UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

WOCKHARDT BIO AG
Petitioner

v.

JANSSEN ONCOLOGY, INC.
Patent Owner

Case IPR2016-01582
U.S. Patent No. 8,822,438

# SUPPLEMENTAL DECLARATION OF ROBERT D. STONER, Ph.D.

I, Robert D. Stoner, hereby declare as follows.

1. I am over the age of eighteen and otherwise competent to make this declaration.

2. I am the same Robert D. Stoner who submitted a declaration on August 10, 2016 in the *inter partes* review proceeding IPR2016-01582. My initial declaration was marked as Wockhardt Exhibit 1077.

3. I understand from counsel that Patent Owner Janssen Oncology, Inc. filed a paper on February 2, 2017, which, in part, objected to certain Exhibits cited to and filed with my initial declaration.

4. I understand from counsel that Janssen objected to Exhibits 1048-1050, 1053, 1054, 1057, 1060-1063, 1065-1074, 1076, 1080, and to Attachments B-1 and B-2 of Exhibit 1077 cited in my initial declaration. I understand that Janssen asserted that the Petition and my Declaration did not establish the origin of those documents or that the documents were a true and correct copy of what they purport to be. I disagree with Janssen's assertions and submit this supplemental declaration in response.

5. Exhibits 1048-1050 and 1054 are copies of background information on prostate cancer from the Mayo Clinic, American Cancer Society (ACS), and American Society of Clinical Oncology (ASCO) websites. Experts in finance and product valuation routinely rely on materials such as Exhibits 1048-1050 and 1054

when analyzing a product's use and market, such as in the opinion set forth in my initial declaration. Exhibit 1049 is a true and correct copy of the ACS web page as published on the ACS's website, www.cancer.org, accessed on August 8, 2016. Exhibit 1050 is a true and correct copy of the ASCO web page as published on ASCO's website, www.cancer.net, accessed on August 9, 2016. Exhibit 1054 is a true and correct copy of the Mayo Clinic web page as published on the Mayo Clinic's website, www.mayoclinic.org, accessed on August 8, 2016. Exhibit 1048 was submitted in IPR2016-00286 as Amerigen Exhibit 1051 by Dr. DeForest McDuff. In IPR2016-00286, Dr. McDuff attested to the authenticity of Amerigen Exhibit 1051 in ¶10 of Amerigen Exhibit 1068.

6. Exhibits 1053 and 1062 are copies of marketing and informational materials from Janssen. Experts in finance and product valuation routinely rely on materials such as Exhibits 1053 and 1062 when analyzing a product's use and characteristics, such as in the opinions set forth in my initial declaration. Exhibit 1053 is a true and correct copy of the Zytiga web page as published on Janssen's Zytiga website, www.zytiga.com, accessed on August 8, 2016. Exhibit 1062 is a true and correct copy of the Zytiga web page as published on Janssen's Zytiga website, www.zytigahcp.com, accessed on August 8, 2016.

7. Exhibits 1057, 1060, 1061, and 1065-1074 are copies of investment reports from Wells Fargo, Cowen & Company, William Blair, Nasdaq,

Medivation, UBS Research, Wedbush Securities, Inc., and RBC Capital Markets. Experts in finance and product valuation routinely rely on investment reports such as Exhibits 1057, 1060, 1061, 1065-1074 when evaluating third-party views of a company's product, such as in the opinions set forth in my initial declaration.

8. Exhibits 1057, 1060, 1061, and 1066-1073 were submitted in IPR2016-00286 as Amerigen Exhibits 1061, 1043, 1062, 1052, 1059, 1060, 1056, 1042, 1044, 1058, and 1063, respectively, by Dr. McDuff. In IPR2016-00286, Dr. McDuff attested to the authenticity of Amerigen Exhibits 1061, 1043, 1062, 1052, 1059, 1060, 1056, 1042, 1044, 1058, and 1063 in ¶¶5-7, 11, 15, and 17-22 of Amerigen Exhibit 1068. I have also supplied a replacement of Exhibit 1070 in this proceeding as an attachment to this declaration. In IPR2016-00286, Dr. McDuff attested to the authenticity of Exhibit 1070 in ¶5 of Amerigen Exhibit 1068.

9. Exhibit 1065 is a true and correct copy of the Nasdaq web page as published on Nasdaq's website, www.nasdaq.com, accessed on August 8, 2016. Exhibit 1074 is a true and correct copy of the Bloomberg web page as published on Bloomberg's website, www.bloomberg.com, accessed on August 9, 2016.

10. Exhibit 1063 is a copy of the dosing and administration information for Jevanta® from the Jevanta® website. Exhibit 1063 is a true and correct copy of the Jevanta® web page as published on the Jevanta® website, www.jevanta.com. Experts routinely rely on web pages such as Exhibit 1063 in evaluating a product's

relevant market, such as in the opinions set forth in my initial declaration. Exhibit 1063 was submitted in IPR2016-00286 as Amerigen Exhibit 1049 by Dr. McDuff. In IPR2016-00286, Dr. McDuff attested to the authenticity of Amerigen Exhibit 1049 in ¶9 of Amerigen Exhibit 1068.

11. Exhibit 1076 is a definition of the term "hurdle rate" from the Investor Words website. Exhibit 1076 is a true and correct copy of the Investor Words web page as published on the Investor Words website, www.investorwords.com, accessed on August 8, 2016. Experts routinely rely on web pages such as Exhibit 1076 when defining terms commonly used in the field of economics.

12. Exhibit 1080 is a compilation of data provided by IMS Institute for Healthcare Informatics ("IMS"). Data from IMS are routinely relied on by experts in the field to determine drug sales, prescriptions, and promotional expenditures for a given product, such as in the opinions set forth in my initial declaration. Exhibit 1080 was submitted in IPR2016-00286 as Amerigen Exhibit 1067 by Dr. McDuff. In IPR2016-00286, Jayesh Bindra, Director of Business Development for Amerigen Pharmaceuticals, attested to the authenticity of Amerigen Exhibit 1067 in Amerigen Exhibit 1070.

13. Exhibit 1077 includes exhibits, B-1 and B-2. Regarding B-1, I compiled B-1 from the publicly available records of the prosecution history at the United States Patent and Trademark Office ("USPTO") for U.S. Patent No.

8,822,438 ("the '438 patent") as a summary of said prosecution history. The prosecution history of the '438 patent is publicly available including from the USPTO's Public Patent Application Information Retrieval ("Public PAIR") website, available at http://portal.uspto.gov/pair/PublicPair. With respect to B-2, I compiled the sales data listed in Exhibit 1080, described above at ¶12.

14. I understand from counsel that Janssen objected to Exhibit 1075 cited in my initial declaration for allegedly being incomplete. I have supplied a replacement of Exhibit 1075 in this proceeding as an attachment to this declaration.

15. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Respectfully submitted,

Robert D. Stoner, Ph.D.

Date: February 14, 2017

5071721

COWEN AND COMPANY

# Biotechnology Quarterly

## Industry Outlook

## Going On Vacation? Don't Forget To Pack Your Stocks

**July 2012**

Analysts
**Phil Nadeau, Ph.D.**
(646) 562-1336
phil.nadeau@cowen.com

**Eric Schmidt, Ph.D.**
(646) 562-1345
eric.schmidt@cowen.com

**Edward Nash**
(646) 562-1385
edward.nash@cowen.com

**Simos Simeonidis, Ph.D.**
(646) 562-1386
simos.simeonidis
@cowen.com

**Please see addendum of this report for important disclosures.**

**www.cowen.com**

This page left blank intentionally.

# Table Of Contents

## Industry Fundamentals And Core Research Universe — Page

This page left blank intentionally.

This page left blank intentionally.

# Dendreon

## Neutral (2)

## Provenge Seeks Redemption

Analysts
**Eric Schmidt, Ph.D.**
(646) 562-1345
eric.schmidt@cowen.com

**Imran Babar, Ph.D.**
(646) 562-1331
imran.babar@cowen.com

### Investment Thesis

Provenge, a personalized immunotherapy for prostate cancer, was approved by the FDA in April 2010. Provenge has demonstrated the ability to prolong survival by 4+ months with very good tolerability in men with minimally symptomatic metastatic castrate-resistant prostate cancer (CRPC). Provenge was launched into a capacity-constrained environment, and hopes were high for a major inflection in sales following the addition of new capacity in mid-2011. However, demand has not materialized as expected, and a number of potential factors may be to blame (reimbursement, physician skepticism, logistical barriers, patient identification). Our research suggests Provenge might eventually reach 25% of the 30-35K patients diagnosed with metastatic prostate cancer each year, supporting peak U.S. sales of $800-900MM. However, even at these sales levels, Provenge's profitability may be modest owing to high COGS. Dendreon filed for EMA approval of Provenge in January 2012. We model a similar sized opportunity for Provenge outside the U.S., but start-up costs associated with E.U. commercialization are expected to be substantial. Based on an NPV-based SOTP valuation for DNDN that ascribes significant success and terminal value to Provenge, we think DNDN shares are modestly undervalued.

| | |
|---|---|
| **DNDN (06/27)** | **$7.42** |
| **Mkt cap** | **$1.1B** |
| Dil shares out | 146.4MM |
| Avg daily vol | 2,580.7K |
| 52-wk range | $5.7-42.0 |
| Dividend | Nil |
| Dividend yield | Nil |
| BV/sh | $2.06 |
| Net cash/sh | $0.26 |
| Debt/cap | 30.0% |
| ROE (LTM) | NA |
| 5-yr fwd EPS growth (Norm) | NA |
| **S&P 500** | **1331.9** |

**Revenue $MM**

| FY Dec | 2011 Actual | 2012E Prior | 2012E Current | 2013E Prior | 2013E Current | 2014E Current | 2015E Current |
|---|---|---|---|---|---|---|---|
| Q1 | 27.0 | — | 82.1A | — | — | — | — |
| Q2 | 48.2 | — | 87.0 | — | — | — | — |
| Q3 | 64.3 | — | 95.0 | — | — | — | — |
| Q4 | 202.1 | — | 105.0 | — | — | — | — |
| Year | **341.6** | **—** | **369.0** | **—** | **560.0** | **775.0** | **975.0** |
| EV/S | — | — | 3.2x | — | 2.1x | 1.5x | 1.2x |

**EPS $**

| FY Dec | 2011 Actual | 2012E Prior | 2012E Current | 2013E Prior | 2013E Current | 2014E Current | 2015E Current |
|---|---|---|---|---|---|---|---|
| Q1 | (0.78) | — | (0.70)A | — | — | — | — |
| Q2 | (0.79) | — | (0.55) | — | — | — | — |
| Q3 | (1.00) | — | (0.48) | — | — | — | — |
| Q4 | 0.26 | — | (0.40) | — | — | — | — |
| Year | **(2.31)** | **—** | **(2.14)** | **—** | **(1.75)** | **(0.80)** | **0.00** |
| P/E | — | — | — | — | — | — | — |

# Provenge Falls (Way) Short Of Expectations

Provenge is a personalized immunotherapy for late-stage prostate cancer. Following a relatively tortuous development and regulatory path, the FDA approved Provenge in April 2010 for the treatment of minimally symptomatic, metastatic prostate cancer. Approval was based upon the Phase III IMPACT study, which demonstrated a 4-month improvement in median survival in patients treated with Provenge relative to placebo (p=0.032). Provenge was launched in the U.S. with a price tag of $93K for a full course of therapy.

The drug was initially made available to 50 of the clinical sites that were involved in Provenge's Phase III studies, with production constrained to 12 hoods at Dendreon's NJ manufacturing facility. However, Dendreon management had guided to 2011 sales of $350-400MM, with a major inflection occurring in H2 following the addition of new manufacturing capacity. Dendreon succeeded in gaining FDA licensure for the remaining 75% capacity at its NJ facility (36 of 48 hoods), as well as new facilities in LA and Atlanta (36 hoods each). However, demand did not materialize at the expected rate, causing the company to withdraw its 2011 revenue guidance. Full year net sales were around $214MM (gross product revenue of $228MM). Sales in 2012 do not appear to be trending much better. Management has guided to low single digit Q/Q growth in the near term, and suggested that sales growth is unlikely to improve until at least Q4.

Dendreon has blamed disappointing adoption on lingering reimbursement concerns, and specifically the "cost density" of unpaid claims at urology practices. In our view, the drug's poor commercial performance likely also reflects lower than expected demand. Dendreon has also referred to challenges in identifying suitable patients, and unique supply chain issues with a personalized therapy. In addition, there are lingering questions regarding Provenge's efficacy and cost. A vocal subgroup of physicians has always been skeptical of Provenge's mechanism, and the drug's clinical profile, including a lack of correlation between surrogate markers of disease (PSA, progression) and survival, and the lack of symptomatic benefit to the patient.

Moreover, according to specialists, the excitement over Provenge is waning in favor of newer drugs like JNJ's Zytiga and MDVN/Astellas's enzalutamide. Based upon numbers supplied by Dendreon, it is clear that the number of patients treated per center has been in steady decline over time, even in advance of the newer drugs being approved in the pre-chemotherapy setting. Our model assumes 30-35K new metastatic CRPC patients per year in the U.S., 85% of whom present with minimally symptomatic disease. We assume Provenge achieves 20% penetration into metastatic CRPC patients within 3-4 years of launch, and more gradual share gains beyond 2014. Our estimate of $750MM in 2016 U.S. sales assumes roughly 7-8K patients per year are treated with Provenge.

**Estimated U.S. Provenge Revenue Build-Up ($MM)**

| | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|
| Incidence of metastatic CRPC | 32.8 | 33.2 | 33.5 | 33.8 | 34.2 | 34.5 |
| *% eligible for Provenge (asymptomatic or minimally symptomatic)* | *85%* | *85%* | *85%* | *85%* | *85%* | *85%* |
| # eligible patients | 27.9 | 28.2 | 28.5 | 28.7 | 29.0 | 29.3 |
| *% penetration into metastatic CRPC* | *9%* | *15%* | *20%* | *23%* | *25%* | *26%* |
| # new patients receiving Provenge | 2,397 | 4,100 | 5,719 | 6,675 | 7,329 | 7,699 |
| | | | | | | |
| Provenge price per patient (000's) | $90 | $90 | $92 | $94 | $96 | $97 |
| **U.S. Provenge sales in CRPC ($MM)** | **$216** | **$369** | **$525** | **$625** | **$700** | **$750** |

Source: Cowen and Company

## Provenge's Profitability Also Falls Short

A second notable disappointment for Dendreon has been the poor margins associated with Provenge. GMs were just 27% in Q1:12, yet management continues to guide to peak gross margins in the 70-80% range. Given a track record of disappointing guidance, it is difficult to have confidence in the margin improvements that underlie this guidance. However, management has said that it plans to focus on automation as a means to decrease COGS. In particular, Dendreon plans to (1) transition from manual to electronic record keeping, implementation expected in 2012; (2) automate the testing of Provenge, implementation expected in 2013; and (3) automate the manufacturing of Provenge, implementation expected in 2014.

In September 2011, Dendreon announced a 500-person workforce reduction (mostly manufacturing, corporate overhead) aimed at allowing the company to achieve cash flow break even status in the U.S. at an approximate $500MM Provenge sales run rate. Yet even this expectation assumes GMs in the range of 50%, substantially higher than current levels. Recently investors have been anticipating a decision from Dendreon on whether or not it will shut down one of its manufacturing plants to further decrease COGS, but a decision has not yet been announced. Provenge's asset value is highly dependent on DNDN's ability to improve GMs toward a level more in sync with other pharmaceuticals.

Our sum-of-the-parts valuation credits Provenge for its long patent life, and the likelihood that generics might never materialize. It also takes into account the discounted value of the company's NOL tax credits, the company's balance sheet, and Dendreon's immunotherapy pipeline and platform. Our conclusions are summarized below. Assuming Provenge achieves peak WW sales in the $1.7B range and using discount rates of 10% (U.S.) and 13% (ex-U.S.), we believe shares are modestly undervalued.

**Sum-Of-The-Parts Value Per Share Summary**

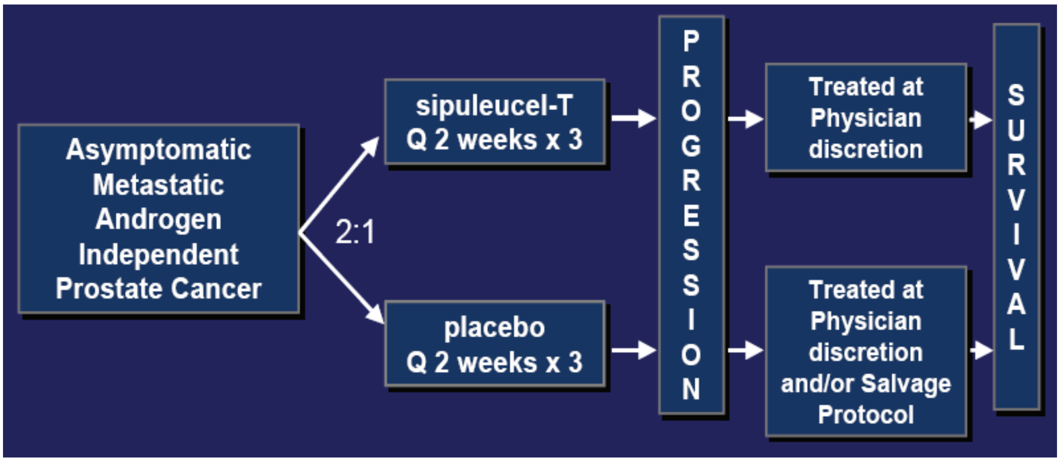
| | |
|---|---|
| U.S. Provenge NPV | $6.48 |
| Ex-U.S. Provenge NPV | $1.36 |
| NOL's NPV | $1.43 |
| Net Cash | $0.30 |
| Sum-Of-The-Parts Value | $9.56 |

Source: Cowen and Company

# A Review Of Provenge's Clinical Program

Dendreon originally filed a BLA for Provenge in 2006 based on data from two similarly designed, randomized, double-blind, placebo-controlled Phase III studies in men with asymptomatic metastatic castrate-resistant prostate cancer (CRPC). Following progression, patients in the placebo arms were permitted to cross over and receive a preserved version of Provenge (prepared from frozen apheresed PBMC's collected at the start of the study for potential crossover use). Patients in both arms of the studies were permitted to receive Taxotere chemotherapy after progression. Both studies had a primary endpoint of time to progression (defined by objective radiographic criteria, clinical progression and pain progression criteria).

**D9901 & D9902A Study Design**




Source: Dendreon Investor Presentation

***D9901***, which enrolled 127 patients (82 received Provenge while 45 received placebo), failed to meet its primary endpoint, demonstrating TTP of 11.0 weeks vs. 9.1 weeks for the Provenge and control arms, respectively (p=0.085). However a 3-year survival analysis performed as part of the follow-up, demonstrated a statistically significant improvement in median survival (25.9 vs. 21.4 months; HR = 0.58; p=0.01). Additional details from the FDA's briefing documents support the notion that Provenge is efficacious in this setting.

***D9902A*** was originally designed to be an identical companion study to D9901. However the negative TTP findings in D9901 led to this study being terminated early. By the time of termination, 98 of a planned 120 patients had been enrolled (65

Provenge, 33 placebo), and results demonstrated trends towards improved TTP (10.9 vs. 9.9 weeks; p=0.72) and overall survival (19.0 months vs. 15.7 months; p=0.331). When a pooled analysis of efficacy data from both studies was done, the overall survival benefit associated with Provenge was statistically significant.

**Pooled Survival Data From D9901 & D9902A**

Summary of PROVENGE Studies

| | Study 1 (D9901) N = 127 | Study 2 (D9902A) N = 98 | Studies 1 and 2 Integrated N = 225 |
|---|---|---|---|
| Median Survival in months: PROVENGE | 25.9 | 19.0 | 23.2 |
| Placebo | 21.4 | 15.7 | 18.9 |
| Median Survival Benefit: % (months) | 21% (4.5) | 21% (3.3) | 23% (4.3) |
| Hazard Ratio | 1.7 | 1.3 | 1.5 |
| p-value (log rank) | p=0.010 | p=0.331 | p=0.011 |
| Hazard Ratio | 2.1 | 1.9 | 1.8 |
| p-value (Cox regression, adj.) | p=0.002 | p=0.023 | p=0.0006 |
| 36-Month Survival: % (patients) PROVENGE | 34% (28) | 32% (21) | 33% (49) |
| Placebo | 11% (5) | 21% (7) | 15% (12) |

Source: Dendreon Investor Presentation

Relative to other cancer therapies, Provenge appeared to be very well tolerated. In the pooled safety data from the two studies, the most common AEs were Grade 1/2 chills, fatigue, fever, and back pain. SAEs were generally equally balanced between the two arms, with the exception of cerebrovascular events (8/147 vs. 0/78 in these studies; 3.9% vs. 2.6% when all other Provenge studies are included).

### Complete Response Letter Caught Investors By Surprise…

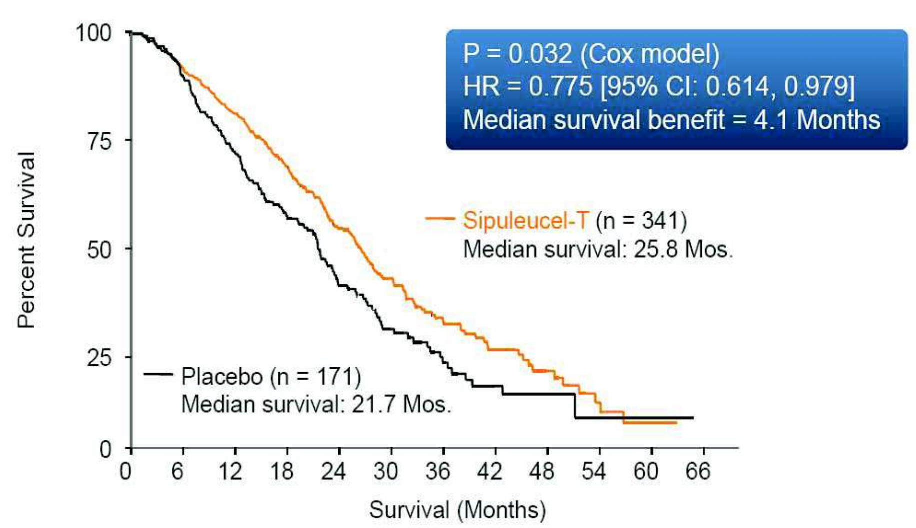
Based on data from D9901 and D9902A, Dendreon filed a BLA with the FDA, which was reviewed at a March 2007 FDA Cellular, Tissue and Gene Therapies advisory panel meeting. The Provenge briefing documents for the meeting concluded that "doubts remain about the persuasiveness of the efficacy data" due to the potential for type I error. Nonetheless, the FDA went on to acknowledge overall survival as the gold standard among cancer endpoints, and did not question the company's analysis of the data.

The advisory panel voted 17 to 0 in favor of the safety of Provenge and 13 to 4 in favor of the drug demonstrating substantial evidence of efficacy in this indication. However, in May 2007 Dendreon received a Complete Response letter requesting additional clinical data in support of the BLA's efficacy claim, as well as additional information regarding the CMC portion of the BLA. With respect to the efficacy claim, the FDA informed Dendreon that it would accept either a positive interim analysis or final analysis of survival from the then-ongoing Phase III IMPACT (D9902B) study.

### …As Did Survival Data From IMPACT

IMPACT (**IM**munotherapy for **P**rostate **A**deno**C**arcinoma **T**reatment study) was a randomized (2:1), double-blind, placebo-controlled Phase III that enrolled 512 men with metastatic CRPC. The study was very similar in design to the previous Phase III studies, with the exception of its primary endpoint (overall survival). Patients enrolled were also stratified for bisphosphonate use, Gleason score, and number of bone metastases. Results of an interim analysis, announced in October 2008, indicated Provenge was associated with a Hazard Ratio for survival of 0.80 (CI: 0.61-1.05), slightly above the threshold needed to hit statistical significance. In April 2009, Dendreon announced that IMPACT had met its primary endpoint at the final data analysis. Full results of the study were presented at the 2009 American Urological Association meeting and published in the *New England Journal of Medicine* in July 2010. Data demonstrated a 4.1 month benefit in median survival (25.8 months vs. 21.7 months; Hazard Ratio = 0.775) achieving a p-value of 0.032, below that pre-specified in the study's protocol (p<0.043, adjusted for a statistical penalty associated with the interim analysis). This was achieved despite 65% of patients in the placebo arm electing to cross over following progression. Consistent with the two previous Phase III studies, TTP was not statistically superior in the Provenge arm (HR=0.95; p=0.63). Safety findings were unremarkable, and consistent with the two earlier studies (most common AEs of chills, pyrexia, headache, usually lasting 1-2 days post-infusion).

**Phase III IMPACT Study: Analysis Of Overall Survival**




Source: Dendreon Investor Presentation

As with the D9901 study, sensitivity analyses demonstrated that the treatment effect was consistent across multiple patient subsets, including when adjusting for use and timing of docetaxel following Provenge. Based on these data, as well as additional CMC work, Dendreon submitted an amended BLA filing to the FDA in November 2009.

## Provenge Approved In 2010

Based on the data from IMPACT, Provenge was approved by the FDA for the treatment of patients with asymptomatic or minimally symptomatic metastatic castrate-resistant prostate cancer in April 2010. Provenge's label includes no contraindications or black-box warnings. Provenge is priced at $31K per infusion, or $93K for a full course of therapy, and was launched in May 2010. Its availability was initially limited to approximately 50 sites, all of which had prior experience with Provenge. Additional manufacturing capacity came online in March 2011 for the New Jersey site and in June and August for the LA and Atlanta sites, respectively. Dendreon has provided an indication of the potential revenue that each facility will be capable of providing when complete. New Jersey (48 workstations at full capacity) is capable of providing $500MM-1B in yearly revenues, while LA and Atlanta (36 workstations each) should be capable of generating $375-750MM each in yearly revenues. In light of Dendreon's struggle to improve GMs, the company is considering whether or not to shut down one of these plants and should come to a decision in H2:12.

## Mobilizing Patients Has Not Been Easy

Dendreon guided to a "step-wise" launch for Provenge. The therapy was initially available at the 50 clinical sites with prior Provenge trial experience. Each of these sites was allocated roughly 2 patient slots per month for a total monthly capacity of approximately 100 treated patients. However, even under this limited capacity scenario, it took several months before Provenge demand exceeded this monthly capacity. Initial headwinds related mostly to reimbursement (see below) and possibly a few logistical kinks. DNDN has also noted difficulty in identifying suitable patients and supply chain issues associated with a personalized therapy, which have limited uptake in the initial stages of the launch. Dendreon reported Provenge sales of $3MM in Q2:10, $20MM in Q3:10, $25MM in Q4:10, $28MM in Q1:11, $49MM in Q2:11, $61MM in Q3:11, $77MM in Q4:11, and approximately $82MM in Q1:12. We suspect the company will eventually achieve demand to support annual U.S. sales of $700-800MM, and we model 2016 U.S. sales of $750MMM. However, based on Dendreon's inability to meet early sales expectations, increasing competition, and a lack of visibility on how to mobilize appropriate patients, we lack conviction in Provenge's peak potential.

### Management Previously Pointed The Finger At Reimbursement…

On several occasions, Dendreon has blamed sluggish sales on uncertainties in the reimbursement process. Given Provenge's high costs, it makes sense that hospital centers or physician practices would demand strong visibility on reimbursement prior to making Provenge broadly available. However, in our view, Provenge's reimbursement outlook has improved substantially over the past year, without little commensurate increase in demand. Questions around reimbursement materialized in June 2010, when CMS surprised the investment community by announcing the initiation of a National Coverage Analysis (NCA) of Provenge for CRPC. Because Medicare coverage is limited to treatments that are deemed "reasonable and necessary", CMS has occasionally initiated an NCA to determine if it should implement a National Coverage Determination (NCD). CMS commissioned an external technology assessment and convened a meeting of the Medicare Evidence Development and Coverage Advisory Committee (MEDCAC) which took place on November 17, 2010. The MEDCAC panel voted (on a scale of 1-5) that there was evidence to support Provenge's benefits on overall survival (score of 3.6) when used on label, but that evidence was lacking to support use in off-label indications (scores of <1.5). On June 30, 2011 CMS issued a final NCD concluding that Provenge was reasonable and necessary as it improves health outcomes for Medicare beneficiaries with asymptomatic or minimally symptomatic metastatic CRPC. CMS further announced that effective July 1, 2011, Provenge will have a specific Q-code allowing more standardized claims reimbursement. Additionally, in November 2011, DNDN announced that CMS would cover infusion costs associated with Provenge treatment.

These advances should have reduced uncertainty concerning which patients are appropriate for therapy and speed up the claims process, respectively, and Dendreon notes that reimbursement concerns are now beginning to fade.

### …But Demand May Be More To Blame

As of Q2:11, Dendreon indicated that Provenge's launch was in "full swing" with utilization limited only by how fast physicians could prescribe the drug. Dendreon ended Q1:11 with 135 "active" Provenge accounts, and the company guided to 225 active accounts by the end of Q2, and roughly 500 active accounts by year end. Dendreon exceeded its guidance for opening new accounts during Q2 with more than 265 sites, but the average number of patients treated per center (0.8/month) was well below expectations (1-2/month). As a result, Dendreon missed its guidance for Q2 sales (reported sales of $49MM versus a target of $54-60MM) and withdrew its 2011 sales projection. At the end of Q4:11 and Q1:12, the number of sites infusing Provenge increased to 595 and 723, respectively, but with similarly low numbers of average patients per site. In our view, visibility into identifying patients who are suitable for Provenge is lacking.

There are several factors that could explain the lower than expected demand. First, patients with minimally symptomatic PRCA are not always closely followed by their physicians, and clinical practices have a difficult time recalling such patients in order to recommend a therapy like Provenge. Second, patients rapidly progress into and out of a metastatic, asymptomatic CRPC state of disease. Unless patients are caught while asymptomatic or minimally symptomatic, it may be too late to offer Provenge. Lastly, JNJ's Zytiga may be gaining traction in Provenge's market. Checks with consultants indicate increasing off-label prescribing in pre-chemotherapy patients, and although Zytiga's mechanism is viewed as complementary to Provenge, it is clear that some physicians are satisfied giving only Zytiga based upon its convenience (oral), rapid onset, and symptomatic benefits. Provenge may also face future competition from Medivation's enzalutamide, which is also being tested in metastatic CRPC. Our consultants expect earlier use of both Zytiga and enzalutamide to pressure Provenge, but overall expect sales growth to "stagnate" as opposed to decline.

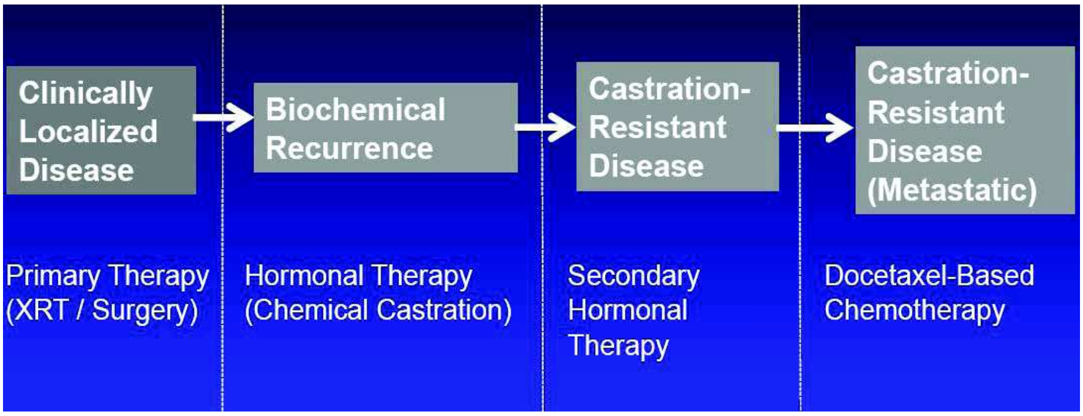
## What Is The Potential Opportunity For Provenge In The U.S.?

According to the American Cancer Society, approximately 217K new cases of prostate cancer were diagnosed in the U.S. in 2010, with an estimated 32K deaths. Patients' disease is usually controlled for many years on anti-androgen therapies, eventually becoming refractory, or "castrate-resistant". We estimate that roughly 30-35K patients in the U.S. develop metastatic CRPC each year.

According to consultants, the large majority (80-90%) of patients with metastatic CRPC initially have few or no symptoms. Most CRPC patients are usually treated initially with a second-line hormonal agent (e.g. Casodex, ketoconazole, estrogens, steroids), and chemotherapy with Taxotere is usually delayed until patients develop symptomatic metastatic disease. Our consultants estimate that about 16K patients with CRPC are treated annually with Taxotere in the U.S., representing about half of all U.S. metastatic CRPC patients.

In general, physicians expect to administer Provenge prior to chemotherapy, based on their view that Provenge takes time to manifest its effect, and because many patients refuse chemotherapy. We note that Dendreon's marketing campaign is primarily directed toward urologists who treat the bulk of earlier-stage CRPC patients. Under ASP+6% economics, Provenge should be profitable to administer.

**Prostate Cancer Treatment Paradigm**

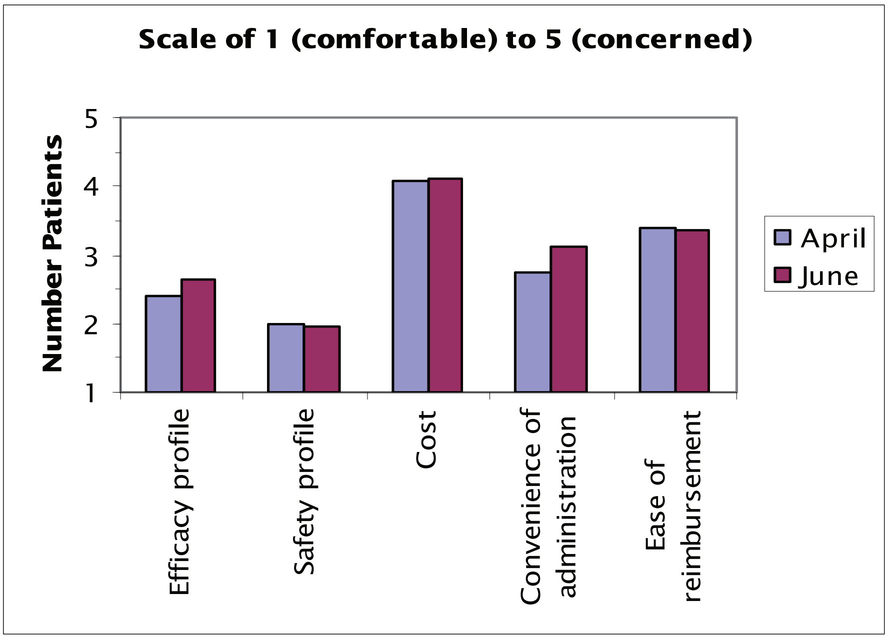



Source: Dendreon 2009 Investor Presentation

## Physicians Recognize Provenge's Benefits…

To better understand physicians' attitudes toward factors that might impact uptake of Provenge, we queried 32 physicians on their level of comfort or concern (with scores ranging from 1 through 5, respectively) with each of five factors relating to Provenge: efficacy, safety, cost, convenience of administration, and ease of obtaining reimbursement. As might be expected, Provenge's cost arose as the top concern, followed by the ease of obtaining reimbursement. Meanwhile, physicians appear comfortable with the various aspects of Provenge's clinical profile.

**Respondent Attitudes Toward Various Aspects of Provenge**

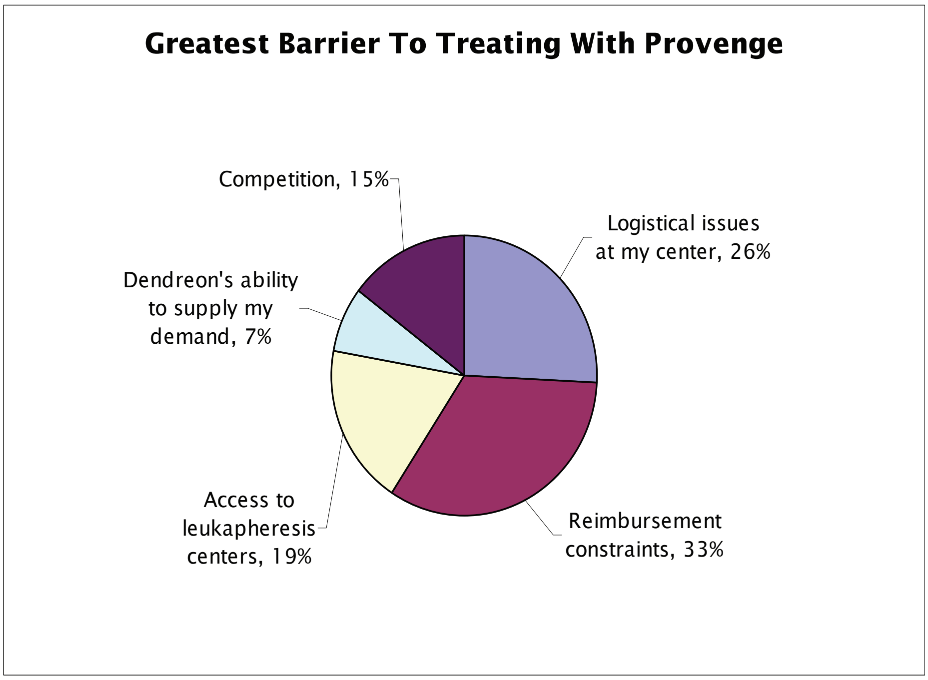



Source: Cowen and Company Provenge Tracking Survey – July 2011

### …But Forecast Somewhat Modest Penetration

Despite strong appreciation for Provenge's efficacy and safety profile, surveyed physicians were surprisingly conservative in their estimate of the drug's peak penetration. In mid 2011, surveyed physicians projected Provenge to penetrate just 24% of the overall mCRPC market in 3 years. Reasons to explain why 75% of patients might not get a life-prolonging therapy are still unclear. However, when asked what in their experience is the greatest near-term barrier to adoption of Provenge, a variety of issues were cited, including reimbursement, logistical issues, and leukapheresis access.

**Respondent Attitudes Toward Various Aspects of Provenge Treatment**




Source: Cowen and Company Provenge Tracking Survey – July 2011

## Not Your Standard Pharmaceutical Model

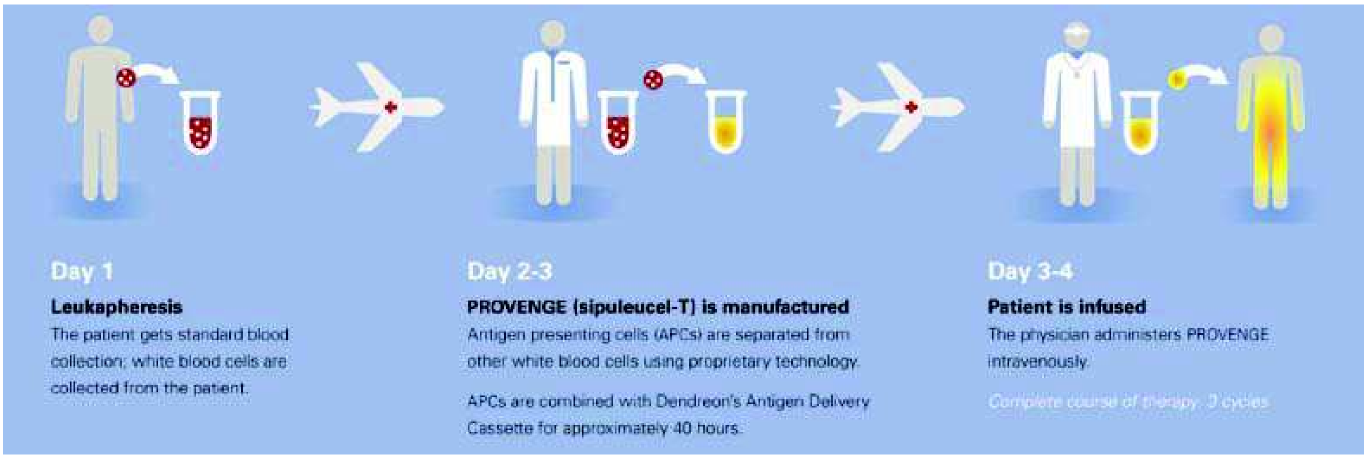
Because Provenge is a personalized, cell-based therapy, its commercialization is atypical for a pharmaceutical or biotech product. Provenge cannot simply be manufactured, inventoried, and shipped as orders come in. Rather Dendreon will need to successfully navigate several logistical issues related to the supply chain in order to maximize Provenge's sales potential. Thus far, selling and manufacturing costs associated with Provenge have exceeded nearly all expectations.

### A Look At The Steps Involved In The Provenge Process

From a patient's perspective, the process of receiving Provenge treatment is relatively simple. After a patient is prescribed Provenge, his/her physician contacts the Provenge call center, which coordinates the process providing the patient with six appointments: three appointments (two or more weeks apart) for blood sample collection at a local apheresis center, and three follow-up appointments for Provenge infusion at the physician's office or infusion center. Each infusion takes place a few days after each blood sample is collected, but the preparation and administration of each infusion involves a sequence of steps that must be precisely coordinated. From a patient's perspective, there will be some scheduling burden, as well as a delay in the time to initiation of therapy.

Overall, none of the individual steps in the process over delivering Provenge is particularly challenging. However, DNDN's operating system has yet to be challenged in a high throughput environment, and costs associated with delivering Provenge have thus far exceeded all expectations with little sign of improving.

**Steps Involved In Provenge's Preparation And Administration**




Source: Dendreon 2009 Analyst Day Presentation

### *Sample Preparation And Transportation*

The patient's blood sample must first undergo leukapheresis (a procedure in which white blood cells are separated from the blood sample). This takes place at an apheresis center. Each of these centers must be validated and contracted by Dendreon. Dendreon estimates there are approximately 600 apheresis centers in the U.S. Approximately 75 of these centers were used in the Phase III IMPACT study, and management estimates roughly 200 sites will be utilized at peak. Dendreon has contracted directly with many leading apheresis service providers (for example, The American Red Cross, the New York blood center), and will motivate centers by reimbursing at higher rates than would be typical for a leukapheresis procedure (typically in the $700-800 range). Management indicated that multiple centers were validated before Provenge's approval, providing sufficient capacity at apheresis centers to meet potential demand.

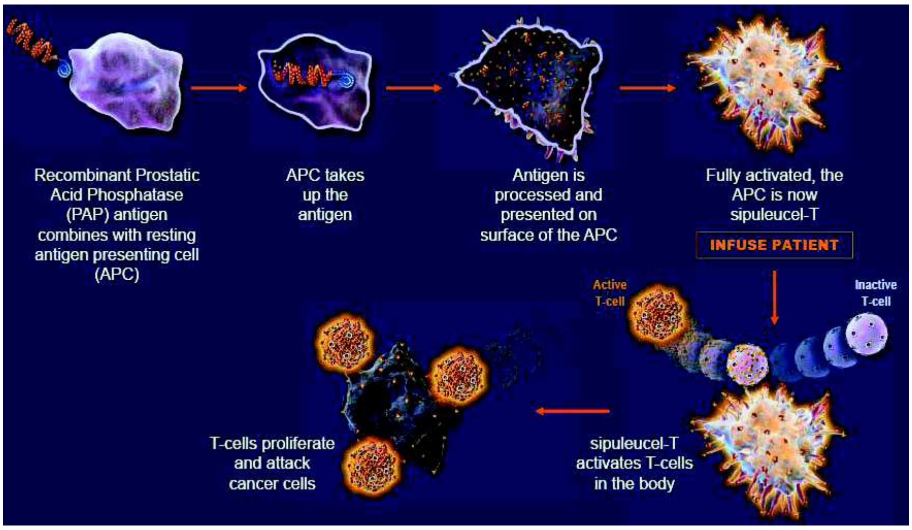
The white cell sample is then shipped by courier to one of Dendreon's three manufacturing facilities (New Jersey, LA, and eventually Atlanta). Dendreon is working with established couriers that specialize in the time sensitive delivery of materials including medical products. Each individual sample is labeled with a bar code that should ensure patient samples are not mixed, and allow each sample to be tracked by a GPS-like computer system.

### *Manufacture Of Provenge*

This is the only step in the process that is managed entirely by Dendreon. The process of making Provenge is depicted on the next page. Specific cells known as

Antigen Presenting Cells (or APCs) are first separated out from a patient's white blood sample. The cells are then incubated with an Antigen-Delivery-Cassette (GM-CSF combined with Prostatic Acid Phosphatase, an antigen that is widely expressed on the surface of prostate cancer cells) for a period of up to two days. This set of "activated" APCs then undergo quality assurance testing including evaluation of surrogate markers for immune stimulation (such as CD54). The sample is then formulated with a buffer to create the final Provenge infusion, which is shipped back to the physician and administered to the patient as a simple one-hour infusion in the physician's office (no pre-medication is needed). The entire cycle is repeated three times, at two-weekly intervals, following which Dendreon collects payment from the physician.

