcements into a battle already lost. Better to exit with financial dignity.

**What games are coming next?** Technology comes in successive waves. Those who have lost out on this wave can position for the next. Conversely, those who have made a killing on this cycle should not become complacent. The ability to profit under increasing returns is only as good as the ability to see what's coming in the next cycle and to position oneself for it – technologically, psychologically, and cooperatively. In high tech, it is as if we are moving slowly on a ship, with new technologies looming, taking shape, through a fog of unknowingness. Success goes to those who have the vision to foresee, to imagine, what shapes these next games will take.

These considerations appear daunting. But increasing-returns games provide large payoffs for those brave enough to play them and win. And they are exciting. Processing, in the service or manufacturing industries, has its own risks. Precisely because processing is low-margin, operations must struggle to stay afloat. Neither world of business is for the fainthearted.

In his book *Microcosm*, technology thinker George Gilder remarked, "The central event of the

twentieth century is the overthrow of matter. In technology, economics, and the politics of nations, wealth in the form of physical resources is steadily declining in value and significance. The powers of mind are everywhere ascendant over the brute force of things." As the economy shifts steadily away from the brute force of things into the powers of mind, from resource-based bulk processing into

knowledge-based design and reproduction, so it is shifting from a base of diminishing returns to one of increasing returns. A new economics—one very different from that in the textbooks—now applies, and nowhere is this more true than in high technology. Success will strongly favor those who understand this new way of thinking.

Reprint 96401

To order reprints, see the last page of this issue.

# NEW METHOD PATENTS

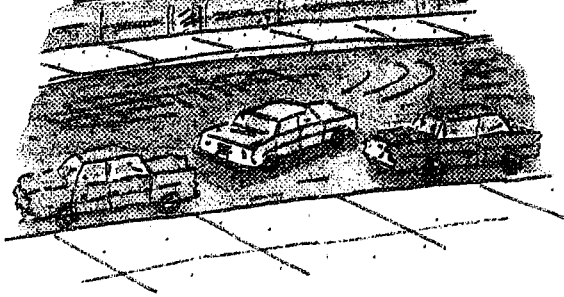
Stan D.'s Comb-over Technique  
U.S. Patent # 6439172642553156



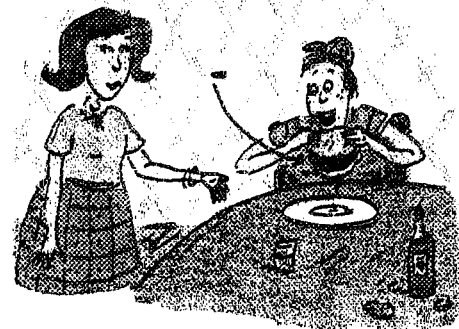
Millie N.'s Jar-Opening Procedure  
U.S. Patent # 53,27912,886,541



Fred S.'s Parallel-Parking Method  
U.S. Patent # 293846113962781



Julie's Never-Fail Child Vegetable-Ingestion Scheme  
U.S. Patent # 5,339,720,160,288



R. Chast