**A Look At The Science Behind Provenge**




Source: Dendreon 2009 Analyst Day Presentation

## Provenge Should Have Longevity

Because Provenge is a one-of-a-kind biologic, comprised of patient-specific, activated APC cells, its revenue stream would appear to be a virtual perpetuity for Dendreon. The company has an array of issued patents covering the Antigen Delivery Cassette used to stimulate APC cells, method of producing antigen, and various manufacturing processes. These patents expire between 2015 and 2022, and we think it likely that multiple additional patents covering Provenge will issue in the intervening years. Additionally, the complexity of the product and the regulatory hurdles created by such complexity are likely to keep generic or biosimilar competition at bay for the foreseeable future. In our view, Provenge represents more of a process than a product. In this regard it is similar to vaccines, which themselves are associated with tremendous franchise longevity. However, relative to vaccines, Provenge is even more complex as its end product is personalized and comprised of multiple living cell types. Moreover, the process of producing Provenge is littered with certain trade secrets (incubations times/temperatures, centrifugation protocols, buffers, etc.) that will make it difficult to replicate. Lastly, the product generated via the Provenge process is difficult to describe in terms of properties that are associated with activity in part because Provenge's mechanism of action is not well understood. Given these attributes, it is difficult to conceive how the FDA would consider approving a biosimilar copy of the drug.

Given the absence of meaningful biosimilar or generic competition, superior branded competition may represent about the only threat to Provenge's longevity. In order to displace Provenge, a new immunotherapy might need to demonstrate better efficacy, safety, or convenience. The only head-on threat to Provenge on our radar screen, comes from Bavarian Nordic's PROSTVAC. PROSTVAC is an off the shelf vaccine comprised of seven monthly injections of two different poxyviruses that overexpress PSA and three immune enhancing co-stimulatory molecules. A 125-patient Phase II trial conducted in a prostate cancer population similar to that of Provenge trials demonstrated that PROSTVAC improved survival relative to placebo control (HR=0.559, p=0.006). Bavarian Nordic has a received an SPA from the FDA for its Phase III study, which is evaluating Prostvac +/- adjuvant GM-CSF vs. placebo in 1,200 metastatic CRPC patients (primary endpoint OS). Enrollment was initiated in November 2011, and we think data might be available in 2014/2015.

We believe our NPV analysis of the profits associated with Provenge sales in the U.S. appropriately credits the company for Provenge's longevity, and reflects a better case outcome for COGS (GMs improving to 60%). Under this scenario, Provenge's U.S. opportunity might be worth roughly $6.50/share to Dendreon.

| | |
|---|---|
| Financial Year End | 12/31/2010 |
| Valuation Date | 6/19/2012 |
| **Discount Rate** | **10.0%** |
| **Perpetual Growth Rate** | **0.0%** |

**Tuesday, June 19, 2012**

| $MM | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **U.S. Provenge Sales** | **216** | **369** | **525** | **625** | **700** | **750** | **780** | **810** | **830** | **855** | **872** | **881** |
| *Y/Y Growth* | *349%* | *71%* | *42%* | *19%* | *12%* | *7%* | *4%* | *4%* | *2%* | *3%* | *2%* | *1%* |
| **U.S. Provenge COGS** | 159 | 243 | 318 | 313 | 314 | 323 | 320 | 324 | 332 | 342 | 349 | 352 |
| *As a % of U.S. Provenge Sales* | *74%* | *66%* | *61%* | *50%* | *45%* | *43%* | *41%* | *40%* | *40%* | *40%* | *40%* | *40%* |
| **U.S. Provenge Sales & Marketing Expense** | 275 | 235 | 250 | 255 | 260 | 265 | 270 | 275 | 280 | 280 | 280 | 280 |
| *As a % of U.S. Provenge Sales* | *127%* | *64%* | *48%* | *41%* | *37%* | *35%* | *35%* | *34%* | *34%* | *33%* | *32%* | *32%* |
| **U.S. Provenge Development Expense** | 50 | 40 | 40 | 40 | 40 | 40 | 40 | 30 | 30 | 30 | 30 | 30 |
| *As a % of U.S. Provenge Sales* | *23%* | *11%* | *8%* | *6%* | *6%* | *5%* | *5%* | *4%* | *4%* | *4%* | *3%* | *3%* |
| **Total Provenge Expenses** | **484** | **518** | **608** | **608** | **614** | **628** | **630** | **629** | **642** | **652** | **659** | **662** |
| **DNDN Operating Income From U.S. Provenge** | **(268)** | **(149)** | **(83)** | **17** | **86** | **123** | **150** | **181** | **188** | **203** | **213** | **218** |
| *Provenge Operating Margin* | *-124%* | *-40%* | *-16%* | *3%* | *12%* | *16%* | *19%* | *22%* | *23%* | *24%* | *24%* | *25%* |
| **Tax-Adjusted EBIT** | **(268)** | **(149)** | **(83)** | **17** | **86** | **123** | **150** | **181** | **188** | **193** | **139** | **142** |
| *Tax rate* | *0%* | *0%* | *0%* | *0%* | *0%* | *0%* | *0%* | *0%* | *0%* | *5%* | *35%* | *35%* |
| Less: Capital Expenditures (related to initial manufacturing facility build-out) | **(150)** | | | | | | | | | | | |
| **Provenge-Related Free Cash Flow** | **(418)** | **(149)** | **(83)** | **17** | **86** | **123** | **150** | **181** | **188** | **193** | **139** | **142** |
| Years | -0.97 | 0.03 | 1.03 | 2.03 | 3.03 | 4.03 | 5.03 | 6.03 | 7.03 | 8.03 | 9.03 | 10.03 |
| Discount Factor | 1.10 | 1.00 | 0.91 | 0.82 | 0.75 | 0.68 | 0.62 | 0.56 | 0.51 | 0.47 | 0.42 | 0.38 |
| **NPV of Provenge Cash flows** | **(459)** | **(148)** | **(75)** | **14** | **64** | **83** | **93** | **102** | **96** | **90** | **59** | **55** |

**Terminal Value Calculation**

| | |
|---|---|
| Final year FCF | 142 |
| Perpetual Growth Rate | 0% |
| **Terminal Value** | **1420** |
| Discount Factor | 0.38 |
| **Present Value of Terminal Value** | **546** |
| **Present Value of Cash Flows** | **433** |
| **Provenge NPV** | **979** |
| Fully Diluted Shares Outstanding | 151.1 |
| **Present Value of Cash Flows Per Share** | **$6.48** |

**Sensitivity Analysis**

| **Term Gr.** | **Discount Rate** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **7%** | **8%** | **9%** | **10%** | **11%** | **12%** | **13%** |
| **5%** | $29 | $19 | $14 | $10 | $8 | $7 | $6 |
| **2.5%** | $15 | $11 | $9 | $8 | $7 | $6 | $5 |
| **0%** | $10 | $9 | $8 | $6 | $6 | $5 | $4 |
| **-2.5%** | $9 | $7 | $6 | $6 | $5 | $4 | $4 |
| **-5%** | $7 | $7 | $6 | $5 | $5 | $4 | $4 |



WCK1130
Wockhardt Bio AG v. Janssen Oncology, Inc.
IPR2016-01582

| | |
|---|---|
| Financial Year End | 12/31/2010 |
| Valuation Date | 6/19/2012 |
| **Discount Rate** | **13.0%** |
| **Perpetual Growth Rate** | **0.0%** |

## Ex-U.S. Provenge NPV Analysis

**Tuesday, June 19, 2012**

| $MM | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ex-U.S. Provenge Sales** | | | | | **35** | **150** | **275** | **400** | **550** | **650** | **720** | **750** | **780** | **811** |
| *Y/Y Growth* | | | | | | | *83%* | *45%* | *38%* | *18%* | *11%* | *4%* | *4%* | *4%* |
| **Ex-U.S. Provenge COGS** | | | | | 21 | 75 | 123 | 162 | 226 | 260 | 288 | 300 | 312 | 324 |
| *As a % of ex-U.S. Provenge Sales* | | | | | *61%* | *50%* | *45%* | *41%* | *41%* | *40%* | *40%* | *40%* | *40%* | *40%* |
| **Ex-U.S. Provenge Sales & Marketing Expense** | | | | 20 | 90 | 150 | 180 | 200 | 220 | 240 | 250 | 260 | 270 | 270 |
| *As a % of ex-U.S. Provenge Sales* | | | | | *257%* | *100%* | *65%* | *50%* | *40%* | *37%* | *35%* | *35%* | *35%* | *33%* |
| **Ex-U.S. Provenge Development Expense** | | | 20 | 20 | 30 | 30 | 30 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| *As a % of ex-U.S. Provenge Sales* | | | | | *86%* | *20%* | *11%* | *5%* | *4%* | *3%* | *3%* | *3%* | *3%* | *2%* |
| **Total Provenge Expenses** | | | **20** | **40** | **141** | **255** | **333** | **382** | **466** | **520** | **558** | **580** | **602** | **614** |
| **DNDN Operating Income From Ex-U.S. Provenge** | | | **(20)** | **(40)** | **(106)** | **(105)** | **(58)** | **18** | **85** | **130** | **162** | **170** | **178** | **197** |
| *Provenge Operating Margin* | | | | | *NA* | *NA* | *NA* | *5%* | *15%* | *20%* | *23%* | *23%* | *23%* | *24%* |
| **Tax-Adjusted EBIT** | | | **(20)** | **(40)** | **(106)** | **(105)** | **(58)** | **18** | **85** | **130** | **162** | **162** | **160** | **177** |
| *Tax rate* | | | | | | *0%* | *0%* | *0%* | *0%* | *0%* | *0%* | *5%* | *10%* | *10%* |
| Less: Capital Expenditures (related to initial manufacturing facility build-out) | | | **(30)** | **(30)** | **(150)** | | | **(150)** | | | | | | |
| **Ex-U.S. Provenge-Related Free Cash Flow** | | | **(50)** | **(70)** | **(256)** | **(105)** | **(58)** | **(132)** | **85** | **130** | **162** | **162** | **160** | **177** |
| Years | | | -0.97 | 0.03 | 1.03 | 2.03 | 3.03 | 4.03 | 5.03 | 6.03 | 7.03 | 8.03 | 9.03 | 10.03 |
| Discount Factor | | | 1.13 | 1.00 | 0.88 | 0.78 | 0.69 | 0.61 | 0.54 | 0.48 | 0.42 | 0.37 | 0.33 | 0.29 |
| **NPV of Provenge Cash flows** | | | **(56)** | **(70)** | **(226)** | **(82)** | **(40)** | **(81)** | **46** | **62** | **69** | **61** | **53** | **52** |

**Terminal Value Calculation**

| | |
|---|---|
| Final year FCF | 160 |
| Perpetual Growth Rate | 0% |
| **Terminal Value** | **1232** |
| Discount Factor | 0.29 |
| **Present Value of Terminal Value** | **362** |
| **Present Value of Cash Flows** | **(156)** |
| **Provenge NPV** | **205** |
| Fully Diluted Shares Outstanding | 151.1 |
| **Present Value of Cash Flows Per Share** | **$1.36** |

**Sensitivity Analysis**

| **Term Gr.** | | | | **Discount Rate** | | | |
|---|---|---|---|---|---|---|---|
| | **10%** | **11%** | **12%** | **13%** | **14%** | **15%** | **16%** |
| **5%** | $8 | $6 | $4 | $3 | $2 | $2 | $1 |
| **2.5%** | $5 | $4 | $3 | $2 | $1 | $1 | $1 |
| **0%** | $3 | $3 | $2 | $1 | $1 | $1 | $0 |
| **-2.5%** | $2 | $2 | $1 | $1 | $1 | $0 | $0 |
| **-5%** | $2 | $1 | $1 | $1 | $0 | $0 | $0 |



WCK1130
Wockhardt Bio AG v. Janssen Oncology, Inc.
IPR2016-01582

## Provenge's Ex-U.S. Strategy Appears Questionable

Based upon an assessment of Provenge's ex-U.S. value, Dendreon management had expected to maintain full rights to Provenge in Europe. The company obtained the go-ahead to file for European approval of Provenge with its existing data and filed in early 2012, suggesting approval is possible in 2013.

However, the slower than expected uptake of Provenge in the U.S. has forced DNDN to reconsider its strategy. The company no longer believes it has the financial capability to make the required investment in E.U. manufacturing and commercialization on its own. Rather management is seeking a contract manufacturer to make the necessary $100-200MM upfront investment in manufacturing. We are skeptical that Dendreon will identify a CRO willing to build an E.U. Provenge facility. In our view, if building such a facility is not a worthy investment for Dendreon, it is unlikely that a less knowledgeable, less incentivized party will put its money at risk to build a facility that holds little value other than producing Provenge. Thus far, Provenge's experience in the U.S. does not support this being a good investment, even in the world's most profitable oncology market.

An analysis of Provenge's value outside the U.S. similar to that conducted for the U.S. suggests E.U. rights might be worth just $1/share given the more fragmented marketplace and need for major upfront investment (see previous page).

### Victrelis Royalty Interest Sold

On its Q3:11 earnings call Dendreon disclosed that it would receive an approximate 5% WW royalty on sales of MRK's Victrelis (boceprevir). Dendreon had acquired the intellectual property related to Victrelis in July 2003 via the $83MM acquisition of Corvas. In Q4:11, Dendreon sold this royalty interest to CPPIB Credit Investments Inc. for $125MM in cash.

## Dendreon Under New Leadership

In February 2011, long-time CEO Dr. Mitch Gold resigned his position in order to pursue other interests. Dendreon Board Member and former Savient and ImClone CEO John Johnson took over leadership of the company. Mr. Johnson, a seasoned pharmaceutical executive, has promised greater focus on improving margins and better communication with the investment community.

**Upcoming Milestones**

| Milestone/Event | Timing |
|---|---|
| EMA issues Day 120 "List of Questions" for Provenge EU filing | Q3:12 |
| Identify potential EU manufacturing partner for Provenge | H2:12 |
| Decision on whether to shut down Atlanta manufacturing plant | H2:12 |
| Updates from ongoing Provenge trials (PROACT, PROTECT, NEOACT) | 2012-2013 |
| Additional Provenge commercial updates | 2012-2013 |
| EU approval of Provenge | 2013 |

Source: Company data, Cowen and Company

## DNDN Quarterly P&L Model ($MM)

| | Q1:11A | Q2:11A | Q3:11A | Q4:11A | 2011A | Q1:12A | Q2:12E | Q3:12E | Q4:12E | 2012E |
|---|---|---|---|---|---|---|---|---|---|---|
| Provenge U.S. Sales | 27.0 | 48.2 | 61.4 | 77.0 | 213.5 | 82.0 | 87.0 | 95.0 | 105.0 | 369.0 |
| Provenge ex-U.S. sales | | | | | | | | | | |
| Other Revenue (Victrelis royalty) | | | 2.9 | 125.2 | 128.1 | 0.1 | | | | |
| **Total Revenue** | **27.0** | **48.2** | **64.3** | **202.1** | **341.6** | **82.1** | **87.0** | **95.0** | **105.0** | **369.0** |
| *Y/Y growth* | | | | | | *204%* | *81%* | *48%* | *-48%* | *8%* |
| COGS | 18.3 | 28.8 | 55.0 | 57.0 | 159.1 | 60.0 | 60.9 | 60.8 | 60.9 | 242.6 |
| *Gross margin* | *32%* | *40%* | *10%* | *26%* | *25%* | *26.8%* | *30%* | *36%* | *42%* | *34%* |
| R&D | 17.6 | 18.6 | 20.4 | 17.7 | 74.3 | 17.3 | 25.0 | 24.0 | 24.0 | 90.3 |
| SG&A | 95.3 | 105.1 | 84.9 | 76.1 | 361.3 | 95.3 | 72.0 | 72.0 | 70.0 | 309.3 |
| Other | | | 38.5 | 0.1 | 38.6 | (0.1) | | | | |
| **Total Expenses** | **131.2** | **152.4** | **198.8** | **150.9** | **633.3** | **172.6** | **157.9** | **156.8** | **154.9** | **642.3** |
| *Operating margin* | | | | | *NM* | | | | | *NM* |
| **Operating Income/Loss** | **(104.2)** | **(104.2)** | **(134.5)** | **51.3** | **(291.7)** | **(90.5)** | **(70.9)** | **(61.8)** | **(49.9)** | **(273.3)** |
| Interest Income | 0.4 | 0.4 | 0.3 | 0.3 | 1.4 | 0.4 | 0.3 | 0.3 | 0.3 | 1.1 |
| Interest Expense | (9.0) | (12.1) | (12.9) | (13.7) | (47.7) | (13.8) | (11.0) | (10.0) | (10.0) | (44.8) |
| Other Income (Expense), net | | | 0.0 | 0.2 | 0.2 | 0.0 | | | | 0.0 |
| **Pre-tax Income/Loss** | **(112.8)** | **(116.0)** | **(147.1)** | **38.1** | **(337.8)** | **(103.9)** | **(81.7)** | **(71.6)** | **(59.7)** | **(317.0)** |
| *Tax rate (%)* | | | | | | | | | | |
| Provision for income taxes | | | | | | | | | | |
| **GAAP Net Income (Loss)** | **(112.8)** | **(116.0)** | **(147.1)** | **38.1** | **(337.81)** | **(103.9)** | **(81.7)** | **(71.6)** | **(59.7)** | **(317.0)** |
| Stock based compensation | 14.7 | 16.8 | 16.7 | 12.1 | 60.3 | 19.5 | 15.0 | 15.0 | 15.0 | 64.5 |
| Other non-GAAP expenses | 11.4 | 14.3 | 49.1 | 16.5 | 91.5 | 33.5 | 13.0 | 13.0 | 13.0 | 72.5 |
| **Non-GAAP Net Income (Loss)** | **(86.7)** | **(84.9)** | **(81.3)** | **66.7** | **(186.0)** | **(51.0)** | **(53.7)** | **(43.6)** | **(31.7)** | **(180.0)** |
| **GAAP EPS** | **($0.78)** | **($0.79)** | **($1.00)** | **$0.26** | **($2.31)** | **($0.70)** | **($0.55)** | **($0.48)** | **($0.40)** | **($2.14)** |
| **Non-GAAP EPS** | **($0.60)** | **($0.58)** | **($0.56)** | **$0.45** | **($1.27)** | **($0.35)** | **($0.36)** | **($0.29)** | **($0.21)** | **($1.21)** |
| Basic Shares | 145.5 | 145.9 | 146.4 | 146.8 | 146.2 | 147.6 | 148.0 | 148.5 | 149.0 | 148.3 |
| Diluted Shares | 145.5 | 145.9 | 146.4 | 150.2 | 146.2 | 151.1 | 151.5 | 152.0 | 152.5 | 151.8 |

*non-GAAP EPS excludes stock-based compensation, D&A, imputed interest expense and 1x items.*

Source: Company data, Cowen and Company estimates

## DNDN Annual P&L Model ($MM)

| | 2011A | 2012E | 2013E | 2014E | 2015E | 2016E |
|---|---|---|---|---|---|---|
| Provenge U.S. Sales | 213.5 | 369.0 | 525.0 | 625.0 | 700.0 | 750.0 |
| Provenge ex-U.S. Sales | 0.0 | 0.0 | 35.0 | 150.0 | 275.0 | 400.0 |
| Other Revenue | 128.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| **Total Revenue** | **341.6** | **369.0** | **560.0** | **775.0** | **975.0** | **1,150.0** |
| *Y/Y growth* | *611%* | *8%* | *52%* | *38%* | *26%* | *18%* |
| COGS | 159.1 | 242.6 | 317.6 | 387.5 | 437.8 | 465.8 |
| *Gross margin* | *25%* | *34%* | *40%* | *50%* | *55%* | *60%* |
| R&D | 74.3 | 90.3 | 93.0 | 97.0 | 100.0 | 108.0 |
| SG&A | 361.3 | 309.3 | 375.0 | 375.0 | 400.0 | 450.0 |
| **Total Expenses** | **633.3** | **642.3** | **785.6** | **859.5** | **937.8** | **1,023.8** |
| *Operating margin* | *NM* | *NM* | *NM* | *NM* | *4%* | *11%* |
| **Operating Income/Loss** | **(291.7)** | **(273.3)** | **(225.6)** | **(84.5)** | **37.2** | **126.3** |
| Interest Income | 1.4 | 1.1 | 1.0 | 2.0 | 3.0 | 3.0 |
| Interest Expense | (47.7) | (44.8) | (40.0) | (40.0) | (40.0) | (2.0) |
| Other Income (Expense), net | 0.2 | 0.0 | | | | |
| **Pre-tax Income/Loss** | **(337.8)** | **(317.0)** | **(264.6)** | **(122.5)** | **0.2** | **127.3** |
| *Tax rate (%)* | *0%* | *0%* | *0%* | *0%* | *0%* | *0%* |
| Provision for income taxes | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| **GAAP Net Income (Loss)** | **(337.8)** | **(317.0)** | **(264.6)** | **(122.5)** | **0.2** | **127.3** |
| Stock based compensation | 60.3 | 64.5 | 62.0 | 63.0 | 65.0 | 65.0 |
| Other non-GAAP expenses | 91.5 | 72.5 | 57.0 | 59.0 | 60.0 | 62.0 |
| **Non-GAAP Net Income (Loss)** | **(186.0)** | **(180.0)** | **(145.6)** | **(0.5)** | **125.2** | **254.3** |
| **GAAP EPS** | **($2.31)** | **($2.14)** | **($1.75)** | **($0.80)** | **$0.00** | **$0.80** |
| **Non-GAAP EPS** | **($1.27)** | **($1.21)** | **($0.96)** | **($0.00)** | **$0.80** | **$1.60** |
| Diluted Shares | 146.2 | 151.8 | 154.5 | 156.5 | 157.5 | 158.5 |

*non-GAAP EPS excludes stock-based compensation, D&A, imputed interest expense and 1x items.*

Source: Company data, Cowen and Company estimate

## Dendreon R&D Pipeline

| Therapeutic Class/Product | Indication | P-C | I | II | III | FILING | MKT | Comments |
|---|---|---|---|---|---|---|---|---|
| **Oncology** | | | | | | | | |
| Provenge | Metastatic CRPC | | | | | | • | Approved April 29, 2010 |
| Provenge | Hormone-sensitive PRCA | | | | • | | | PROACT (P-11) Study |
| Provenge | Neoadjuvant PRCA | | | • | | | | Ph II NEOACT (P07) initiated Aug '08 |
| Provenge | Metastatic CRPC | | | • | | | | Evaluating Provenge with concurrent vs. sequential Abi/Prednisone |
| Neuvenge | Breast Cancer | | • | | | | | Phase I data published JCO Aug 2007 |
| Neuvenge | Bladder cancer | • | | | | | | Phase II to start in 2012 |
| CA9 program | Kidney/colon/cervical ca | • | | | | | | Ph I/II in mRCC could begin in 2012 |
| DN24-02 | Urothelial Carcinoma | | | • | | | | Neu-ACT study evaluating DN24-02 as an adjuvant in HER2+ UC patients |
| CEA program | Breast/colon/lung cancers | • | | | | | | Clinical studies could begin in 2012 |
| TRMP8 small molecule program | Solid tumors | | • | | | | | Lead D3263 in Phase I |
| **Total Drugs In Development** | | **3** | **2** | **3** | **1** | **1** | **1** | |

**Seattle, WA** **Investor Relations Contact: Nicole Soley (206) 455-12384**

Source: Cowen and Company

# Medivation

## Neutral (2)

## Enzalutamide Hits Its Target

Analysts
**Eric Schmidt, Ph.D.**
(646) 562-1345
eric.schmidt@cowen.com

**Imran Babar, Ph.D.**
(646) 562-1331
imran.babar@cowen.com

### Investment Thesis

Medivation is developing a single product candidate, enzalutamide (formerly, MDV3100), an androgen receptor antagonist that has successfully completed Phase III development in chemotherapy-refractory castrate-resistant prostate cancer (CRPC). Impressive data from the AFFIRM trial were presented at ASCO GU in February 2011. Enzalutamide improved median OS by 4.8 months versus placebo and reduced the risk of death by 37% (HR=0.631). Safety data indicated that enzalutamide was well tolerated and that the risk of seizures was modest. U.S. and E.U. regulatory filings for this indication were submitted in May and June 2012, respectively. A second pivotal trial (PREVAIL) examining enzalutamide's effects in the pre-chemotherapy metastatic CPRC population is ongoing. We believe the likelihood of success in this related indication is high and expect data in 2013. We estimate enzalutamide could achieve peak worldwide sales of $3B in the post-chemotherapy and pre-chemotherapy markets. Medivation and partner Astellas will split U.S. profits on enzalutamide 50/50 and Medivation is entitled to royalties (estimated in the high teens to low 20%s) on ex-U.S. sales. Our valuation analysis is highly sensitive to changes in enzalutamide's peak potential, but suggests shares are fairly valued for peak sales of $3B.

| | |
|---|---|
| **MDVN (06/27)** | **$88.74** |
| **Mkt cap** | **$3.1B** |
| Dil shares out | 34.9MM |
| Avg daily vol | 415.8K |
| 52-wk range | $14.3-90.4 |
| Dividend | Nil |
| Dividend yield | Nil |
| BV/sh | $-0.11 |
| Net cash/sh | $4.68 |
| Debt/cap | 0.0% |
| ROIC (LTM) | NA |
| 5-yr fwd EPS growth (Norm) | NA |
| **S&P 500** | **1331.9** |

**Revenue $MM**

| FY Dec | 2011 Actual | 2012E Prior | 2012E Current | 2013E Prior | 2013E Current | 2014E Current | 2015E Current |
|---|---|---|---|---|---|---|---|
| Q1 | 14.7 | — | 36.8A | — | — | — | — |
| Q2 | 15.8 | — | 41.5 | — | — | — | — |
| Q3 | 14.9 | — | 26.4 | — | — | — | — |
| Q4 | 14.9 | — | 35.9 | — | — | — | — |
| Year | **60.4** | — | **140.6** | — | **189.0** | **332.5** | **499.0** |
| EV/S | — | — | 10.4x | — | 7.7x | 4.4x | 2.9x |

**EPS* $**

| FY Dec | 2011 Actual | 2012E Prior | 2012E Current | 2013E Prior | 2013E Current | 2014E Current | 2015E Current |
|---|---|---|---|---|---|---|---|
| Q1 | (0.24) | — | 0.01A | — | — | — | — |
| Q2 | (0.27) | — | 0.07 | — | — | — | — |
| Q3 | (0.29) | — | (0.42) | — | — | — | — |
| Q4 | (0.31) | — | (0.35) | — | — | — | — |
| Year | **(1.11)** | — | **(0.66)** | — | **(0.55)** | **2.64** | **4.07** |
| P/E | — | — | — | — | — | 33.6x | 21.8x |

*EPS estimates include stock-based compensation expense, exclude one-time charges

# AFFIRMation Of Enzalutamide

Enzalutamide is a small molecule androgen receptor (AR) antagonist for the treatment of castration-resistant prostate cancer (CRPC). Medivation in-licensed enzalutamide along with a series of AR antagonists from UCLA, and following compelling Phase I/II proof-of-concept data, signed a major development collaboration with Astellas. Enzalutamide's mechanism of action is unique, and highly relevant to CRPC. Moreover, data from the company's first pivotal trial (AFFIRM) indicate solid efficacy and very good tolerability, supporting optimism from consultants that enzalutamide could become a major if not dominant player in the CRPC market. In May 2012, Medivation/Astellas submitted an NDA to the FDA for enzalutamide in chemotherapy-refractory castrate-resistant prostate cancer. The company does not expect an advisory panel for the drug and continues to expect priority (6-month) review, positioning enzalutamide for a late 2012 launch. Enzalutamide was submitted for EMA approval in June.

## No Devil In The Details

In November 2011, Medivation reported highly positive top-line efficacy from its pivotal AFFIRM trial testing enzalutamide in patients with chemotherapy-resistant prostate cancer. The 4.8-month median overall survival advantage (p<0.0001, 18.4 months vs. 13.6 months for placebo) is at least on par with that reported from a pivotal trial on JNJ's Zytiga. Full data for AFFIRM were released at ASCO GU in February 2012. The new data confirm enzalutamide's efficacy across multiple secondary endpoints and point to a very well tolerated side effect profile. The only blemish on enzalutamide is a slightly higher incidence of seizures (0.6% on enzalutamide vs. 0% on placebo). While the rate of seizures was lower than some had feared, it leaves the door open for counter-detailing from competitor JNJ. Nonetheless, the presenting authors conclude that enzalutamide will be a first-line drug in chemotherapy-resistant PRCA.

**Enzalutamide's Impressive Phase III Results In AFFIRM**

| | MDV3100 | Placebo |
|---|---|---|
| Median Survival* | 18.4 months | 13.6 months |

*p<0.0001; HR=0.631

- MDV3100 prolonged survival by 4.8 months compared to placebo
- MDV3100 reduced risk of death by 37% compared to placebo
- Survival benefit was consistently seen across all patient subgroups

Source: Medivation

Data from AFFIRM are summarized as follows:

* enzalutamide met all secondary efficacy endpoints including radiographic progression-free survival (8.3 months vs. 2.9 months for placebo, p<0.0001, HR=0.404), soft tissue response rate (28.9% vs. 3.8%, p<0.001), and time to PSA progression (8.3 months vs. 3.0 months, p<0.0001, HR=0.249).

* PSA declines of 50% and 90% were also more common in the enzalutamide group versus placebo.

**Enzalutamide Achieved All Secondary Endpoints In AFFIRM**

| | MDV3100 | Placebo |
|---|---|---|
| Radiographic PFS | 8.3 months | 2.9 months |
| Soft tissue objective response rate (PR + CR) | 28.9% | 3.8% |
| Time to PSA progression | 8.3 months | 3 months |
| PSA reductions of 50% or greater from baseline | 54% | 1.5% |
| PSA reductions of 90% or greater from baseline | 24.8% | 0.9% |

Source: Medivation

* Enzalutamide was generally well tolerated. Common side effects were fatigue, diarrhea and hot flashes. SAEs and AEs causing dose discontinuation were all lower in the enzalutamide group. Grade 3 or greater AEs occurred in 45.3% of enzalutamide patients and 53.1% of PBO patients. Considering a longer adverse reporting period for enzalutamide relative to placebo (patients monitored for 30 days post dosing), lower adverse event rates are remarkable.

* Seizures occurred in five enzalutamide patients versus zero on placebo. Given the 2:1 randomization, AFFIRM suggests that enzalutamide is associated with a modest risk of seizure. Clearly enzalutamide's overall risk benefit is still highly favorable, and Medivation does not expect any driving restrictions on enzalutamide's label. However, our consultants had suggested that this level of seizure risk could be borderline concerning to physicians, and might even out the playing field with JNJ's Zytiga (which is itself constrained by concomitant steroid dosing).

**Grade 3 Or Greater AEs Of Interest in Enzalutamide's AFFIRM Trial**

| | MDV3100 | Placebo |
|---|---|---|
| Fatigue | 6.3% | 7.3% |
| Cardiac Disorders* | 0.9% | 2.0% |
| Myocardial Infarction | 0.3% | 0.5% |
| Seizure | 0.6% | 0% |
| LFT Abnormalities | 0.4% | 0.8% |

Source: Medivation

Based upon the strength of these results, investors have begun to credit enzalutamide for sales in the much larger pre-chemotherapy market (PREVAIL results expected in 2013 or 2014). Our best guess is that enzalutamide has worldwide potential of roughly $3B, and shares appear priced to reflect this level to slightly more. We note a wide range of outcomes for enzalutamide is still possible and that our valuation analysis is highly sensitive to changes in enzalutamide's peak potential.

## Quality Of Life Data Presented At ASCO 2012

In June 2012 at ASCO, Medivation presented data on secondary quality of life endpoints from AFFIRM. As determined by the Functional Assessment of Cancer Therapy – Prostate (FACT-P) questionnaire, patients treated with enzalutamide had a significantly higher response rate (p<0.0001) in health-related quality of life compared to placebo (43.2% vs. 18.3%). The questionnaire examined 27 core items to evaluate overall patient function (for example level of pain, ability to work, level of energy, ability to cope with illness, etc), and recorded a 10 point or greater improvement in score as a health-related quality of life response. Additionally, in patients treated with enzalutamide, the median time to first skeletal-related event was 16.7 months vs. 13.6 months in the placebo group (p=0.0001; HR=0.688).

## Background On The AFFIRM Trial

The AFFIRM trial evaluated the 160mg dose of enzalutamide administered QD compared to placebo (2:1 randomization). This dose was selected from the Phase I/II dose escalation study (testing doses of 30mg to 600mg) as one that offered optimal levels of tolerability and efficacy. The trial began enrolling patients with castration-resistant prostate cancer (CRPC) refractory to docetaxol-based chemotherapy in September 2009 and completed enrollment at 1,199 patients in November 2010. The trial design of AFFIRM is nearly identical to Zytiga's successful Study 301 which included 1,195 post-chemotherapy patients.

**AFFIRM Trial Design**


Source: Medivation

### Study Halted After Interim Analysis

AFFIRM was a large, well-designed trial with a relatively low hurdle to success (good statistical power, placebo controlled). In addition, the event rate in AFFIRM appeared slower than expected, suggesting that enzalutamide might be having the desired effect on survival. Nonetheless, top-line data released in November 2011 topped nearly all expectations. AFFIRM's interim analysis, which was triggered once 80% of the planned events (520 out of 650 deaths) occurred, easily achieved statistical significance ($p<0.0001$) in the primary overall survival endpoint. Enzalutamide was associated with a 4.8-month median OS advantage. Median OS for men treated with enzalutamide was 18.4 months compared to 13.6 months for the placebo group. After considering the safety profile, the IDMC informed Medivation that the risk-to-benefit ratio was favorable, and recommended that the trial be stopped and patients receiving placebo be offered enzalutamide. According to Medivation, no further analysis of the data will be conducted.

The FDA has granted enzalutamide Fast Track designation, enabling MDVN to request priority (six month) review. Medivation and partner Astellas still plan to meet with the FDA in early 2012 to discuss regulatory timelines, and we anticipate a filing in H1:12. MDVN will receive a $10MM milestone upon acceptance of the NDA, and additional filing milestones for Europe and Japan. As expected, Medivation announced that it has exercised its 50/50 co-promotion option on U.S. rights to enzalutamide.

## Astellas A Solid Partner For Enzalutamide

In October 2009, Medivation entered into a collaboration with Astellas to co-develop and commercialize enzalutamide for the treatment of early and late-stage prostate cancer. Under the terms of the agreement, Medivation received an up-front payment of $110MM and potential development and regulatory milestones of up to $335MM and commercial milestones of up to $320MM. Medivation and Astellas will split U.S. development and commercialization costs as well as profits 50/50. Medivation exercised its option to co-promote enzalutamide in the U.S. market, contributing 20-50% of the sales reps. Outside of the U.S., Astellas will bear full responsibility for the development and commercialization of enzalutamide: Medivation will receive tiered double-digit royalties on sales outside the U.S. (we estimate 15-22%).

Astellas has a strong commercial presence in urology, selling Flomax for benign prostatic hyperplasia (BPH) and Vesicare for overactive bladder on a global basis, so should be a good sponsor for enzalutamide in urology circles.

## JNJ's Zytiga Is The Main Competitor

There are a number of new entrants in the prostate cancer (PRCA) marketplace (Zytiga, carbazitaxel, Provenge) and several promising therapies waiting in the wings (Alpharadin, cabozantinib). However, JNJ's Zytiga appears to be the main initial threat to enzalutamide given their related mechanisms of action and potential to be used in patients who are not yet truly androgen independent.

JNJ's Zytiga (abiraterone) demonstrated a 3.9 month median overall survival benefit at the interim analysis and a 4.6 month benefit at the final analysis Study 301 (chemotherapy-refractory CRPC). Patients in the abiraterone/prednisone arm had a median survival of 14.8 months, vs. 10.9 months for the control arm (HR 0.65; $p<0.0001$). Zytiga also improved time to disease progression (10.2 vs. 6.6 months; $p<0.0001$). Hence the efficacy of Zytiga and MDV3011 appears comparable.

In March, Zytiga's pivotal COU-AA-302 trial in chemotherapy-naive prostate cancer patients was halted due to convincing efficacy of Zytiga relative to the standard-of-care control arm. The interim data through December 2011 showed a statistically significant improvement in radiographic progression-free survival (rPFS) in patients receiving Zytiga plus prednisone compared to the placebo plus prednisone in the control group ($p<0.0001$). The median rPFS in the control group was 8.3 months, but as of the December 2011 interim look, median rPFS had not yet been reached in the Zytiga treatment group (n=150 progression events in the Zytiga treatment group vs. 251 in the control group; HR = 0.43; 95% CI = [0.35, 0.52]; $p<0.0001$). Zytiga showed a strong trend in favor of an overall survival benefit, but did not achieve statistical significance on OS at the interim look. (HR=0.75, p=0.0097 vs. prespecified p-value of 0.0008).

Zytiga works through a similar anti-androgen-based mechanism, and at this stage there is little data on its combinability with enzalutamide. There is also some anecdotal evidence to suggest that sequential therapy may not be ideal. In other words, once a patient becomes resistant to one, he is more likely resistant to the other. As a result, it was important that enzalutamide produce survival data that are at least on par with Zytiga, allowing Medivation and Astellas to position enzalutamide as a therapy that could be used ahead of Zytiga.

Zytiga was approved by the FDA in April 2011 and EMA in September 2011. JNJ posted $301MM WW sales in 2011. Enzalutamide's major advantage relative to Zytiga appears to be its superior tolerability. Zytiga requires co-administration of prednisone, a steroid which over time is difficult to tolerate. This is less of an issue in the post-chemotherapy setting where many men receive prednisone with Taxotere and never come off. However, in the pre-chemotherapy setting, where patients could stay on drug for multiple years, enzalutamide could have a natural advantage. Enzalutamide (once daily, no food restrictions) may also be somewhat more convenient to administer than Zytiga (BID, with food).

### Consultants Give Enzalutamide The Nod Over Zytiga

In May we hosted a physician consultanting call on prostate cancer. The two clear "winners" from our call were MDVN's enzalutamide and JNJ's abiraterone (Zytiga).

Our specialists view JNJ's Zytiga (abiraterone) and Medivation's enzalutamide as having comparable efficacy in the treatment of post-chemotherapy CRPC. However, they noted that the two drugs have differences in their activity on androgens and that a differential mechanism is reflected in the side-effect profiles. While both drugs are viewed as far safer than chemotherapy, clinicians indicated that enzalutamide may have a modestly better side-effect profile, terming it "amazingly clean", and suggesting its low seizure risk was a bit more acceptable than Zytiga's requirement for concomitant steroid dosing. In terms of seizures, doctors indicated that those observed were mild, and not necessarily all drug related. They think the important thing is not to use enzalutamide in combination with other agents that lower seizure risk. One of our specialists has been dosing Zytiga with a lower ("sub-physiologic") dose of prednisone (5mg QD rather than 5mg BID), and has seen no change in the side-effects related to Zytiga but an improvement in the prednisone-related side effects. Still this consultant believes that enzalutamide's overall tolerability profile remains superior to that of Zytiga.

## Comparison Between MDVN's Enzalutamide And JNJ's Abiraterone

**MDV3100 VS. ZYTIGA POST-CHEMO CRPC COMPARISON**

| Key Endpoints Measured | MDV3100 160mg QD | placebo | p-values (HR 95% CI) | Zytiga 1,000mg QD + 5mg prednisone BID | placebo + prednisone | p-values (HR 95% CI) |
|---|---|---|---|---|---|---|
| | (n=799) | (n=400) | | (n=797) | (n=398) | |
| **Overall survival (OS)** | **18.4 months** | **13.6 months** | **p<0.0001 (HR=0.631)** | **15.8 months** | **11.2 months** | **p<0.0001 (HR=0.74)** |
| **net improvement in OS vs. PBO** | **4.8 months** | | | **4.6 months** | | |
| radiographic PFS (rPFS) | 8.3 months | 2.9 months | p<0.0001 (HR=0.404) | 5.6 months | 3.6 months | p<0.001 |
| time to PSA progression (TTPP) | 8.3 months | 3.0 months | p<0.0001 (HR=0.249) | 10.2 months | 6.6 months | p<0.001 |
| soft tissue/PSA response rate | 28.9% | 3.8% | p<0.001 | 29.1% | 5.5% | p<0.001 |
| PSA declines of ≥50% | 54.0% | 1.5% | p<0.0001 | | | |
| PSA declines of ≥90% | 24.8% | 0.9% | p<0.0001 | | | |
| **Safety Profile** | | | | | | |
| Common Grade 3 or greater AEs: | 45.3% | 53.1% | | | | |
| seizure incidence | 0.6% (5 patients) | 0.0% | | 0.0% (0 patients) | | |
| fatigue | 6.3% | 7.3% | | --- | --- | |
| cardiac disorders | 0.9% | 2.0% | | --- | --- | |
| myocardial infarction | 0.3% | 0.5% | | --- | --- | |
| liver function test abnormalities | 0.4% | 0.8% | | --- | --- | |
| joint swelling/discomfort | --- | --- | | 29.5% | 4.2% | |
| muscle discomfort | --- | --- | | 26.2% | 3.0% | |
| edema | --- | --- | | 26.7% | 1.9% | |
| hypertension | --- | --- | | 8.5% | 1.3% | |
| diarrhea | --- | --- | | 17.6% | 0.6% | |
| dyspepsia | --- | --- | | 6.1% | 0.0% | |
| urinary tract infections | --- | --- | | 11.5% | 2.1% | |
| upper respiratory tract infections | --- | --- | | 5.4% | 0.0% | |
| arrhythmias | --- | --- | | 7.2% | 1.1% | |
| chest pain or discomfort | --- | --- | | 3.8% | 0.5% | |
| cardiac failure | --- | --- | | 2.3% | 1.9% | |
| **Label Warnings** | TBD (currently seeking approval in post-chemo CRPC) | | | - mineralocorticoid excess<br>- adrenocortical insufficiency<br>- hepatotoxicity<br>- food effect | | |
| **Pregnancy Category** | TBD (currently seeking approval in post-chemo CRPC) | | | Category X (not indicated for use in pregnant women) | | |

*Source: Cowen and Company; Company Data*

Source: Cowen and Company, ASCO Abstract 2010

## Post-Chemotherapy Setting Is Relatively Modest

Between 30-35K U.S. patients succumb to metastatic PRCA each year. Roughly 60% of these patients are believed to receive Taxotere and are therefore likely to be eligible for enzalutamide. We assume that Zytiga's entrenched status and enzalutamide's less differentiated profile in this setting will allow these drugs to share this market roughly equally. However, the more successful enzalutamide becomes in the much larger pre-chemotherapy setting (see below), the less likely it will be used in post-chemotherapy patients as patients are unlikely to get two courses of the drug. We model U.S. sales of $75MM in 2013 and $225MM in 2017 in the post-chemotherapy CRPC setting. We view Europe as a similar sized opportunity on which we estimate MDVN will receive escalating royalties in the 15-22% range.

**U.S. Post Chemotherapy Revenue Model**

| United States | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|
| **Post-chemotherapy Setting** | | | | | |
| Incidence of metastatic CRPC | 31.0 | 31.3 | 31.7 | 32.0 | 32.4 |
| *Penetration of Taxotere* | *60%* | *60%* | *60%* | *60%* | *60%* |
| # patients who receive first-line chemotherapy | 18.6 | 18.8 | 19.0 | 19.2 | 19.4 |
| *Penetration of MDV 3100 into second-line* | *27%* | *48%* | *37%* | *26%* | *23%* |
| # patients who receive MDV 3100 post-chemotherapy | 5.0 | 9.0 | 7.0 | 5.0 | 4.5 |
| MDV 3100 price per patient | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 |
| **MDV3100 sales in post-chemotherapy setting ($MM)** | **$250** | **$450** | **$350** | **$250** | **$225** |

Source: Cowen and Company

## Pre-Chemotherapy Is Where The Money Lies

Many more patients fail androgen deprivation therapy and develop metastatic CRPC each year than receive Taxotere. We estimate the pre-chemotherapy market might be 5x larger than the post chemotherapy market. In theory, both MDV3011 and Zytiga are likely to work in this patient subset. Both drugs rely on the fact that there are several ways through which prostate tumors become resistant to androgen deprivation therapy. These include through the overproduction of adrenal androgens, the development of androgen receptor mutations, amplifications in the androgen receptor, and the outgrowth of cells no longer sensitive to androgens. Neither Zytiga nor enzalutamide will work in patients who harbor tumors that are truly insensitive to androgens. However Zytiga and enzalutamide ought to work in patients who overproduce androgens, or develop androgen receptor mutations/ amplifications that make their cancer cells more sensitive to low concentrations of testosterone for growth.

### PREVAIL Fully Enrolled

In September 2010, Astellas and Medivation initiated the Phase III PREVAIL trial evaluating enzalutamide in CRPC patients who are chemotherapy naïve. PREVAIL is testing the hypothesis that enzalutamide improves OS and PFS in these patients. In June 2012 Medivation announced that this trial is fully enrolled. PREVAIL is modeled off Zytiga's Study 302 (for which JNJ received an SPA), but is designed to enroll more patients (1,700 vs. 1,000). The inclusion of more patients could allow events to accrue faster, enabling Medivation and Astellas to close some of the gap in timing. JNJ's Study 302 began in April 2009.

**PREVAIL Trial Design**


Source: Medivation

The trial costs for PREVAIL will be allocated to Medivation and Astellas in the same proportion as the AFFIRM trial (roughly one third : two thirds). The PREVAIL study is enrolling patients in North America, Europe, Australia, and Israel. The co-primary endpoints of the study are overall survival and progression-free survival (PFS). The secondary endpoints in the PREVAIL study include time to first skeletal-related event and time to initiation of chemotherapy.

## We Have High Expectations For Enzalutamide In Pre-Chemotherapy

Biologically there appears to be limited differences between a CRPC patient who has failed chemotherapy and one who has not. Moreover, enzalutamide's Phase I/II experience suggests the drug is highly active in pre-chemotherapy patients. In addition, as noted above, pre-chemotherapy is the setting in which enzalutamide's tolerability advantages relative to Zytiga could have the most benefit. Lastly, there is some scientific basis to support why enzalutamide may have greater efficacy relative to Zytiga in this setting (though in the end, the proof will be in the pudding).

While Phase III data in the pre-chemotherapy setting are unlikely before 2013 or 2014, we are comfortable adding significant estimates for enzalutamide in pre-chemotherapy patients to our model. At peak, we believe U.S. sales of enzalutamide in this setting could approach $1.5B. We also model ex-U.S. sales of nearly $1.5B in this setting.

**U.S. Pre Chemotherapy Revenue Model**

| United States | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|
| **Pre-chemotherapy setting** | | | | | |
| # patients on hormonal therapy (000) | 675.0 | 682.4 | 689.9 | 697.5 | 705.2 |
| *% patients who progress on hormonal therapy* | *23%* | *23%* | *23%* | *23%* | *23%* |
| # patients who fail hormonal therapy | 155.3 | 157.0 | 158.7 | 160.4 | 162.2 |
| *% patients who advance to second-line hormonal therapy* | *50%* | *50%* | *51%* | *53%* | *55%* |
| # patients who receive second-line hormonal therapy | 77.6 | 78.5 | 80.9 | 85.0 | 89.2 |
| *Penetration of MDV 3100 into second-line hormonal therapy* | *1%* | *1.6%* | *7%* | *11%* | *13%* |
| # patients who receive MDV 3100 in pre-chemotherapy setting | 0.6 | 1.3 | 5.6 | 9.4 | 11.3 |
| MDV3100 price per patient | $80,000 | $80,000 | $80,000 | $80,000 | $80,000 |
| **MDV3100 sales in HRPC ($MM)** | **$50** | **$100** | **$450** | **$750** | **$900** |

Source: Cowen and Company (All figures estimated.)

## Enzalutamide Predicted To Gain Use Ahead Of Abiraterone…

Our consultants expect the PREVAIL trial of enzalutamide in pre-chemotherapy patients to demonstrate a statistically significant improvement on the PFS endpoint. They pointed to the Phase II experience with enzalutamide (12 month median PFS) as providing strong support for this optimism. While they expect enzalutamide's efficacy to be similar to that of abiraterone, they did note trial design risk in the form of crossover (control arm patients receiving drugs like abiraterone) as a potential source of risk to PREVAIL's ability to achieve its OS endpoint. To the degree that insurers allow, oncologists expect to use enzalutamide in pre-chemotherapy CRPC patients when it becomes available for post-chemotherapy use early next year.

Assuming enzalutamide succeeds in the PREVAIL trial and produces similar data to that of abiraterone, our consultants expect to use enzalutamide ahead of abiraterone in their pre-chemo CRPC patients. While doctors say there is a lack of data to guide this decision and would like to have more information, their bias is to use enzalutamide ahead of abiraterone. This given enzalutamide's favorable ease-of-use (no prednisone co-administration), simpler administration, durable responses, and its ability to be combined with Provenge. While combination of the two agents is a theoretical possibility, our consultants do not feel comfortable with the approach in the absence of data. However, a straw poll of 26 oncologists taken at a recent meeting was more split in its view, with one third indicating they would prefer to use enzalutamide before abiraterone, one third abiraterone before enzalutamide, and one third in combination.

Their expectation is that 40-50% of patients who fail enzalutamide might respond to abiraterone and stay on that drug for a moderate period of time. Hence, despite likely use after enzalutamide, physicians anticipate renewed growth in abiraterone scrips as the drug moves into the pre-chemotherapy setting.

## …And Even Displace Casodex

MDVN's Phase II TERRAIN trial is enrolling 370 patients who have failed first-line hormone therapy, but do not necessarily have metastatic disease, to enzalutamide versus bicalutamide (Casodex, the most commonly used anti-androgen). The primary endpoint is progression free survival. The study began in March 2011 and a Phase II trial in hormone-naïve patients opened in May 2011.

**TERRAIN Trial Design**


Source: Medivation

Even without data from this trial, physicians are very comfortable prescribing enzalutamide in patients who have failed ADT ahead of Casodex. They view Casodex as a poor man's anti-androgen, that will fall by the wayside in place of drugs like enzalutamide (in particular) as well as Zytiga. Given the large number of new patients with CRPC (roughly 50K/year in the U.S.) and prolonged duration of response (12+ months?), the market for enzalutamide in this setting could be multi-billion.

Doctors also commented favorably on Zytiga's 58-patient Phase II trial in neo-adjuvant prostate cancer (Zytiga + Lupron, data to be presented at ASCO). The study shows that Zytiga plus Lupron outperformed Lupron alone in localized, high-risk prostate cancer, and was well-tolerated. However, doctors noted that much larger trials were required to confirm either Zytiga's or enzalutamide's efficacy in this earlier-stage population. Physicians were also less sanguine about enzalutamide's ability to move into the first-line hormone setting and displace androgen deprivation therapy. While they thought this was theoretically possible, they believe it would require very large and long-term trials.

## We Project $3B Enzalutamide's Peak Potential

We project a 2013 U.S. and ex-U.S. launch. Assuming $3B in peak WW sales (Casodex achieved peak sales of $1.3B), we think the NPV of enzalutamide is in the $70-75 range.

## NPV For Enzalutamide ($3B Peak Sales)

| | |
|---|---|
| Financial Year End | 12/31/2010 |
| Valuation Date | 6/21/2012 |
| **Discount Rate** | **10.0%** |
| **Perpetual Growth Rate** | **0.0%** |

Thursday, June 21, 2012

| $MM | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **US MDV3100** | | | | | | | | | | | | | | | | | |
| **MDV3100 Sales** | 0 | 300 | 550 | 800 | 1,000 | 1,125 | 1,226 | 1,312 | 1,378 | 1,433 | 1,476 | 1,505 | 1,520 | 1,536 | 1,551 | 1,241 | 248 |
| *%Y/Y Growth* | | | *83%* | *45%* | *25%* | *13%* | *9%* | *7%* | *5%* | *4%* | *3%* | *2%* | *1%* | *1%* | *1%* | *-20%* | *-80%* |
| **MDV3100 COGS** | 0 | 45 | 55 | 80 | 90 | 90 | 98 | 105 | 110 | 115 | 118 | 120 | 122 | 123 | 124 | 99 | 20 |
| *Gross Margin* | | *85%* | *90%* | *90%* | *91%* | *92%* | *92%* | *92%* | *92%* | *92%* | *92%* | *92%* | *92%* | *92%* | *92%* | *92%* | *92%* |
| **R&D Expense In Support of Prostate Cancer indication** | 70 | 60 | 50 | 40 | 40 | 40 | 30 | 20 | 20 | 20 | 20 | 20 | 20 | 0 | 0 | 0 | 0 |
| *% of sales* | | *20%* | *9%* | *5%* | *4%* | *4%* | *2%* | *2%* | *1%* | *1%* | *1%* | *1%* | *1%* | *0%* | *0%* | *0%* | *0%* |
| **SG&A In Support of Prostate Cancer Indication** | 80 | 150 | 250 | 275 | 300 | 310 | 320 | 325 | 325 | 325 | 325 | 325 | 300 | 275 | 250 | 100 | 10 |
| *% of sales* | | *50%* | *45%* | *34%* | *30%* | *28%* | *26%* | *25%* | *24%* | *23%* | *22%* | *22%* | *20%* | *18%* | *16%* | *8%* | *4%* |
| **MDVN's Share of US MDV3100 Operating Profit** | **(75)** | **23** | **98** | **203** | **285** | **343** | **389** | **431** | **461** | **487** | **506** | **520** | **539** | **569** | **588** | **521** | **109** |
| *% Overall JV's US MDV3100 Operating Margin* | *na* | *na* | *35%* | *51%* | *57%* | *61%* | *63%* | *66%* | *67%* | *68%* | *69%* | *69%* | *71%* | *74%* | *76%* | *84%* | *88%* |
| **MDVN's US MDV3100 Milestones** | **40** | **10** | **5** | **20** | **20** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** |
| **Ex-US MDV3100** | | | | | | | | | | | | | | | | | |
| **MDV3100 Sales** | 0 | 50 | 300 | 550 | 800 | 1,000 | 1,130 | 1,232 | 1,318 | 1,384 | 1,439 | 1,482 | 1,512 | 1,527 | 1,222 | 489 | 195 |
| *%Y/Y Growth* | | | *500%* | *83%* | *45%* | *25%* | *13%* | *9%* | *7%* | *5%* | *4%* | *3%* | *2%* | *1%* | *-20%* | *-60%* | *-60%* |
| **MDVN's Ex-US MDV3100 Royalties** | **0** | **8** | **45** | **99** | **160** | **210** | **249** | **271** | **290** | **304** | **317** | **326** | **333** | **336** | **269** | **88** | **29** |
| *% Royalty Rate* | | *15%* | *15%* | *18%* | *20%* | *21%* | *22%* | *22%* | *22%* | *22%* | *22%* | *22%* | *22%* | *22%* | *22%* | *18%* | *15%* |
| **MDVN's Ex-US MDV3100 Milestones** | **5** | **20** | **15** | **20** | **10** | **20** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** |
| **WW MDV3100** | | | | | | | | | | | | | | | | | |
| **MDVN's Total MDV3100 WW Operating Income** | **(30)** | **60** | **163** | **342** | **475** | **573** | **638** | **702** | **751** | **791** | **823** | **846** | **872** | **905** | **857** | **609** | **138** |
| **Tax Adjusted EBIT** | **(30)** | **60** | **106** | **222** | **309** | **372** | **414** | **456** | **488** | **514** | **535** | **550** | **567** | **588** | **557** | **396** | **90** |
| *Tax rate* | *0%* | *0%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* | *35%* |
| **Approximate MDV3100-related Free Cash Flow** | **(30)** | **60** | **106** | **222** | **309** | **372** | **414** | **456** | **488** | **514** | **535** | **550** | **567** | **588** | **557** | **396** | **90** |
| *% Y/Y Growth* | | | *76%* | *110%* | *39%* | *21%* | *11%* | *10%* | *7%* | *5%* | *4%* | *3%* | *3%* | *4%* | *-5%* | *-29%* | *-77%* |
| Years | 0.02 | 1.02 | 2.02 | 3.02 | 4.02 | 5.02 | 6.02 | 7.02 | 8.02 | 9.02 | 10.02 | 11.02 | 12.02 | 13.02 | 14.02 | 15.02 | 16.02 |
| Discount Factor | 1.00 | 0.91 | 0.82 | 0.75 | 0.68 | 0.62 | 0.56 | 0.51 | 0.47 | 0.42 | 0.38 | 0.35 | 0.32 | 0.29 | 0.26 | 0.24 | 0.22 |
| **NPV of MDV3100 Cash flows** | **(30)** | **54** | **87** | **166** | **210** | **231** | **233** | **234** | **227** | **218** | **206** | **192** | **180** | **170** | **146** | **95** | **20** |

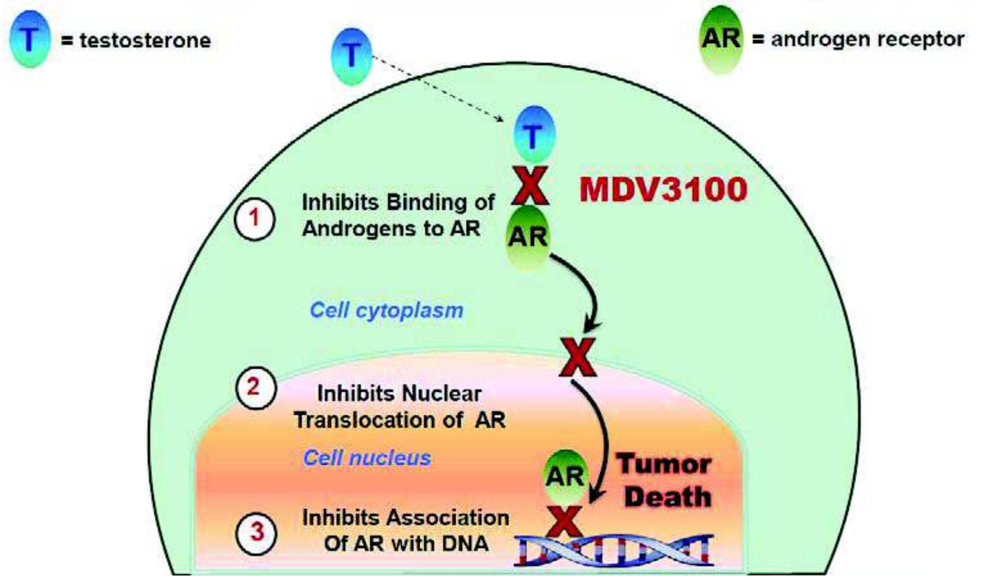
| | |
|---|---|
| Final year FCF | 0 |
| Perpetual Growth Rate | 0.0% |
| **Terminal Value** | 0 |
| Discount Factor | 0.47 |
| **Present Value of Terminal Value** | **0** |
| **Present Value of Cash Flows** | **2,640** |
| **Present Value of MDVN's MDV3100 Cash Flows** | **2,640** |
| Net cash | 121 |
| Fully Diluted Shares Outstanding | 38.6 |
| **Present Value of Cash Flows Per Share** | **$71.44** |

Source: Cowen and Company

# Background On Enzalutamide

An article published in the April 9, 2009 issue of *Science Express* highlighted the mechanism and androgen receptor binding properties of enzalutamide. The article was authored by Charles Sawyers, who discovered the over-expression of the androgen receptor in prostate cancer cells on which the enzalutamide mechanism is based. The paper compares enzalutamide to AstraZeneca's leading oral anti-androgen therapy bicalutamide (Casodex, another AR receptor blocker) and offers scientific evidence that enzalutamide inhibits the androgen receptor with enhanced binding affinity compared to bicalutamide. According to the article, enzalutamide also blocks DNA binding and nuclear translocation, a critical process required for androgen-dependent prostate cancer cell growth, and induces castration-resistant tumor cell death. Enzalutamide has exhibited activity in patients whose tumors are not responsive to bicalutamide and other prostate cancer treatments. The enhanced binding affinity and unique mechanism of action may offer advantages to other prostate cancer therapies marketed and in development. Enzalutamide also could have promise as an adjunctive therapy given its differentiated mechanism.

**Enzalutamide Blocks Androgen Signaling In Three Ways**




Source: Medivation

# Phase I/II Results Were Promising

Data from Medivation's Phase I/II study for enzalutamide were published in The Lancet in April 2010. Efficacy endpoints included circulating tumor cell counts, serum prostate specific antigen (PSA) measurements, soft tissue and bony metastases, and duration of treatment. Efficacy highlights are as follows:

**Anti-Tumor Activity Observed in Enzalutamide's Phase I/II Trial**

- 140 pre-chemo and post-chemo patients enrolled 7/07 to 12/08
- Doses 30-600 mg/day; well tolerated up to 240 mg/day
- Anti-tumor activity shown across multiple endpoints

| | Pre-chemo | Post-chemo |
|---|---|---|
| ≥50% PSA decline | 62% | 51% |
| Stabilization of soft tissue metastases | 80% | 65% |
| Stabilization of bone metastases | 63% | 51% |
| Favorable CTC Conversion | 75% | 37% |

Source: Medivation

The data on CTC conversion deserves special mention. Circulating tumor cells (CTC) is a biomarker for overall survival in patients with metastatic prostate cancer. In retrospective analyses post-treatment CTC counts has been a predictive measure of survival rates and may be a better indicator of overall survival than PSA response. 75% of chemotherapy-naïve patients and 37% of post-chemotherapy patients with unfavorable CTC counts of 5 or higher showed favorable CTC counts less than 5 after treatment. While ongoing trials including AFFIRM will look at CTC conversion on a prospective basis, opinion leaders place much weight on the conversion rates seen to date with enzalutamide.

On the safety side, at the higher enzalutamide doses of 360mg and 600mg, two patients experienced seizures. However, it is unclear whether these were attributable to enzalutamide, since both patients were on other meds. Another patient taking 480mg experienced symptoms consistent with a seizure. Fatigue appears to be the most commonly reported AE on enzalutamide, but the drug was well tolerated at doses less than 240mg.

## Long-Term Follow Up Confirm Sustained Efficacy

Medivation's long-term Phase 1/2 data were presented in February 2011 at ASCO's Genitourinary Cancer's Symposium and were summarized in a poster "Anti-tumor activity of enzalutamide in pre- and post-docetaxel advanced prostate cancer: long-term follow-up of the Phase I/II study."

The key data include median time to prostate-specific antigen (PSA) progression analyzed in three ways defined in the following table: (1) criteria as specified in the Phase I/II study protocol; (2) the Prostate-Specific Antigen Working Group 1 (PSAWG1) criteria; and (3) the published Prostate Cancer Clinical Trials Working Group 2 (PCWG2) criteria.

**PSA Response Criteria**

| 3 METHODS DEFINED FOR PROGRESSION FREE SURVIVAL | | |
|---|---|---|
| **Method** | **Approach** | **Definition** |
| (1) MDV3100 Phase I/II protocol | **Liberal** | ·25% increase in PSA above baseline |
| (2) Prostate-Specific Antigen Working Group 1 (PSAWG1) | **Intermediate** | ·50% increase in PSA above nadir for patients with PSA decline ≥50%<br>·25% increase in PSA above nadir for patients with PSA decline <50%<br>·25% increase in PSA above baseline for patients without PSA decline |
| (3) Prostate Cancer Clinical Trials Working Group 2 (PCWG2) | **Conservative** | ·25% increase in PSA above nadir |

Source: Cowen and Company

The additional data confirm enzalutamide's sustained efficacy profile in advanced prostate cancer patients as measured by median times to prostate-specific antigen (PSA) progression, radiographic progression, and circulating tumor cell counts in 18/140 patients that remained on active therapy (n=16 chemotherapy naïve and 2 post-chemotherapy). In addition, circulating tumor counts showed that 91% (70/77) of patients that had favorable pretreatment counts of <5 cells/7.5mL blood remained favorable in a post-treatment setting and 49% (25/51) of patients converted from unfavorable pre-treatment counts to favorable post-treatment counts.

**Longer-Term Phase I/II Follow Up Data From ASCO GU 2011**

| **Key Long-Term Data Presented At ASCO** | **MDV3100**<br>**chemo-naïve**<br>n=65 | <br>**post-chemo**<br>n=75 |
|---|---|---|
| Median time to PSA progression | not reached | 316 days (45 weeks) |
| Prostate-Specific Antigen Working Group 1 (PSAWG1) | 420 days (60 weeks)*<br>812 days (116 weeks)** | 166 days (24 weeks) |
| Prostate Cancer Clinical Trials Working Group 2 (PCWG2) | 281 days (40 weeks) | 148 days (21 weeks) |
| Median times to radiographic progression | 392 days (56 weeks) | 175 days (25 weeks) |

*all chemotherapy-naïve patients; **chemotherapy-naïve patients who were also ketoconazole-naïve

Source: Cowen and Company

## Enzalutamide Trial In Breast Cancer

In addition to the prostate cancer indication, Medivation announced in August 2011 that enzalutamide blocks both androgen- and estrogen-mediated breast cancer cell proliferation in pre-clinical studies. Astellas and Medivation are likely to begin a Phase I/II trial in breast cancer in 2012.

## Enzalutamide Series Licensed From UCLA

Enzalutamide was in-licensed from UCLA. A group of academics synthesized approximately 170 small molecule compounds that targeted the androgen receptor (the MDV 300 series). Medivation has licensed all rights to these compounds and holds exclusive license to multiple pending and issued patent applications. The U.S. composition of matter patent runs into 2027 in the U.S., 2026 in Europe and 2026 in Japan. UCLA is owed a low single digit royalty on sales, which will be covered by Astellas outside the U.S.

## Upcoming Milestones

| Milestone | Timing |
|---|---|
| Data from enzalutamide's 67-patient study in hormone naïve prostate cancer | Mid-2012 |
| U.S. approval and launch of enzalutamide for post-chemo CRPC | H2:12 |
| E.U. approval and launch of enzalutamide for post-chemo CRPC | 2013 |
| PREVAIL results in CRPC (pre-chemo) | 2013 |

Source: Cowen and Company

## Medivation Quarterly P&L Model ($MM)

| | Q1:11A | Q2:11A | Q3:11A | Q4:11A | 2011A | Q1:12A | Q2:12E | Q3:12E | Q4:12E | 2012E |
|---|---|---|---|---|---|---|---|---|---|---|
| MDV3100 50% of U.S. Gross Profits | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| MDV3100 ex-U.S. Royalties | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Milestones and License Fees | 14.7 | 15.8 | 14.9 | 14.9 | 60.4 | 36.8 | 41.5 | 26.4 | 35.9 | 140.6 |
| **Total Revenue** | **14.7** | **15.8** | **14.9** | **14.9** | **60.4** | **36.8** | **41.5** | **26.4** | **35.9** | **140.6** |
| COGS | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| R&D | 17.6 | 19.1 | 18.7 | 18.0 | 73.4 | 20.0 | 22.0 | 23.0 | 25.0 | 90.0 |
| SG&A | 6.2 | 7.0 | 7.7 | 9.0 | 29.9 | 15.7 | 18.0 | 20.0 | 25.0 | 78.7 |
| **Total Expenses** | **23.8** | **26.2** | **26.4** | **26.9** | **103.3** | **35.7** | **40.0** | **43.0** | **50.0** | **168.7** |
| **Operating Income/Loss** | **(9.1)** | **(10.4)** | **(11.5)** | **(12.0)** | **(42.9)** | **1.1** | **1.5** | **(16.6)** | **(14.1)** | **(28.1)** |
| Non-Operating Income | (0.3) | (0.1) | 0.1 | 0.1 | (0.2) | (0.7) | 0.1 | 0.3 | 0.3 | 0.1 |
| **Pre-tax Income/Loss** | **(9.4)** | **(10.5)** | **(11.4)** | **(11.9)** | **(43.2)** | **0.5** | **1.6** | **(16.3)** | **(13.8)** | **(28.0)** |
| *Tax rate (%)* | *NM* | *NM* | *NM* | *NM* | *NM* | *NM* | *NM* | *NM* | *NM* | *NM* |
| Provision for income taxes | (0.9) | (1.0) | (1.4) | (1.1) | (4.3) | 0.0 | (1.0) | (1.0) | (1.0) | (3.0) |
| **Net Income (Loss) From Operations** | **(8.5)** | **(9.5)** | **(10.0)** | **(10.9)** | **(38.8)** | **0.4** | **2.6** | **(15.3)** | **(12.8)** | **(25.0)** |
| **GAAP EPS** | **($0.24)** | **($0.27)** | **($0.29)** | **($0.31)** | **($1.11)** | **$0.01** | **$0.07** | **($0.42)** | **($0.35)** | **($0.66)** |
| Diluted Shares | 34.7 | 34.9 | 34.9 | 35.4 | 35.0 | 38.6 | 38.8 | 36.5 | 36.8 | 37.7 |

Source: Cowen and Company

## Medivation Annual P&L Model ($MM)

| | 2011A | 2012E | 2013E | 2014E | 2015E | 2016E |
|---|---|---|---|---|---|---|
| MDV3100 50% of U.S. Gross Profits | 0.0 | 0.0 | 127.5 | 247.5 | 360.0 | 455.0 |
| MDV3100 ex-U.S. Royalties | 0.0 | 0.0 | 7.5 | 45.0 | 99.0 | 160.0 |
| **Total Revenue** | **60.4** | **140.6** | **189.0** | **332.5** | **499.0** | **645.0** |
| COGS | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| R&D | 73.4 | 90.0 | 105.0 | 110.0 | 120.0 | 125.0 |
| SG&A | 29.9 | 78.7 | 110.0 | 110.0 | 165.0 | 180.0 |
| **Total Expenses** | **103.3** | **168.7** | **215.0** | **220.0** | **285.0** | **305.0** |
| **Operating Income/Loss** | **(42.9)** | **(28.1)** | **(26.0)** | **112.5** | **214.0** | **340.0** |
| **Pre-tax Income/Loss** | **(43.2)** | **(28.0)** | **(25.0)** | **114.0** | **215.5** | **342.0** |
| *Tax rate (%)* | *NM* | *NM* | *NM* | *NM* | *15%* | *35%* |
| **Net Income (Loss) From Operations** | **(38.8)** | **(25.0)** | **(21.0)** | **116.0** | **183.2** | **223.0** |
| **GAAP EPS** | **($1.11)** | **($0.66)** | **($0.55)** | **$2.64** | **$4.07** | **$4.85** |
| Diluted Shares | 35.0 | 37.7 | 38.0 | 44.0 | 45.0 | 46.0 |

Source: Cowen and Company

## Medivation R&D Pipeline

| Therapeutic Class/Product | Indication | P-C | I | II | III | FILING | MKT | Comments |
|---|---|---|---|---|---|---|---|---|
| **Oncology** | | | | | | | | |
| Enzalutamide | CRPC Post-Chemo | | | | | • | | AFFIRM interim results positive, trial halted; Filed in May 2012 |
| Enzalutamide | CRPC Pre-Chemo | | | | • | | | PREVAIL trial is enrolling |
| Enzalutamide | PRCA patients who have progressed on LNRH | | | • | | | | TERRAIN study, head-to-head comparison with bicalutamide. |
| Enzalutamide | PRCA patients with no prior hormonal therapy | | | • | | | | Hormone-Naïve study |
| Enzalutamide | Breast Cancer | • | | | | | | Encouraging pre-clinical results |
| **Central Nervous System** | | | | | | | | |
| Dimebon | Alzheimer's Disease | | | | | | | Failed CONNECTION Ph. III and CONCERT studies; Discontinued. |
| Dimebon | Huntington's Disease | | | | | | | Failed Ph. III HORIZON study; Discontinued. |
| **Total Drugs In Development** | | **1** | **0** | **2** | **1** | **1** | **0** | |

**San Francisco, CA** **Investor Relations Contact: Pat Machado (415) 829-4101**

Source: Medivation, Cowen and Company

between the responsible agency, the Fair Trade Commission, and the Ministry of International Trade and Industry, which is more inclined to favor exemptions from procompetitive rules. At the multinational level, the 1957 Treaty of Rome, which created the European Economic Communities, contained explicit competition policy provisions. They have for the most part been vigorously enforced.

### The Interweaving of Analysis and Policy

One can scarcely understand the rationale for such policy interventions as antitrust and regulation without comprehending how market performance is related to structure and conduct. The links between structure, conduct, and performance are in turn illuminated by careful study of important real-world policy cases. In this edition, therefore, we integrate our approach to what might otherwise be considered separate topics. Chapter 5 combines an analysis of merger motives and consequences with an evaluation of leading merger antitrust cases. After Chapters 6 through 8 lay an analytic foundation, Chapter 9 reviews antitrust policies toward price fixing and related restraints. Further foundation-building sets the stage for the analysis in Chapter 12 of alternative approaches to structural monopoly conditions, including both antitrust and some forms of regulation. Price discrimination behavior and its legal status are discussed together in Chapter 13. In Chapters 14 and 15, the complex policies governing pricing relationships between vertically linked (for example, seller and buyer) firms are investigated. No attempt is made to provide a systematic survey of traditional public regulation institutions, which merit the book-length treatment they receive elsewhere. However, several propositions central to understanding the logic of public regulation will be taken up as we develop a theory of how various types of markets function.

# 2 The Welfare Economics of Competition and Monopoly

Competition has long been viewed as a force that leads to an ideal solution of the economic performance problem, and monopoly has been condemned through much of recorded history for frustrating attainment of the competitive ideal. To Adam Smith, the vital principle underlying a market economy's successful functioning was the pursuit of individual self-interest, channelled and controlled by competition. As each individual strives to maximize the value of his own capital, said Smith, he

> . . . necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. . . . [H]e intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention.[1]

Smith's "invisible hand" is the set of market prices emerging in response to competitive forces. When these forces are thwarted by "the great engine of . . . monopoly," the tendency for resources to be allocated "as nearly as possible in the proportion which is most agreeable to the interests of the whole society" is frustrated.[2]

Much of Smith's detailed analysis is obsolete. Yet his arguments on the efficacy of free competition remain intact, a philosophical lodestar to nations relying upon a market system of economic organization. Economists have, to be sure, amended their view of competition since Smith's time, and they have developed more elegant models of how competitive markets do their job of allocating resources and distributing income. One objective of this chapter is to survey these modern views. In addition, we shall examine some of the qualifications and doubts that have led to the partial or complete rejection of Smith's gospel in many parts of the world.

## *Competition Defined*

We must begin by making clear what is meant by *competition* in economic analysis. Two broad conceptions, one emphasizing the conduct of sellers and buyers and the other emphasizing market structure, can be distinguished. Adam Smith's widely scattered comments, dealing with both conduct and structural features, typify the dominant strain of economic thought during the eighteenth and nineteenth centuries.[3] On the conduct side, Smith considered the essence of competition to be an *independent striving* for patronage by the various sellers in a market. The short-run structural prerequisites for competitive conduct were left ambiguous. Smith observed that independent action might emerge with only two sellers, but it was more likely (that is, collusion among the sellers was less likely) with

1. Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (New York: Modern Library edition, 1937), p. 423.
2. Smith, *Wealth of Nations*, pp. 594–595. See also pp. 61, 147, and 712.
3. For admirable surveys of the development of economic thought on the nature of competition, see George J. Stigler, "Perfect Competition, Historically Contemplated," *Journal of Political Economy*, vol. 65 (February 1957), pp. 1–17; J. M. Clark, *Competition as a Dynamic Process* (Washington: Brookings, 1961), Chapters 2 and 3; Paul J. McNulty, "A Note on the History of Perfect Competition," *Journal of Political Economy*, vol. 75, Part 1 (August 1967), pp. 395–399; and *idem*, "Economic Theory and the Meaning of Competition," *Quarterly Journal of Economics*, vol. 82 (November 1968), pp. 639–656.

twenty or more sellers.[4] Competition in Smith's schema also had a long-run dimension that could be satisfied, despite short-run aberrations, as long as it was possible for resources to move from industries in which their returns were low to those in which they could earn comparatively high returns. This in turn depended upon a structural condition: the absence of barriers to resource transfers. Recognizing that resources were often fairly immobile in the short run, Smith and his followers conceded that the full benefit of competitive market processes might be realized only in the long run.

As mathematical reasoning began to penetrate economics during the nineteenth century, a different, essentially structural concept of competition came to the forefront. In modern economic theory, a market is said to be competitive (or more precisely, purely competitive) when the number of firms selling a homogeneous commodity is so large, and each individual firm's share of the market is so small, that no individual firm finds itself able to influence appreciably the commodity's price by varying the quantity of output it sells. In mathematical jargon, price is a *parameter* to the competitive seller — it is determined by market forces and not subject to the individual seller's conscious control. The parametric character of price to the competitive firm is fundamentally a subjective phenomenon. If market demand curves are smooth and continuous, it is not strictly true that a small seller's output changes have *no* effect on the market price. They simply have such a minute effect that the influence is *imperceptible* to the seller, who can therefore act as if the effect were in fact zero.[5]

This technical definition of competition differs markedly from the usage adopted by businesspeople who, following Adam Smith's lead, are apt to view competition as a conscious striving against other business firms for patronage, perhaps on a price basis but possibly also (or alternatively) on nonprice grounds. Failure to recognize these implied semantic distinctions has often led to confusion in policy discussions. To keep such confusion at a minimum, we adopt the term "rivalry" to characterize much of the activity businesspeople commonly call "competition." The essence of rivalry is a striving for potentially incompatible positions (for example, if Firm A sells 100 units of output to Jones, Firm B cannot satisfy that part of Jones' demand) combined with a clear awareness by the parties involved that the positions they seek to attain may be incompatible. Under this dichotomy, it is possible for there to be vigorous rivalry that cannot be called pure competition. The jockeying for position in the automobile market among General Motors, Ford, and Honda is an obvious example. At the same time, there can be pure competition without rivalry. For instance, two Iowans growing corn on adjacent farms are pure competitors but not rivals in the sense implied here. Since the market for corn is so large relative to the two farmers' potential supply, it can readily absorb their offerings with scarcely a ripple in the Chicago Board of Trade price. Neither farmer can consider the neighbor's output decisions as having any adverse impact on his or her own economic position.

Violations of the principal structural preconditions for pure competition give rise to a rich variety of sellers' market types. For present purposes it suffices to identify the six most important types, using the two-way classification based upon the number of sellers and the nature of the product presented in Table 2.1. The distinction between homogeneity and differentiation in this classification hinges



**Table 2.1** Principal Seller's Market Structure Types

| | Number of Sellers | | |
|---|---|---|---|
| | One | A Few | Many |
| Homogeneous products | Pure monopoly | Homogeneous oligopoly | Pure competition |
| Differentiated products | Pure multiproduct monopoly | Differentiated oligopoly | Monopolistic competition |

on the degree of substitutability among competing sellers' products. Homogeneity prevails when, in the minds of buyers, products are perfect substitutes. Products are differentiated when, owing to differences in physical attributes, ancillary service, geographic location, information, and/or subjective image, one firm's products are clearly preferred by at least some buyers over rival products at a given price. The distinguishing trait of a differentiated product is the ability of its seller to raise the product's price without sacrificing its entire sales volume. Obviously, infinite gradations in the degree of product differentiation may exist, and it is difficult in practice to draw a precise line where homogeneity ends and differentiation begins. Similarly, although pure monopoly ends and oligopoly begins when the number of sellers rises from one to two, it is difficult to specify exactly where oligopoly shades into a competitive market structure. The key to the distinction is subjective — whether or not the sellers consider themselves conscious rivals in the sense defined earlier. If the sellers are sufficiently few in number to have each believe (a) that its economic fortunes are perceptibly influenced by the market actions of other individual firms, and (b) that those firms are in turn affected significantly by its own actions, then the market can be said to be oligopolistic.

Pure monopolists, oligopolists, and monopolistic competitors share a common characteristic: each recognizes that its output decisions have a perceptible influence on price, or in other words, each can increase the quantity of output it sells under given demand conditions only by reducing its price. All three types possess some degree of power over price, and so we say that they possess *monopoly power* or *market power*.

Homogeneity of the product and insignificant size of individual sellers and buyers relative to their market (that is, *atomistic* market structure) are sufficient conditions for the existence of pure competition, under which sellers possess no monopoly power. Several additional structural conditions are added to make competition in economic theory not only "pure" but "perfect."[6] The most important is the absence of barriers to the entry of new firms, combined with mobility of resources employed, or potentially employable, in a market. Conversely, significant

4. Smith, *Wealth of Nations,* p. 342.

5. This definition is given for the sellers' side of an industry. The definition of buyers' competition is symmetric. Pure competition exists among buyers when the number of entities buying a homogeneous product is so large, and each buyer's share of the market so small, that each buyer believes variations in the quantity it buys have an imperceptible effect on the market price. When some buyer can perceptibly influence price, *monopsony* is said to exist.

6. This distinction is essentially the one adopted by Edward H. Chamberlin in *The Theory of Monopolistic Competition* (Cambridge: Harvard University Press, 1933), Chapter 1.

entry barriers are the *sine qua non* of monopoly and oligopoly, for as we shall see in later chapters, sellers have little or no enduring power over price when entry barriers are nonexistent. A newer extension of this concept is pertinent when firms can enter a market and then *exit* easily, in the precise sense that their investments can be liquidated without loss more rapidly than producers within the market can react to the entering-and-exiting firms' decisions. When this exit condition is satisfied, markets are said to be *contestable* even if the conditions for pure competition are not met.[7]

Other conditions sometimes associated with perfect competition include perfect knowledge of present and future market conditions and continuous divisibility of inputs and outputs. These are less important, as well as less realistic, for their violation does not necessarily alter the main conclusions generated by the theoretical model of a purely and perfectly competitive market system's operation.

One final terminological point deserves mention, because it is a common source of confusion. The power over price possessed by a monopolist or oligopolist depends upon the firm's size *relative to* the market in which it is operating. It is entirely possible for a firm to be very small in absolute terms, but to have considerable monopoly power. The physician in an isolated one-doctor town is an excellent example. So was the Besser Manufacturing Company, which was found guilty in 1951 of illegally monopolizing the concrete block machinery industry, even though it employed only 465 persons at the time and had sales of less than $15 million.[8] On the other hand, a firm may be enormous in absolute terms, but possess little monopoly power in its principal markets. An example is the Sun (Oil) Company, which had sales of $8.7 billion in 1987 but accounted for only about 4 percent of U.S. petroleum refining and marketing. To postulate a 1-to-1 relationship between monopoly power and absolute size is like confusing pregnancy with obesity. Some superficial manifestations may be similar, but the underlying phenomena could hardly differ more.

## The Case for Competition

We proceed now to the principal questions on our agenda. Why is a competitive market system held in such high esteem by statesmen and economists alike? Why is competition the ideal in a market economy, and what is wrong with monopoly?

### Political Arguments

We begin with the political arguments, not merely because they are sufficiently transparent to be treated briefly, but also because when all is said and done, they, and not the economists' abstruse models, have tipped the balance of social consensus toward competition. One of the most important arguments is that the atomistic structure of buyers and sellers required for competition decentralizes and disperses power. The resource allocation and income distribution problem is solved through the almost mechanical interaction of supply and demand forces on the market, and not through the conscious exercise of power held in private hands (for example, under monopoly) or government hands (that is, under state enterprise or government regulation). Limiting the power of both government bodies and private individuals to make decisions that shape people's lives and fortunes was a fundamental goal of the men who wrote the U.S. Constitution, which in turn has served as a model for many other nations. As James Madison wrote (under the pseudonym Publius) in Federalist Paper No. 10, nothing was more important to a well-constructed union than avoiding the imposition on all citizens of measures favored by narrow factions.[9] Factions, continued Madison, arise most frequently from the unequal distribution of property, pitting the wishes of "a landed interest, a manufacturing interest, a mercantile interest, a moneyed interest, with many lesser interests" against the common good. The best way to avoid faction-dominated outcomes, said Madison, was to keep the individual factions so small and diverse that they would be "unable to concert and carry into effect schemes of oppression."

A closely related benefit is the fact that competitive market processes solve the economic problem *impersonally*, and not through the personal control of entrepreneurs and bureaucrats. There is nothing more galling than to have the achievement of some desired objective frustrated by the decisions of an identifiable individual or group. Who, on the other hand, can work up much outrage about a setback administered by the impersonal interplay of competitive market forces?

A third political merit of a competitive market is its freedom of opportunity. When the no-barriers-to-entry condition of perfect competition is satisfied, individuals are free to choose whatever trade or profession they prefer, limited only by their own talent and skill and by their ability to raise the (presumably modest) amount of capital required.

### The Efficiency of Competitive Markets

Admitting the salience of these political benefits, our main concern nonetheless will be with the economic case for competitive market processes. Figure 2.1(*b*) reviews the conventional textbook analysis of equilibrium in a competitive industry, and Figure 2.1(*a*) portrays it for a representative firm belonging to that industry. Suppose we begin observing the industry when the short-run industry supply curve is $S_1$, which embodies the horizontal summation of all member firms' marginal cost curves. The short-run market equilibrium price is $OP_1$, which is viewed as a parameter or "given" by our representative firm, so the firm's subjectively-perceived demand curve is a horizontal line at the level $OP_1$. The firm maximizes its profits by expanding output until marginal cost (MC) rises into equality with the price $OP_1$. It produces $OX_1$ units of output and earns economic profits — that is, profits above the minimum return required to call forth its capital investment — equal to the per-unit profit $GC_1$ times the number of units of output $OX_1$. Because economic profits are positive for the representative firm, this cannot be a long-run equilibrium position. New firms attracted by the profit lure will enter the industry, adding their new marginal cost functions to the industry's supply curve, and existing firms will expand their capacity, so the industry supply curve shifts to the right. Entry and expansion will continue, augmenting output and driving the price down, until price has fallen into equality with average total cost (ATC) for

7. See William J. Baumol, John C. Panzar, and Robert D. Willig, *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt Brace Jovanovich, 1982).

8. *U.S.* v. *Besser Mfg. Co.*, 96 F. Supp. 304 (1951), affirmed 343 U.S. 444 (1952).

9. *The Federalist Papers*, Mentor Book edition (New York: New American Library, 1961), pp. 77–84.

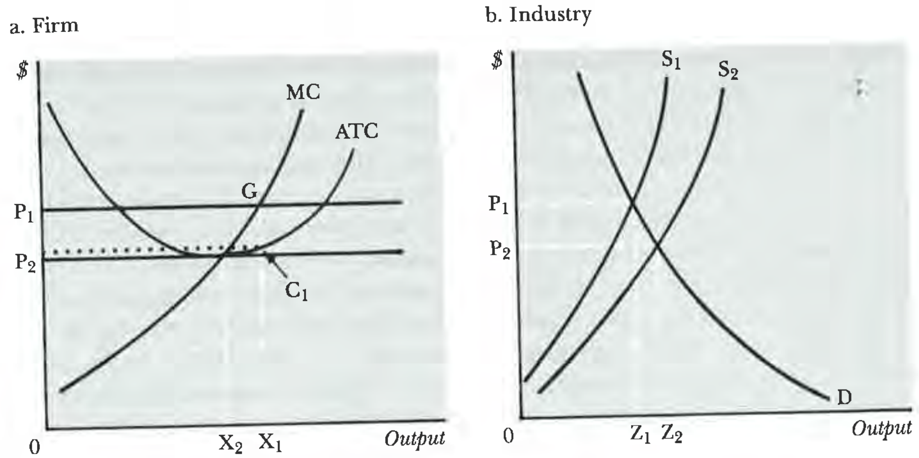



**Figure 2.1**
Equilibrium under Pure Competition

the representative firm.[10] In the figures shown, this zero-profit condition emerges with the short-run supply curve $S_2$, yielding the market price $OP_2$. The representative firm maximizes its profits by equating marginal cost with new price $OP_2$, barely covering its unit costs (including the minimum necessary return on its capital) at the output $OX_2$.

The long-run equilibrium state of a competitive industry has three general properties with important normative implications:

**a.** The cost of producing the last unit of output — the marginal cost — is equal to the price paid by consumers for that unit. This is a necessary condition for profit maximization, given the competitive firm's perception that price is unaffected by its output decisions. It implies efficiency of resource allocation in a sense to be explained momentarily.

**b.** With price equal to average total cost for the representative firm, economic (that is, supra-normal) profits are absent. Investors receive a return just sufficient to induce them to maintain their investment at the level required to produce the industry's output efficiently. Avoiding a surplus return to capital is considered desirable in terms of the equity of income distribution.

**c.** In long-run equilibrium, each firm is producing its output at the minimum point on its average total cost curve. Firms that fail to operate at the lowest unit cost will incur losses and be driven from the industry. Thus, resources are employed at maximum production efficiency under competition.

One further benefit is sometimes attributed to the working of competition, although with less logical compulsion. Because of the pressure of prices on costs, entrepreneurs may have especially strong incentives to seek and adopt cost-saving technological innovations. Indeed, if industry capacity is correctly geared to demand at all times, the *only* way competitive firms can earn positive economic profits is through innovative superiority. We might expect therefore that techno-



logical progress will be more rapid in competitive industries. However, doubts concerning the correctness of this hypothesis will be raised in a moment.

## The Inefficiency of Monopoly Pricing

Monopolists and monopolistic competitors differ from purely competitive firms in only one essential respect: They face a downward-sloping demand curve for their output. Given this, the firm with monopoly power knows that to sell an additional unit (or block) of output, it must reduce its price to the customer(s) for that unit; and if it is unable to practice price discrimination (as we shall generally assume, unless otherwise indicated),[11] the firm must also reduce the price to all customers who would have made their purchases even without the price reduction. The net addition to the nondiscriminating monopolist's revenue from selling one more unit of output, or its *marginal revenue*, is equal to the price paid by the marginal customer, minus the change in price required to secure the marginal customer's patronage multiplied by the number of units that would have been sold without the price reduction in question.[12] Except at prices so high as to choke off all demand, the monopolist always sacrifices something to gain the benefits of increased patronage: the higher price it could have extracted had it limited its sales to more eager customers. When demand functions are continuous and smooth, marginal revenue under monopoly is necessarily less than price for finite quantities sold. When the monopolist's demand function can be represented by a straight line, marginal revenue for any desired output is given by the ordinate of a straight line intersecting the demand curve where the latter intersects the vertical axis, and with twice the slope of the demand curve, as illustrated in Figures 2.2(*a*) and 2.2(*b*).[13] We will normally use straight-line demand curves in subsequent illustrations because they make it easier to get the geometry of their associated marginal revenue curves exactly right.

Now the profit-maximizing firm with monopoly power will expand its output only as long as the net addition to revenue from selling an additional unit (the marginal revenue) exceeds the addition to cost from producing that unit (the marginal cost). At the monopolist's profit-maximizing output, marginal revenue equals marginal cost. But with positive output, marginal revenue is less than price, and so the monopolist's price exceeds marginal cost. This equilibrium condition for firms with monopoly power differs from that of the competitive firm. For the competitor, price equals marginal cost; for the monopolist, price exceeds marginal cost. This difference has important implications to which we shall return in a moment.

10. We assume perfect imputation of all factor scarcity rents here. If the imputation process is imperfect, only the marginal firm — the firm just on the borderline between entering and not entering — will realize zero economic profits.

11. The logic of price discrimination will be explored in Chapter 13.

12. Generally, for the monopolist price is a function $P = f(Q)$ of the quantity Q sold. Total sales revenue $R = PQ$. Marginal revenue is the change in total revenue associated with a unit change in quantity sold, thus, $MR = dR/dQ = P + Q(dP/dQ)$. P in the MR expression is the price paid by marginal consumers; $dP/dQ$ is the change in price necessary to attract them (usually with a negative sign); and Q corresponds approximately to the quantity that would be sold without the price reduction.

13. Proof: Let the demand curve have the equation $P = a - bQ$, where Q is the quantity demanded. Total revenue $R = PQ = aQ - bQ^2$. Marginal revenue $dR/dQ = a - 2bQ$. At $Q = 0$, $P = MR$. The slope $(-2b)$ of the marginal revenue function is twice the slope $(-b)$ of the demand curve.

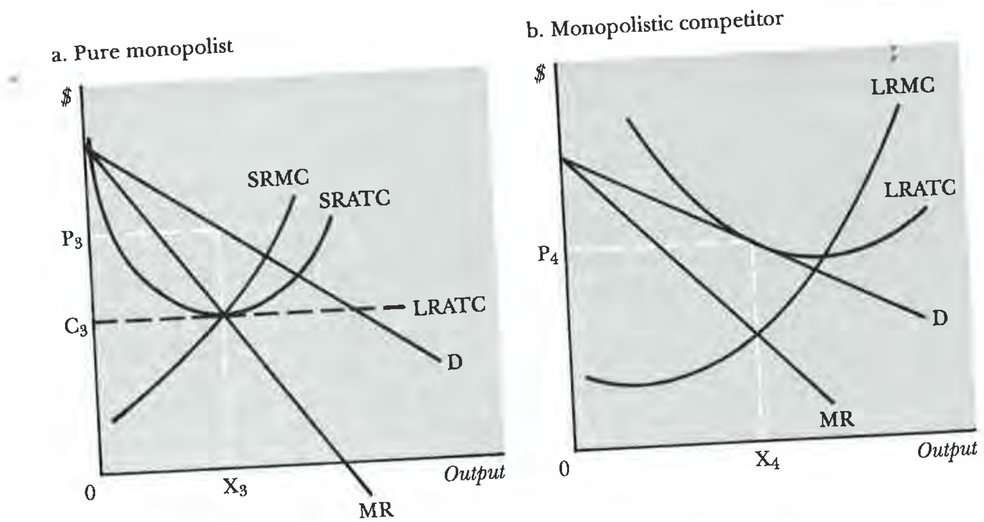



**Figure 2.2**
Equilibrium under Monopoly

The competitive enterprise earns zero economic profit in long-run equilibrium. Is the firm with monopoly power different? Perhaps, but not necessarily. Figure 2.2(*a*) illustrates one of the many possible cases in which positive monopoly profits are realized: specifically, the per-unit profit margin $P_3C_3$ times the number of units $OX_3$ sold. As long as entry into the monopolist's market is barred, there is no reason why this profitable equilibrium cannot continue indefinitely. Figure 2.2(*b*), on the other hand, illustrates the standard long-run equilibrium position of a monopolistic competitor.[14] The crucial distinguishing assumptions are that monopolistic competitors are small relative to the market for their general class of differentiated products and that entry into the market is free. Then, if positive economic profits are earned, new firms will squeeze into the industry, shifting the typical firm's demand curve to the left until, in long-run equilibrium, it is tangent to the firm's long-run unit cost function LRATC. The best option left for the firm then is to produce output $OX_4$, where marginal revenue equals marginal cost (as in any monopolistic situation) and the average revenue or price $OP_4$ is barely sufficient to cover unit cost. Thus, while firms with monopoly power *may* secure monopoly profits, they need not, especially under the plausible conditions of monopolistic competition.

We found earlier that in long-run equilibrium, the purely and perfectly competitive firm produces at minimum average total cost. Is this true also of the monopoly? Many textbooks imply that it is not, or that it will be true only by accident. Again consider Figure 2.2(*a*). It assumes that the monopolist operates under constant long-run cost conditions; that is, that plants (or plant complexes) designed to produce at high outputs give rise to roughly the same cost per unit as those designed to produce at low outputs. We shall see in Chapter 4 that many real-world cost functions exhibit this property over substantial output ranges. If so, the firm will invest in a plant complex characterized by the short-run cost function SRATC, with minimum short-run unit costs identical to the minimum long-



run cost $OC_3$ at the profit-maximizing output $OX_3$. We conclude that it is quite possible theoretically and empirically for monopolists, like their competitive brethren, to operate in such a way as to minimize average total cost. However, this is also not necessary. Figure 2.2(*b*) presents the most widely discussed exception. Since the monopolistic competitor in Chamberlinian equilibrium operates with its demand curve tangent to its LRATC curve, and since the demand curve is downward sloping, the LRATC curve must also have a negative slope at the equilibrium output. It follows that average cost is *not* minimized, for lower unit costs could be realized by expanding the firm's output. The monopolistic competitor does not do so because price (read off the demand curve) falls more rapidly than unit cost beyond the Chamberlinian equilibrium output, so that a higher output would spell negative profits.

In sum, firms with monopoly power may deviate from the zero-profit and minimum-cost conditions associated with purely and perfectly competitive equilibrium, but they need not do so.[15] The only distinction necessarily implied by the theories of pure competition and pure monopoly is that the monopolist's price exceeds marginal cost, while the competitor's price equals marginal cost. This seemingly technicality, so trivial at first glance, is the basis of the economist's most general condemnation of monopoly: it leads to an allocation of resources that is inefficient in the sense of failing to satisfy consumer wants as completely as possible.

To see this, we must think more deeply about the meaning of price as it affects the decisions of a consumer just on the margin between buying one more unit of a product and not buying it. A numerical illustration is especially helpful, so let us consider Figure 2.3(*a*). It assumes that the production of a composite commodity "manufactured goods" with the demand curve $D_M$ is monopolized. The industry is assumed (for simplicity) to produce under constant cost conditions, with long-run average total cost and marginal cost equal to $5.00 per unit at any output level chosen. The manufactured goods monopolist maximizes its profits by setting marginal cost equal to marginal revenue, which for the assumed cost and demand conditions requires producing 2 million units and setting a market-clearing price of about $9.70 per unit.

Now in setting this price, the monopolist chokes off the demand of consumers who would have been willing to purchase units (or additional units) at prices below $9.70. Consider some consumer who would purchase an extra unit at $9.60, but not at $9.70. We say that $9.60 is her *reservation price* -- the price just low enough to overcome her reservations about purchasing an extra unit. She buys the extra unit at $9.60 because it is worth that much to her; she refrains from purchasing at $9.70 because she considers the unit not worth the higher price. The consumer's reservation price for any incremental unit of consumption indicates in monetary terms how much that unit is worth to her. It is an index of the value of an extra unit

14. Chamberlin, *Monopolistic Competition*, supra note 6, Chapter 5.

15. But firms with monopoly power cannot normally be free of both deviations simultaneously. If they earn zero or negative profits, they will necessarily find it optimal to operate at higher than minimum average total cost. And (ignoring some dynamic complications to be introduced in Chapter 10) if they find it optimal to operate at minimum average cost, they will earn positive monopoly profits.

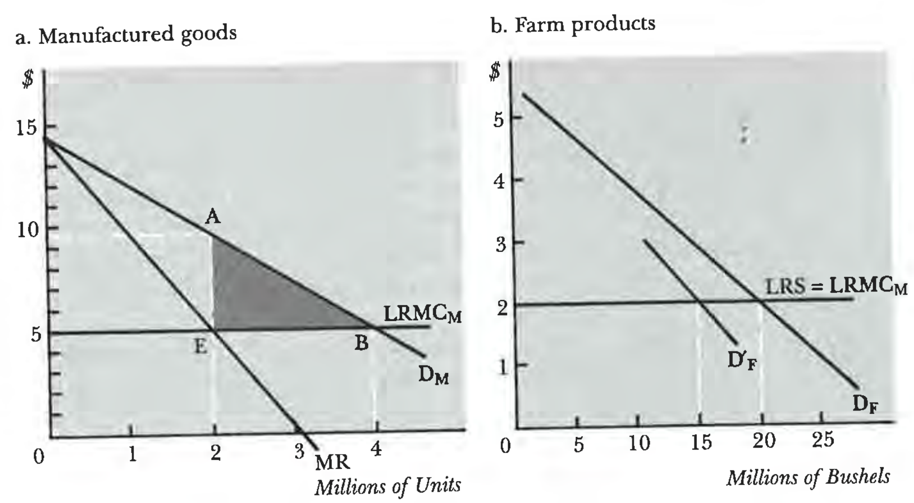



**Figure 2.3**
Resource Allocation with Competition and Monopoly

of consumption from the consumer's viewpoint and hence, in a social system honoring consumer sovereignty, from the viewpoint of society.[16]

The extra unit of manufactured goods required to satisfy the demand of this marginal consumer can be produced with resources costing $5.00. The marginal social value of the extra unit is $9.60. Marginal value exceeds marginal cost, so it would appear eminently worthwhile to produce that unit. The same can be said for all other units of manufactured goods that would be demanded at prices from $9.60 down to $5.00; their value to marginal consumers (read off the demand curve) exceeds their marginal cost, so they ought to be produced. They are not produced – that is, output is unduly restricted – because the monopolist is unwilling to sacrifice the profits it can secure by charging the higher price ($9.70) and selling fewer units.

For virtually all units of manufactured goods the monopolist does supply, the value to consumers of those units (measured as the demand curve ordinate for any given unit consumed) exceeds the monopolist's $9.70 price. On all but the two-millionth (that is, marginal) unit supplied and demanded, therefore, there is a surplus of value to consumers over the price paid. This is called *consumers' surplus*.[17] With a monopoly price of $9.70, the total consumers' surplus realized is defined by the triangular area in Figure 2.3(*a*) bounded by the vertical axis, a horizontal line at the $9.70 level, and the demand function from its vertical intercept to point *A*. By analogy, the monopolist's profit is called a producer's surplus. It is measured as quantity sold times unit profit, or in Figure 2.3(*a*) as the rectangular area between zero and two million units and between the $9.70 price and the $5.00 cost. If, contrary to its profit-maximizing instincts, the monopolist reduced its price to $5.00, its profit or producer's surplus would be converted into consumers' surplus on the two million units that would have been consumed even at the $9.70 price. This is essentially a redistribution of income. But in addition, two million more



units will be demanded, virtually all (given the demand function's slope) at reservation prices exceeding $5.00. Satisfying that demand would add consumers' surplus equal to the triangular area ABE between the demand curve and the $5.00 price line. At the $9.70 monopoly price, this surplus is realized neither by the monopolist nor by consumers. It is in effect lost, and therefore it is called a *dead-weight welfare loss*. It provides a first indication of the inefficiencies associated with monopolistic output restriction. Only through an expansion of manufactured good output to 4 million units, where price equals marginal cost, does some segment of society realize the surplus ABE.

Of course, the resources needed to expand production of manufactured goods must come from somewhere, which (assuming full employment) means that consumption of some other end product must be reduced. The problem of efficient resource allocation is a general equilibrium problem, involving the balance of all sectors in the economy. The analysis of general equilibrium takes us between Scylla and Charybdis: the rigorous models lack intuitive appeal and the intuitive models lack rigor. Because it is so important to understand the common sense of monopoly resource allocation, we opt for an intuitive approach here.[18]

Suppose the economy consists of only two industries, a monopolized manufactured goods industry and a competitive farm products industry. Figure 2.3 shows these two industries in general equilibrium, given the assumed demand functions $D_M$ and $D_F$, constant-cost production conditions in each industry, and the assumed market structures. (Note that the diagrams are not drawn to the same scale.) The output of manufactured goods is (as before) two million units per year; the output of farm products is twenty million bushels per year. Now suppose we could arrange to transfer resources valued at $100 from the farm products industry to the manufactured goods industry. Since the marginal cost of manufactured goods is $5.00, it will be possible to produce twenty extra units with these resources. To sell the extra production, the price will have to be reduced infinitesimally – for example, to $9.699. The value of the extra manufactured output from the viewpoint of marginal consumers barely willing to pay this new, lower price is 20 units × $9.699 = $194. However, the loss of farm products owing to the resource transfer must be weighed against this gain. Since the marginal cost of a bushel of farm produce is $2.00, the transfer of $100 in resources forces society to sacrifice fifty bushels of output. To choke off the demand for this output (at least, as a first approximation), the price of farm products must be raised slightly – for example, to $2.001 per bushel. This reduction in quantity demanded comes at the

16. Indirectly, the reservation price measures the utility of a marginal unit to consumers, for when utility-maximizing consumers are in equilibrium, the price of each commodity included in their market baskets equals the marginal utility of that commodity divided by the marginal utility of money.

17. The terminology here follows Alfred Marshall, *Principles of Economics*, 8th ed. (London: Macmillan, 1920), pp. 124 ff. and 467 ff.

18. A more rigorous exposition was included as an appendix to the first and second editions of this book. For a geometric approach using transformation functions but otherwise paralleling the argument here, see Robert Dorfman, *Prices and Markets* (Englewood Cliffs: Prentice-Hall, 1967) pp. 120–135. For other approaches, see William J. Baumol, *Welfare Economics and the Theory of the State* (rev. ed.; Bell, 1965), Chapters 1–6; Francis Bator, "The Simple Analytics of Welfare Maximization," *American Economic Review*, vol. 47 (March 1957), pp. 22–59; and Tjalling Koopmans, *Three Essays on the State of Economic Science* (New York: McGraw-Hill, 1957), pp. 4–104.

expense of consumers who were willing to buy an extra bushel at $2.00 but not at $2.001. Since they would rather abstain from consuming that bushel than pay $2.001, the value of the farm products foregone at the margin must be about $2.00 per bushel. The total value of farm products sacrificed as a result of the resource transfer is approximately $2.00 × 50 bushels = $100.00. Recapitulating, consumers have benefited from the resource reallocation by a net increase in output value of approximately $194 − $100 = $94.

If it is possible through such a reallocation to increase the value of the overall output bundle, it must follow that the value of output was not maximized in the original (monopoly) equilibrium. Too few resources were allocated to the monopolistic sector, and too many to the competitive sector, relative to the allocation that maximizes the value of output to society. Because it leads to an allocation of resources that fails to maximize the value of the overall output bundle, we say that monopoly misallocates resources, or that it leads to an inefficient allocation of resources.

The same point can be shown in terms of consumers' and producers' surplus. The value of the farm output transferred here is less than $2.001 per bushel. Since before the transfer farm output was sold at a price of $2.00 per bushel, consumers' surplus for the marginal output bundle must have been virtually nil. Because price equals marginal cost in a competitive industry, producers' surplus must also have been nil. But after reallocation, the $100 of resources yield additional manufactured goods valued at $194, so either producers' or consumers' surplus of approximately $94 must be generated by the transfer. If it is possible through reallocation to increase surplus in this way, it must follow that the sum of consumers' plus producers' surpluses was not maximized in the original (monopoly) equilibrium. Failure to maximize the value of the output bundle and failure to maximize the sum of consumers' plus producers' surpluses are conceptually identical manifestations of monopolistic resource misallocation.[19] Although we shall use the surplus concept again later, the output bundle maximization approach is somewhat more convenient for illustrating the relatively drastic changes implied by our two-sector example, so we emphasize the latter here.

If significant value gains can be had by reallocating $100 worth of resources, additional gains must come from carrying the process farther. Let us go all the way, breaking up the manufactured goods monopoly into numerous independent production units and eliminating any barriers to the entry of new resources. With the manufactured goods price initially well above the cost of production, resources will flow (or be drawn) into manufacturing, where the lure of positive profits beckons, and out of farming, where a zero-profit competitive equilibrium prevailed. It might seem that the price of farm products must rise above marginal cost as resources are pulled away and output contracts. This is true as a first approximation, but not as a second, for two reasons. First, a competitive industry simply cannot be in long-run equilibrium if price exceeds marginal cost. Something must give to restore the equality between price and cost. Second, as the price of manufactured goods is reduced to sell an expanding output, a substitution effect in favor of manufactured goods and adverse to farm products is induced. Assuming for the moment that the price of farm products hovers near the marginal cost of $2.00, the *ceteris paribus* (other prices equal) assumption on which the manufactured goods



sector's demand function was constructed remains valid, so there will be no shift in $D_M$. But because of the fall in the manufactured goods price there must be a leftward shift in the farm products demand curve — for example, to $D'_F$. Ignoring some complications temporarily, let us assume that $D'_F$ represents the final farm products demand curve after all adjustments have occurred, and $D_M$ the manufactured products demand curve. To be in final equilibrium, each competitive industry must have price equal to long-run marginal cost. This implies an output of four million units of manufactured goods with a price of $5.00 per unit and an output of 15 million bushels of farm products at a price of $2.00 per bushel. Resources originally valued at $10 million have been transferred from farm products to manufactured goods production, increasing the value of the aggregate output to society by a substantial amount — specifically, by the triangular area ABE in Figure 2.3(*a*).

Now let us attempt a further reallocation of resources. If we transfer resources valued at $100 from farm products to manufactured products, we sacrifice fifty bushels of farm output. These would have been bought by consumers with reservation prices of $2.00 or slightly higher, so that the value of farm output sacrificed is at least $100 and perhaps a bit more. We gain twenty extra units of manufactured goods saleable only at prices slightly less than $5.00, so the value of the additional manufactured output is less than $100. The value of the output gained is less than the value of the output sacrificed, so the transfer reduces the overall value of output to society. If we transfer $100 of resources in the opposite direction, we obtain fifty more bushels of farm output saleable only at prices slightly less than $2.00 for a gain of less than $100. We give up twenty manufactured units that would have been bought by consumers with reservation prices of $5.00 or higher, implying a value sacrifice exceeding $100. The value of the output added in the farm sector is less than the value of the manufactured goods sacrificed, and so this transfer also reduces the total value of output. Thus, a transfer of resources in either direction away from the competitive equilibrium allocation reduces output value. It follows that the value of output must have been at a (local) maximum in competitive equilibrium. Quite generally, when all sectors of an economy are in competitive equilibrium, with price equal to marginal cost for each firm, the total value of the output, measured in terms of each commodity's equilibrium price, is at a maximum. It is impossible to make any small resource reallocations that yield a higher output value. Because it maximizes output value in this sense, a fully competitive market system is said to allocate resources efficiently.[20] Conversely, a system shot through with monopoly elements is inefficient because it fails to do so. This, in a nutshell, is the heart of the economist's case for competition and against monopoly.

19. Cf. Arnold C. Harberger, "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay," *Journal of Economic Literature*, vol. 9 (September 1971), pp. 785–797; Robert D. Willig, "Consumer's Surplus without Apology," *American Economic Review*, vol. 66 (September 1976), pp. 589–597; and Jerry A. Hausman, "Exact Consumer's Surplus and Deadweight Loss," *American Economic Review*, vol. 71 (September 1981), pp. 662–676.

20. It is remarkable how acute Adam Smith's insight was on this point, when he observed that the individual producer in a competitive economy necessarily labors to render "the exchangeable value of the whole annual produce . . . as great as possible." *Wealth of Nations*, p. 423.

The analysis thus far has ignored a few complications. To describe the final equilibrium, we need a third approximation. One loose end is that the fall in manufactured goods prices increases the real income of consumers. This income effect will shift both sector demand curves to the right (unless one of the commodities happens to be an inferior good). Monopoly profits are also wiped out, freeing part of money transactions holdings to support those increases in demand.[21] The increased demand for products will be transmitted into increased demand for productive inputs, whose wages will be bid up.[22] This leads to upward shifts in the industry cost functions. With the present model, it is not possible to specify exactly where the final equilibrium will occur after all these effects have worked their way through the system, and therefore the shifted curves are not shown in Figure 2.3. One thing is certain, however. In each sector price will be equal to marginal cost for every producer, and so no further resource transfers can increase the ultimate aggregate value of output. Efficient resource allocation will have been achieved.

While this end result of eliminating monopoly is clearly desirable, another effect is more difficult to assess. Income will have been redistributed, with former monopoly profit recipients losing and other claimants (such as workers) gaining. Whether this is good or bad cannot be determined without a value judgment over which reasonable persons may disagree. There are at least two reasons for thinking that the competitive equilibrium may be preferred over the monopolistic one on equitable grounds, but the case is not airtight. First, society may object to monopoly profits as unearned gains and place ethical value on seeing them eliminated. One trouble with this argument is that the monopoly's original creators may already have reaped their gains by selling out their ownership interests at high capitalized values, leaving secondary and tertiary stock buyers, who before the monopoly's dissolution were receiving no more than a normal return on their money investment, with severe capital losses. Second, of the roughly half of all industrial enterprise common stocks held in 1983 by individual U.S. citizens, 72 percent of the stocks by value were held by families in the top 10 percent of the income distribution.[23] If all families of given size have similar income utility functions, the marginal utility of income must be higher for the multitudes who supply only their labor services than for the wealthy few with substantial monopoly shareholdings. A redistribution of income away from monopolists and toward labor suppliers will therefore add to the sum of utility for all citizens. Yet however appealing this may appear on intuitive grounds, there is no scientific way of making the interpersonal utility comparisons required to support it. Therefore, we tread warily when we say that competition is beneficial not only because it allocates resources efficiently, but also in terms of income distribution equity.

This completes the case based upon orthodox economic theory. Some other criticisms of monopoly can be mentioned more briefly. Monopolists' price-raising propensities may stimulate imports and worsen individual nations' terms of trade.[24] Lacking competitive pressure, firms may not exercise diligence in controlling their costs and therefore waste resources. As Adam Smith observed, "Monopoly . . . is a great enemy to good management."[25] For similar reasons, monopolists may display a lethargic attitude toward technological innovation, although contrary suggestions will be considered shortly. And, finally, enterprises



with monopoly power, or those that seek it, may devote excessive resources to advertising, legal stratagems, and the maintenance of excessive capacity; or they may use pricing systems that encourage inefficient geographic locations and unnecessarily high transportation costs. These alleged flaws, we shall find, may be even more serious than the resource misallocation problem. It is only for reasons of orderly presentation that we defer a more detailed examination until later.

## Qualifications and Doubts

General equilibrium analysis reveals the superiority of a competitive market system in solving society's resource allocation and income distribution problems under certain assumptions. But can we expect real-world economies to conform to the assumptions of the general equilibrium model? Might there be violations of assumptions stated explicitly or implicitly, or additional considerations not taken into account, that would cause us to modify our judgment? Several qualifications and doubts come to mind.

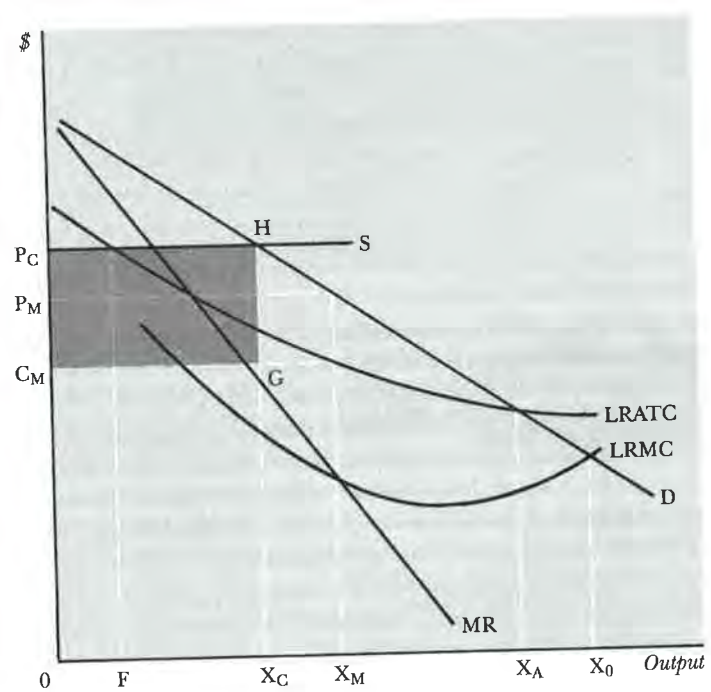
For one, the whole concept of efficient resource allocation is built upon the fundamental belief that the consumer is sovereign — that individual preferences are what count in the ledger of social values.[26] If, for example, consumers freely choosing in the market demonstrate that they would prefer at the margin to give up fifty bushels of grain to get an additional twenty hair shirts, we conclude that society is really better off because of the shift. Yet in practice our respect for consumer sovereignty is by no means universal — not, in any event, for infants, convicted criminals, dope addicts, the mentally ill, and others whose preferences cannot be trusted to generate rational choices. And in this age of widespread neuroses and psychoses, the line between rationality and irrationality is not all that easy to draw. One might even entertain doubts about the soundness of consumption decisions made by presumably normal, rational adults whose tastes (assumed in the standard theory of consumer behavior to be stable) have been remolded under a barrage of advertising messages. Further qualms intrude when we recognize that there are external diseconomies in consumption, for example, that the purchase of a new hair shirt by Mr. Willoughby may not only increase his utility, but simultaneously reduce the utility of envious neighbors. All this warns us that the theorems of welfare economics are erected upon sandy foundations. This does not

21. Under the previous (second) approximation, payments to all income claimants were \$50 million, compared to \$59.4 million when the manufacturing sector was monopolized. For monetary equilibrium to be restored, there must either be input and output price increases or a contraction in the money stock.
22. Unless supply functions are perfectly inelastic, the rise in real wages will also call forth increased input supplies, which in turn will permit a general expansion of output, ceteris paribus.
23. See F. M. Scherer, "Corporate Ownership and Control," in John R. Meyer and James M. Gustafson, eds., *The U.S. Business Corporation: An Institution in Transition* (Cambridge: Ballinger, 1988), pp. 46–48.
24. See Lawrence J. White, "Industrial Organization and International Trade: Some Theoretical Considerations," *American Economic Review,* vol. 64 (December 1974), pp. 1013–1020.
25. Smith, *Wealth of Nations,* p. 147.
26. On the consumer sovereignty question, see Tibor Scitovsky, "On the Principle of Consumers' Sovereignty," and Jerome Rothenberg, "Consumers' Sovereignty Revisited and the Hospitality of Freedom of Choice," both in the *American Economic Review,* vol. 52 (May 1962), pp. 262–290; Scitovsky, *The Joyless Economy* (New York: Oxford University Press, 1976); and the debate among G. L. Bach, Steven Hymer, Frank Roosevelt, Paul Sweezy, and Assar Lindbeck in the *Quarterly Journal of Economics,* vol. 86 (November 1972), pp. 635–636, 648–650, 661–664, and 672–674.

mean that their conclusions are wrong. The demonstration of a competitive system's allocative efficiency makes considerable sense even when complications related to advertising, ignorance, and the like are introduced. But blind faith is also uncalled for.

A second assault on the economist's conventional wisdom holds that under conditions known as "natural monopoly," firms can be large enough to realize all economies of scale only if they are monopolists. Under monopolistic (or in weaker cases, oligopolistic) organization, then, costs are lower than they would be if an industry includes many small-scale producers. The consequences are illustrated in Figure 2.4. The long-run average total cost curve available to large or small firms is LRATC, with associated marginal cost curve LRMC. If the industry were atomistically structured, each member firm would operate a small plant designed to produce OF units of output at a unit cost of $OP_C$. Then the long-run supply curve of the competitive industry would be $P_CS$, and total output would be $OX_C$. Alternatively, the monopolist would consider its marginal cost possibilities LRMC, choose to produce $OX_M$ units (where marginal cost equals marginal revenue) at an average total cost per unit of $OC_M$, and let the market clear at price $OP_M$. Output under monopoly, with its low costs, is higher than output under atomistic competition, with its high costs. Clearly, it cannot be said that consumers are poorly served by the monopoly, even though one might ideally prefer that the monopoly expand its output to $OX_A$ or even (as we shall see in a later chapter) $OX_O$. Furthermore, on the output $OX_C$ supplied by the small competitors, unit costs are higher than those of the monopoly by $P_C C_M$ per unit, or by area $P_CHGC_M$ in total. Excess cost, in this case caused by the too-small scale of production, is a dead-weight loss to society just as the failure to satisfy demand (by restricting output below the level at which price and marginal cost are equated) is. As Figure 2.4 is drawn, monopoly is unambiguously superior to competition, having lower costs and higher output, and hence smaller dead weight losses with respect to each. In a more complex case to be addressed in Chapter 5, monopoly may have lower costs but also supply less output, requiring a tradeoff between the lower dead weight losses associated with its cost structure and the higher dead weight losses stemming from its output restriction. Whether monopolists or oligopolists actually enjoy cost advantages as great as those shown in Figure 2.4, or indeed any at all, is an empirical question. We shall deal with it thoroughly in Chapter 4.

**Figure 2.4**
Natural Monopoly






Previously it was suggested that monopolists, sheltered from the stiff gale of competition, might be sluggish about developing and introducing technological innovations that reduce costs or enhance product quality. Yet some economists, following the late Joseph A. Schumpeter, have argued exactly the opposite: firms need protection from competition before they will bear the risks and costs of invention and innovation, and a monopoly affords an ideal platform for shooting at the rapidly and jerkily moving targets of new technology.[27] If this is true, then progress will be more rapid under monopoly than under competition. And, Schumpeter argued, it is the rate of technical progress, not the efficiency of resource allocation at any moment in time, that in the long run determines whether real incomes will be high or low. Suppose, as a hypothetical illustration, that real gross national product this year could be $5 trillion under pure and perfect competition, but that the misallocation caused by monopoly elements reduces it at every moment in time by 10 percent – that is, to $4.5 trillion this year. Suppose furthermore that a purely and perfectly competitive economy can sustain real growth of 3 percent per year, so that in five years the GNP under competition will be $5.81 trillion. How much more rapid must growth be under monopoly to catch up to the competitive potential in five years, starting from the lower monopoly base of $4.5 trillion? The answer is, a monopolistic economy growing at 5 percent will catch up in five years, and it will surpass the competitive system by an increasing margin from then on. Or if a monopolistic economy starting from a 10 percent static allocation handicap could grow at the rate of 3.5 percent per year, it would overtake the competitive system (grown to $9.1 trillion) in twenty years. If in fact growth is more rapid under monopoly than under competition, sooner or later the powerful leverage of compound interest will put the monopolistic system in the lead, despite any plausible starting disadvantage owing to static misallocation. The question is, of course, was Schumpeter right? Is technological progress really more rapid under monopoly? Orderly presentation demands that we leave the issue unsettled, returning to it in Chapter 17.

Since growth could conceivably bring overpopulation, depletion of natural resources, pollution, and generally reduced standards of living, some have

27. Joseph A. Schumpeter, *Capitalism, Socialism and Democracy* (New York: Harper, 1942), especially pp. 88 and 103.

questioned whether a high rate of economic growth is desirable.[28] We cannot resolve that value conflict here, but two observations are warranted. First, for given levels of population, available resources will be stretched farther and real income per capita will almost surely be higher with more rather than less (nonmilitary) technical progress. Second, the logic of monopoly resource misallocation has special twists when applied to natural resources available only in fixed amounts. In certain cases (for example, with zero extraction cost and unchanging demand elasticities over time), the profit-maximizing prices, and hence rates of resource depletion, are the same under monopoly as under competition.[29] In other more plausible cases, monopolies tend to charge higher prices than a competitive industry in the early years, but this means that there will be more of the resource, and hence lower prices, in later years.[30] Thus, if one believes that society is excessively myopic in its rate of natural resource use (which is tantamount to rejecting market-determined interest rates as a guide to dynamic resource allocation), the conservationist bias of monopolists may be applauded. To be sure, the same conservationist result might be achieved with less objectionable income distribution implications through appropriate public policy interventions, for example, through taxes or government-owned land leasing decisions.

It is also possible that monopolistic industry organization might be more conducive to the macroeconomic stability of employment, listed in Chapter 1 as an important performance goal. The hair-trigger price adjustments of purely and perfectly competitive markets could intensify tendencies toward instability, making it more difficult to combat cyclical unemployment through fiscal and monetary measures. The issue is too important to be left to the macroeconomists, who in any event have made little progress addressing it. Yet boundaries must be drawn, and we choose to leave the question for another forum.[31]

The discussion of allocative efficiency thus far has emphasized the monopoly and monopolistic competition cases, deliberately ignoring oligopolistic market structures. We now ask, how much competition is necessary to bring prices into rough equality with marginal cost? Will rivalry among the few, or oligopoly, suffice? This turns out to be an extraordinarily difficult question, for the theory of oligopoly pricing does not yield the neat, confident generalizations derivable under the theories of pure monopoly or monopolistic competition. We shall spend several chapters exploring the theory and evidence before answers can be ventured. In a similar vein is Professor Galbraith's contention that countervailing power — the power of a few large buyers dealing with monopolistic sellers — offers an effective surrogate for competition.[32] Power on the buyer's side will be a focus of Chapter 14.

Particularly important qualifications are required by the conditions associated with monopolistic competition. Because its product is differentiated physically or through other distinctive attributes, the monopolistic competitor faces a downward-sloping demand curve. It therefore seeks an equilibrium in which price exceeds marginal cost, and in addition, production occurs at less than minimum average total cost. But the consumer gets something in exchange: greater variety in the available bundle of goods and services. Again, a tradeoff is required. And as Professor Chamberlin concluded in his pioneering work, "Differences in tastes, desires, incomes, and locations of buyers, and differences in the uses which they



make of commodities all indicate the need for variety and the necessity of substituting for the concept of a 'competitive ideal,' an ideal involving both monopoly and competition."[33] This tradeoff problem will be addressed in Chapter 16. We shall see that under monopolistic competition, too much variety may emerge under certain identifiable conditions and too little under others.

## The Problem of Second Best

There are many reasons why, in the real world, it is impossible or undesirable to satisfy all the assumptions of the purely competitive general equilibrium model. Given that competition cannot be pure and perfect in all sectors, what policy should be pursued toward the remaining sectors? Is the second-best solution one that encourages maximum conformity to the rules of competition, whenever and wherever possible? The answer suggested by the theory of second best is quite possibly no, but it is difficult to say, because the answer depends upon complex circumstances peculiar to each case.

To provide a preliminary perspective, we return to Figure 2.3. We assume an economy consisting of two sectors, manufactured goods and farm products. The manufactured goods sector is monopolized; there is misallocation of resources, and we want to improve matters. Suppose, however, that there is no feasible way to break up the monopoly, for example, because of scale economies or a reticent Supreme Court. If something is to be done, it must be done in the farm products sector. What to do? Let us, out of desperation, organize the farmers into a monopoly. The farm price will be raised. This will set off a chain of repercussions. Owing to the change in relative prices, a substitution effect will shift the manufactured goods demand curve to the right, drawing resources out of the farming sector. Income effects may induce further shifts and price changes; monopoly profits will rise; wages of some productive inputs will probably fall, shifting marginal cost curves downward; and so on. Suppose, after all the necessary adjustments have been made, the economy settles down into the equilibrium illustrated in Figure 2.5. The price of manufactured goods is \$7.00, with 4.25 million units supplied at a cost of \$4.00 per unit; and the price of farm products is \$2.80, with 11 million bushels supplied at a cost of \$1.60 per unit.[34]

28. For various views see E. J. Mishan, *Technology and Growth: The Price We Pay* (New York: Praeger, 1969); D. H. Meadows et al., *The Limits to Growth* (New York: Universe, 1972); and Julian Simon and Herman Kahn, eds., *The Resourceful Earth: A Response to Global 2000* (Oxford: Blackwell, 1984).

29. See Joseph E. Stiglitz, "Monopoly and the Rate of Extraction of Exhaustible Resources," *American Economic Review*, vol. 66 (September 1976), pp. 655–661; the critiques by Tracy R. Lewis et al. and Gordon Tullock, *American Economic Review*, vol. 69 (March 1979), pp. 227–233; and Joseph Stiglitz and Partha Dasgupta, "Market Structure and Resource Depletion," *Journal of Economic Theory*, vol. 28 (October 1982), pp. 128–164.

30. For simulation analyses on petroleum, bauxite, and copper, see Robert S. Pindyck, "Gains to Producers from the Cartelization of Exhaustible Resources," *Review of Economics and Statistics*, vol. 60 (May 1978), pp. 238–251.

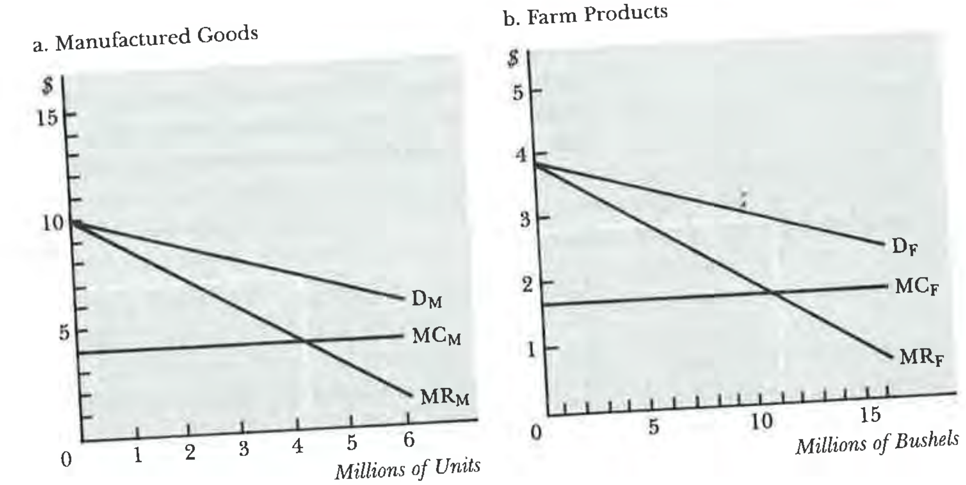
31. The first two editions of this text devoted a full chapter to the questions, but the material appears to have been little read. The (rare?) interested reader is referred to the 1980 edition.

32. John Kenneth Galbraith, *American Capitalism: The Concept of Countervailing Power* (Boston: Houghton Mifflin, 1952), Chapter 9.

33. Chamberlin, *Monopolistic Competition*, supra note 6, pp. 214–215.

34. Marginal costs have fallen by 20 percent, reflecting a commensurate decline in input wages. The diagrams assume a 13 percent average decline in the quantity of inputs supplied.

**Figure 2.5**
Resource Allocation in a World of Monopolies




Now let us try, as before, to increase the aggregate value of the economy's output by further resource transfers. Suppose we transfer resources valued at $160 out of the farm monopoly into manufacturing. This entails a sacrifice of one hundred bushels of farm produce (since the marginal cost is $1.60) and a gain of forty manufactured good units. The additional manufactured goods can be sold only to consumers with reservation prices below $7.00, and so the value gain must be slightly less than $280. So far, so good. But the farm produce given up has a value to consumers of $2.80 or slightly more per unit, implying a total value sacrifice exceeding $280. The value sacrifice exceeds the gain, and hence the transfer is not worthwhile. It can be seen by similar reasoning that a transfer of resources from manufacturing into farming will reduce the value of output slightly. It follows that the aggregate value of output must be at a (local) maximum in the equilibrium attained by monopolizing both sectors! By abandoning our effort to maintain a world of competition and moving instead to a world of monopolies, we have secured an equilibrium with output value implications closely resembling those of competitive equilibrium.[35]

This happens to be a special case. The data underlying Figure 2.5 were rigged so that the ratio of the monopoly equilibrium price to marginal cost in the manufactured goods sector (7 to 4) is the same as the price/marginal cost ratio in farming (2.80 to 1.60). Or to put the point in a form that will be more convenient in later chapters, prices are marked up above marginal cost so that the monopoly equilibrium *price – cost margins*, defined as (P – MC)/P, in the two industries are identical:

$$\frac{P - MC}{P} = \frac{7 - 4}{7} = \frac{2.80 - 1.60}{2.80} = 0.4286 \quad (2.1)$$



A simple but precise rule of thumb for profit maximization under monopoly is for the monopolist to set its price – cost margin equal to the inverse of its price elasticity of demand:[36]

$$PCM = \frac{P - MC}{P} = \frac{1}{e} \quad (2.2)$$

This in turn implies that equal price – cost margins will be set voluntarily by monopolists if, and only if, each firm encounters the same price elasticity of demand in the neighborhood of its equilibrium output. Quite generally, when the supply of production inputs is fixed, and when all producers sell their output directly and only to consumers, a completely efficient (first-best) allocation of resources can be achieved by equating price/marginal cost ratios across all industries. This *rule of proportionality* was proposed by some economists during the 1930s as a possible solution to the monopoly problem, to be enforced by direct government intervention in pricing decisions.[37]

Intervention would in fact be required, since the odds against every producer facing the same demand elasticity values, and hence voluntarily choosing the same price – cost margins, are astronomical. Moreover, enforcing the equal (P – MC)/P rule through public authority will not lead to an efficient allocation of resources under certain highly probable conditions. For one, some goods or services are almost necessarily sold under competitive conditions, frustrating the attainment of a "world of monopolies" solution. The most important example is leisure. The price of leisure to the typical worker-consumer is the opportunity cost he or she incurs by not supplying additional labor services – that is, the wage for additional hours of work. For firms as employers, the marginal cost of labor is the wage paid, unless the labor market is monopsonistic, in which case it exceeds the wage paid. For the world of monopolies solution to be achieved, the price of leisure must be raised above the marginal cost of labor services to firms in the same proportion as all other commodity prices exceed their marginal costs. It is extremely difficult to find a practical way of doing this; direct per-hour subsidy payments by the government to each and every worker obviously fail to meet the practicality test. But unless the price of leisure is raised relative to the wage paid by firms, or unless all workers' labor supply curves are completely inelastic, workers will consume too much leisure and supply too little labor relative to the quantities required for an efficient allocation of resources. Thus, labor inputs, and hence the output bundle available for consumption, will be smaller under a world of monopolies approach to product pricing than it would be if all products were supplied under competitive conditions.

35. Cf. Joan Robinson, *The Economics of Imperfect Competition* (London: Macmillan, 1934), Chapter 27.

36. Proof: From note 12 above, marginal revenue MR = P[1 + (dP/dQ)(Q/P + Q(dP/dQ). This can be manipulated to P[1 + (dP/dQ)(Q/P)]. Price elasticity of demand e is defined as –[(dQ/dP)(P/Q)] = –[1/(dP/dQ)(Q/P)]. Substituting this last expression into the expression for MR, we obtain MR = P(1 – 1/e) = P – P/e. With MC = MR, we have MC = P – P/e, which rearranges to (P – MC)/P = 1/e.

37. See R. F. Kahn, "Some Notes on Ideal Output," *Economic Journal*, vol. 45 (March 1935), pp. 1–35; and Abba P. Lerner, "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies*, vol. 1 (June 1934), pp. 157–175.

Another complication stems from the elaborate vertical and horizontal interrelationships characterizing any modern economy. Many products are not sold exclusively to final consumers; they are also used in whole or in part as intermediates by other firms. Consider the case of common salt, and assume (merely to simplify matters) that the only scarce basic resource is labor, which receives an equilibrium wage of $5.00 per hour. Suppose that a hundredweight of salt can be produced using one hour of labor, and that monopolistic salt producers sell their product at $10 per hundredweight, so that the ratio of price to marginal cost is 2. Suppose in addition that the production of a barrel of pickles requires one hundredweight of salt (for which $10 is paid) and one hour of labor (for which the pickle-maker pays $5.00). The combined marginal cost, from the pickle producer's perspective, is $15 per barrel. If the rule of proportionality were enforced, the pickles will be sold at a price twice their marginal cost, that is, at $30 per barrel. But since both salt and pickles enter directly into the market baskets of consumers, a distortion emerges. Two hours of labor will produce salt valued at $20 by the marginal consumer or pickles valued at $30. A reallocation of resources (labor) from the production of salt for final consumption to the production of pickles (requiring, of course, more intermediate salt) will raise the aggregate value of output. Thus, applying the rule of proportionality fails to maximize the value of output when products serve both as intermediates and final consumption goods. Analogous difficulties arise when a particular intermediate good enters into the various end products' manufacturing processes at different stages, or when the intermediate good accounts for varying proportions of the total cost of different end products.[38]

These and similar problems are assaulted in the general theory of second best.[39] The starting point is recognition that one or more of the conditions necessary for a first-best optimum simply cannot be satisfied, for example, because some sectors of the economy are unavoidably monopolistic and others competitive. Violations of the first-best optimum conditions are introduced as constraints upon the problem of maximizing the desired criterion function — that is, of maximizing the aggregate value of output or the sum of consumers' plus producers' surpluses. The second-best problem is to find new decision rules for the sectors a policy-maker *can* control, given the imposed constraints.

The typical result is a set of formidably complex decision rules in place of the simple "price equals marginal cost" conditions customary in first-best problems. The only positive generalization reached by Lipsey and Lancaster in their pioneering article was that, when some first-best optimum condition could not be attained, it was no longer desirable to fulfill the other first-best optimum conditions. An important contribution by Abram Bergson provided insight into the directions a second-best policy might take under certain circumstances.[40] He assumes a competitive labor supply sector, one or more uncontrollable industries that elevate price above marginal cost, and one or more industries controlled by the government in which prices are to be set to achieve an economywide second-best optimum. When the government-controlled output is a substitute in consumption for the output of uncontrolled and distorted industries, the second-best solution typically entails raising the controlled output's price above marginal cost.



The magnitude of the elevation — that is, the ratio of the controlled output's price to marginal cost — turns out to be the weighted harmonic mean of price/cost ratios in the uncontrolled sectors (including labor supply), with the weights being the equilibrium quantity of each uncontrolled industry's output times its elasticity with respect to the controlled output's price/cost ratio. The greater the price/cost distortion for any given uncontrolled industry, the more important that industry is in the economy, and the more that industry's output expands as the substitute-controlled output's price is raised,[41] the higher will be the second-best controlled price/cost markup, ceteris paribus. When the controlled output is a complement in consumption to the uncontrolled output, the controlled output's price may be depressed below its marginal cost in the second-best solution. In a world of mixed substitutes and complements, the direction of the second-best price adjustment depends upon the magnitudes of the uncontrolled industries' price distortions, outputs, and demand cross elasticities.

This is, needless to say, a formidably tough assignment. No central authority could conceivably obtain the masses of information on demand elasticities, cross elasticities, cost functions, and prices needed to devise fully articulated second-best pricing strategies for even a single major industry richly interconnected with other sectors. Even if the data were available, it is doubtful whether second-best actions in that industry alone would do much to offset distortions pervading the rest of the economy. Because it indicates that maintaining competitive pricing whenever possible is not necessarily optimal, but offers little clear guidance toward improved policies in the absence of information seldom if ever attainable, the theory of second best is a counsel of despair.

A "third-best" approach might be to choose among alternative *general* policies, trying to adopt the policy that *on average* has the most favorable resource allocation implications. In the framework of our present concern, the main alternatives boil down to letting monopoly increase in sectors not already monopolized, versus attempting to enforce as much competition as one can consistent with economies of scale, product differentiation, and so forth. When the issue is put this way, the case for competition gains renewed appeal. On the positive side, if one has little or no prior information concerning the direction in which second-best optima lie, eliminating avoidable monopoly power is about as likely to improve welfare as is encouraging new monopoly distortions where none existed previously. And on the negative side, it is easy for a policy that is permissive toward monopoly to get out

**38.** See Lionel McKenzie, "Ideal Output and the Interdependence of Firms," *Economic Journal*, vol. 61 (December 1951), pp. 785–803; and Richard E. Just and Darrell L. Hueth, "Welfare Measures in a Multimarket Framework," *American Economic Review*, vol. 69 (December 1979), pp. 947–954.

**39.** The standard reference is R. G. Lipsey and Kelvin Lancaster, "The General Theory of Second Best," *Review of Economic Studies*, vol. 24, no. 1 (1956), pp. 11–32. An almost simultaneous and similar formulation is Marcel Boiteux, "Sur le gestion des monopoles publics astreints a l'equilibre budgetaire," *Econometrica*, vol. 24 (January 1956), pp. 22–40. For a generalization, see Edward Foster and Hugo Sonnenschein, "Price Distortion and Economic Welfare," *Econometrica*, vol. 38 (March 1970), pp. 281–297. A two-commodity, one-input proof of the main propositions here is contained in the Appendix to Chapter 2 of the 1980 edition of this text, pp. 599–600.

**40.** Abram Bergson, "Optimal Pricing for a Public Enterprise," *Quarterly Journal of Economics*, vol. 86 (November 1972), pp. 519–544.

**41.** This last relationship is probably a special case stemming from Bergson's assumption that price/cost markups in the uncontrolled sector are not influenced by changes in the controlled sectors' price/cost markups.

of hand. Too much monopoly distortion may emerge in formerly competitive sectors, especially in view of the fact that some important sectors (such as the labor-leisure market) will remain competitive. Given the difficulties, a generally procompetitive policy seems more apt to be superior in a third-best sense.[42]

Nevertheless, one might conclude that the whole question of allocative efficiency is so confused and uncertain, once second-best considerations are introduced, that policymakers would be well advised to give up trying to achieve the best possible allocation of resources. Rather, heavier weight might be accorded other criteria such as equity of income distribution, compatibility with political beliefs, conduciveness to production efficiency, and speed of technological progress. It is worth repeating that if this view is adopted, a procompetitive policy is likely to be favored under the first two criteria. Judgment is reserved on the production efficiency and progress issues until the relevant evidence can be examined more carefully.

## *Doubts Concerning the Profit Maximization Hypothesis*

The conclusions of economic theory concerning the allocation of resources and distribution of income under competition and monopoly are based, among other things, upon the assumption that consumers maximize their subjective satisfaction and firms maximize profits. However, the profit maximization assumption has been challenged vigorously on several fronts.[43] The argument, in brief, is that profit maximization is at best unappealing and at worst meaningless to business decision makers operating in an environment of uncertainty, organizational complexity, and conflicting goals. Since these changes could require modifications in performance predictions based upon orthodox price theory and in judgments about the desirability of competition, we must pay them careful heed.

### The Effects of Uncertainty

Nearly all the interesting decisions made by business firms require predictions about uncertain future events. Decision makers simply cannot know precisely how strong and how elastic demand will be in the next period, let alone ten years hence, or how far labor unions will carry their struggle for higher wages in forthcoming negotiations, or how rival sellers will react to a price change, or how the long-term bond rate will move in coming years. How should they behave in the face of such uncertainties? Economic theory usually assumes that managers formulate definite expectations about the future values of relevant variables and then plug the expected values into their profit-maximizing decision rules. The expectations presumably include at least an estimate of the most likely, or best-guess, value and perhaps also some notion about the probability of more pessimistic and optimistic outcomes. This is already assuming a lot. Critics have noted that many businesspeople are poorly informed about business conditions in general, know almost nothing about the axioms of probability,[44] and understand only crudely the logic of profit maximization, that is, what variables (such as marginal cost and marginal revenue) must be taken into account and how they must be related. It is scarcely realistic to expect that profit maximization will sprout magically from such barren soil, the skeptics continue. In a famous defense of orthodoxy, Fritz Machlup countered that business managers have an intuitive understanding of



what is required to maximize profits, even though they cannot articulate rules matching the economist's price-equals-marginal-cost condition, just as automobile drivers who have never taken a course in differential equations can intuitively solve the problem of passing another car on a two-lane highway. He went on to stress a subjective interpretation of the variables that are manipulated by entrepreneurs:

> It should hardly be necessary to mention that all the relevant magnitudes involved — costs, revenue, profit — are subjective — that is, perceived or fancied by the men whose decisions or actions are to be explained . . . rather than "objective.". . . Marginal analysis of the firm should not be understood to imply anything but subjective estimates, guesses and hunches.[45]

This defense comes close to saying that whatever managers choose to do can be called profit maximization, however remotely it resembles the policies an omniscient maximizer would set. Carried so far, the theory of profit maximization becomes little more than tautology.

Yet even if we assume a close correspondence between business expectations and objective reality, further dilemmas appear. Imagine a decision maker weighing two alternative policies, one offering a best-guess profit expectation of $1 million with a 5 percent chance of bankrupting the firm (whose net worth is currently $4 million), the other an expected profit of $2 million with a 15 percent chance of disaster. Which is the rational choice? It is impossible to say without further information on the attitudes of the firm's owners toward increases in wealth versus total loss of their equity.

Further complications are posed by the interactions among uncertainty, risk aversion, and the decision maker's time horizon. Long-run profit maximization implies maximizing the discounted present value of an enterprise's current and future profit stream. But at what rate shall future profits be discounted? Finance texts advise that the discount rate be derived from the firm's weighted average cost of capital, with the cost of each financing mode evaluated at the rate currently prevailing in securities markets. Yet market rates fluctuate widely from time to

42. For similar arguments, see L. Athanasiou, "Some Notes on the Theory of Second Best," *Oxford Economic Papers,* vol. 18 (March 1966), pp. 83–87; and Y. K. Ng, "Towards a Theory of Third Best," *Public Finance,* vol. 32, no. 1 (1977), pp. 1–15.

43. The literature is enormous. Seminal contributions include Fritz Machlup, "Marginal Analysis and Empirical Research," *American Economic Review,* vol. 36 (September 1946), pp. 519–554; Armen A. Alchian, "Uncertainty, Evolution, and Economic Theory," *Journal of Political Economy,* vol. 58 (June 1950), pp. 211–221; Andreas G. Papandreou, "Some Basic Problems in the Theory of the Firm," in B. F. Haley, ed., *A Survey of Contemporary Economics,* vol. 2 (Homewood: Irwin, 1952), pp. 183–222; Herbert Simon, "Theories of Decision-Making in Economics and Behavioral Science," *American Economic Review,* vol. 49 (June 1959), pp. 253–283; Robin Marris, "A Model of the 'Managerial' Enterprise," *Quarterly Journal of Economics,* vol. 77 (May 1963), pp. 185–209; R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm* (Englewood Cliffs: Prentice-Hall, 1963); Oliver E. Williamson, *The Economics of Discretionary Behavior* (Englewood Cliffs: Prentice-Hall, 1964); R. J. Monsen and Anthony Downs, "A Theory of Large Managerial Firms," *Journal of Political Economy,* vol. 73 (June 1965), pp. 221–236; Michael C. Jensen and William H. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure," *Journal of Financial Economics,* vol. 3 (1976), pp. 305–360; Herbert Simon's Nobel Prize address, "Rational Decision Making in Business Organizations," *American Economic Review,* vol. 69 (September 1979), pp. 493–513; Richard R. Nelson and Sidney G. Winter, *An Evolutionary Theory of Economic Change* (Cambridge: Harvard University Press, 1982); and Harvey Leibenstein, *Inside the Firm: The Inefficiencies of Hierarchy* (Cambridge: Harvard University Press, 1987).

44. For evidence, see Mark J. Machina, "Choice under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives,* vol. 1 (Summer 1987), pp. 121–154.

45. "Marginal Analysis and Empirical Research," supra note 43, pp. 521–522.

time.[46] What if management is convinced that the current market rate is too low or too high and will soon be overtaken by events? Should expectations of future rates be substituted? How far into the future? How these questions are answered — and unanimous agreement is lacking even in treatises on managerial economics — can have a substantial impact on business firms' conduct.

Consider, for example, the problem of a firm with monopoly power deciding upon its pricing strategy. One possibility is to price so as to realize the highest possible profits today, regardless of future consequences. Alternatively, the firm may choose not to exploit its monopoly power fully today, hoping that a low current price will cement customer loyalties and deter the emergence of new competitors, increasing the probability that the firm will weather whatever storms the uncertain future will bring. The "right" choice in this and more complex dynamic pricing problems, to be explored in Chapter 10, is far from obvious.

That managerial goals can differ strikingly in different market environments is shown by a comparative survey of top managers in 1,031 Japanese and 1,000 U.S. industrial corporations during the early 1980s.[47] The managers were asked how important eight diverse goals were to their organizations. The average rankings for the two panels were as follows:

| Goal | American Managers | Japanese Managers |
|---|---|---|
| Return on investment | 1 | 3 |
| Higher stock prices | 2 | 8 |
| Increased market share | 3 | 2 |
| Improving products and introducing new products | 4 | 1 |
| Streamlining production and distribution systems | 5 | 4 |
| High net worth ratio | 6 | 5 |
| Improvement of social image | 7 | 6 |
| Improving working conditions | 8 | 7 |

Achieving high returns on investment and higher stock prices — arguably, indicia of an emphasis on profit maximization — were ranked much higher by the American managers than by the Japanese. Yet it is far from clear that the Japanese were neglecting profits in an equally meaningful sense, for the introduction of superior new products, though possibly reducing current profits, is likely to enhance the firm's future profit-earning potential. And although they may be gained only through patient struggle, larger market shares, we shall see in Chapter 11, are strongly and systematically associated with greater profitability. Whether these differences in stated managerial goals stem from differences in the cost of capital (because of their stronger propensity to save, the Japanese have lower interest rates), differences in business organization (for example, the longer employment tenure of Japanese managers or their closer relationships with banks), or differences in the deeper-seated aspects of culture is a question that warrants careful research.



## Organizational Complexity

Another feature of the modern business enterprise that may prevent it from behaving in strict conformity to the profit maximization hypothesis is its organizational complexity. Most large corporations have multiple operating units supervised by one or more corporate-level executives. Within the corporate staff, and often mirrored at the operating level, there are functional components specializing in production, sales, materials procurement, finance, accounting, research and development, and so forth. An elaborate chain of command extends from workers at the operating level to top management and the board of directors — the latter presumably representing stockholder interests. This structure must be tied together by a communications network so that decisions taken at various levels and in the diverse functional specialties mesh. It is here that breakdowns occur. Conflicts among functional groups are bound to arise. David Halberstam documents, for example, the hard-fought goal disparities between MBAs on Ford Motor Company's finance staff, who stressed strategies with clear, quantifiable, profit-maximizing prospects, and the engineering-oriented "product" people, who urged an accelerated pace of design improvement and were more inclined to accept added cost to improve product quality.[48] Such conflicts must often be passed up to higher, and perhaps top, management for resolution. But the information transmission process is subject to attenuation. Top managers cannot possibly digest all the knowledge possessed by every operating-level employee, and so some grasp of special circumstances affecting particular cases must be sacrificed. Furthermore, the content of messages is often distorted to suit the prejudices and fears of both senders and receivers. The more hierarchical filters through which information passes and the sharper the conflict is among specialists, the more distorted the message is likely to become, and the greater is the chance that incorrect choices will be made. Or in the reverse flow from top management to operating levels, the more likely instructions will be misinterpreted or deliberately ignored.

Given this organizational complexity, it may prove very difficult for top management to arrive at and enforce decisions that maximize profits. Complicating matters is the fact that operating-level personnel often care little about profit maximization. Even the best-designed employee bonus and profit-sharing systems seldom instill much zeal for profit maximization below the middle management level. Operating-level employees see little correlation between their individual actions and the size of the profit pie in which they will share, just as individual firms in a competitive industry consider their output decision to have an imperceptible effect on the market price and hence their profits. At the same time, they have many goals that conflict with profit maximization by any criterion. Division chiefs fearing dismissal or playing for time gloss over operating problems until the

46. See, for example, Robert J. Shiller, "Do Stock Prices Move Too Much To Be Justified by Subsequent Changes in Dividends?" *American Economic Review*, vol. 71 (June 1981), pp. 421–436. Compare Terry A. Marsh and Robert C. Merton, "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices," *American Economic Review*, vol. 76 (June 1986), pp. 483–498.

47. Tadao Kagawa et al., "Strategy and Organization of Japanese and American Corporations," summarized in M. J. Peck, "The Large Japanese Corporation," in Meyer and Gustafson, *The U.S. Business Corporation*, supra note 23, pp. 35–36.

48. David Halberstam, *The Reckoning* (New York: William Morrow, 1986), Chapters 11, 13, 28, 33, 34, and 41.

situation has deteriorated beyond repair. Research and development engineers seek technical sophistication and product refinement for their own sake, even when they add more to cost than to revenues. Production and staff supervisors find make-work jobs for redundant personnel because firing people is unpleasant, and so on.

It is the classic responsibility of top management to ferret out these deviations and to establish a system of controls and incentives that ensures internal conformity with the firm's profit maximization goal. Organizational complexity can frustrate such efforts. But an equally significant obstacle arises out of the very character of the modern business corporation. Like the divisional and functional specialists they command, top managers may pursue the profit goal with less than complete diligence.

One reason for this is said to be the increasing separation between ownership and control of industrial enterprises — a phenomenon alluded to by Lenin[49] and first studied intensively during the early 1930s by Berle and Means.[50] There was a time more than a century ago when individuals managing even the largest U.S. companies held, if not a majority, at least a substantial minority equity interest in their enterprises. Gradually this has changed. Big business has become bigger; the lion's share of industrial output is produced by corporations large under any plausible definition of the word. The ownership of large corporations has been dispersed among thousands of stockholders, no one of whom may own a sufficiently large fraction of the outstanding shares to exercise a significant controlling role. AT&T had 2.8 million common stockholders of record in 1986, Exxon 740,000, and IBM 793,000. The median (250th) firm in terms of stockholder numbers on *Fortune's* list of the 500 largest industrial corporations for 1956 (the last year such data were reported) had more than nine thousand shareholders.

In their pioneering study, Berle and Means found eighty-eight of the 200 largest American nonfinancial corporations to be "management controlled" in 1929 because no individual, family, corporation, or group of business associates owned more than 20 percent of all outstanding voting stock, and because evidence of control by a smaller ownership group was lacking.[51] Only twenty-two of the corporations were judged to be privately owned or controlled by a group of stockholders with a majority interest. Replicating the Berle and Means methodology, but with a lower 10 percent ownership share threshold, Robert Larner found that by 1963, 161 of the 200 largest nonfinancial corporations had come to be management controlled. He concluded that the "managerial revolution" identified by Berle and Means was "close to complete."[52]

This view of the world has been challenged by Phillip Burch, Edward Herman, Thomas Dye, and others for ignoring the power positions of financial intermediaries and the possibility that family or other interest groups might exercise effective control from a position well below a 10 or 20 percent share ownership threshold.[53] Proceeding under an assumption that control could be exercised with a family stock position of 4 to 5 percent accompanied by extended representation on the board of directors, Burch concluded that in 1965, only 41 percent of the leading 300 industrial corporations and 28 percent of the top fifty merchandising corporations were "probably management controlled."[54] He found that 43 percent of



the industrials and 58 percent of the merchandising corporations were "probably family controlled."

Thus, one's conclusion as to whether the separation of ownership and control has advanced nearly to its limit, or less than half of the way, depends upon subtle quantitative distinctions over which reasonable scholars can disagree.

When managers themselves are not major stockholders, the task of representing stockholder interests falls upon the board of directors. A path-breaking study by Myles Mace showed that the role of stock-holding family representatives in U.S. corporations varied widely.[55] Sometimes, Mace found, family members ceased exercising active control in the boardroom when their stock ownership dropped below 50 percent. In other cases they continued to exercise de facto control despite holding only a small fraction of the company's stock. Much evidently depends upon such idiosyncracies as the business ability and interest of family members. When family members did take an active part, it was manifested among other things in bringing family influence to bear on the choice of other board members.

When the board of directors is more clearly controlled by management, the inside chief executive officer (usually the chairman or president) typically selects outside board members. A reputation for not rocking the boat receives heavy weight in selection decisions; few CEOs wish to repeat the experience of General Motors with board member H. Ross Perot, who criticized company policies both within the boardroom and publicly.[56] Nominations are normally validated by an overwhelming margin as stockholders docilely assign their proxies. Except in crisis situations, board members in turn rubber-stamp management recommendations on new managerial appointments and other policies. Thus, through reciprocal self-selection, the management group maintains its control. Mace's 1971 study revealed that outside directors tended to be ill-informed about their corporations' operations and that conventional boardroom ethics discouraged independent action, or even the posing of embarrassing questions.[57] There is reason to believe that outside directors' participation became more active during the 1980s

**49.** V. I. Lenin, *Imperialism: The Highest Stage of Capitalism* (New York: International Publishers, 1939), p. 59 (originally published in 1917).
**50.** Adolf A. Berle and Gardiner Means, *The Modern Corporation and Private Property* (New York: Macmillan, 1932). See also the June 1983 issue of *The Journal of Law & Economics* (vol. 26), containing papers from a conference on the fiftieth anniversary of the Berle and Means book's publication; and the special issue of the *Journal of Financial Economics,* vol. 20, January/March 1988.
**51.** Berle and Means, *The Modern Corporation,* pp. 90–118. Other early studies reaching similar conclusions include Raymond W. Goldsmith, *The Distribution of Ownership in the Largest 200 Nonfinancial Corporations,* Temporary National Economic Committee Monograph no. 29 (USGPO, 1940); and Robert Aaron Gordon, *Business Leadership in the Large Corporation* (Washington: Brookings, 1945).
**52.** Robert J. Larner, *Management Control and the Large Corporation* (Cambridge: Dunellen, 1970), pp. 9–24.
**53.** Philip H. Burch, Jr., *The Managerial Revolution Reassessed* (Lexington: Heath, 1972); Thomas R. Dye, "Who Owns America: Strategic Ownership Positions in Industrial Corporations," *Social Science Quarterly,* vol. 64 (December 1983), pp. 865–867; and Edward S. Herman, *Corporate Control, Corporate Power* (Cambridge: Cambridge University Press, 1981), pp. 54–65.
**54.** Burch, *The Managerial Revolution,* pp. 68 and 96.
**55.** Myles L. Mace, *Directors: Myth and Reality* (Boston: Harvard Business School Division of Research, 1971), pp. 154–174.
**56.** "GM Plans Offer To Pay $700 Million To Buy Out Its Critic H. Ross Perot," *Wall Street Journal,* December 1, 1986, pp. 1, 13.
**57.** Mace, *Directors,* pp. 43–71 and 94–101.

through fear of multimillion-dollar liability judgments against directors who approved, after only superficial consideration, actions that adversely affected stockholder welfare.[58] Thus, in a 1987 Harris poll of 400 U.S. corporations' chief executive officers, 36 percent reported that their directors had become "more assertive" during the past five years, while only 2 percent said that the board had become less assertive.[59] Still it almost surely remains true that under noncrisis conditions, boards of directors do not impose much restraint on management's operating discretion.

Although the legal mandates affecting board of directors' makeup and functioning vary widely from nation to nation, there is surprising uniformity in the extent to which, when strong outside ownership interests are absent, inside management calls the tune. In Japan, most corporations' boards consist almost entirely of inside management members.[60] A few members are appointed by banks or affiliated companies, but once they join, they tend to become full-time employees of the company on whose board they sit. In West Germany, large corporations are required by law to have two different boards — an operating board comprising inside management members and a supervisory board with strong representation from both outside interests and delegates selected by the company's workers.[61] In principle, the supervisory board has the power to steer management in directions it might otherwise not choose. In practice, however, its ability to do so is limited by management's control of information and by the tendency of worker representatives to become co-opted as they spend more and more time on corporate policy matters and less with the workers who elected them.

## The Multitude of Managerial Goals

Within certain imprecisely defined bounds, then, management may be free to pursue goals not necessarily consistent with maximizing returns to stockholders. What are those goals? How seriously do they conflict with profit maximization?

One possibility is for managers to seek a placid, comfortable, risk-free existence. As J. R. Hicks put it in a much-quoted quip, "The best of all monopoly profits is a quiet life."[62] Such a characterization does considerable justice to the archetypal British business leader of past generations. Whether it reflects the psychology of American, Japanese, and Continental European counterparts, or even the newest generation of British managers, is more doubtful.

A goal emphasized by the leaders of twelve large, successful U.S. corporations is "long-term corporate survival."[63] Donaldson and Lorsch, the interviewers, found that corporate heads often perceived significant conflicts between maximizing short-run profits and their longer-term survival goal. They complained about "the tyranny of quarter-to-quarter profit growth standing in the way of investment for the long term."[64] Since the proper time frame for profit maximization is ambiguous and since the alternative to survival, even for stockholders, can be unpleasant, it is arguable that behavior emphasizing survival serves owner interests too. Nevertheless, as we shall see in Chapter 5, conflicts can arise — for example, if managers of companies in declining industries seek to perpetuate their enterprises by channeling stockholders' funds into the acquisition of firms whose operations are quite different from their own, requiring know-how and managerial skills they may not have.



Managers seek survival not only of the companies they direct but also of their own jobs. Desiring security of job tenure does not necessarily conflict with profit maximization; the manager who keeps the profits rolling in is, after all, a good person to have around. Yet conflicts can arise, especially in decision making under uncertainty. As William Fellner has pointed out, there tends to be an asymmetry in the rewards to hired managers.[65] If risky decisions turn out badly, stockholders lose their assets and managers lose their jobs.[66] If they turn out well, the manager may be promoted or receive a bonus, but the rewards are seldom commensurate with the stockholders' profit gains. Faced with this asymmetry, the hired manager may sacrifice higher expected profits for lower risk than an owner-manager would under otherwise identical circumstances. Also, because the stock market reacts sensitively to changes, managers concerned about their tenure exhibit an apparent preference for lower but more stably growing earnings over higher but fluctuating earnings. One manifestation is the use of accounting discretion to smooth reported earnings — for example, by shifting the reporting of deferable costs from bad to good years.[67] This is most likely innocuous, but other stability-enhancing behavior can have more negative implications. To protect themselves against attention-drawing setbacks, managers tend to accumulate *organizational slack* — inessential resources that are shed only when operating unit profits are under severe pressure. From a study of three divisions of large corporations during the 1970s, Schiff and Lewin estimated that the amount of slack built into divisional budgets averaged 20 to 25 percent of the units' operating expenses.[68] Top managers were said to be aware that such slack existed; however, they were unable to purge it because they lacked the detailed knowledge needed to set cost and profit budgets tight enough to maximize profits, but not so tight as to jeopardize operations.

58. See, for example, Jeremy Bacon and James K. Brown, *The Board of Directors: Perspective and Practices in Nine Countries* (New York: Conference Board, 1977), Chapters 1 and 8; Charles A. Anderson and Robert N. Anthony, *The New Corporate Directors: Insights for Board Members and Executives* (New York: Wiley, 1986); and Winthrop Knowlton and Ira Millstein, "Can the Board of Directors Help the American Corporation Earn the Immortality It Holds So Dear?" in Meyer and Gustafson, eds., *The U.S. Business Corporation,* supra note 23, pp. 169–191.
59. "Changing Roles," *Business Week,* October 23, 1987, p. 28.
60. Peck, "The Large Japanese Corporation," in Meyer and Gustafson, *The U.S. Business Corporation,* supra note 23, pp. 21–22.
61. Bacon and Brown, *The Board of Directors,* supra note 58.
62. Hicks, "Annual Survey of Economic Theory: The Theory of Monopoly," *Econometrica,* vol. 3 (January 1935), p. 8.
63. Gordon Donaldson and Jay W. Lorsch, *Decision Making at the Top* (New York: Basic Books, 1983), pp. 7–8.
64. Ibid., p. 170.
65. William Fellner, *Competition Among the Few* (New York: Knopf, 1949), pp. 172–173.
66. For evidence that adverse deviations from trend profitability lead to a shortening of corporate presidents' tenure, see W. M. Crain, Thomas Deaton, and Robert Tollison, "On the Survival of Corporate Executives," *Southern Economic Journal,* vol. 43 (January 1977), pp. 1372–1375; Michael S. Weisbach, "Outside Directors and CEO Turnover," *Journal of Financial Economics,* vol. 20 (January/March 1988), pp. 431–460; and Jerold B. Warner, Ross L. Watts, and Karen Wruck, "Stock Prices and Top Management Changes," *Journal of Financial Economics,* vol. 20 (January/March 1988), pp. 461–492.
67. See Jacob Y. Kamin and Joshua Ronen, "The Effects of Corporate Control on Apparent Profit Performance," *Southern Economic Journal,* vol. 45 (July 1978), pp. 181–191.
68. Michael Schiff and Arie W. Lewin, "Where Traditional Budgeting Fails," *Financial Executive,* vol. 36 (May 1968), pp. 50–62; and *idem,* "The Impact of People on Budgets," *Accounting Review,* vol. 45 (1970), pp. 259–268. See also Williamson, *Discretionary Behavior,* supra note 43, pp. 85–126. Two of the slack-laden companies studied by Williamson, it is worth noting, were controlled by dominant family ownership groups.

There is reason to believe that U.S. corporations' behavior changed during the early 1980s. Many firms experienced severe pressure on profits from a combination of the deepest recession since the 1930s and sharply rising import competition, stimulated in turn by unusual (and unsustainable) strength of the U.S. dollar vis-à-vis other currencies. They reacted by implementing extensive, deep staff cutbacks.

Another important change in the 1980s was the large number of "going private" transactions, under which whole publicly-owned corporations or parts thereof restructured themselves financially and reverted to essentially private ownership status. Between 1980 and 1987, the total recorded volume of such transactions exceeded $120 billion. Usually, most of the financing for these new organizations came from borrowing, and inside managers tended to hold a substantial share of the relatively small residual common stock interest. Case studies and statistical analyses reveal that operating efficiency improved significantly following the restructurings[69] – presumably because of strengthened incentives attributable to management's ownership stake and the pressure of heavy interest payment obligations. It remains unclear whether less desirable tradeoffs may also have been precipitated by such "leveraged buyouts." In particular, there is evidence that capital investments and research and development spending were cut back, possibly weakening the new organizations' long-run strength.

Being human, most hired managers derive considerable satisfaction from achieving personal prestige and power. Both appear to be correlated more closely with the volume of company sales, employment, and assets than with the volume or rate of profits it earns. There have also been indications that the compensation of top executives is more closely associated with corporate size than with profitability. In view of this, it has been argued that hired managers are more concerned with enhancing sales or the growth of sales than with increasing profits.[70] In the static theory of the monopolistic or oligopolistic firm, maximizing sales revenue is incompatible with maximizing profits except under improbable circumstances (for example, when marginal cost is zero). In more dynamic models, the conflict between sales growth and profitability is less pronounced. Nevertheless, increases in growth beyond some point must impose profit sacrifices as management's ability to control is overstrained, as low payoff investment projects are approved, and as high-cost sources of capital are tapped.

## Constraints on Departures from Profit Maximization

In sum, there is no shortage of ways business managers *may* deviate from profit maximization. However, there are also constraints on management's freedom to deviate.

One is the threat of takeover. According to hypotheses first advanced by Robin Marris and Henry Manne, failure to maximize profits will depress company stock prices below their potential value.[71] This may induce some outside entrepreneur to bid for a controlling interest, remove the incumbent management, and redirect the company's energies toward increasing profits and hence raising stock values. During the 1960s, such hostile takeovers evolved from rare to commonplace events in the United States and United Kingdom. After a hiatus, takeover activity exploded during the 1980s and expanded beyond Anglo-American frontiers.

There has been extensive quantitative research on this theory of takeovers and managerial constraint. Early British studies and more recent U.S. analyses show



that the stock prices of firms singled out as takeover targets were much lower in relation to the accounting or replacement cost value of assets than firms not subjected to takeover attempts.[72] Analyzing 371 companies on *Fortune*'s list of the 500 largest U.S. industrial corporations for 1980, Mørck et al. found that the ratio of stock prices to assets per share averaged 0.524 for forty takeover targets, compared to 0.848 for nontarget sample members.[73] This can be interpreted as evidence of a selection mechanism disciplining lax managers. However, parallel studies reveal that the takeover targets were at worst only slightly less profitable than their industrial peers in the years preceding initiation of the takeover attempt.[74] The profitability evidence has at least two quite different rationalizations. For one, depressed stock prices may reflect investors' expectations of declining future profitability rather than unsatisfactory past profits. Needless to say, such a hypothesis is difficult to test when a takeover alters the target company's control structure. An alternative explanation is that valuation errors on the stock market, not managerial errors leading to deficient profits, are what nominates companies for a takeover attempt. Those with the most undervalued shares become prime targets. This view is also difficult to substantiate or refute empirically; without a good way of ascertaining future earnings absent the takeover attempt, it is hard to tell whether the company's stock was in fact undervalued. It would not be surprising to learn that reality embodies a mixture of these cases: some takeovers occur because of management shortcomings and some because of misvaluation by the stock market.

**69.** See Steven Kaplan, "Management Buyouts: Efficiency Gains or Value Transfers?" paper presented at a New York University Salomon Brothers Center conference on *Financial-Economic Perspectives on the High-Yield Debt Market* (December 1988); the testimony of F. M. Scherer before the House of Representatives Committee on Ways and Means, March 1989; and Mike Wright and John Coyne, *Management Buyouts* (London: Croom Helm, 1985).

**70.** For various approaches, see Edith T. Penrose, *The Theory of Growth of the Firm* (Oxford: Blackwell, 1959); William J. Baumol, *Business Behavior, Value, and Growth* (rev. ed.; New York: Harcourt, Brace and World, 1967), Chapters 5–10; Marris, "A Model of the 'Managerial' Enterprise," supra note 43; J. Williamson, "Profit, Growth and Sales Maximization," *Economica*, vol. 34 (February 1966), pp. 1–16; John Lintner, "Optimum or Maximum Corporate Growth under Uncertainty," and Robert M. Solow, "Some Implications of Alternative Criteria for the Firm," in Robin Marris and Adrian Wood, eds., *The Corporate Economy* (Cambridge: Harvard University Press, 1971), pp. 172–241 and 318–342; Richard Schramm, "Profit Risk Management and the Theory of the Firm," *Southern Economic Journal*, vol. 40 (January 1974), pp. 353–363; George K. Yarrow, "Growth Maximization and the Firm's Investment Function," *Southern Economic Journal*, vol. 41 (April 1975), pp. 580–592; and Robert E. Wong, "Profit Maximization and Alternative Theories: A Dynamic Reconciliation," *American Economic Review*, vol. 65 (September 1975), pp. 689–694.

**71.** Marris, "A Model of the 'Managerial' Enterprise," supra note 43; Marris, *The Economic Theory of "Managerial" Capitalism* (London: Macmillan, 1964); and Henry G. Manne, "Mergers and the Market for Corporate Control," *Journal of Political Economy*, vol. 73 (April 1965), pp. 110–120. For a more recent exchange of views, see the symposium in the *Journal of Economic Perspectives*, vol. 2 (Winter 1988), pp. 3–82.

**72.** See especially Ajit Singh, *Take-overs: Their Relevance to the Stock Market and the Theory of the Firm* (Cambridge University Press, 1971); Douglas Kuehn, *Takeovers and the Theory of the Firm* (London: Macmillan, 1975); and (for an early U.S. test) Brian Hindley, "Separation of Ownership and Control in the Modern Corporation," *Journal of Law & Economics*, vol. 13 (April 1970), pp. 185–221. On the extent to which managers can depart from profit maximization before triggering a takeover, see Robert Smiley, "Tender Offers, Transaction Costs and the Theory of the Firm," *Review of Economics and Statistics*, vol. 58 (February 1976), pp. 22–32.

**73.** Randall Mørck, Andrei Shleifer, and Robert W. Vishny, "Characteristics of Targets of Hostile and Friendly Takeovers," in Alan J. Auerbach, ed., *Corporate Takeovers: Causes and Consequences* (University of Chicago Press, 1988), p. 118.

**74.** David J. Ravenscraft and F. M. Scherer, "Life After Takeover," *Journal of Industrial Economics*, vol. 36 (December 1987), pp. 149–150. See also Edward S. Herman and Louis Lowenstein, "The Efficiency Effects of Hostile Takeovers," in John Coffee et al., eds., *Knights, Raiders, and Targets* (New York: Oxford University Press, 1988), pp. 211–240; Arthur T. Andersen and T. Crawford Honeycutt, "Management Motives for Takeovers in the Petroleum Industry," *Review of Industrial Organization*, vol. 3 (1987), pp. 1–12; and Timothy Hannan and Stephen A. Rhoades, "Acquisition Targets and Motives: The Case of the Banking Industry," *Review of Economics and Statistics*, vol. 69 (February 1987), pp. 67–74.

An analysis of unusually rich data by Ravenscraft and Scherer raises doubts concerning the efficacy of takeovers in remedying managerial failures.[75] Nine years on average after sixty-two U.S. companies experienced 1960s- and 1970s-vintage takeovers, the operating profitability of the acquired units, appropriately adjusted for accounting valuation changes, had neither improved nor regressed relative to pretakeover profit rates. This evidence runs contrary to the hypothesis that takeover occurs to *raise* profitability. Leaders of the acquiring firms may have overestimated their ability to improve matters, but if so, that too speaks poorly for the role of takeovers as a disciplinary mechanism.

Against the still unconfirmed hope that takeovers discipline management and force profit maximization, there is also the possibility of negative consequences. For one, until the 1980s, large corporations were much less likely to be taken over than smaller firms.[76] Recognition of this induced some firms to enter hastily brokered mergers to increase their size and reduce the probability of takeover. However, financial innovations during the 1980s brought even multibillion dollar corporations under siege, reducing the attractiveness of such defensive strategies. In the 1980s a new criticism surfaced. On the assumption that stock market investors are often short-sighted, it was claimed that the threat of takeover induced managers to emphasize short-run profit maximization over longer-term profit-seeking, with adverse consequences for research and development spending and other far-sighted activities. The evidence on this point has thus far been weak and inconclusive.[77]

When forced into the trenches on the question of whether firms maximize profits, economists resort to the ultimate weapon in their arsenal: a variant of Darwin's natural selection theory.[78] Over the long pull, there is one simple criterion for the survival of a business enterprise: profits must be nonnegative. No matter how strongly managers prefer to pursue other objectives, and no matter how difficult it is to identify profit-maximizing strategies in a world of uncertainty and high information costs, failure to satisfy this criterion means ultimately that a firm will disappear from the economic scene. Profit maximization is therefore promoted in two ways. First, firms departing too far from the optimum, either deliberately or by mistake, will disappear. If the process of economic selection continues long enough, the only survivors will be firms that did a tolerably good job of profit maximization. Second, knowledge that only the fit will survive provides a potent incentive for all firms to channel their behavior in profit-maximizing directions, learning whatever skills they need and emulating organizations that excel at the survival game.

To be sure, the selection process operates imperfectly. The environment is constantly changing, altering the behavior required for survival, so that adaptations learned today may not serve tomorrow. And adaptation by industry members may be sufficiently slow to permit firms performing less than optimally to keep their heads above water for a long time.

Despite these qualifications, it seems reasonable to believe that the natural selection process is a stern master in a competitive environment. That it works equally well under monopoly does not follow. When firms with monopoly power are shielded by entry barriers, product differentiation, government favoritism, or



the like, threats to their survival may be sufficiently blunted that the organizations can survive for decades without maximizing profits or minimizing costs. The crucial question remains, How much protection from the forces of natural selection do real-world enterprises enjoy? How far from profit-maximizing norms can they stray and still remain viable?

## Quantitative Research

There has been much quantitative research on managerial incentives to conform to, or depart from, profit maximization and on the conditions under which departures occur. The work has three main emphases: managerial compensation patterns, returns on new corporate investment, and the relationship between profitability and corporate ownership structure.

Many studies have investigated whether executive salaries and bonuses are correlated with their companies' profitability. The stronger the correlation, the more one might expect managers' incentives to parallel those of owners. An immediate problem is that greater corporate size undoubtedly requires superior managerial skills, and so a positive association between size and compensation is expected. A related problem is that *levels* of sales or assets and levels of profits are highly correlated, making it difficult to tell whether managers strive for size per se or the higher profits that accompany greater size. Close statistical fit[79] and also some a priori theory[80] suggest a nonlinear size–compensation relationship of logarithmic form. When the size effect is so controlled, the weight of evidence suggests that higher or rising profits imply higher executive compensation, although the regulated electric power and telephone companies may be an exception.[81] Using stock price changes rather than reported profits as his indicator of how well managers served stockholders, Kevin Murphy found for a sample of 461 U.S.

**75.** Ravenscraft and Scherer, "Life After Takeover," pp. 150–154.

**76.** See Singh, *Take-overs,* supra note 72, pp. 139–144; and Singh's review of Kuehn's book in the *Journal of Economic Literature,* vol. 14 (June 1976), pp. 505–506.

**77.** See, for example, Herbert I. Fusfeld, "Corporate Restructuring — What Impact on U.S. Industrial Research?" *Research Management,* vol. 30 (July-August 1987), pp. 10–17; and James F. Mathis and Arthur B. Hill, "How 'Speculator's Capitalism' Affects R&D in the Chemical Processing Industries," *Research-Technology Management,* vol. 31 (May-June 1988), pp. 44–49.

**78.** See especially Alchian, "Uncertainty," Winter, "Natural Selection;" and Nelson and Winter, *An Evolutionary Theory,* supra note 43.

**79.** See, for example, George K. Yarrow, "Executive Compensation and the Objectives of the Firm," in Keith Cowling, ed., *Market Structure and Corporate Behaviour* (London: Gray–Mills, 1972), pp. 149–173; and W. J. Boyes and Don E. Schlagenhauf, "Managerial Incentives and the Specification of Functional Forms," *Southern Economic Journal,* vol. 45 (April 1979), pp. 1225–1232.

**80.** Herbert A. Simon, "The Compensation of Executives," *Sociometry*, vol. 20 (March 1957), pp. 32–35.

**81.** See, for example, Yarrow, "Executive Compensation," supra note 79; Geoffrey Meeks and Geoffrey Whittington, "Directors' Pay, Growth and Profitability," *Journal of Industrial Economics,* vol. 24 (September 1975), pp. 1–14; William A. McEachern, *Managerial Control and Performance* (Lexington: Heath, 1975); David H. Ciscel and Thomas Carroll, "The Determinants of Executive Salaries," *Review of Economics and Statistics,* vol. 62 (February 1980), pp. 7–13; the comment by James A. Dunlevy, *Review of Economics and Statistics,* vol. 67 (February 1985), pp. 171–174; John R. Deckop, "Determinants of Chief Executive Officer Compensation," *Industrial and Labor Relations Review,* vol. 41 (January 1988), pp. 215–226; A. D. Cosh and A. Hughes, "The Anatomy of Corporate Control: Directors, Shareholders and Executive Remuneration in Giant U.S. and U.K. Corporations," *Cambridge Journal of Economics,* vol. 11 (December 1987), pp. 285–313; and Edward A. Dyl, "Corporate Control and Management Compensation," *Managerial and Decision Economics,* vol. 9 (March 1988), pp. 21–26. On regulated firms, compare Mark Hirschey and James Pappas, "Regulatory and Life Cycle Influences on Managerial Incentives," *Southern Economic Journal,* vol. 48 (October 1981), pp. 327–334; and Thomas M. Carroll and David H. Ciscel, "The Effects of Regulation on Executive Compensation," *Review of Economics and Statistics,* vol. 64 (August 1982), pp. 505–509.

executives over the years 1964–1981 that annual percentage changes in compensation were positively correlated with stock price changes.[82] Managers who led companies experiencing stock price declines of 30 percent or more saw their pay decrease by 1.2 percent per year on average; those whose stock increased in value by 30 percent or more averaged 8.7 percent annual compensation gains. The shifts are modest, but perhaps at least point managers in a profit-maximizing direction.

It is also important to ask whether the level of executive compensation and the extent of its correlation with profitability are greater under some ownership structures than under others. The research on this point has yielded equivocal results. Using data from the 1920s and 1930s, Stigler and Friedland found no significant correlation between executive compensation levels and a variable (taken from Berle and Means) distinguishing management-controlled from other corporations, taking into account also company size in terms of asset values.[83] A study of U.S. data for 1980 showed that compensation levels were significantly higher in corporations with widely dispersed stock ownership (implying control by inside managers), holding company size and profit rates equal,[84] while a British study obtained the opposite result.[85] With data on forty-eight companies in three U.S. industries, McEachern discovered that compensation was more closely correlated with profitability in corporations controlled by an outside ownership group than in companies controlled by either an inside owner-manager group or by inside managers alone.[86]

Even in very large and diffusely held corporations, stock ownership can be an important component of top managers' wealth. To be sure, managers may own only a tiny fraction of their employer's outstanding shares, but a small share of a large sum can be a lot of money. Studying fifty leading U.S. manufacturing corporations, William Lewellen found that on average, dividends, capital gains on company stock holdings, and changes in the value of stock options averaged 1.36 times fixed dollar (salary plus bonus) compensation for the companies' top three executives over the period 1940–1963.[87] More recent confirming evidence on the compensation and stock holdings of 461 high-level U.S. corporation executives comes from Kevin Murphy.[88] On average, the executives' stock holdings averaged $4.7 million – thirteen times their annual salary plus bonus. Moving from the lowest decile in terms of observed returns to common stockholders to the top quintile, the value of the managers' *own* stock holdings increased on average by some $5 million. Even though heroic efforts by the managers might not achieve such dramatic stock return increases, it seems clear that the typical top executive must have been keenly aware that his or her own wealth was linked to the fortunes of the company's shareholders. However, the link may have been less close in management-controlled corporations, for which McEachern found inside managerial stock holdings to be lower than in companies with strong inside or outside ownership groups.[89]

Although executive salaries and bonuses are correlated with both firm size and the level of profits, the size correlations tend to be the stronger of the two. Consequently, executives might be willing to sacrifice profitability at the margin to enhance sales and hence compensation. One way of doing this is to plow back earnings into growth-oriented investments yielding at the margin a lower return than



stockholders could achieve reinvesting more generous dividends in other firms' stocks. Baumol and associates found evidence of such behavior in a statistical study of several hundred U.S. corporations.[90] The average rate of return on plowback investments was estimated to be in the range of 3.0 to 4.6 percent, compared to 14 to 21 percent on new equity capital. Subsequent clarifications revealed that the subset of companies issuing new equity capital realized sizable returns on their plowback too, and that the main locus of depressed plowback returns was firms issuing no new common stock.[91] Further light was shed on the matter by Grabowski and Mueller. They argued that most corporations pass through a life cycle. In the early years, when owner control is often strongest, lucrative investment opportunities call for external financing. But as the firm and its products mature, cash flows increase while investment opportunities dwindle. Investment of internally generated funds may be maintained at high levels despite falling profitability because managers seek to expand in accustomed ways and because earnings reinvested internally are taxed more lightly than earnings paid out as dividends and then reinvested. Testing their hypotheses, Grabowski and Mueller classified 759 U.S. corporations into mature (that is, with the majority of their sales in pre-1940 products) and nonmature categories. They found that the nonmature companies realized significantly higher returns on invested capital in each of three periods from 1957 to 1970. From this they concluded that "managers of mature corporations in technologically unprogressive industries re-invest too large a percentage of their internal funds. Their shareholders would apparently be better off with higher payouts. . . ."[92]

If management-controlled companies reinvest their funds in low-yield projects or otherwise fail to maximize profits, there should be a discernible difference in

**82.** Kevin J. Murphy, "Corporate Performance and Managerial Remuneration," *Journal of Accounting and Economics,* vol. 7 (April 1985), pp. 11–42. For similar earlier findings, see Robert T. Masson, "Executive Motivations, Earnings, and Consequent Equity Performance," *Journal of Political Economy,* vol. 79 (November/December 1971), pp. 1278–1292.

**83.** George J. Stigler and Claire Friedland, "The Literature of Economics: The Case of Berle and Means," *Journal of Law & Economics,* vol. 26 (June 1983), pp. 237–268.

**84.** Rexford E. Santerre and Stephen P. Neun, "Stock Dispersion and Executive Compensation," *Review of Economics and Statistics,* vol. 68 (November 1986), pp. 685–693. Compare Dennis C. Mueller, *Profits in the Long Run* (Cambridge University Press: 1986), pp. 159–161, who found statistically insignificant stockholder control coefficients for a large U.S. sample.

**85.** John Cubbin and Graham Hall, "Directors' Remuneration in the Theory of the Firm," *European Economic Review,* vol. 20 (January 1983), pp. 345–346.

**86.** McEachern, *Managerial Control,* supra note 81, pp. 77–84.

**87.** Wilbur G. Lewellen, *The Ownership Income of Management* (New York: Columbia University Press, 1971), pp. 79–103.

**88.** Murphy, "Corporate Performance," supra note 82, pp. 26–27.

**89.** McEachern, *Managerial Control,* supra note 81, pp. 82–83.

**90.** William J. Baumol, Peggy Heim, Burton Malkiel, and Richard Quandt, "Earnings Retention, New Capital and the Growth of the Firm," *Review of Economics and Statistics,* vol. 52 (November 1970), pp. 345–355. Results for the United Kingdom are similar; see, for example, Geoffrey Whittington, "The Profitability of Alternative Sources of Finance," *Review of Economics and Statistics,* vol. 60 (November 1978), pp. 632–634. For contrasting Canadian results, see Daniel M. Shapiro, William A. Sims, and Gwenn Hughes, "The Efficiency Implications of Earnings Retentions: An Extension," *Review of Economics and Statistics,* vol. 65 (May 1983), pp. 327–331.

**91.** See Irwin Friend and Frank Husic, "Efficiency of Corporate Investment," *Review of Economics and Statistics,* vol. 55 (February 1973), pp. 122–127, with a reply by Baumol et al., pp. 128–131; and Steven Fazzari, R. G. Hubbard, and Bruce Petersen, "Financing Constraints and Corporate Investment," *Brookings Papers on Economic Activity* (1988, no. 1), pp. 141–195; and James M. Griffin, "A Test of the Free Cash Flow Hypothesis: Results from the Petroleum Industry," *Review of Economics and Statistics,* vol. 70 (February 1988), pp. 76–82.

**92.** Henry G. Grabowski and Dennis C. Mueller, "Life-Cycle Effects on Corporate Returns on Retentions," *Review of Economics and Statistics,* vol. 57 (November 1975), pp. 400–416.

profitability or stock market returns between such firms and enterprises with more powerful owner control groups. Whether the owner group remains outside or participates actively in management, in which case it might sacrifice profits to enhance its own managerial perquisites, could affect the comparison. So also could the degree of monopoly power, since the managers of firms operating in an intensely competitive environment may have little or no discretion to divert profits to their personal ends. Numerous studies have tested some subset of these hypotheses, but only one has attempted to control for the entire set of plausible behavioral variables. The results have varied widely, depending upon the data analyzed, the range of variables taken into account, and the statistical methodology employed.[93] The most comprehensive study (Mueller's) included as explanatory variables the proportion of common shares held by the top five managers, the control position of outside owners, company market share, company size, diversification, industry structure, industry advertising, and industry innovative propensities.[94] Given these variables, Mueller found that corporations with a strong inside owner group had significantly *lower* 1950–1972 profit trajectories, while the existence of a strong outside owner group made no discernible difference. Whether these results will generalize for other nations, company samples, and analytic techniques remains to be seen.

**Implications** The last word has by no means been uttered on how assiduously modern industrial corporations strive to maximize their profits. From the voluminous and often inconsistent evidence, it appears that the profit maximization assumption at least provides a good first approximation in describing business behavior. Deviations, both intended and inadvertent, undoubtedly exist in abundance, but they are kept within more or less narrow bounds by competitive pressures, the self-interest of stock-owning managers, and the threat of managerial displacement by important outside shareholders or takeovers. To the extent that deviations do occur, they are apt to be larger when competition is weak and when management is strongly entrenched, either by virtue of its own stockholdings or because no coherent outside ownership coalition exists. If firms with monopoly power do tolerate extensive organizational slack, pay princely managerial salaries, and the like, economic inefficiency must be higher than one would predict under the assumption of strict profit maximization.

## Workable Competition

We return now to our original question: how valid is the competitive ideal as a prescription for economic policy? Given all the qualifications and doubts unearthed in the foregoing pages, extreme confidence is hardly appropriate. We may even experience an impulse to return to the womb — to Adam Smith's crude vision of how the market economy does its job. Smith was wrong in numerous details, but details of the system may be less important than the broad scheme of operation. If one stands back and gazes astigmatically at competitive market systems without worrying about the fine points, one sees that they do display generally greater responsiveness to consumer demands and generate more potent incentives for the frugal use of resources than do monopoly market structures. This,



rather than the satisfaction of all optimal conditions in a general equilibrium system containing 43 quadrillion equations, may be the core of the case for competition.

Doubts concerning the competitive model's utility as a policy guide prompted a search for more operational norms of "workable competition." The phrase was coined by J. M. Clark, who observed in his seminal article that perfect competition "does not and cannot exist and has presumably never existed" and that the competitive model of theory affords no reliable standard for judging real-world market conditions.[95] Clark went on to argue that some departures from the purely and perfectly competitive norm are not as harmful from a longer-run perspective as was commonly supposed, and he formulated certain minimal criteria for judging the workability of competition. The criteria he chose were influenced by the depression psychosis of the times and are less important than the impact Clark's work had in stimulating other economists.

The result was an explosion of articles on workable competition, many in substantial disagreement with one another. We will not attempt to review the literature here, since the job has been done admirably by Stephen Sosnick.[96] Using Sosnick's general schema, the criteria of workability suggested especially frequently by diverse writers can be divided into structural, conduct, and performance categories.

*Structural* criteria:

- The number of traders should be at least as large as scale economies permit.
- There should be no artificial inhibitions on mobility and entry.
- There should be moderate and price-sensitive quality differentials in the products offered.

*Conduct* criteria:

- Some uncertainty should exist in the minds of rivals as to whether price initiatives will be followed.
- Firms should strive to attain their goals independently, without collusion.
- There should be no unfair, exclusionary, predatory, or coercive tactics.
- Inefficient suppliers and customers should not be shielded permanently.
- Sales promotion should be informative, or at least not be misleading.
- There should be no persistent, harmful price discrimination.

93. Contributions since the second edition of this text appeared include James L. Bothwell, "Profitability, Risk, and the Separation of Ownership from Control," *Journal of Industrial Economics,* vol. 28 (March 1980), pp. 303–311; Stigler and Friedland, "The Literature of Economics," supra note 83, pp. 254–258; Harold Demsetz and Kenneth Lehn, "The Structure of Corporate Ownership: Causes and Consequences," *Journal of Political Economy,* vol. 93 (December 1985), pp. 1155–1177; Randall Mørck, Andrei Shleifer, and Robert Vishny, "Management Ownership and Market Valuation," *Journal of Financial Economics,* vol. 20 (January/March 1988), pp. 293–315; and Wi-Saeng Kim and Esmeralda O. Lyn, "Excess Market Value, Market Power, and Inside Ownership Structure," *Review of Industrial Organization,* vol. 3 (Fall 1988), pp. 1–26.

94. *Profits in the Long Run,* supra note 84, pp. 149–157.

95. J. M. Clark, "Toward a Concept of Workable Competition," *American Economic Review,* vol. 30 (June 1940), pp. 241–256. For an extension, see Clark's *Competition as a Dynamic Process,* supra note 3, especially Chapters 2–4.

96. Stephen Sosnick, "A Critique of Concepts of Workable Competition," *Quarterly Journal of Economics,* vol. 72 (August 1958), pp. 380–423.

*Performance* criteria:

- Firms' production and distribution operations should be efficient and not wasteful of resources.
- Output levels and product quality (that is, variety, durability, safety, reliability, and so forth) should be responsive to consumer demands.
- Profits should be at levels just sufficient to reward investment, efficiency, and innovation.
- Prices should encourage rational choice, guide markets toward equilibrium, and not intensify cyclical instability.
- Opportunities for introducing technically superior new products and processes should be exploited.
- Promotional expenses should not be excessive.
- Success should accrue to sellers who best serve consumer wants.

Critics of the workable competition concept have questioned whether the approach is as operational as its proponents intended. On many of the individual variables, difficult quantitative judgments are required. How price sensitive must quality differentials be? When are promotional expenses excessive, and when are they not? How long must price discrimination persist before it is persistent? And so on. Furthermore, fulfillment of many criteria is difficult to measure. For instance, to determine whether firms' production operations have been efficient, one needs a yardstick calibrated against what is possible. Finally and most important, how should the workability of competition be evaluated when some, but not all, of the criteria are satisfied? If, for example, performance but not structure conforms to the norms, should we conclude that competition is workable, since it is performance that really counts in the end? Perhaps not, because with an unworkable structure there is always a risk that future performance will deteriorate. If stress *is* placed on performance, what conclusion can be drawn when performance is good on some dimensions but not on others? Here a decision cannot be reached without introducing subjective value judgments about the importance of the various dimensions. And as George Stigler warned with characteristic irony, embarrassing disagreements may result:

> To determine whether any industry is workably competitive, therefore, simply have a good graduate student write his dissertation on the industry and render a verdict. It is crucial to this test, of course, that no second graduate student be allowed to study the industry.[97]

To investigate these weighting and consensus problems, Steven Cox obtained from forty-two economists, marketing professors, and business writers responses to questionnaires eliciting evaluations of the quality of fourteen major U.S. industries' 1960–1969 performance, both overall and on four subdimensions – product pricing (defined to approximate the criteria of allocative efficiency), technological progressiveness, cost minimization, and wage-price inflation.[98] There was a moderately high level of agreement among the panelists in their overall performance judgments, with the strongest consensus emerging among the panelists claiming greatest knowledge of the fourteen industries. In evaluations of overall performance, it was clear from a factor analysis that panelists placed by far the greatest



weight on whether an industry was technically progressive: the more progressive it was, the higher its performance was ranked. Among the academic economist panelists, high scores on the product pricing subdimension and low advertising expenditures as a percentage of sales (the latter calculated from nonquestionnaire data) also led to significantly higher overall performance ranks. But among the business specialists (that is, business journal writers and marketing professors) the opposite propensity held: industries that spent a *large* fraction of their sales dollar on advertising were ranked more favorably, as were industries yielding what appeared to be supra-normal profits. This shows that the most severe stumbling block in evaluating industrial performance is likely to be securing agreement on what is considered good or bad attributes of performance. Conflicting value judgments concerning performance attributes and their weights undoubtedly underlie many disputes as to the proper public policy toward monopolistic business enterprises.

## Conclusion

Readers seeking a precise, certain guide to public policy are bound to be disappointed by this survey, for we have found none. The competitive norm does seem to serve as a good first approximation, but it is difficult to state in advance how much competition is needed to achieve desirable economic performance, nor can we formulate hard and fast rules for identifying cases in which a departure from competition is desirable. We therefore begin our journey with only a primitive road map to guide us. Let us hope that we can avoid going too far astray and end with experience useful in drawing a better map.

97. George J. Stigler, "Report on Antitrust Policy – Discussion," *American Economic Review*, vol. 46 (May 1956), p. 505.

98. Steven R. Cox, "An Industrial Performance Evaluation Experiment," *Journal of Industrial Economics*, vol. 22 (March 1974), pp. 199–214.

## 17

# Market Structure, Patents, and Technological Innovation

Making the best use of resources at any moment in time is important. But in the long run, it is dynamic performance that counts. As we observed in Chapter 2, an output handicap amounting to 10 percent of gross national product owing to static inefficiency is surmounted in twenty years if the output growth rate can be raised through more rapid technological progress from 3.0 to 3.5 percent. Or if the growth rate can be increased to 4.0 percent, the initial disadvantage is overcome in 10.6 years.

To Adam Smith, the pin factory was the epitome of static efficiency.[1] Through mechanization and the division of labor, an average output per worker of 4,800 pins per day could be achieved in the 1770s. Two centuries later, thanks to countless technological changes, output per worker had risen to 800,000 pins per day.[2] The increase seems astounding at first glance, but it implies an average annual work-day productivity growth rate of only 2.56 percent per year—a rate matched by many industries over substantial periods of time.

From the time of Smith's successor David Ricardo, and especially after the Neoclassical breakthrough of the 1870s, until well into the twentieth century, the mainstream of (non-Marxian) economic theory exhibited remarkably little sensitivity to the importance of compound productivity growth through technological innovation. Emphasis was on the result of combining labor and capital with production functions of an essentially static character. Not until the 1950s did technological change become more than a side-show attraction. For the shot that signaled a revolution in economic thought, Robert Solow received the Nobel Prize.[3] He set out to measure the extent to which increases in the amount of capital employed were responsible for the rise of U.S. nonfarm output per labor hour between 1909 and 1949. To the surprise of economists mired in the static tradition, he found that increased capital intensity accounted for only 12.5 percent (later corrected to 19 percent) of the measured growth in output per work hour. The rest of the observed 1.79 percent average annual productivity gain was evidently attributable to improvements in production practices and equipment (technological change in the strictest sense) and to the increased ability of the labor force. In a subsequent extension, Edward Denison estimated that 13 percent of the gain in output per worker between 1929 and 1982 could be credited to increased capital intensity, 34 percent to improved work force education, 22 percent to the greater realization of scale economies, and 68 percent to advances in scientific and technological knowledge, broadly construed.[4] Although one can quibble with the detailed estimates, it is hard to dispute the main thrust of Solow's and Denison's conclusion: that the growth of output per worker in the United States (and also in other industrialized lands[5]) has come predominantly from the

1. Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776), Book I, Chapter I.
2. Clifford F. Pratten, "The Manufacture of Pins," *Journal of Economic Literature*, vol. 18 (March 1980), pp. 93–96.
3. Robert M. Solow, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, vol. 39 (August 1957), pp. 312–320.
4. Edward F. Denison, *Trends in American Economic Growth, 1929–1982* (Washington: Brookings, 1985) p. 30. The percentages add to more than 100 because there are also negative factors, for example, a decrease in hours worked by the average employee (−25 percent) and the effects of government regulation (−4 percent).
5. See, for example, Edward F. Denison, *Why Growth Rates Differ* (Washington: Brookings, 1967).

application of new, superior production techniques by an increasingly skilled work force. It seems clear too that imperfectly understood declines in the pace of technological innovation were responsible in part for a fall in the growth of U.S. business-sector productivity from 2.92 percent per year over 1947–1973 to 1.01 percent per year over 1973–1987.[6]

The introduction of better production methods is one main arm of technological advance. The other is the creation of new and superior consumer goods. Because of measurement difficulties, it is doubtful whether the full effect of consumer product innovations is captured by available productivity growth statistics. Consequently, the impact of technological change on consumer well-being is probably understated by those statistics.

Several further effects of technological change can be identified. For one, advanced industrialized nations like the United States, Japan, and Sweden derive much of their comparative advantage in international trade not from the land, labor, and capital endowments stressed in neoclassical economic theory, but from superiority in developing and producing technologically advanced products such as aircraft, machine tools, electronic appliances, pharmaceuticals, and computers.[7] Second, international differences in the ability to develop and apply modern technology have a crucial impact on the balance of military power. Third, process innovation alters the structure of labor demands, most likely strengthening the demand for skilled workers and weakening demand for the unskilled, with troubling implications for income distribution. Finally, technological change affects market structure, for major innovations often bring new firms to the fore and displace laggards, defining the structural conditions within which price and other more static forms of rivalry are conducted for decades to come.

Here we are concerned largely with a possible causal flow in the opposite direction: from market structure to innovation. Is progress faster or slower under monopolistic conditions, or does it make no difference? A leader in stressing the important role technological change plays in capitalistic economies was Joseph A. Schumpeter. In a widely read and controversial book, he argued that market structure does make a difference. Despite their restrictive pricing behavior, he asserted, large monopolistic firms are ideally suited for introducing the technological innovations that benefit society:

> What we have got to accept is that [the large-scale establishment or unit of control] has come to be the most powerful engine of [economic] progress. . . . In this respect, perfect competition is not only impossible but inferior, and has no title to being set up as a model of ideal efficiency.[8]

Whether or not this "Schumpeterian" view is correct is the question we tackle in this chapter. More precisely, we explore several narrower issues: the incentive role of patents, the links between market structure and innovation, and the climate for innovation in large, diversified firms.

## Industrial Innovation

Before addressing these questions, we must pause and examine more carefully what the process of technological change is all about.



### Organized Industrial R&D

Technical innovations do not fall like manna from heaven. They require effort—the creative labor of invention, development, testing, and introduction into the stream of economic life. To some extent innovative effort is a haphazard thing, conducted by individuals and firms as a digression from routine workaday activities. But to an increasing degree, the task of creating and developing new products and processes has been institutionalized through the establishment of formal research and development (R&D) laboratories. This phenomenon evolved gradually. In the 1770s and 1780s, the firm of Boulton & Watt had the equivalent of an R&D laboratory for work on steam engines.[9] The genesis of the modern R&D laboratory in America is commonly traced to 1876, when Thomas Edison opened his famed laboratory in Menlo Park and Alexander Graham Bell established a similar facility in Boston. Wherever the starting point is placed, the idea spread rapidly until research and development came to be big business. In 1987, U.S. companies expended some $91 billion on research and development activities—2.4 percent of business-sector gross national product.[10] Real (that is, inflation-adjusted) total industrial R&D spending rose at an average rate of 6.6 percent per year between 1953 and 1970 and at the slower 3.4 percent rate from 1970 to 1987. Thirty-six percent of 1987 industrial R&D outlays were financed under government contracts, mostly covering military and space systems development, while for the balance, company funds were directly at risk. Our emphasis here is on the company-financed component.[11]

The intensity with which companies pursue privately financed research and development varies widely from industry to industry. The manufacturing sector conducts 97 percent of all industrial R&D and hence is the prime mover in generating technological progress. Among 238 U.S. manufacturing industries in 1977, the median industry devoted 0.8 percent of sales to company-financed R&D. The eight R&D/sales leaders were:[12]

| | |
|---|---|
| Ethical drugs | 10.2% |
| Electronic computing equipment | 8.9 |

6. See Martin Baily, "What Has Happened to Productivity Growth?" *Science*, October 24, 1986, pp. 443–451; and F. M. Scherer, "The World Productivity Growth Slump," in Rolf Wolff, ed., *Organizing Industrial Development* (Berlin: de Gruyter, 1986), pp. 15–27. The data here are from the *Economic Report of the President* (Washington: USGPO, 1988), p. 300.

7. See Edward M. Graham, "Technological Innovation and the Dynamics of the U.S. Comparative Advantage in International Trade," in Christopher T. Hill and James M. Utterback, *Technological Innovation for a Dynamic Economy* (New York: Pergamon, 1979), pp. 118–160; Leo Sveikauskas, "Science and Technology in United States Foreign Trade," *Economic Journal*, vol. 93 (September 1983), pp. 542–554; Kirsty S. Hughes, "Exports and Innovation," *European Economic Review*, vol. 30 (April 1986), pp. 383–399; David B. Audretsch and Hideki Yamawaki, "R&D Rivalry, Industrial Policy, and U.S.-Japanese Trade," *Review of Economics and Statistics*, vol. 70 (August 1988), pp. 438–447; and U.S. National Science Board, *Science & Engineering Indicators: 1987* (Washington: USGPO, 1988), pp. 15, 124–136, and 313–326.

8. Joseph A. Schumpeter, *Capitalism, Socialism, and Democracy* (New York: Harper, 1942), p. 106.

9. See F. M. Scherer, *Innovation and Growth. Schumpeterian Perspectives* (Cambridge: MIT Press, 1984), Chapter 2.

10. *Science & Engineering Indicators*, supra note 7, p. 294.

11. On structure-performance links in federally supported programs, see William Burnett and F. M. Scherer, "The Weapons Industry," in Walter Adams, ed., *The Structure of American Industry* (8th ed.; New York: Macmillan, 1989).

12. Federal Trade Commission, *Statistical Report: Annual Line of Business Report, 1977* (Washington: April 1985), p. 21. The FTC Line of Business reports for 1974 through 1977 are the only reliable source of R&D spending data disaggregated to the four-digit level of detail.

WCK1130
Wockhardt Bio AG v. Janssen Oncology, Inc.
IPR2016-01582

| | |
|---|---|
| Aircraft engines and parts | 8.4 |
| Calculating and accounting machines | 7.3 |
| Photographic equipment and supplies | 6.3 |
| Semiconductors | 6.1 |
| Photocopying equipment | 5.7 |
| Optical instruments and lenses | 5.5% |

## What R&D Laboratories Do

Approximately 3 percent of U.S. industry's R&D effort goes into basic research, defined by the National Science Foundation as "original investigation for the advancement of scientific knowledge, without specific commercial objectives." Most basic research by this definition is done not in industry, but by academic and nonprofit institutions (56 percent of the total in 1985) and government laboratories (24 percent). Industry's forte is applied research and the development of new or improved products and processes. From a count of the patents resulting from the 1974 R&D efforts of 443 large U.S. corporations, the distribution of industrial inventions by intended user orientation was as follows:[13]

| | |
|---|---|
| Production processes for internal company use | 26.2% |
| Capital goods for use by other industries | 44.8 |
| Same, with consumer usage too | 7.8% |
| Materials for use by other industries | 21.6 |
| Same, with consumer usage too | 8.7% |
| Consumer goods only | 7.4% |

What is particularly striking is the revealed emphasis on inventions by one industry that flow to other industries to improve the recipients' products and production processes.[14] Thus, a new fiber from du Pont improves the fabrics sold by textile weavers to garment makers, which only then enter the consumer's wardrobe; or a new aircraft engine from General Electric is sold to the airlines to let their planes operate more reliably, quietly, and economically. A surprisingly small fraction of industrial R&D is oriented toward improving products that will serve solely as consumer goods.

Another useful way of viewing what goes on in industrial R&D laboratories is to extend a trichotomy originally proposed by Schumpeter into a five-function schema: invention, entrepreneurship, investment, development, and diffusion.[15] Invention is the act of insight by which a new and promising technical possibility is worked out (at least mentally, and usually also physically) in its essential, most rudimentary form. Development is the lengthy sequence of detail-oriented technical activities, including trial-and-error testing, through which the original concept is modified and perfected until it is ready for commercial introduction. The entrepreneurial function involves deciding to go forward with the effort, organizing it, obtaining financial support, and cultivating the market. Investment is the act of risking funds for the venture. These creative functions need not be performed by the same person or even by the same organization. Indeed, they are often organizationally separate. Finally, diffusion (or imitation) is the process by which an



innovation comes into widespread use as one producer after another follows the pioneering firm's lead.

## Two Illustrations

To illustrate the first four functions, it is helpful to consider two brief examples. No case is completely typical. The histories presented here were chosen because they reflect unusually important and ambitious technical changes and because the stages and functions stand out with particular clarity. One, the Watt-Boulton steam engine venture of the 1770s, is ancient; the other, xerography, relatively new.

It is generally accepted that James Watt "invented" his improved steam engine in 1765, when he repaired a Newcomen steam engine model owned by Glasgow University and perceived that its efficiency could be greatly enhanced by condensing the steam outside the operating cylinder. He wrote later that "In three days, I had a model at work nearly as perfect . . . as any which have been made since that time."[16] But much remained to be done before he could supply a machine useful in industrial practice. Full-scale models had to be built, condenser concepts had to be devised and tested, valves designed, methods of machining and sealing the operating cylinder perfected, and so forth. All this required time and money, and for want of both financial support and entrepreneurial initiative, Watt twice abandoned the venture to work as a salaried engineer. Not until Matthew Boulton came forward to provide these missing ingredients was a full-scale model completed, and the first commercially useful Watt-Boulton steam engine was installed only in 1776, eleven years after the original invention. Expenditures preparing the way for operating the first commercial engine amounted to the equivalent of at least 60 person-years of skilled labor.

While working as a patent attorney, Chester Carlson was impressed by the difficulty of copying documents efficiently.[17] For several years he spent much of his spare time after work mulling over the problem and browsing in potentially relevant technical literature, eventually (in 1938) conceiving the basic idea of xerography. With the assistance of an unemployed physicist, he successfully tested his concept through an extremely crude model. After numerous fruitless attempts to interest industrial firms in helping him develop a commercially practical copying system, he enlisted in 1944 the cooperation of the Battelle Development Company, a subsidiary of the nonprofit Battelle Institute. From two years' labor by a Battelle research physicist and his aides came two key inventions building upon Carlson's original principle: the use of a selenium-coated plate to store the electrostatic image, and the corona discharge method for sensitizing the plate and applying

13. Scherer, *Innovation and Growth*, supra note 9, p. 36.

14. Strictly defined, R&D aimed at improving one's own productivity is *process* R&D. When a new machine is developed for sale to enhance another industry's productivity, the work is *product* R&D for its performer.

15. See Joseph A. Schumpeter, *The Theory of Economic Development*, trans. by Redvers Opie (Cambridge: Harvard University Press, 1934), Chapter 2; and Scherer, *Innovation and Growth*, supra note 9, Chapters 1 and 2.

16. Quoted in Scherer, *Innovation and Growth*, p. 10.

17. This example is drawn from John Jewkes, David Sawers, and Richard Stillerman, *The Sources of Invention*, 2nd ed. (New York: Norton, 1969), pp. 321–323; and Erwin Blackstone, "The Economics of the Copying Machine Industry" (Ph.D. diss., University of Michigan, 1968).

ink to the copying paper. With these inventions, xerography began to show distinct signs of commercial promise.

At this point the Haloid Corporation (later renamed the Xerox Corporation) took over developmental responsibility. By 1950 its engineers had completed a prototype system useful primarily for making offset lithography masters. This was a cumbersome, three-machine contraption, however, with limited market potential. The company then devoted its resources to developing a single-unit console copier—a task that required surmounting a difficult lens design problem and numerous lesser engineering challenges. The result was the 914 copier, which took the world by storm after its introduction in 1959. By that time, Xerox had accumulated a portfolio of nearly 300 issued and pending patents on its copying machine inventions.

During the more than two decades preceding the 914 copier's debut, formulation of the basic xerographic copying concept consumed the inventor's energies part-time for a very few years. The selenium plate and corona discharge inventions came from an only slightly greater increment of effort. After Haloid entered the picture, it expended roughly $4 million on research and development up to 1953, when it redirected attention to devising a console model. To attain that goal, further R&D outlays estimated at $16 million, severely straining the company's financial resources, were committed.

Several generalizations can be extracted from these and other innovation case studies. First, the initial invention that precipitates a major innovative effort is typically inexpensive, both relatively and absolutely. Its money cost is often so modest that almost any well-prepared imaginative individual thrown into contact with the problem might achieve the essential insight. The inventive challenge may be recognized as a result of formal work assignments, as a by-product of work or leisure pursuits, or from any of the 101 experiences a person has each day. If industrial R&D laboratories enjoy comparative advantage in generating inventions, it is because they are more likely to put together the critical combination of a fertile mind, a challenging problem, and the will to solve it.

Second, there is a substantial random component in fundamental invention. Thousands of persons may recognize an unsolved problem or unmet need, but only a fraction will be sufficiently intrigued to devote serious thought to it, and an even smaller fraction will have the ingenuity and good luck to gain a correct insight by viewing the problem in exactly the right way—that is, in the proper gestalt.[18] After the insight is achieved, the solution may seem obvious, but before the fact invention is largely unpredictable. If this were not so, every problem, once recognized, would be solved quickly.

Third, supporting inventions of greater or lesser creative magnitude may be required before the innovation begins to look technically and economically viable. Once the original insight is attained, however, it forms a gestalt within which such supporting inventions will tend inevitably to emerge if good minds are focused on the problem.

Fourth, when the necessary conceptual advances have occurred and when their essential correctness has been demonstrated, typically through crude model tests requiring only a modest resource investment, the uncertainties associated with innovation are transformed qualitatively and quantitatively. The question, "Is there something interesting and technically feasible here?" is no longer a serious



issue. Uncertainty centers on such questions as: What will the detailed configuration be? How well can it be made to work? How much will perfecting it cost? How long will it take? At what price can it be sold? What will be the market demand at that price? These are not negligible uncertainties, but usually they are not overwhelming or outside the bounds of entrepreneurial experience. The vast majority of all industrial R&D projects, it should be noted, begin at this stage, since they embody no fundamental new concepts, or build upon insights achieved elsewhere.

Finally, once the sequence has progressed this far, outlays much greater than those incurred during the early conceptual stages are necessary before an innovation is brought to the point of commercial utility. The investment decision at this juncture entails committing possibly substantial resources in the face of moderate technological uncertainties. It is quite different from the earlier stage, where the amounts of money at risk are small, but the technical uncertainties are great. If inexpensive conceptual work has not reduced the degree of technical uncertainty to tolerable levels, a decision to move into full-scale development will be taken only under the most unusual pressures (as in the atomic and hydrogen bomb programs[19]). Normally, conceptual work will be continued at a low spending level until the main technical uncertainties have been resolved.

## The Costs and Risks of R&D

The costs of individual industrial R&D projects vary widely. Most projects entail fairly modest outlays. Since 1963 the magazine *Research & Development* (earlier, *Industrial Research*) has conducted an annual competition to name the one hundred most significant technical advances of the year. On average, the winners in the years 1984 to 1986 had an average R&D cost of $2.47 million, with a range from $1,000 to $110 million.[20] However, the survey excludes some of the most complex and costly innovations. The original version of the F-18 fighter aircraft cost roughly $1 billion to develop during the late 1970s, and the B-2 Stealth bomber, first flown in 1989, absorbed $22 billion of R&D funds. Such military projects are so expensive and risky that they almost always go forward only with more or less full financing under government contracts. The most expensive projects undertaken under private initiative have been commercial airliners. In the late 1980s, developing a completely new long-range subsonic airliner was estimated to cost nearly $3 billion.

It is useful to think about R&D project costs in terms of a frequency distribution. A census would undoubtedly reveal the distribution to be highly skewed. The spectacularly costly projects that receive the most press attention are few in number, forming the distribution's long thin tail.[21] Smaller projects are much more numerous, giving rise to a peak or mode in a spending range near $1 million

18. This interpretation of the inventive act is based upon Abbott P. Usher, *A History of Mechanical Inventions*, rev. ed. (Cambridge: Harvard University Press, 1954), Chapter 4; N. R. Hanson, *Patterns of Discovery* (Cambridge: Cambridge University Press, 1958); and Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

19. See Richard Rhodes, *The Making of the Atomic Bomb* (New York: Simon & Schuster, 1986).

20. "Developers of 100 Significant Products Feted," *Research & Development*, October 1986, p. 55. For earlier cost data, see F. M. Scherer, *The Economic Effects of Compulsory Patent Licensing* (Monograph 1977-2, New York University Graduate School of Business Administration, 1977), p. 15, pulling together the results of several surveys by Edwin Mansfield and associates.

21. The distribution's extreme skewness is indicated by the wide range of values reported for the most costly of the 100 annual *Research & Development* citation winners: $40 million in 1981, $8 million in 1982, $960 million in 1983, $30 million in 1984, $50 million in 1985, and $110 million in 1986.

at 1987 price levels. The parameters of this distribution have been shifting over time; that is, the modal R&D project today is more expensive in constant-dollar terms than its counterpart thirty years ago.

However, the sizes of business enterprises have also been rising. In 1987, the two hundred fiftieth firm on *Fortune*'s list of the 500 largest industrials had sales of $1.39 billion. The average manufacturing corporation spent roughly 2.3 percent of its sales on R&D in that year. These two statistics imply an annual R&D budget of $40 million. Assuming an average R&D project to entail total expenditures of $2.5 million spread over three years,[22] our two hundred fiftieth industrial company could maintain a portfolio of roughly forty-eight projects. Thus, unless it inhabits an industry in which "big ticket" developments are the norm, the typical medium-sized corporation could hedge its risks with a fairly sizable and well-diversified R&D project portfolio.

The riskiness of individual corporate R&D projects has been illuminated in a series of studies by Edwin Mansfield and colleagues. In one such survey, they discovered that among projects carried out during 1963 to 1965, the average fraction fulfilling their technical objectives was 70 percent in seven chemical laboratories, 32 percent in five drug companies, 73 percent in three electronics organizations, and 50 percent in the laboratories of four petroleum companies.[23] The reason for the characteristically high rate of technical success has been brought out already: firms do not as a rule begin new product or process development until the principal technical uncertainties have been whittled down through inexpensive research, conducted either by their own personnel or by outsiders. Still the possibility of technical failure is not the only risk borne in industrial R&D. Consumer reactions and/or the size of the cost reductions achieved with new production processes may also be misjudged. Analyzing the fate of projects undertaken in the laboratories of sixteen chemical, pharmaceutical, electronics, and petroleum companies, Mansfield and associates identified three different success probabilities: (1) the probability that technical goals would be achieved; (2) the probability that, conditional upon technical success, the resulting product or process would be commercialized; and (3) given commercialization, the probability that the project yielded a return on investment at least as high as the opportunity cost of the firm's capital.[24] For all sixteen firms combined, the average conditional probabilities were:

| | |
|---|---|
| Technical success | 0.57 |
| Commercialization, given technical success | 0.65 |
| Financial success, given commercialization | 0.74 |

The firms' overall track record is assessed by multiplying the three conditional probabilities. Thus, on average, 27 percent of the projects initiated ultimately achieved financial success. For the prototypical *Fortune* 250 firm whose portfolio was characterized above, this means that its R&D laboratories would deliver four financial successes in a typical year. To be sure, the averages derived by Mansfield et al. were attended by considerable variation, so some of the firms studied did even better than average while others reaped barren harvests.



## The Logic of Patent Protection

To reward those who invest their time and money in technological invention and innovation, and thus to encourage such investment, has been the classic function of invention patents since the first patents were granted in fifteenth-century Italy.[25] Modern patents, awarded to the first inventor of some new and useful product or process, give their holder the right to exclusive use of the invention for a specified period of time. This right includes the right to enjoin others from using the invention, or to license as few or as many other users as the patent holder chooses.

Most nations conferring patent rights accept broad guidelines stemming, with subsequent amendments, from the Paris Convention of 1883. In 1987, the Paris Convention had been adopted by ninety-seven member states, including thirty industrialized western nations, twelve eastern (that is, communist) nations, and fifty-five less-developed nations. Under the prevailing law of most western European nations, which was harmonized by the European Patent Convention in 1977, patents are issued to the first person to file an application validly claiming to have made an invention. The exclusive right they confer lasts for twenty years from the day of application. Under U.S. law, the normal life is seventeen years following the date a patent is issued. Priority under the U.S. system (emulated only by Canada and the Philippines) goes to the "first to invent," whose determination in contested cases entails a complex admixture of three criteria—who was first to conceive the idea, who was first to reduce it to practice, and whether the party exercised reasonable diligence in reducing the invention to practice. Between the application and issue dates is an examination period averaging roughly two years, but sometimes lasting as long as fifteen years, during which the Patent Office decides whether the claimed invention is indeed new, inventive, and useful, and, in cases of dispute, who the true inventor is.

Between 1980 and 1986, the U.S. Patent Office issued an average of 64,580 patents per year.[26] From the late 1960s until 1986, there was a declining trend in the total number of patents issued. There has also been a drop in the fraction of U.S. patents gained by domestic inventors. During the 1950s, foreign inventors received only 7 percent of U.S. patents. By 1980–1986, their share had increased to 42.6 percent, with Japanese inventors alone achieving a 15 percent share. A trend that predated the increasing internationalization of U.S. patenting was the shift from individual to corporate invention. When an employee makes an invention in the course of normal duties, the inventor typically "assigns" his or her rights

22. The mean outlay assumed here exceeds the mode because of the disproportionate influence of the largest projects.

23. Edwin Mansfield et al., *Research and Innovation in the Modern Corporation* (New York: Norton, 1971), pp. 34–35. See also Mansfield, *Industrial Research and Technological Innovation* (New York: Norton, 1968), pp. 56–61.

24. Edwin Mansfield, Samuel Wagner, et al., *The Production and Application of New Industrial Technology* (New York: Norton, 1977), pp. 22–32.

25. For new insights into the early history of patent grants and a thorough exploration of the underlying economic logic, see Erich Kaufer, *The Economics of the Patent System* (Chur: Harwood, 1988). An earlier classic is Fritz Machlup, *An Economic Review of the Patent System*, Study No. 15 of the Senate Committee on the Judiciary, Subcommittee on Patents, Trademarks, and Copyrights (Washington: USGPO, 1958).

26. Recent data are summarized in *Science & Engineering Indicators*, supra note 7, pp. 302–303. Longer time series on a slightly different classification basis are found in the annual *Statistical Abstract of the United States*. The most comprehensive source on other nations' patent grants is the annual statistical report in the World Intellectual Property Organization's journal, *Industrial Property*.

to the employing corporation, in whose name the patent is then issued. Thus, a distinction is made between corporate and "individual" patent grants. The share of all U.S. patents issued to individual inventors was 91 percent in 1901, 72 percent in 1921, 42 percent in 1940, 25 percent during the 1960s, and 18.5 percent during the early 1980s.

### The Basic Logic

The funds supporting invention and the commercial development of inventions are front-end "sunk" investments; once they have been spent, they are an irretrievable bygone. To warrant making such investments, an individual inventor or corporation must expect that once commercialization occurs, product prices can be held above postinvention production and marketing costs long enough so that the discounted present value of the profits (or more accurately, quasi rents) will exceed the value of the front-end investment. In other words, the investor must expect some degree of protection from competition, or some monopoly power. The patent holder's right to exclude imitating users is intended to create or strengthen that expectation. Patents also confer a property right which the original patent holder can sell, recouping its original investment and letting another entity exclusively commercialize the patented subject matter. Partial "sale" is also possible, for example, when the patent holder licenses others to exploit the invention and charges a royalty for the right.[27]

The simplest case of a product innovation covered by patent protection is shown in Figure 17.1(*a*). If the product is really new and useful, it creates a wholly new demand curve $D_1$ — one that did not exist previously. With an exclusive right to make and sell its product, the patent holder is a monopolist. It derives its marginal revenue $MR_1$, equates marginal revenue with marginal production and distribution cost MC, and sets price $OP_1$, realizing "monopoly" profits in the amount of rectangular area $P_1AXM$. These are not pure profits, however, because the innovator's sunk R&D costs must be taken into account. To make that one-time lump sum consistent with Figure 17.1(*a*), which is expressed in annual "flow" terms, let us assume that the innovator finances its R&D investment by taking out a seventeen-year mortgage whose annual payment obligation is given by the area of the inset rectangle IJKL.[28] If the patent monopoly lasts for seventeen years, the annual "profit" $P_1AXM$ will more than cover the annual R&D debt service cost, and the innovator will be well compensated for its efforts. It is not true, however, that the monopoly innovator is the only one to gain. The ordinates of demand curve $D_1$ array the values diverse consumers place upon having the new product to consume. The product's availability on monopolized terms generates not only producer's surplus $P_1AXM$, but also consumers' surplus $BAP_1$. With linear demand and constant marginal production and distribution costs, as shown in Figure 17.1(*a*), the monopolist is said to "appropriate" to itself only two-thirds of the total surplus its product creates.[29] The remaining third goes to consumers.

Suppose, however, that there were no patent protection and no other barriers to the imitation of the innovator's invention. Then a scenario like the one shown in panel (b) of Figure 17.1 might unfold. Soon after the new product appears, competing firms will introduce their imitating products, squeezing the demand schedule left for the original innovator to $D_2$. With less residual demand, the innovator must derive a new marginal revenue function $MR_2$ and set a new, lower price



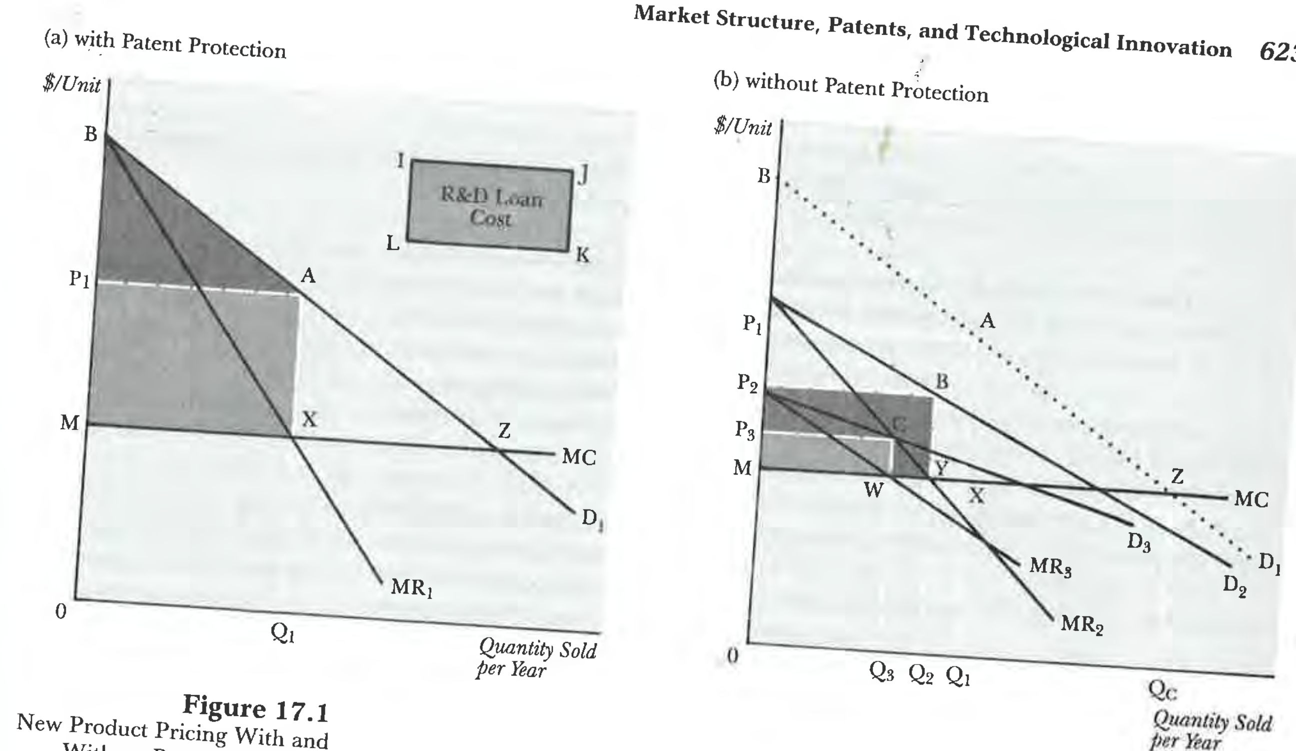



**Figure 17.1**
New Product Pricing With and Without Patent Protection

$OP_2$, which yields profit rectangle $P_2BYM$ barely covering the innovator's R&D debt service cost. However, the imitator firms may have had to incur little or no R&D cost on their own, "free riding" on the innovator's R&D, and thus, with unit costs of only OM, they will realize supra-normal profits at price $OP_2$. More competitors will be drawn in by this price and profit lure, squeezing the innovator's residual demand curve further to $D_3$. The innovator must reoptimize again, setting profit-maximizing price $OP_3$ and capturing "profit" rectangle $P_3CWM$, which is now smaller than the continuing R&D debt service obligations. If this

27. On the economics of licensing, see John S. McGee, "Patent Exploitation: Some Economic and Legal Problems," *Journal of Law & Economics*, vol. 9 (October 1966), pp. 135–162; Richard E. Caves, Harold Crookell, and J. Peter Killing, "The Imperfect Market for Technology Licenses," *Oxford Bulletin of Economics and Statistics*, vol. 45 (August 1983), pp. 249–267; and Morton I. Kamien, Y. Tauman, and I. Zang, "Optimal License Fees for a New Product," *Mathematical Social Sciences*, vol. 16 (August 1988), pp. 77–106.

28. Where RD is the original R&D investment, the yearly mortgage payment will be given by:

$$M(t) = \frac{RD}{\frac{1 - (1 + i)^{-17}}{i}}$$

with i the annual rate of interest and 17 the number of years over which the mortgage is paid down. If the R&D investment is $1 million and the interest rate i is 10 percent, the annual debt service will be $124,664.

29. Proof: From footnote 13, Chapter 2, MR falls by twice as much as $D_1$ over output range $OQ_1$. Thus, $BM = 2\ BP_1$ and $BP_1 = P_1M$. With height and base equal to those of rectangle $P_1AXM$, triangle $BAP_1$ has half the rectangle's area.

diffusion process were to unfold rapidly and if the innovator correctly foresaw its course, the innovator would perceive that its R&D costs will not be recouped and would therefore choose not to invest in the R&D. Consumers will be deprived of a valuable new product—one that, even under pure monopoly conditions, could yield them a sizable consumers' surplus.

Pursuing the analysis a step further, we see a kind of dilemma. *If* the R&D investment were incurred and the innovation made, imitative entry might, absent patent protection, continue until the price is driven all the way down to the competitive level, ignoring the innovator's front-end costs—that is, to OM. If this happens, surplus $P_1AXM$, originally captured by the innovator, will be transformed into consumers' surplus. In addition, the competitive expansion of output to $OQ_C$ leads to the emergence of still more consumers' surplus, measured by triangle AZX. In this limiting case, the innovator appropriates none of the (now larger) surplus its invention has created—consumers get it all. If the innovator is allowed to monopolize the new product's sale, its profit-maximizing output restriction means that total surplus will be less than it might ideally be by dead-weight loss triangle AZX. In this sense, granting patent monopolies imposes a cost upon society. Seeing this, consumers might urge that the government renege on its patent monopoly grant so they can have the best of all worlds—the new product, competitive pricing, and maximum surplus. But if this occurs with any frequency, would-be innovators will expect rapid imitation to erode their surpluses, causing them to lose money on their R&D investments, so they will not invest in additional new products. The technological well will run dry. The patent system makes a deliberate tradeoff, accepting during the patent grant's life dead-weight surplus losses in order to ensure that new products and processes, along with the surpluses they create, will not be discouraged by fear of rapid imitation. Only after the patent expires, when competitive imitation can run its full course, are consumers able to have their new product along with the extra surplus competitive pricing brings.[30]

## Complications

Although devised to solve an important incentive problem, the patent system is a crude and imperfect instrument. Because of diverse real-world complications, the patent protection given an innovator may be too little, too much, or of the wrong kind.

The protection provided is often weak because there can be many viable solutions to a technical problem, so other firms can "invent around" a given patented solution. Individual patents that solidly protect a whole field of product or process technology are rare, and when such cases occur, the credit is frequently due as much to the skill of the patent attorney as to the breadth of the inventor's vision. To be sure, companies often seek to fence in their technological domain by patenting every conceivable variation on a product or process.[31] Yet fences are also permeable. Du Pont, for example, took out hundreds of patents on variants of its nylon synthetic fiber technology. But even in the directly applicable polyamide molecule family it left a gap into which Germany's I.G. Farben moved with Perlon L, and other companies invented competitive fibers using polyester and polyolefin molecules.

Further complications emerge because the growth of technology is cumulative and richly interactive. Company B may patent an improvement on Company A's



invention, or Companies C and D may each hold patents on diverse features, all of which a state-of-the-art product should ideally incorporate. Each might if it wishes block the other from using desired complementary technology. Under Chinese patent law and Japanese practice, firms holding improvement patents on others' inventions can demand a license to the original patent by reciprocally offering to license their improvement. This logjam-breaking provision, absent from U.S. and European law, engendered considerable conflict between U.S. and Japanese industry leaders during the 1980s.[32] In the United States and Europe, the holders of complementary patents often agree voluntarily to cross license each other.[33] This enables all to achieve state-of-the-art technology but lessens the exclusionary power of patents. However, such cross licensing agreements have sometimes been used as a fulcrum for industrywide price-fixing and entry-excluding cartels that suppress competition more than would have been possible if each firm independently exploited its own patented technology. In the United States, cartelization through reciprocal licensing of patents has for the most part been dealt with harshly under the antitrust laws.[34]

For smaller and especially less-developed countries (LDCs), patent holders' power to block use of their inventions by others poses a special problem. Multinational corporations commonly patent their most important inventions in dozens of national jurisdictions. The quest for scale economies leads them to produce in one

30. Extending this logic, William D. Nordhaus developed a pioneering model to determine the optimal patent life tradeoff—that is, one that maximizes the discounted present value of consumers' plus producers' surpluses. See his *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change* (Cambridge: MIT Press, 1969), Chapter 5. He shows that the optimal patent life is longer: (1) the less price-elastic demand is (and hence the smaller the dead-weight loss AYX incurred until the patent expires); (2) the smaller the overall benefit from invention is in relation to the chosen R&D outlay; and (3) the more responsive the amount of invention is to an increase in R&D expenditures. In extensions, it has been shown that the optimal patent life is longer than Nordhaus' computations when patents provide imperfect protection against imitation or when there is compulsory licensing at controlled (that is, low) royalty rates. It is shorter when preinnovation competition raises R&D costs and reduces the net profitability of innovation. See, for example, Scherer, *Innovation and Growth*, Chapter 7; Pankaj Tandon, "Optimal Patents with Compulsory Licensing," *Journal of Political Economy*, vol. 90 (June 1982), pp. 470–486; D. G. McFetridge and M. Rafiquzzaman, "The Scope and Duration of the Patent Right and the Nature of Research Rivalry," with a comment by Roger Beck, in *Research in Law and Economics*, vol. 8 (Greenwich, CT; JAI Press, 1986), pp. 91–129; and Brian D. Wright, "The Economics of Invention Incentives," *American Economic Review*, vol. 73 (September 1983), pp. 691–707.

31. Under U.S. antitrust precedents, this is not in itself illegal. See *Automatic Radio Mfg. Co. v. Hazeltine Research, Inc.*, 339 U.S. 827, 834 (1950). However, when collusive or abusive practices were used in conjunction with the accumulation of patents, or when the accumulation was achieved through merger rather than one's own R&D, licensing of the accumulated patents has been required. See, for example, U.S. Senate, Committee on the Judiciary, Subcommittee on Patents, Trademarks, and Copyrights, staff report, *Compulsory Patent Licensing under Antitrust Judgments* (Washington: USGPO, 1960); and F. M. Scherer, *The Economic Effects of Compulsory Patent Licensing*, supra note 20, pp. 59–78.

32. See "An American Views Japan's Copycat Culture," *Wall Street Journal*, July 12, 1988, p. 33.

33. In "The Learning Curve, Technology Barriers to Entry, and Competitive Survival in the Chemical Processing Industries," *Strategic Management Journal* (1989), Marvin B. Lieberman reports that the licensing of technology to new entrants was more common in unconcentrated than in concentrated industries. On the licensing practices of West German firms, see Klaus Grefermann and K. C. Roethlingshöfer, *Patentwesen und technischer Fortschritt*, Teil II, "Patent- und Lizenzpolitik der Unternehmen" (Göttingen: Schwartz, 1974).

34. See U.S. Senate, *Compulsory Patent Licensing*, supra note 31; Floyd L. Vaughan, *The United States Patent System* (Norman: University of Oklahoma Press, 1956); and Louis Kaplow, "The Patent-Antitrust Intersection," *Harvard Law Review*, vol. 97 (June 1984), pp. 1813–1892. During the 1980s some backpedaling occurred, following arguments first emphasized by Ward S. Bowman, *Patent and Antitrust Law: A Legal and Economic Appraisal* (Chicago: University of Chicago Press, 1973), especially Chapter 10. See also Roger B. Andewelt, "Analysis of Patent Pools under the Antitrust Laws," *Antitrust Law Journal*, vol. 53, no. 3 (1985), pp. 611–638.

or a few preferred locations, exporting elsewhere and preventing the local exploitation of their technology in export markets by asserting their patent rights. For an LDC, this typically means that high prices will be paid for imported patented products, while opportunities to build a home industry using first-line technology are restricted. Many LDCs and some more industrialized nations have therefore included in their patent laws provisions requiring that the technology on which they issue patents be "worked" domestically within a few years or be subjected to compulsory licensing. Such provisions have been a focus of conflict between multinational corporations, stressing the rationale of patent incentives and efficiently centralized production, and LDCs, emphasizing their need to escape from backwardness by building dynamic modern industries.[35]

## Alternative Protection from Imitation

The patent grant is a tradeoff, but if an invention would be made and commercialized without patent protection, the terms are altered. The power to impose monopolistic restrictions remains, but what consumers get in the bargain would have been available on less restrictive terms. There are several reasons why competitive imitation might be impeded even without patents, leaving sufficient incentive for investments in research and development.

For one, to imitate one must know about the innovation and its advantages, and knowledge is almost always imperfect. Firms often protect their technological advances by keeping them secret for as long as they can.[36] Even when patent protection is sought, there is a lag from the time of invention to the time the patent (or under European and Japanese procedures, its application) is published. And once a new technology is made public, it takes time for potential imitators to learn about it and decide whether it is worth copying. Some are quicker than others in this respect. Studies of the diffusion process reveal that adoption spreads, first slowly and then more rapidly, in a pattern characterized well by an ogive curve such as the cumulative logistic.[37] For a sample of twelve significant innovations, lags ranging from one to twenty years and averaging ten years occurred before 60 percent of all relevant producers adopted the new technology.[38] Edwin Mansfield found the speed of imitation to be positively correlated with the profitability of adopting the new technology. For product innovations in particular, this means that the pace of imitation is a variable under the innovator's control. Companies pricing their new products to make a quick killing will encourage rapid imitation, while those pursuing dynamic limit-pricing strategies like those analyzed in Chapter 10 may be able to retain sizable market positions for a considerable period.

Second, free riding on an innovator's technical contribution is often far from free. An appreciable but varying fraction of the original R&D may have to be replicated. At one extreme in this respect are airliners. One can inspect a rival's design, but to build a similar aircraft, one must generate detailed engineering drawings for each part, program machine tools to produce the parts, build prototypes, and subject them to static and dynamic testing to ensure that unnoticed design flaws do not lead to catastrophe. Only the innovator's most basic conceptual work and wind tunnel testing can be circumvented. At the other extreme are new prescription drugs. To introduce a new drug to the U.S. market during the late 1980s entailed research, development, and testing costs of $50 to $100 million. Most of these costs were incurred discovering molecules with desirable therapeutic effects in humans and proving through extensive clinical testing that the



substances were effective and safe. Once these formidable information-generating hurdles are surmounted, it typically costs only a few hundred thousand dollars for an able biochemist to develop production methods. Thus, if there is no patent (or regulatory) barrier,[39] imitators can free ride on most of the innovator's investment.

Through a broad-ranging survey of company research and development decision makers, Richard Levin and colleagues learned that the R&D costs of duplicating a major *unpatented* new product exceeded 50 percent of the original innovator's R&D costs in 86 percent of the 127 covered industries.[40] Duplication costs exceeded 75 percent of the original costs in 40 percent of the industries. Patenting does, however, affect the cost of duplication. Levin et al. estimated that patent protection increased imitation costs by 40 percentage points in pharmaceuticals (relative to a scale on which equally costly imitation is 100), by 25 points for typical chemical products; by 7 to 15 points for semiconductor, communications equipment, and computer products; and by an average of 17 percentage points for machine tools, pumps, and compressors.[41] Timely duplication of a major patented new product was reported to be impossible in only twelve of the 127 surveyed industries.[42]

Third, as we have seen in Chapter 16, being the first to bring a new product onto the market, with or without patent protection, often confers a substantial reputational advantage over imitators, permitting the innovator to maintain elevated prices while defending a sizable market share. Also, as Chapter 10 revealed, the first mover has a head start in the race down learning curves, gaining cost advantages which, if exploited sufficiently aggressively, can be used to deter entry and enjoy supra-normal profits until the relevant technology matures.

35. For various views, see Edith T. Penrose, "International Patenting and the Less Developed Countries," *Economic Journal*, vol. 83 (September 1973), pp. 768–786; the articles by S. J. Patel, Pedro Roffe, and Peter O'Brien in *World Development*, vol. 2 (September 1974), pp. 2–36; Michael Berkowitz and Y. Kotowitz, "Patent Policy in an Open Economy," *Canadian Journal of Economics*, vol. 15 (February 1982), pp. 1–17; and Helena Stalson, *Intellectual Property Rights and U.S. Competitiveness in Trade* (Washington: National Planning Association, 1987).

36. The number of patents obtained per million dollars of R&D varies widely from industry to industry, in part because of differing emphases on secrecy. See F. M. Scherer, "The Propensity to Patent," *International Journal of Industrial Organization*, vol. 1 (March 1983), pp. 107–128. In an analysis of the data underlying that paper, Swarthmore College student Laurie Laird found that the propensity to patent declined significantly with the importance surveyed corporate R&D officials attached to keeping their inventions secret. On the survey, see note 40 infra.

37. See William L. Baldwin and John T. Scott, *Market Structure and Technological Change* (Chur: Harwood, 1987), pp. 128–144; and Colin Thirtle and Vernon W. Ruttan, *The Role of Demand and Supply in the Generation and Diffusion of Technical Change* (Chur: Harwood, 1987), pp. 77–124. Recognition appears to occur much more rapidly than implementation. See Edwin Mansfield, "How Rapidly Does New Technology Leak Out?" *Journal of Industrial Economics*, vol. 34 (December 1985), pp. 217–223. Empirical studies reveal fairly consistently that large firms tend to imitate new technologies more rapidly than small firms, all else equal. Baldwin and Scott conclude (p. 143) that diffusion is likely to proceed more rapidly in industries of relatively low seller concentration, but find (p. 132) that the relationship is "blurred."

38. Estimated (with assistance from the author) from Edwin Mansfield, *Industrial Research and Technological Innovation* (New York: Norton, 1968), pp. 134–135.

39. On regulatory barriers, see Edmund W. Kitch, "The Patent System and the New Drug Application," in Richard L. Landau, ed., *Regulating New Drugs* (Chicago: University of Chicago Center for Policy Study, 1973), pp. 81–109.

40. Richard C. Levin, Alvin Klevorick, Richard R. Nelson, and Sidney G. Winter, "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity* (1987, no. 3), p. 809. Altogether, the survey encompassed 650 executives.

41. Ibid., p. 811. The mechanical product average is derived by us from the Levin et al. raw data. Their results are similar to those obtained from a study of forty-eight specific new products by Edwin Mansfield, Mark Schwartz, and Samuel Wagner, "Imitation Costs and Patents: An Empirical Study," *Economic Journal*, vol. 91 (December 1981), pp. 907–918.

42. See also Mansfield et al., "Imitation Costs," p. 913, who found that 60 percent of the successful patented inventions in their sample were imitated within four years.

Finally, imitation may be held back by the very structure of the market. For all but innovations that define a completely new field,[43] the most likely early imitators are companies already operating in the industry to which the innovation pertains. Lack of production facilities, managerial experience, and channels of distribution impedes the entry of outsiders. If in addition the market is moderately or tightly concentrated, postimitation pricing discipline may remain firm for a considerable period. The innovator can then expect to retain at least its historical share of the industry's enhanced profits, and if innovation confers first-mover advantages, an already sizable share may be augmented.

In sum, competitive elimination of innovators' profits is often delayed because of natural secrecy and recognition lags, imitators' need to duplicate some or all of the innovator's R&D effort, first-mover advantages accruing to the innovator, and the protection an oligopolistic market structure affords. As a result, the profit expectations associated with a prospective innovation may be sufficient to warrant going ahead even when no patent protection is anticipated.

This conclusion is supported by the findings from several surveys of R&D executives, revealing quite uniformly that in most industries, patents are not very important compared to other incentives for innovation. In the most recent and comprehensive effort, Levin et al. asked 650 U.S. R&D executives to evaluate on a scale of from 1 ("not at all effective") to 7 ("very effective") the effectiveness of alternative means of protecting the competitive advantages from new and improved products and processes.[44] Averaging across 130 industries, the scores on six questionnaire items were as follows:

| Method of Appropriating the Benefits from Innovation | New and Improved Product Average | New and Improved Process Average |
|---|---|---|
| Patents to prevent duplication | 4.33 | 3.52 |
| Patents to secure royalty income | 3.75 | 3.31 |
| Secrecy | 3.57 | 4.31 |
| Being first with an innovation | 5.41 | 5.11 |
| Moving quickly down the learning curve | 5.09 | 5.02 |
| Superior sales or service efforts | 5.59 | 4.55 |

For both products and processes, the nonpatent strategic advantages from being an innovator were found to be substantially more important than patent protection.[45] In only twenty-five of the 130 industries did the average score for product patents as a means of preventing duplication exceed 5 on a 7 point scale (that is, between "moderately effective" and "very effective"). R&D executives placed greatest stress on product patent protection in the pharmaceuticals industry, agricultural chemicals (for example, pesticides and herbicides, subject to analogous federal testing regulations), and industrial organic chemicals.

Pursuing a different survey thrust, Edwin Mansfield asked the chief R&D executives of one hundred U.S. firms what proportion of the inventions they developed during 1981 through 1983 would not have been developed had they been



unable to obtain patent protection. The results for eight surveyed industry groupings were as follows:[46]

| | |
|---|---|
| Pharmaceuticals | 60% |
| Other chemicals | 38 |
| Petroleum | 25 |
| Machinery | 17 |
| Fabricated metal products | 12 |
| Electrical equipment | 11 |
| Primary metals | 1 |
| Instruments | 1 |

In four other groups, office equipment, motor vehicles, rubber products, and textiles, the average score was 0 percent. Weighting the responses according to 1982 company-financed R&D expenditures in the reporting groups, the aggregate loss of inventions without patent protection would have been roughly 14 percent of those actually made.[47]

Thus, a world without patents quite clearly would not be a world without innovation. Other incentives for innovation would fill most gaps. Some inventions would be lost, especially when the output of R&D is mostly information on whether the product works, upon which free riding is easy, and not on the details of product design and manufacturing. Patent protection appears also to be relatively important where the coverage is an all-or-nothing affair, for example, on particular chemical molecules or the features of a unique mechanism. The absence of patents might have particularly serious negative effects for independent inventors and small fledgling firms, who were not included in the surveys reported here. On this the evidence is more sparse, but offsetting tendencies can be identified. Small firms are at a severe disadvantage trying to claim patent rights, or enforce them against large rivals better able to sustain the multimillion-dollar costs of a protracted patent litigation.[48] Thus, even though they need patent

43. For field-creating innovations, new outside entry is more likely, but only after a period of slow entry averaging fourteen years in an analysis of forty-six new product histories by Michael Gort and Steven Klepper. "Time Paths in the Diffusion of Product Innovations," *Economic Journal*, vol. 92 (September 1982), pp. 640–642. The slow-entry stage was found to have declined over time to five years on average for products introduced after 1940.

44. "Appropriating the Returns," supra note 40, pp. 794–796. Similar earlier studies include F. M. Scherer, S. E. Herzstein, Alex Dreyfoos, et al., *Patents and the Corporation* (rev. ed.; Boston: privately published, 1959); and C. T. Taylor and Z. A. Silberston, *The Economic Impact of the Patent System* (Cambridge: Cambridge University Press, 1973).

45. Among eight reasons why the effectiveness of patents in protecting new technology was limited, the ability of competitors legally to "invent around" patented inventions was accorded by far the greatest weight. Levin et al., "Appropriating the Returns," supra note 40, p. 803.

46. Edwin Mansfield, "Patents and Innovation: An Empirical Study," *Management Science*, vol. 32 (February 1986), p. 175. Earlier surveys yielding similar results include, for the United Kingdom, Taylor and Silberston, *The Economic Impact*, supra note 44, pp. 195–199; for Germany, Klaus Grefermann, K. H. Oppenländer, et al., *Patentwesen und technischer Fortschritt*, Teil I, "Die Wirkung des Patentwesens im Innovationsprozess" (Göttingen: Schwartz, 1974), pp. 47–52 and Appendix Tables 75–78; and Mansfield et al., "Imitation Costs," supra note 41, pp. 915–916.

47. The weights are from U.S. National Science Foundation, *National Patterns of Science and Technology Resources: 1984* (Washington: USGPO, 1984), p. 52.

48. In a rarer "David slays Goliath" case, a former graduate student inventor, supported by a venture capital firm, won basic laser patent rights after an eighteen-year legal struggle against Bell Telephone Laboratories. See "Now the Father of the Laser Can Get Back To Inventing," *Business Week*, February 17, 1986, p. 98; "Patlex Wins Suit on Laser Patent," *New York Times*, November 6, 1987, p. D4; and "Ex-Astronaut To Lead Small Laser Company," *New York Times*, June 16, 1988, p. D4.

protection more than well-established companies, the protection they actually receive is more fragile. Wherever the balance lies for both small and large corporations, it is clear that we must seek deeper insight into the stimuli for innovation in a world where patents provide at best partial protection against rapid imitation, and where other barriers to imitation are at least as important as patents. To that task we proceed.

## *The Links Between Market Structure and Innovation*

Under the logic of the patent grant, the expectation that successful innovation will lead to a monopoly position induces firms to invest in R&D. Now we turn the tables to examine a series of hypotheses inspired by Joseph A. Schumpeter.[49] They suggest that the *possession* of monopoly power is conducive to innovation in the kind of turbulent environment associated with, and often caused by, technological change. The chain of causation is said to run from existing market structure to the pace of innovation, though it must be recognized at once that there are feedback effects from innovation to market structure.

One hypothesis is that profits accumulated through the exercise of monopoly power are a key source of funds to support costly and risky innovation. The predictions of economic theory on this point are ambiguous,[50] so the matter can only be resolved empirically. Carrying out clear-cut statistical tests is difficult because, while profits may lead to innovation, innovation also leads to profits, and it is crucial to get both the timing and the shapes of the lags right.[51] Also, increases in demand are believed to "pull" increased inventive effort,[52] and since outward demand shifts normally increase short-run profits, one must disentangle demand-pull from financing influences. There is considerable evidence that augmented R&D activity follows increases in profitability with typically short lags,[53] but in view of the conceptual problems, it is hard to be sure that the cause was enhanced financing ability. Especially in the United States, the growth of a thriving venture capital industry channeling investment into new high-technology firms shows that past monopoly profits are no *sine qua non* for supporting innovation.[54]

### Monopoly and Oligopolistic Rivalry

Economic theory has yielded numerous insights on firms' incentives to innovate under market structures ranging from pure monopoly with blockaded entry to pure competition, with diverse empirically important oligopoly variants in between. We focus here on incentives for *product* innovation, on which in the United States roughly three dollars are spent for each dollar of internal *process* R&D.[55]

Introducing a product innovation is usefully viewed as tapping into a market with some potential for earning profits, which means that there must be a surplus of sales over production and distribution costs so that front-end R&D costs can be recouped. Precision of terminology is important here, so despite the awkwardness, let us call that transitory surplus by its correct name, a *quasi rent*. The quasi-rent potential has a significant time dimension. It is approximated by the line b(t) in Figure 17.2 as being constant over time, for example, $800,000 per year, year after year. If demand-pull influences were at work, the quasi-rent potential b(t) would instead be rising steadily over time or jumping discontinuously at certain moments in time (for example, when an energy shock occurs). How innovators



and their rivals tap into the stream's potential is crucial to the analysis of incentives.[56] Absent government or similar subsidies, no gains can be realized until research and development are completed and the product is commercially introduced. This introduction date is shown as T* in Figure 17.2. Usually it takes time to build up to the new product's full quasi-rent potential, so the innovator's annual gain *if it subsequently monopolizes the new product market* is shown by the rising line $v_M$. However, as we have seen, neither patents nor being first with a new product are normally sufficient to exclude competitive imitation. Suppose another firm imitates at time $T_{ii}$. Then the rival begins chewing away some of the quasi rents the innovator would otherwise capture. The effect on the innovator's quasi-rent realizations is shown by the dotted line $v_{ii}$. The distance between $v_M$ and $v_{ii}$ is the quasi rent realized by the imitator, assuming that the total profit pool is not shrunk by price competition; so the distance under $v_{ii}$ is what remains for the innovator. Suppose alternatively that the imitator delays its product introduction even more to time $T_{iii}$. This will presumably strengthen the innovator's first-mover advantage. If so, the imitator will not only tap the market later, but will capture a

49. *Capitalism, Socialism, and Democracy*, supra note 8, especially Chapter VIII.

50. See Morton I. Kamien and Nancy L. Schwartz, "Self-Financing of an R&D Project," *American Economic Review*, vol. 68 (June 1978), pp. 252–261.

51. See David J. Ravenscraft and F. M. Scherer, "The Lag Structure of Returns to Research and Development," *Applied Economics*, vol. 14 (December 1982), pp. 603–620.

52. See Jacob Schmookler, *Invention and Economic Growth* (Cambridge: Harvard University Press, 1966); and for a survey of subsequent studies, many critical, see Thirtle and Ruttan, *The Role of Demand*, supra note 37, especially pp. 6–11.

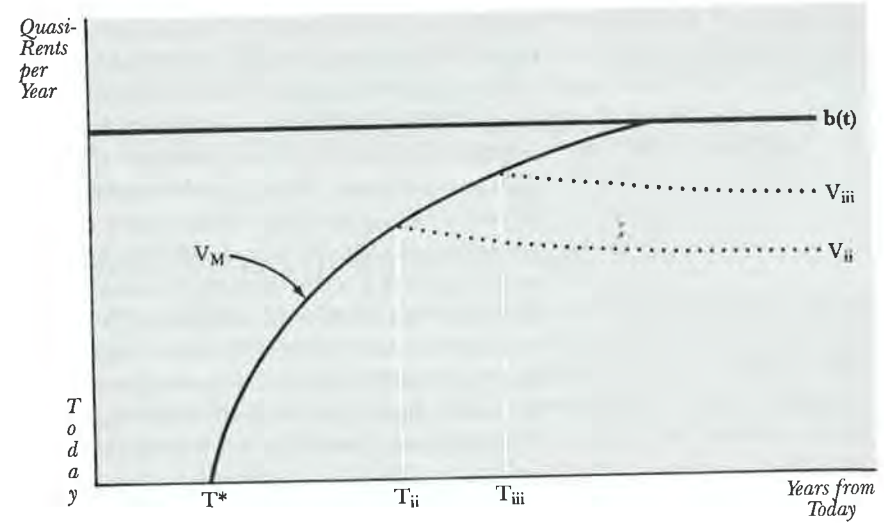
53. The most careful attempts to estimate both forward and backward lag structures are Ben S. Branch, "Research and Development Activity and Profitability," *Journal of Political Economy*, vol. 82 (September/October 1974), pp. 999–1011; and Ariel Pakes, "On Patents, R&D, and the Stock Market Rate of Return," *Journal of Political Economy*, vol. 93 (April 1985), pp. 390–409. See also Dennis C. Mueller, "The Firm Decision Process: An Econometric Investigation," *Quarterly Journal of Economics*, vol. 81 (February 1967), pp. 71–73; Albert N. Link, "An Analysis of the Composition of R&D Spending," *Southern Economic Journal*, vol. 49 (October 1982), pp. 343–349 (suggesting that basic and applied research are more responsive to prior profitability than development); and Lorne Switzer, "The Determinants of Industrial R&D," *Review of Economics and Statistics*, vol. 66 (February 1984), pp. 163–168.

54. For data on the growth of technology-oriented venture capital activity, see U.S. National Science Board, *Science Indicators: The 1985 Report* (Washington: USGPO, 1985), pp. 260–261. Updated expenditures data and other current reports appear in a periodical, the *Venture Economics Journal*. For an international view, see OECD, *Venture Capital: Context, Development and Policies* (Paris: 1986).

55. In addition to the approach taken here, there are two main alternative lines of theorizing. One, focusing on outlays that shift the firm's cost curve (that is, process invention), began with Kenneth J. Arrow, "Economic Welfare and the Allocation of Resources for Invention," in the National Bureau of Economic Research conference volume, *The Rate and Direction of Inventive Activity* (Princeton: Princeton University Press, 1962), pp. 609–625. A recent extension with comprehensive references to earlier Arrow progeny is Thomas F. Cosimano, "The Incentive To Adopt Cost Reducing Innovation in the Presence of a Non-Linear Demand Curve," *Southern Economic Journal*, vol. 48 (July 1981), pp. 97–104. Another line invokes the Dorfman-Steiner theorem (see Chapter 15, note 3; and Chapter 16, notes 64 and 65) to determine an overall R&D budget (undifferentiated by project types) that maximizes profits. See Douglas Needham, "Market Structure and Firms' R&D Behavior," *Journal of Industrial Economics*, vol. 23 (June 1975), pp. 241–255; and Takeo Nakao, "Product Quality and Market Structure," *Bell Journal of Economics*, vol. 13 (Spring 1982), pp. 133–142.

56. The analysis here follows F. M. Scherer, "Research and Development Resource Allocation under Rivalry," *Quarterly Journal of Economics*, vol. 81 (August 1967), pp. 359–394, and two related articles reproduced in Scherer, *Innovation and Growth*, supra note 9, Chapters 4 and 6. Rich variations and extensions are presented in Morton I. Kamien and Nancy L. Schwartz, "On the Degree of Rivalry for Maximum Innovative Activity," *Quarterly Journal of Economics*, vol. 90 (May 1976), pp. 245–260; Kamien and Schwartz, "Potential Rivalry, Monopoly Profits and the Pace of Inventive Activity," *Review of Economic Studies*, vol. 45 (October 1978), pp. 547–557; Jennifer Reinganum, "A Dynamic Game of R&D: Patent Protection and Competitive Behavior," *Econometrica*, vol. 50 (May 1982), pp. 671–688; Marion B. Stewart, "Noncooperative Oligopoly and Preemptive Innovation without Winner-Take-All," *Quarterly Journal of Economics*, vol. 98 (November 1983), pp. 681–694; and Michael L. Katz and Carl Shapiro, "R&D Rivalry with Licensing or Imitation," *American Economic Review*, vol. 77 (June 1987), pp. 402–420.

**Figure 17.2**
Innovator's Quasi-rents Without and With Imitation




smaller share of the total quasi-rent potential, leaving the innovator with a gain function like $v_{iii}$ in Figure 17.2.

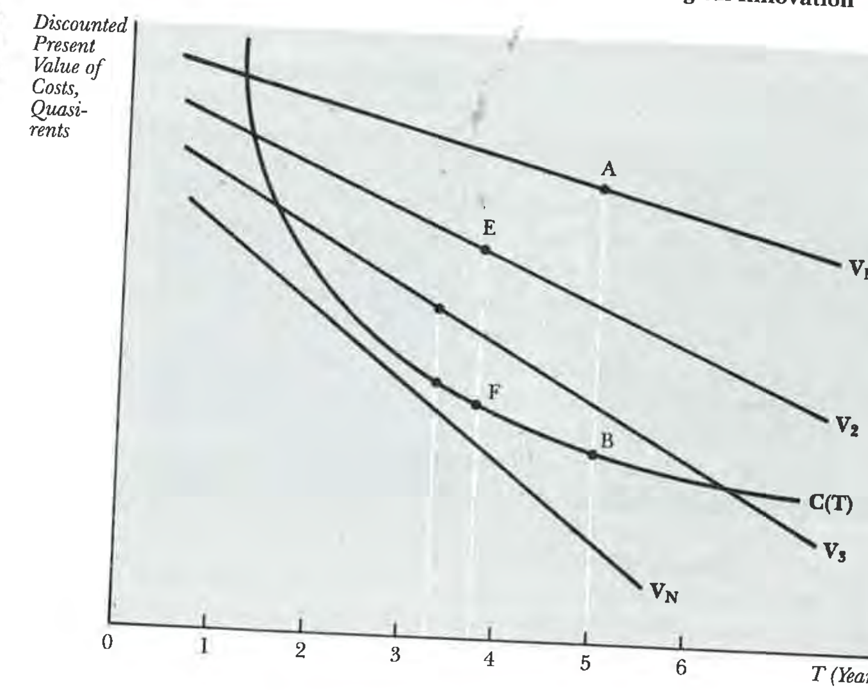
The question remains, How rapidly should a firm that believes it will (or in a more uncertain world, might) be the innovator conduct its R&D project? In other words, what determines the position of T* in Figure 17.2?

To find out, we must first introduce the cost side of the picture. This is done through the *time-cost tradeoff function* C(T) in Figure 17.3. It shows the cost of conducting the R&D effort over a continuous array of alternative time schedules, assuming that the effort begins at time 0 (that is, today). Accelerating the pace of R&D, that is, achieving lower values of T, increases the cost of R&D for three reasons: because one may have to pursue parallel experimental approaches to find a good solution quickly under uncertainty, because moving on to design and production tooling steps before early experiments have yielded their information saves time but increases the number of false starts, and because of conventional diminishing marginal returns in allocating talent to a given technical assignment.[57] The more uncertain the technological environment is, the more curvature the time-cost tradeoff function is likely to have.[58]

The line $V_M$ in Figure 17.3 integrates (literally) the data in Figure 17.2 to show the discounted present value of the *total* quasi rents expected over time by an innovating firm on the assumption that it will be a monopolist, experiencing no competition from imitators. It is downward sloping because the longer the firm takes to carry out its development, the longer it will have to wait to tap the stream of potential quasi rents, and the more heavily discounted the benefits it anticipates will be. To maximize profits, the firm must find a development schedule that leaves the maximum vertical distance between quasi-rent function $V_M$ and R&D cost function C(T). This is achieved where the slopes of the two functions are equal, that is, with a planned project time of five years, leaving a surplus AB of quasi rents over R&D costs.



**Figure 17.3**
Profit-maximizing R&D Schedule Choices




Let us now relax our assumption that the innovator monopolizes its product's sales after innovating. Moving step by step, we consider first the case in which innovative rivalry is duopolistic. If one firm is an imitator, reaching the market with its new product later than the innovator, an uneven division of quasi rents like those shown by curves $v_{ii}$ or $v_{iii}$ in Figure 17.2 will occur. But if the firms start from symmetric structural positions, there is no reason to believe that one rival will consciously choose to accept second place. Instead, each may struggle to be the first mover. We saw in Chapter 6 that Cournot-Nash rivalry in price-setting duopolies is implausible; at best, the parties may jump without intervening moves to an equilibrium position. But each new product rivalry is *sui generis*, not a repeated game, and there are lags between observation, reaction, and further reaction. It is much more realistic, therefore, to assume that something like Cournot reactions will ensue.

Figure 17.4(*a*) shows the R&D schedule reaction pattern in a completely symmetric duopolistic new product rivalry. The horizontal axis measures Firm 1's R&D time, the vertical axis Firm 2's time. Quasi-rent functions like those shown in Figure 17.2 and cost functions like C(T) in Figure 17.3 give rise to reaction functions like $R_1$ (for Firm 1) and $R_2$ in Figure 17.4(*a*). Suppose Firm 1 starts the new product development at the leisurely pace implied by point A. Seeing its rival

57. For statistical estimates, see Edwin Mansfield et al., *Research and Innovation*, supra note 23, Chapters 6 and 7; and *Technology Transfer, Productivity, and Economic Policy* (New York: Norton, 1982), pp. 87–96.

58. Scherer, *Innovation and Growth*, supra note 9, Chapter 4.

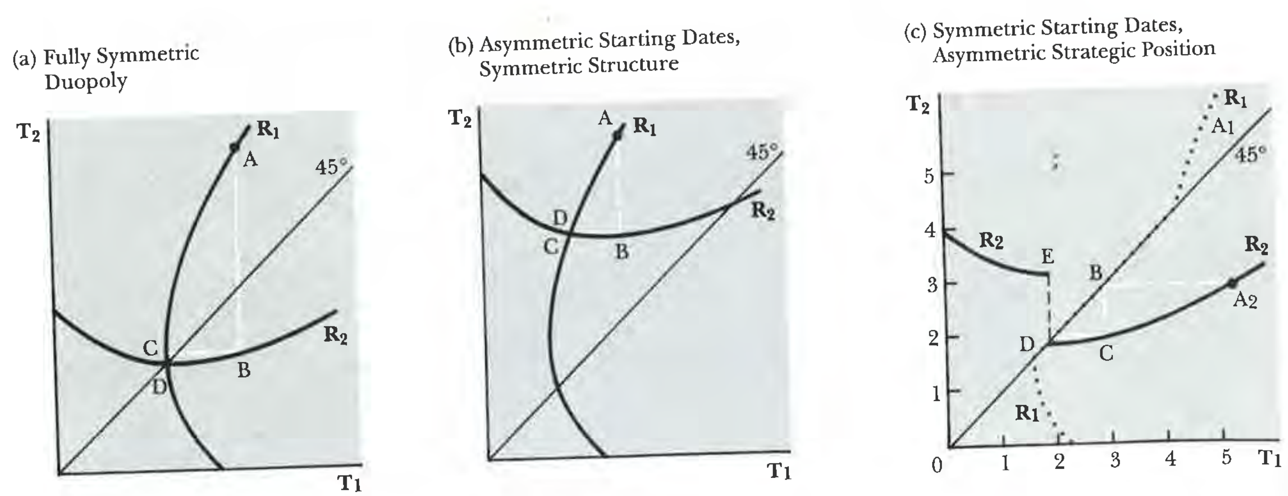



**Figure 17.4**
R&D Rivalry Reaction Functions

ahead (that is, on the northwest side of the 45° equal-times line), Firm 2 gets going and moves to point B on its reaction function. Seeing Firm 2 pull ahead, Firm 1 accelerates to point C. Firm 2 reacts to Cournot equilibrium point D, where both firms are symmetrically pursuing much faster R&D schedules than they would have chosen in the absence of rivalry. Except to the extent that chance factors intervene, they will enter the market with their new products in a dead heat.

The effects of two-firm rivalry can be shown more summarily for a representative firm through the discounted total quasi-rent line $V_2$ in Figure 17.3. It slopes downward more steeply than monopoly function $V_M$ because of the stimulus rivalry provides. With steeper slope, the maximum distance between $V_2$ and C(T) occurs at a more compressed R&D schedule – for example, in Figure 17.3, 3.75 years, with total quasi rents exceeding R&D costs by the distance EF. $V_2$ is nearer the origin than $V_M$ because the gains from innovation are shared by two firms, each of which appropriates less than a single monopolist would for any given market introduction date.

Symmetry in Figure 17.4(*a*) implies two tacit assumptions: equal natural strengths in the market, reflected for instance in equal market shares for the most closely related existing products, and equal (or insignificantly different) R&D project starting times. Panel (*b*) of Figure 17.4 relaxes the second but not the first assumption. Now Firm 1 gets a jump on its rival, scheduling its project at A. When Firm 2 wakes up, it moves to B on its own reaction function. But with its head start, Firm 1 can accelerate to C, putting Firm 2 at a hopeless disadvantage. Laggards like Firm 2 under these conditions commonly turn *submissive*, slowing down to point D on their reaction function to reduce their otherwise high R&D costs and giving up a bit more market share in compensation.[59]



Now we relax the symmetry assumption in a more important way. Suppose that, by virtue of well-established reputation, distribution channels, and other attributes in related markets, Firm 1 would win 80 percent of the new product's sales and Firm 2 only 20 percent if each rolled out the fruits of their R&D on the same day. Figure 17.4(*c*) shows the reactions, but to understand them it is useful to jump ahead to Figure 17.5. The size of the rectangle shows the total amount of quasi rents to be divided between the rivals if both had the same innovation date.[60] The vertical line shows the division of quasi rents between the two with identical product introduction dates. Suppose further that a year's first-mover advantage permits one firm to capture more or less permanently one-fourth of what would otherwise be its rival's quasi rents. Then for small Firm 2, the cannibalization gain from being first is the large darkest shaded rectangle, while for dominant Firm 1, the gain is the much smaller shaded rectangle. The small firm has more to gain from being first than the large firm. Turning back to Figure 17.4(c), this shows up in different reaction function positions on that side of the 45° equal-times diagonal where a firm leads its rival. If Firm 1 saw no competition on the horizon, it would choose a point like $A_1$ on its reaction function (marked by large dots), corresponding to a 4.75 year R&D schedule. With more to gain by being first, smaller Firm 1 will choose $A_2$, implying a 2.8 year schedule. If both firms choose strategies at the same time, Firm 2 will set the pace. But now Firm 1 realizes that it has much to lose – the darkest shaded rectangle in Figure 17.5 – if it lags significantly behind Firm 2. It does not want to force the pace, but if forced, it reacts along the 45° line to point B. Firm 2, perceiving that hope is not yet lost, accelerates to C. But Firm 1 responds in kind by accelerating to D, in effect preempting its challenger. It is possible, depending upon the circumstances, that Firm 2, seeing it cannot be first, then reacts submissively to a point like E.

This asymmetric model provides insight into the frequently observed tendency for market-dominating firms to be slow in developing important new products, but to roar back like tigers when smaller rivals – often new entrants with no historical market share at all – challenge their dominance. Examples include Gillette's lag behind Wilkinson on stainless steel razor blades in 1962–1963; IBM's slow start and (sometimes) fast finish in developing its first digital computers and their transistorized, super-, mini-, and microcomputer descendants;[61] AT&T's record in developing microwave radio relay systems and communications satellites; the reactions of Kellogg and General Mills to small firms' success in selling granola-type cereals; NCR's shift from mechanical cash registers to electronic point-of-sale devices; Boeing's reaction to European "Airbus" competition; and Texas Instruments' response to the pace-forcing efforts of Japanese dynamic random-access semiconductor memory makers, among others.[62] However, variations in the assumptions can change the outcome in important ways. In particular, we have

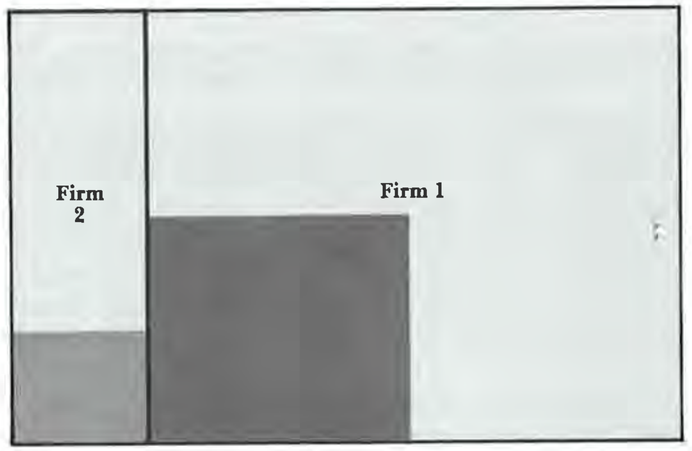
59. See Scherer, "Research and Development Resource Allocation," supra note 56, p. 379; and Gene M. Grossman and Carl Shapiro, "Dynamic R&D Competition," *Economic Journal*, vol. 97 (June 1987), pp. 376–379.

60. The rectangle's size is of course larger, the earlier both innovate.

61. For a similar but more richly structured model explaining IBM's leased computer development incentives, see Gerald W. Brock, *The U.S. Computer Industry* (Cambridge: Ballinger, 1975), pp. 211–215.

62. For references on some of these illustrations, see the second edition of this work, pp. 433–435.

**Figure 17.5**
Division of Quasi-Rents with Equal R&D Times and Shifts when One Firm Gains a First-Mover Advantage




Total Quasi-Rents (Area)

assumed thus far that both rivals must operate on identical time-cost tradeoff functions. But if a laggard can shift its time-cost tradeoff function inward by hesitating, observing, and avoiding the first movers' false technical starts, and if small interlopers initially penetrate the market only slowly, dominant firms may find it more profitable to pursue a "fast second" strategy—not so fast that they must blaze the trail, but sufficiently fast that smaller pioneers achieve only modest inroads into their dominant positions before their aggressive response occurs.[63]

The asymmetric market positions logic helps us understand what happens when we relax the assumption that only two rivals compete with new products. For small Firm 2 in Figure 17.5, the gains it can capture *if* it can reach the market first do not really depend upon whether the (large) rest of the market is held by a single dominant firm or by several firms—for example, four others, each with 20 percent related market shares like its own. If being first mover confers lasting market share gains, the larger the share of the market *all* other firms would command if they exactly matched Firm 2's new product introduction date, the more Firm 2 has to gain by being first and capturing some of what otherwise would be rivals' quasi rents. Or to accentuate the negative, the smaller a company's share will be if it only ties others to the market, the more it has to gain by leading. If firms' preinnovation market positions are symmetric, the more rivals there are, the smaller will be a representative firm's simultaneous-innovation market share, and so the larger will be the gains from being first mover for any given head start over the others. Returning to Figure 17.3, this means that the more symmetric firms there are, the steeper a representative firm's discounted quasi-rent function V will be.[64] Thus, $V_3$ (for three symmetric rivals) has a steeper slope than $V_2$, whose slope is steeper than with the blockaded monopolist $V_M$. The steeper the slope, the quicker is the profit-maximizing R&D schedule. An increase in the number of symmetric rivals accelerates product research and development!

As Figure 17.3 reveals, there is a hitch. As the number of would-be innovators increases to some critical value N, the representative firm's discounted quasi-rents function is not only steeper, but lies at all points below the time-cost tradeoff function. If R&D is to be undertaken at all, it will be done quickly. But it is not profitable, and so if firms correctly see what is happening, there will be a market failure. Each firm will refrain from investing in R&D in the expectation that if all invest, the market will be divided up into so many pieces that each firm's discounted quasi rents are too small to cover R&D costs.



This problem is reinforced by another that we have kept in abeyance. As the market structure moves from monopoly to duopoly to looser oligopoly, the ability of firms to hold prices at monopoly levels will break down. When that happens, the quasi rents per firm will be too small to cover R&D costs not only because the new product market is divided into so many slices, but also because price competition has eroded the market's profitability.

We find therefore a clash of structural propensities, giving rise to a dualism in the links between market structure and incentives for innovation. Up to a point, increased fragmentation stimulates more rapid and intense support of R&D. This influence is called the *stimulus factor*. But when the number of firms becomes so large that no individual firm can appropriate quasi rents sufficient to cover its R&D costs, innovation can be slowed or even brought to a halt. This influence can be called the *market room* factor.

## Dynamics and Welfare Implications

To carry the analysis further, we need a broader dynamic conception of the process of technological advance. The profitability of innovation depends most fundamentally upon demand and supply conditions. On the demand side, as population grows, per capita incomes rise, and (less certainly) input cost ratios change, the quasi-rent potential b(t) in Figure 17.2 increases over time, leading in Figure 17.3 to upward shifts in discounted total quasi-rent functions V for any given number of firms. On the supply side, as scientific and technological knowledge advances, what may be impossible today will be feasible but costly next year and easy several years hence. This means that time-cost tradeoff functions, recalibrated periodically in current time, shift toward the origin of Figure 17.3 (that is, in a southwesterly direction) as knowledge accumulates. An innovation becomes profitable when, as a result of changes in knowledge and/or demand, the discounted total quasi-rent function comes to lie at least partially above the time-cost tradeoff function. Innovations induced mainly by advances in knowledge are called *technology-push innovations*. Those rendered attractive by rising demand are called *demand-pull inventions*.

63. See William L. Baldwin and G. L. Childs, "The Fast Second and Rivalry in Research and Development," *Southern Economic Journal*, vol. 36 (July 1969), pp. 18–24. Later contributions show that the dominant firm will be more inclined to preempt than lag, the more perfect patent protection is, the more the innovation threatens its rents in existing markets, the less uncertain R&D completion time is, the less likely licensing of the challenger's technology is, and the less likely multiple challenges are. See Richard J. Gilbert and David Newberry, "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, vol. 72 (June 1982), pp. 514–526; Jennifer Reinganum, "Uncertain Innovation and the Persistence of Monopoly," *American Economic Review*, vol. 73 (September 1983), pp. 741–748; Jonathan Cave, "A Further Comment on Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, vol. 75 (March 1985); Christopher Harris and John Vickers, "Patent Races and the Persistence of Monopoly," *Journal of Industrial Economics*, vol. 33 (June 1985), pp. 461–481; and Partha Dasgupta, "The Theory of Technological Competition," in Joseph Stiglitz and G. Frank Mathewson, eds., *New Developments in the Analysis of Market Structure* (London: Macmillan, 1986), pp. 519–547.

64. A formal proof is found in Scherer, "Research and Development," supra note 56, pp. 389–390.

Suppose now that the advance of knowledge and increases in demand occur smoothly and continuously. At some moment in time, an innovation that was not profitable before will suddenly become profitable for a pure monopolist as the shifting $V_M$ quasi-rent and C(T) cost functions fleetingly become tangent to one another. At this same moment, the development would not yet be profitable if the market were divided among oligopolists, each of whom (assuming imperfect patent protection) anticipates having to share the new product's sales with rivals. If innovation is to occur as rapidly as is practically feasible under the conditions postulated, monopoly is essential!

The story is different when the functions shift discontinuously by large amounts, as when a scientific breakthrough (for example, the discovery of the transistor effect, or low-temperature superconductivity, or gene splicing) occurs, *or* when there are significant lags in the recognition of profitable innovation opportunities. Then the quasi-rent functions may lie above the time-cost tradeoff function even for firms in relatively atomistic industries. If so, firms' behavior is dominated by the stimulus factor, and the pace of innovation will almost surely be faster when there is rivalry than under a securely monopolized market structure.

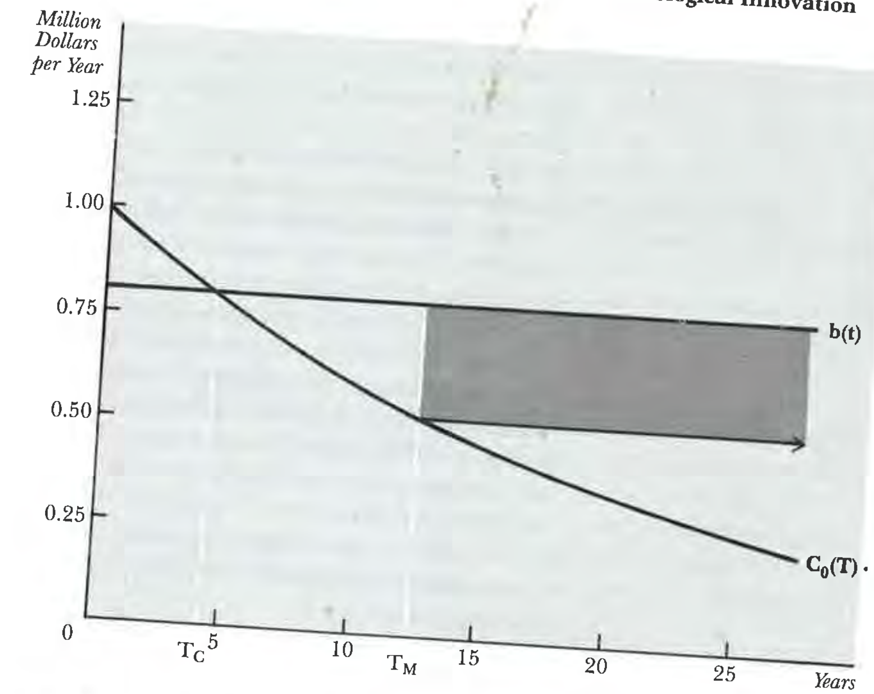
Although a monopolist *could* go faster when demand and knowledge advance continuously, it remains to be seen whether its incentives lead it actually to do so. Also, we have tacitly assumed that faster is better, but that assumption must be probed. Both of these questions are illuminated by the pioneering contribution of Yoram Barzel.[65]

Barzel's innovators inhabit a world similar to the one we postulated in our initial exploration of the patent grant's logic. Once a patent is gained, it is assumed to provide its owner perfect protection. Among other things, inventing around the patent is assumed impossible. But before invention occurs, a wide array of market structures, ranging from monopoly with blockaded entry to a form of all-out competition, can exist. Dynamics are introduced by assuming (in our adaptation) that the cost of developing a new product or process falls steadily over time. If the development were carried out today, it would cost $10 million. The advance of science reduces that cost at a rate of 5 percent per year. Thus, the cost of development follows the time trajectory $C_o(T) = \$10^7 \times e^{-.05T}$. We ignore demand-pull influences, assuming that the stream of tappable quasi rents has a constant depth of $800,000 per year. The question is, how long should firms wait before innovating? And how does market structure matter?

The problem can be characterized in simple graphic form with a further assumption: that when the development is carried out, it is financed by taking out a loan at an interest rate *r* of 10 percent per year. To keep the geometry and algebra simple, we assume further that both the resulting patent and the loan have perpetual life.[66] Then the yearly cost of carrying the development loan beginning at time T and continuing in perpetuity will be $0.1 \times C_o(T)$. This annual cost declines exponentially over time as R&D is postponed, as shown by the declining solid curve in Figure 17.6. If the development were undertaken today (at year 0), loan carrying costs would exceed annual quasi rents b(T), so any profit-maximizing firm would wait. After 4.5 years, $C_o(T)$ has declined sufficiently with the advance of knowledge that break-even is achieved; loan carrying costs equal annual quasi rents. But a profit-maximizing monopolist would not innovate at this time if it anticipates a further fall in development costs, because if it waits, it can make a



**Figure 17.6**
Innovation Date Choices and Profits Under Competition and Monopoly




profit. Even though the profit *margin* rises continuously over time with waiting, the monopolist will not wait forever because early profits are preferred to later profits. The discounted present value of its profits under the assumptions of Figure 17.6 is maximized when it innovates at year $T_M = 12.5$.[67] Then its annual quasi rents will start and continue at $800,000 per year. Incurring a one-time R&D cost reduced to $5.35 million, its annual financing costs will be $535,000 per year, and it will realize the darker-shaded profit rectangle in perpetuity.

This is the solution for the extreme case of a monopolist so secure in its position that it can wait until the profit-maximizing moment to innovate. How does the opposite extreme—pure competition—compare? Barzel invokes here the classic equilibrium definition: competition is characterized by such free entry that expected profits (after covering fixed R&D costs) are zero. This can only occur at the

65. "Optimal Timing of Innovations," *Review of Economics and Statistics*, vol. 50 (August 1968), pp. 348–355.

66. This is unrealistic, but its impact on the results is transparent. The shorter the patent life, the longer firms will delay innovating.

67. Generalization: Let b be the annual quasi rent realized after innovation occurs, r the rate of interest, $C_o$ the one-time cost of development at time t = 0, and p the rate at which development costs fall over time. Then the cost of development at time T is $C_o e^{-pT}$. The monopolist maximizes the discounted present value of quasi rents less costs:

(1) $$\pi = \int_T^{\infty} b\, e^{-rt}\, dt - C_o\, e^{-(p+r)T};$$

(2) $$\pi = \frac{b\, e^{-rT}}{r} - C_o\, e^{-(p+r)T},$$

Differentiating with respect to T, the first-order condition for a maximum is:

(3) $$b\, e^{-rT} = (p+r)\, C_o\, e^{-(p+r)T}.$$

break-even point $T_C$ = 4.5 years.[68] Just how the competitive process evolves to achieve this zero-profit equilibrium need not concern us in detail yet. It cannot be through postinnovation price competition, for a perfect patent permits monopoly pricing then. So the competition is at the preinnovation stage, for example, in Barzel's original schema, as a single firm preempts its less bold rivals with an R&D project conducted so early that its costs equal discounted quasi rents.

Thus, pure competitors in the Barzel sense innovate sooner than secure monopolists. The question remains, what is best for society? A natural criterion for evaluating society's interest is that the sum of all surpluses, consumers' plus producers', should exceed development (and other fixed) costs by as much as possible. Normally, we saw in our discussion of Figure 17.1,[69] innovations generate both consumers' and producers' surplus. Let us define a coefficient of appropriation k, measuring the ratio of the total surplus from innovation, producer's plus consumers', to the producer's surplus alone.[70] Under the assumptions of Figure 17.1, consumers' surplus was half as great as the monopolist producer's surplus, so k would have a value of 1.5. If instead $k = 1$, the monopolist's gains coincide with society's gains, and so the monopolist's timing choice is socially optimal. In the more realistic case where $k > 1$, the monopolist proceeds too slowly, the more so, the larger k is.[71]

How well do competitive innovators perform in comparison? A useful way of proceeding is to ask, Under what conditions does break-even competition in the Barzel sense innovate at exactly the social welfare-maximizing date? It is not difficult to show that given the assumptions accepted here, the competitive and social welfare-maximizing choices coincide when k equals the sum of the rate at which development costs fall (which we call p) plus the interest rate r, divided by r.[72] Figure 17.7 relates the values of k consistent with social optimality to the range of

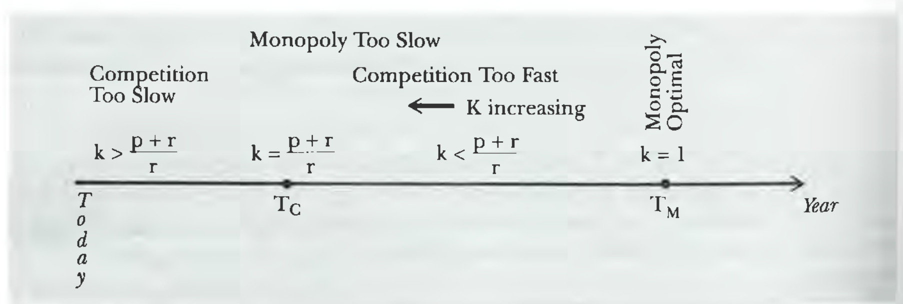



**Figure 17.7**
How the Optimality of R&D Plans Varies with Appropriability Coefficient k under Competition and Monopoly



possible monopoly and competitive innovation date choices. When $k = (p+r)/r$, the competitive break-even date $T_C$ is optimal.[73] Leisurely innovation dates between $T_M$ and $T_C$ are socially optimal for $1 < k < (p+r)/r$; that is, when the innovator appropriates a large fraction of the total surplus. Dates faster than the pace set by competition are optimal when k is large, that is, when there are large external benefits not captured by the innovator. In this case, competitive innovators' inability to appropriate a sufficient share of the benefits from their R&D causes a market failure delaying the introduction of the new technology. When the degree of appropriability is low (k is large), the competitive break-even solution tends to be optimal when the rate of advance in knowledge is fast, that is, p is relatively high. For high appropriability, competition tends to be optimal when knowledge advances slowly.

Needless to say, k will exactly equal $(p+r)/r$, and break-even competition will be socially optimal, only by happenstance. There is no invisible hand assuring that competition will set the right pace for innovation. This result parallels our finding in Chapter 16 that monopolistic competition with free entry might lead to either too much or too little product variety.[74] Still pure monopoly with blockaded entry has no better claim; indeed, it is always too slow when appropriability is

68. From equation (2) of note 67 above, break-even occurs where:

(1) $$\frac{b\,e^{-rT}}{r} = C_o\,e^{-(p+r)T}$$

Multiplying both sides by r, we obtain a condition analogous to profit maximization condition (3) of note 67:

(2) $$b\,e^{-rT} = r\,C_o\,e^{-(p+r)T},$$

except that the premultiplier on the right-hand side is r rather than $(p+r)$. To satisfy equation (2) here, T must be lower than it needs to be to satisfy equation (3) in note 67. Thus, the break-even competitive $T_C$ is shorter than the monopoly $T_M$.

69. And also in our discussion of optimal product variety in Chapter 16, whose parallels with the present analysis should become evident.

70. Some important problems must be dodged here. In principle, k should depend on both market structure and pricing conduct. Under monopoly pricing (absent first-degree price discrimination), total surplus is less than it would be under competitive pricing. If k is approximated as a parameter, should the total surplus measured for the numerator of k be the total actually realized, or the total possible under competitive pricing? Since without postinnovation monopoly pricing there could be no innovation, the solution must be second best, and so the proper measure of total surplus is the surplus actually realized.

71. Proof: A social planner would maximize the difference between discounted total surplus and R&D costs. By definition, total surplus is k times the innovator's surplus (from equation [2] of note 67, $be^{-rT}$). Assuming that the discount rate is the same for innovators and society at large, we wish to maximize:

(1) $$k\,(be^{-rT}) / r - C_o\,e^{-(p+r)T}$$

with respect to T. The first-order condition is:

(2) $$k\,(be^{-rT}) = (p+r)\,C_o\,e^{-(p+r)T},$$

which differs from profit-maximizing condition (3) in footnote 67 only by the k premultiplier. If $k = 1$, the two solutions are identical. The more k exceeds unity, the more T must be reduced relative to the monopolist's profit-maximizing $T_M$ to satisfy the equality.

72. T must satisfy equation (3) of note 67 and equation (2) of note 71 simultaneously. Let $Z = C_o\,e^{-(p+r)T}$ and $X = b\,e^{-rT}$. Then we must simultaneously have:

(1) $$X = rZ, \text{ so } Z = X/r, \text{ and}$$

(2) $$kX = (p+r)\,Z.$$

Substituting (1) for Z in (2), we obtain:

(3) $$kX = (p+r)\,X/r;$$

(4) $$k = (p+r) / r.$$

73. In a study of seventeen innovations, Edwin Mansfield and associates found the median social/private surplus ratio to be on the order of 2.25. "Social and Private Rates of Return from Industrial Innovations," *Quarterly Journal of Economics*, vol. 91 (May 1977), pp. 221–240.

74. The parallelism holds even more strongly. The high surplus appropriation associated with cartelized pricing tips the balance toward excessive product variety, just as preinnovation "competitors" are likely to innovate excessively quickly when a high fraction of the postinnovation benefits will be appropriated.

incomplete, as it must be if consumers are to enjoy the fruits of technological progress. Barzel's model is silent on the intermediate oligopoly case. From our previous analysis, oligopolists (or monopolists reacting to preempt smaller firms' R&D challenge) are likely to innovate more quickly than Barzel's secure monopolist unless the number of rivals is so large that profit expectations turn negative. Profitable oligopoly is most apt to sustain an innovation pace preferable to that of monopoly and competition in the range of Figure 17.7 appropriability values where competition is too fast.

## Further Theoretical Insights

Economists have extended the above theoretical skeleton with a rich profusion of variants—perhaps too rich to satisfy the criteria for optimal model variety. Through an astute choice of assumptions, virtually any market structure can be shown to have superior innovative qualities. This may only mirror the complexity of the real world. Yet some strong generalizations persist across a wide range of assumptions.

Thus far, we have assumed innovations to be one-time events. However, most industries experience a continuing stream of innovations over time, and in many cases, each completed new product or process sets an agenda focusing improvement work for the next technological generation. This has two consequences.[75] Since the next generation often makes previous innovations obsolete, the degree to which any innovation can expect to appropriate the future surpluses it generates is lessened, that is, k rises. This weakens incentives to innovate and extends the range of outcomes (in Figure 17.7) over which even competitive markets proceed too slowly. Second, to the extent that subsequent R&D is focused and informed, becoming more productive than it otherwise would have been, an innovation at time T confers benefits upon innovators at stages $T+1$, $T+2$, etc., in the chain. Except in the most extraordinarily tightly monopolized industries, it is unlikely that the firms generating such benefits will capture them exclusively in their own future R&D, and so the benefits spill over as positive externalities to other firms. This too extends the range of outcomes over which competitive markets innovate too slowly or in excessively modest qualitative steps, although, since Barzel competition may proceed too rapidly for other reasons, it yields no precise guidance as to the optimal market structure.

Barzel's competitive case was offered in the spirit of an extreme among alternatives; and his explanation of how zero profits emerged, through a particularly aggressive firm's preemption of potential rivals before they started their own R&D efforts, ignored alternative but plausible possibilities. These lacunae stimulated several further contributions.[76]

A common thread is the assertion that the zero-profit outcome is not simply an extreme case, but the natural state toward which research and development rivalry gravitates in the absence of artificial entry barriers. Thus, the market structure *affecting* R&D decisions is not a given, but is endogenously determined by technology and competition. If the cost per firm of conducting R&D is small relative to the size of appropriable quasi rents, many firms will join in, and an atomistic zero-profit equilibrium will emerge. If project costs are large relative to quasi rents, there will be few participants, but the number will be large enough again to drive profits to zero. This view is plausible as a gross approximation. Within a year of



the first announcement of high-temperature superconductivity, hundreds of firms had commenced R&D projects to explore its perceived high potential—far more than the number of firms working to devise a more easily opened sardine can. Yet the approximation is indeed rough. Firms are limited by managerial know-how, channels of distribution, production capabilities, and patiently accumulated R&D staff skills in the innovation challenges to which they can respond. In a dynamic world, these change, and can be changed, much more slowly than the rate at which the technological knowledge base and demand conditions shift. Especially when the quasi-rent potential of an innovation could support far more firms than the number possessing the capabilities needed to exploit it, convergence to a zero-profit R&D equilibrium seems highly implausible. A zero-profit result is more readily imagined when there are more firms with relevant capabilities than the quasi-rent potential can accommodate. But even then, history teaches, some firms recognize an opportunity more quickly than others, and the first movers characteristically gain advantages that prevent imitators from driving the industry to a symmetric zero-profit equilibrium.[77]

A more significant contribution of the "endogenous market structure" analyses is their extension of the Barzel schema to alternative quasi-rent eroding scenarios. In particular, when the success of a given R&D project is highly uncertain,[78] but the payoffs are large relative to the cost of a single project, something resembling a lottery may occur. Multiple firms will undertake one or more projects, and in the limiting case, the number of projects will be such that the expected value of the winning ticket(s) equals the sum of all firms' project costs. If the first or most successful participants acquire patents, and if the patents provide strong protection, there may be many losers and only one winner, or a very few winners. That in itself is uninteresting, unless firms are risk-averse. But in addition, under some conditions, the lessened probability of being a winner as the number of rivals rises may cause individual firms to reduce the intensity of their R&D efforts.[79] This does not mean that the rate of innovation is retarded by competition, since the increase in the number of rivals normally raises the joint probability of an early success, more than offsetting the reduced probability that any single firm will be successful.

75. See especially C. C. von Weizsäcker, *Barriers to Entry* (Berlin: Springer, 1980), Chapters 8 and 9; and Jennifer F. Reinganum, "Innovation and Industry Evolution," *Quarterly Journal of Economics*, vol. 100 (February 1985), pp. 81–99.

76. See Glenn C. Loury, "Market Structure and Innovation," *Quarterly Journal of Economics*, vol. 93 (August 1979), pp. 395–410; Tom Lee and Louis Wilde, "Market Structure and Innovation: A Reformulation," *Quarterly Journal of Economics*, vol. 94 (March 1980), pp. 429–436; Partha Dasgupta and Joseph Stiglitz, "Uncertainty, Industrial Structure, and the Speed of R&D," *Bell Journal of Economics*, vol. 11 (Spring 1980), pp. 1–28; Dasgupta and Stiglitz, "Industrial Structure and the Nature of Innovative Activity," *Economic Journal*, vol. 90 (June 1980), pp. 266–293; Pankaj Tandon, "Rivalry and the Excessive Allocation of Resources to Research," *Bell Journal of Economics*, vol. 14 (Spring 1983), pp. 152–165; and Tandon, "Innovation, Market Structure, and Welfare," *American Economic Review*, vol. 74 (June 1984), pp. 394–403.

77. Simulation results showing how R&D affects market structure in the long run, assuming varying degrees of appropriability and rates of exogenous (or cumulative) knowledge growth, are presented in Richard R. Nelson and Sidney G. Winter, *An Evolutionary Theory of Economic Change* (Cambridge: Harvard University Press, 1982), Chapters 13 and 14. See also Carl A. Futia, "Schumpeterian Competition," *Quarterly Journal of Economics*, vol. 94 (June 1980), pp. 675–695.

78. This is more the exception than the rule, Mansfield's research shows. See note 23 supra.

79. Compare Loury, "Market Structure," with Lee and Wilde, "A Reformulation," supra note 76.

Even more importantly, if the prospect of large gains to a successful innovator induces many firms to undertake R&D, there may be extensive duplication of R&D approaches. This is not per se wasteful. When the success of any single project is uncertain, running duplicated projects hastens success unless the rivals conduct exactly identical experiments, which is unlikely. The greater the social gains from a successful innovation, the larger is the optimal number of parallel but uncertain approaches.[80] But if firms independently proliferate R&D approaches until profits are driven to zero, individual projects may be reduced to inefficiently small scales, and total costs may rise to inefficient levels. A coordinated (more monopolistic) approach might use the R&D resources more efficiently.[81]

From this it does not necessarily follow that competition proceeds too rapidly, for as we have seen, the economic welfare analysis yields no simple generalizations except that secure monopoly proceeds too slowly. Favoring competitive duplication are two further considerations not yet incorporated into the formal analyses of R&D rivalry. First, any single firm (that is, monopolist) coordinating parallel but uncertain R&D approaches is likely to have perceptual blind spots, overlooking some promising avenues and putting too much stress on a committee's favorites. By propagating a greater diversity of approaches, competition often evokes winning solutions at lower cost despite seemingly inefficient duplication.[82] Second, the multiple solutions emerging from duplicated R&D are usually not identical, and among other things, they may add desirable product variety. Product variety, like speed in reaching the market, can of course be carried too far. No simple generalizations are possible, but these considerations strengthen the case for at least some degree of rivalry in research and development.

## *The Evidence*

The theory of how market structure affects the vigor of technological innovation provides a rich array of predictions, some conflicting. To sort out the most likely tendencies, qualitative and especially quantitative evidence must be marshaled. Considerable progress has been made toward this objective. In reviewing the evidence, we consider first relationships at the industry level and then investigate how firm size, both within an industry and spanning multiple industries, matters.

Several general problems must be conquered to study market structure-innovation links statistically. For one, ways must be found to measure the vigor or success of innovative activity. Early investigations focused, for want of better data, on *inputs* into the innovative process—for example, in the first studies, counts of scientists and engineers employed by companies or industries, and later, on expenditures devoted to research and development. Other work, usually possible only after new data development efforts had been completed, analyzed such measures of innovative *output* as the number of invention patents received, tallies of new products and processes introduced by industry members, and the growth of productivity. Here we emphasize the results of the typically more recent studies taking advantage of superior data sources.[83]

Second, to avoid biased inferences, it is necessary to take into account variables other than market structure that affect the pace of innovation. Particularly important in this respect is some measure of what has come to be called *technological*



*opportunity*—that is, the rate at which more or less exogenous and cumulative advances in science and technology generate profitable new innovative possibilities.

As always, one must get the direction of causation right. Our theories reveal that market structure can affect the pace of innovation, but innovation in turn can shape market structure. The structure-to-innovation linkage probably operates over a much shorter time span than the innovation-to-structure linkage, but especially in industries blessed with rich technological opportunities, powerful links of the second type also exist, and so, as we shall see, controlling for opportunity takes on added importance.

### The Role of Market Concentration

In the long run, improved standards of living track productivity growth, that is, the growth of real output per hour of work. At first glance, it might seem surprising to postulate a link between market structure and productivity growth. In a modern economy, there are rich interrelationships among industries. Some industries specialize in producing technologically advanced machines, components, and materials which, when purchased by other industries, enhance the buying industry's productivity.[84] Concentrated or atomistic, for an industry not to take advantage of such externally supplied advances would be like refusing a free lunch. However, closer scrutiny reveals that while *non*manufacturing industries do rely almost exclusively upon manufacturers for their capital goods, a majority of the *special-purpose* production equipment used by U.S. manufacturers is internally developed.[85] Thus, the stronger the incentives for internal development are, the more rapid manufacturers' productivity growth should be.

A positive and statistically significant correlation between productivity growth and seller concentration ratios has been found for U.S. manufacturing industries over time periods ranging from 1919 through 1978.[86] However, when industry expenditures on product and process research and development per dollar of sales were included as additional explanatory variables, the R&D variables took away the concentration indices' explanatory power, reducing them to statistical insignificance. Thus, the chain of causation appears to run from higher R&D spending, which is correlated with seller concentration, to higher productivity growth. The question remains, what is the nature of the R&D-concentration relationship?

80. See Scherer, *Innovation and Growth*, supra note 9, Chapter 4.
81. An alternative to monopoly could be a research and development joint venture. See, for example, Barry Bozeman, Albert Link, and Asghar Zardkoohi, "An Economic Analysis of R&D Joint Ventures," *Managerial and Decision Economics*, vol. 7 (December 1986), pp. 263–266; John T. Scott, "Diversification versus Co-operation in R&D Investment," *Managerial and Decision Economics*, vol. 9 (June 1988), pp. 173–186; Alexis Jacquemin, "Co-operative Agreements in R&D and European Antitrust Policy," *European Economic Review*, vol. 32 (March 1988), pp. 551–560; and Thomas M. Jorde and David J. Teece, "Innovation, Cooperation, and Antitrust," paper presented at a University of California, Berkeley, conference on Antitrust, Innovation, and Competitiveness, October 1988.
82. This is the pervasive message of Burton Klein, *Dynamic Economics* (Cambridge: Harvard University Press, 1977).
83. For more thorough surveys of earlier empirical research, see the first and second editions of this text, Chapter 15, and Baldwin and Scott, *Market Structure*, supra note 37, pp. 64–113.
84. See p. 614 supra; Scherer, *Innovation and Growth*, supra note 9, Chapters 3 and 15; and Albert Link, "Alternative Sources of Technology: An Analysis of Induced Innovations," *Managerial and Decision Economics*, vol. 4 (March 1983), pp. 40–43.
85. Scherer, *Innovation and Growth*, p. 250.
86. See *Innovation and Growth*, pp. 250–252; Douglas F. Greer and Stephen A. Rhoades, "Concentration and Productivity Changes in the Long and Short Run," *Southern Economic Journal*, vol. 43 (October 1976), pp. 1031–1044; and Louis Amato and J. Michael Ryan, "Market Structure and Dynamic Performance in U.S. Manufacturing," *Southern Economic Journal*, vol. 47 (April 1981), pp. 1105–1110.

Our earlier theoretical analysis predicts that more rivalry, approximated by lower concentration indices, invigorates R&D spending up to a point, but that too atomistic a market structure discourages R&D by causing would-be innovators to appropriate an insufficiently large share of the ensuing benefits to expect positive profits from their innovations. Multiseller rivalry is more apt to stimulate R&D spending when advances in the underlying science and technology base occur quickly and unexpectedly, generating large quasi-rent opportunities for the tapping, than when the pace of advance is slow and continuous. Both predictions have received statistical support.

Most studies for the United States and other leading nations reveal a positive correlation between concentration and industry R&D/sales ratios, or cruder proxies for that ratio.[87] A test for nonlinearities using 1960 U.S. employment data showed an "inverted-U" relationship, with peak R&D/sales ratios occurring at average four-digit industry, four-seller concentration indices of 50 to 55.[88] Industries with four-firm shares below 15 percent appeared to have fatally defective incentives for supporting R&D.

Since then, the inverted-U hypothesis has been tested repeatedly as richer data have become available, especially from the Federal Trade Commission's Line of Business statistical surveys covering the years 1974 to 1977. Working with FTC data aggregated to the industry level, Richard Levin and associates found strong initial support for the inverted U, with the maximum R&D/sales ratio occurring at a four-firm concentration ratio of 52.[89] From their survey of research and development executives, Levin et al. had parallel, subjectively measured, indices of the rate at which new products and processes had been introduced into 130 manufacturing industries during the 1970s. Concentration was found to influence those measures in ways nearly identical to those for the R&D/sales ratios. Disaggregating FTC R&D/sales data to the level of 3,388 individual lines of business, John Scott also observed an inverted U in two-variable regressions, with the maximum intensity of R&D at (adjusted) four-firm concentration ratios of 64.[90]

However, the inverted U hypothesis fares less well when additional variables are introduced to account for technological opportunity and other innovation-affecting influences. Using simple dummy variables at the two-digit industry level to control for interindustry differences, Levin et al. found their results virtually unchanged; the inverted U persisted. They then introduced a battery of survey-derived indices assessing the relevance of diverse scientific fields to the industries' R&D efforts, the extent to which outside R&D performers contributed to industry technological progress, and the strength of various innovative reward appropriation mechanisms (such as patents, secrecy, and lead time). With these variables added, the concentration coefficients fell to insignificant values, and the existence of an inverted U could no longer be inferred. Similarly, the "U" disappeared when Scott added to his 3,388-line analysis twenty two-digit industry dummy variables and 437 dummy variables permitting each sample company to have its own best-fitting R&D/sales relationship.[91] Evidently, the U-shaped concentration influence captured by relatively simple statistical analyses is correlated with a set of more complex industry and firm effects. Whether those effects influ-



ence concentration causally, are affected by it, or are spuriously correlated with it, is unclear, so one cannot be certain whether the inverted U is a phenomenon important in its own right or an accident of the data. That the underlying theory points to its existence suggests that the phenomenon should not be dismissed too quickly.

Interindustry differences in the richness of technological opportunities might affect concentration-R&D relationships in another way. Inability to appropriate a sufficient share of an innovation's quasi rents because of excessive rivalry is more likely when the relevant science base is advancing slowly and predictably than when it moves forward rapidly and discontinuously, that is, in breakthroughs. Therefore, we expect R&D/sales ratios to be more strongly correlated with seller concentration indices, the less rich an industry's technological opportunities are. Typically, the role of technological opportunity is investigated by classifying industries either dichotomously or into more elaborate subdivisions reflecting the principal technologies exploited. Most U.S. and European studies that have attempted such a test support the differential correlation hypothesis.[92] A particularly rich categorization was possible using data on 1974 company-financed R&D/sales ratios from the FTC's Line of Business survey. The industry technology groupings, the number of industries in each, and the simple correlation

87. See again Baldwin and Scott, *Market Structure*, supra note 37. Examining the composition of 108 relatively large U.S. companies' R&D portfolios, Edwin Mansfield found the percentage of company expenditures devoted to basic research, long-term projects, and entirely new products and processes to be inversely correlated with seller concentration in industries occupied by the firms. "Composition of R and D Expenditures: Relationship to Size of Firm, Concentration, and Innovative Output," *Review of Economics and Statistics*, vol. 63 (November 1981), pp. 610–613. For similar results, see Albert N. Link, "An Analysis of the Composition of R&D Spending," supra note 53.

88. Scherer, *Innovation and Growth*, supra note 9, p. 246. See also Thomas M. Kelly, "The Influences of Size and Market Structure on the Research Efforts of Large Multiple-Product Firms," Ph.D. diss., University of Oklahoma, 1969, pp. 85–86; and (on the food processing industries) John D. Culbertson, "Should Antitrust Use the Schumpeterian Model?" in Robert L. Wills et al., ed., *Issues After a Century of Federal Competition Policy* (Lexington: Heath, Lexington, 1987), pp. 106–107.

89. Richard C. Levin, Wesley M. Cohen, and David C. Mowery, "R&D, Appropriability, and Market Structure: New Evidence on Some Schumpeterian Hypotheses," *American Economic Review*, vol. 75 (May 1985), pp. 20–24.

90. John T. Scott, "Firm versus Industry Variability in R&D Intensity," in Zvi Griliches, ed., *R&D, Patents, and Productivity* (Chicago: University of Chicago Press, 1984), pp. 233–240. See also Reinhard Angelmar, "Market Structure and Research Intensity in High-Technological-Opportunity Industries," *Journal of Industrial Economics*, vol. 34 (September 1985), pp. 69–79, who analyzed disaggregated PIMS data on 160 relatively high-technology business units and found an inverted U, with maximum R&D/sales ratios at four-digit four-firm concentration ratios of 44. For evidence of a logistic relationship for Finland peaking at a three-firm concentration ratio of roughly 80, see B. Wahlroos and M. Backström, "R&D Intensity with Endogenous Concentration," *Empirical Economics*, vol. 7 (1982, no. 1/2), pp. 13–22.

91. Because company diversification occurs largely through merger and high-R&D companies tend to acquire companies with similarly high R&D/sales ratios, such company dummy variables are undoubtedly mirroring complex industry effects. See David J. Ravenscraft and F. M. Scherer, *Mergers, Sell-offs, and Economic Efficiency* (Washington: Brookings, 1987), p. 51.

92. Supporting studies include Scherer, *Innovation and Growth*, supra note 9, pp. 241–246; William S. Comanor, "Market Structure, Product Differentiation, and Industrial Research," *Quarterly Journal of Economics*, vol. 85 (November 1967), pp. 524–531; Ronald Shrieves, "Market Structure and Innovation: A New Perspective," *Journal of Industrial Economics*, vol. 26 (June 1978), pp. 329–347; John Lunn, "An Empirical Analysis of Process and Product Patenting: A Simultaneous Equation Framework," *Journal of Industrial Economics*, vol. 34 (March 1986), pp. 319–328; John Lunn and Stephen Martin, "Market Structure, Firm Structure, and Research and Development," *Quarterly Review of Economics and Business*, vol. 26 (Spring 1986), pp. 31–44; and (for the Netherlands) Alfred Kleinknecht and Bart Verspagen, "R&D and Market Structure: The Impact of Measurement and Aggregation Problems," *Small Business Economics*, vol. 1 (December 1989). For supporting Canadian and French studies and a contrary Belgian analysis, see the second edition of this text, p. 437, note 112.

coefficients between industry R&D/sales ratios and 1972 four-firm concentration indices were as follows:

| Technology Grouping | Number of Industries | Correlation Coefficient |
|---|---|---|
| All industries | 236 | 0.347* |
| Traditional technologies | 78 | 0.305* |
| General and mechanical | 106 | 0.404* |
| Organic chemicals | 6 | 0.210 |
| Other chemicals | 12 | 0.101 |
| Metallurgical | 12 | 0.165 |
| Electronics | 9 | 0.362 |
| Electrical | 13 | −0.158 |

Positive and statistically significant correlations (denoted by asterisks) are found for all industries together, the least progressive "traditional" technologies (for example, dairies and brick making), and the general and mechanical technologies. For industries rooted more firmly in fast-moving chemical and electrical technologies, the correlations are small and in one case negative. Only the electronics industry stands out as a contradiction to the hypothesis, with the second-highest correlation. However, that correlation is strongly influenced by a single concentrated industry's values (for computers), and with only nine industries in the subsample, the correlation falls far short of being statistically significant. Thus, there appears to be a rough tendency for concentration to be more conducive to technological vigor in relatively slow-moving fields. In no case, it must be recognized, are the concentration-R&D correlations strong. A conclusion that emerges from every such study is that interindustry differences in technological opportunity, however measured, have much greater power in explaining varying R&D or innovation intensities than differences in such market structure indices as concentration.

A more novel approach to the problem of relating market structure and technological opportunity has been taken by Paul Geroski.[93] He used extraordinarily rich data tallying 1,203 product and process innovations emerging from seventy-three British manufacturing industries (defined at the three-digit level) over the period 1970 to 1979. Splitting the sample into two time segments, he used the level of innovative activity in an industry during one time period as a predictor of innovation in the other period. When this approach to controlling for technological opportunity was adopted, he found higher seller concentration (and increases in other monopoly-related variables) to have a significant *negative* impact on the emergence of innovations. When no such controls were included, the impact of concentration was positive but statistically insignificant. However, in addition to the negative direct influence of concentration with opportunity controls, Geroski discovered that greater monopoly power led to larger time-lagged profit margins, and the expectation of those higher margins had a *positive* influence on innovation.



In this indirect way, monopoly favored innovation. But when the direct and indirect influences were combined, the negative direct effect substantially outweighed the positive indirect influence (operating through expected profitability). Thus, on balance, Geroski's results indicate that high market concentration was more likely to retard innovation than to stimulate it.

It would be premature to conclude that this new U.K. evidence overturns the weight of prior evidence suggesting a modest positive influence for concentration, especially in low-opportunity industries. Much remains to be learned.

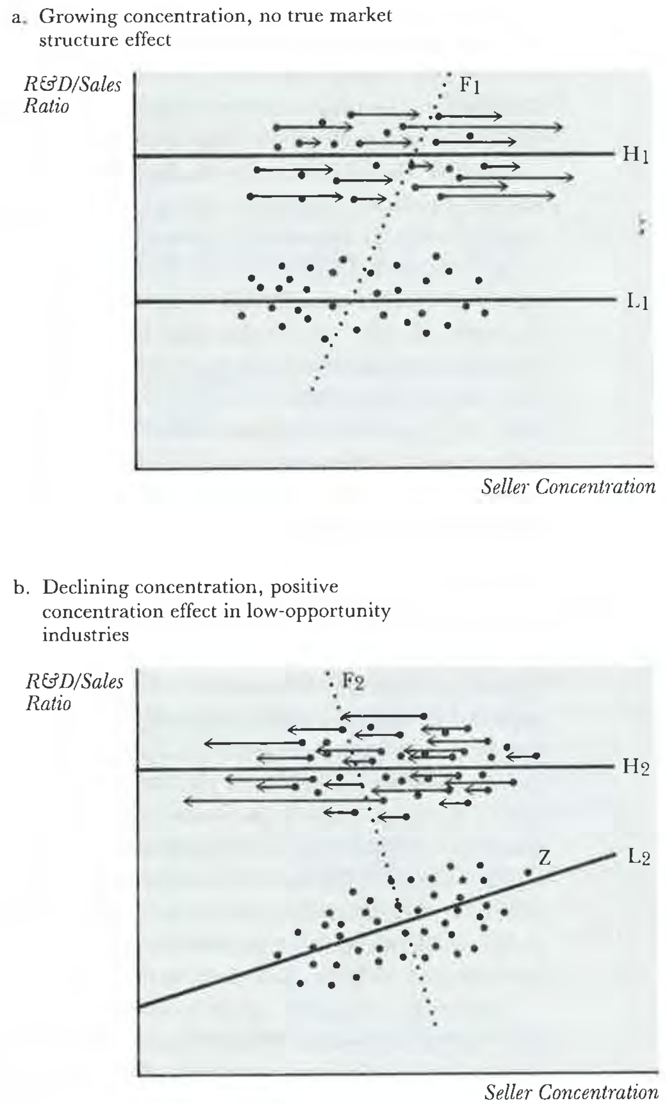
A particularly important loose end is the chain of causation leading from innovation to market structure. If vigorous innovation induces appreciable market structure changes, and if the need for concentration to appropriate quasi rents depends upon technological opportunity, opportunity and structural feedback effects might be confounded. Figure 17.8 lays out the connections in simplified form. In Figure 17.8(*a*), two technological opportunity regimes are postulated. Seller concentration (measured on the horizontal axis) is assumed to have no causal influence on the intensity of innovation (vertical axis), so the "true" relationships between concentration and innovation are given by the upper line $H_1$ (for high-opportunity industries) and $L_1$ (for low-opportunity industries). Random deviations about the true tendency lines are shown by the scatter of dots, one per industry. However, if vigorous innovation leads to increasing concentration, the industry observations in high-opportunity fields will be shifted over time in a rightward direction, as shown by the arrows. As the center of gravity for the high-opportunity industry observations shifts to the right, a regression line fitted to the shifted data without controls for opportunity will approximate dotted line $F_1$, showing a weak but spurious positive concentration-innovation association. Figure 17.8(*b*) illustrates a case more consistent with theory: concentration has no effect on innovation in high-opportunity fields, but a positive effect when opportunities are weak. If (reversing our previous assumption) rapid innovation leads to *lower* concentration, a regression analysis that fails to control for opportunity will show a weak but spurious *negative* concentration-innovation association: the opposite of reality in low-opportunity industries.[94]

To avoid biases under these circumstances, a well-specified simultaneous equation system is needed. If differences in technological opportunity are not controlled properly, if concentration is favorable to innovation, and if rapid innovation raises concentration levels over time, a "virtuous circle" dynamics will ensue, and cross-sectional regression analyses will underestimate concentration's beneficial long-run effects. Or, alternatively, if innovation thrives in relatively atomistic markets and rapid innovation *reduces* concentration, the virtuous circle dynamics will again be underestimated even when there are perfect controls for opportunity.

93. P. A. Geroski, "Innovation, Technological Opportunity, and Market Structure," *Oxford Economic Papers*, forthcoming in 1989.

94. Other biases can result if opportunity is improperly inferred from high innovation intensity. For example, point Z in Figure 17.8(*b*) is in the low-opportunity set, but has relatively high innovation because of favorable market structure. If Z is put in a medium-opportunity class because of its observed performance, the effect of concentration will be underestimated. This bias could occur with Geroski's control technique.

**Figure 17.8**
Biases Resulting from Innovation-Induced Concentration Changes




Innovation could be concentration-increasing if successful innovators rise to market dominance and can defend themselves successfully from imitators,[95] or if vigorous innovation increases the variance of firms' growth rates under many versions of Gibrat's Law.[96] However, the standard Gibrat formulations assume no entry. If innovation stimulates the entry of new competitors, concentration could tend to fall in high-opportunity industries.

Historical studies reveal that high concentration in such American and European industries as synthetic fibers, synthetic rubber, synthetic dyestuffs and derivative organic chemicals, electric lamps, telephone equipment, aircraft engines,



and photographic supplies was caused in part by vigorous innovation combined with patent and/or know-how barriers to imitation. However, it is less clear that these tendencies have survived into more recent times. Using the British innovation count data described earlier, Geroski found a strong tendency for concentration to fall during the 1970s with more vigorous innovative activity.[97] A U.S. study using less satisfactory data also found a tendency toward falling concentration in high-opportunity industries.[98] John Lunn discovered that U.S. market concentration tended to rise with innovation when innovative activity emphasized internal production process changes, but to fall when new product work predominated.[99] In the United States, it will be recalled, roughly three-fourths of industrial R&D is oriented toward products and one-fourth toward processes. Lunn's findings are reinforced by those of Gort and Konakayama, who found that new entry rates were much higher than exit rates in the early life cycle stages of major product innovations.[100] Although much remains to be learned on this important question, the weight of existing evidence favors a conclusion that innovation under late twentieth-century conditions has tended to be more concentration-reducing than the opposite. This in turn implies possible underestimation of concentration's R&D-supporting role if the statistical controls for technological opportunity are inadequate.

## The Advantages of Large and Small Firms

We move now from the industry to the firm level. Within a given industry, are relatively large or relatively small firms the more potent innovators? And do diversified companies—those that span multiple industries—pursue research and development more aggressively than those that specialize in a single line of business?

Theory offers ambiguous predictions on one aspect of the firm size question. As we have seen, firms with relatively small or (for new entrants) zero market shares have incentives to force the innovative pace when they can anticipate gaining first-mover advantages and capturing substantial chunks of market share. However, dominant firms subjected to such threats are motivated to respond aggressively, minimizing the fringe firms' lead or even preempting them. Which pattern is more common can only be ascertained empirically.

95. This is the scenario emphasized by Nelson and Winter in *An Evolutionary Theory*, supra note 77, Chapters 13 and 14.
96. See pp. 141–146 supra.
97. "Innovation," supra note 93, footnote 11; and P. A. Geroski and R. Pomroy, "Innovation and the Evolution of Market Structure," London Business School Centre for Business Strategy, working paper no. 36 (November 1987).
98. Arun Mukhopadhyay, "Technological Progress and Change in Market Concentration in the U.S., 1963–77," *Southern Economic Journal*, vol. 52 (July 1985), pp. 141–149. See also Micha Gisser, "Price Leadership and Dynamic Aspects of Oligopoly in U.S. Manufacturing," *Journal of Political Economy*, vol. 92 (December 1984), pp. 1035–1048, who found rising productivity to be accompanied by Galtonian tendencies toward rising concentration in low-concentration industries and falling concentration in high-concentration industries.
99. "An Empirical Analysis," supra note 92, pp. 324–328. In an elaborately structured model, Richard C. Levin and Peter C. Reiss found greater R&D efforts to induce *higher* concentration. See "Cost-Reducing and Demand-Creating R&D with Spillovers," National Bureau of Economic Research working paper no. 2876 (March 1989), and for similar results with more aggregated data, "Tests of a Schumpeterian Model of R&D and Market Structure," in Griliches, ed., *R&D, Patents and Productivity*, supra note 90, pp. 175–202.
100. Michael Gort and Akira Konakayama, "A Model of Diffusion in the Production of an Innovation," *American Economic Review*, vol. 72 (December 1982), pp. 1111–1119. Compare note 43 supra.

Large companies have noteworthy advantages in supporting research and innovation. Their size permits them to maintain a diversified portfolio of R&D projects, hedging the risks that any given project will fail. The ability to exploit scale economies is another potential advantage. A large laboratory can justify purchasing highly specialized equipment such as wind tunnels, supercomputers, differential scanning calorimeters, and much else. It can employ specialists in many disciplines to cross-fertilize one another and lend temporary assistance when a team working on some project bogs down on a technical problem outside its normal sphere of competence.[101] Scale economies may also accrue in other parts of the large firm's operations. As we have seen in Chapter 4, large corporations can attract capital at lower cost than their smaller cousins and may therefore be better able to finance ambitious R&D undertakings. This has probably become less important over time, at least in the United States, with the growth of venture capital firms seeking to invest in small high-technology enterprises. Large corporations usually have well-established marketing channels and may realize scale economies in advertising and other promotional activities (such as the "detailing" of new drugs by field salespersons). Such promotional advantages permit them to penetrate markets more rapidly with new products, enhancing the products' expected profitability. And finally, large producers have stronger incentives to develop internal process improvements. A new process that reduces costs by a given percentage margin yields larger total savings, the larger the developing firm's affected output is.[102]

Against this impressive array of actual and conjectured advantages, the disadvantages of corporate size must be weighed. For one, research in large laboratories can become overorganized. If too many people are involved in a project, they spend a disproportionate amount of their time writing memoranda to each other at the expense of more creative endeavor. Also, the quickest path to higher status and pay in a large firm's R&D establishment often entails giving up work at the bench and becoming a member of the management team. Although some companies have tried to combat this tendency by creating well-paid positions for senior research fellows, it is still commonplace to find the most able people in a laboratory devoting nearly all their time to supervising others. This is not the way truly creative work gets done.

Even more important, small firms may be more adept at risk taking. Their decisions to go ahead with an ambitious project typically are made by a handful of people who know one another well. In a large corporation, on the other hand, the decision must filter through an elaborate chain of command—the person with the idea, his or her section chief, the laboratory manager, the vice president for research, and if substantial financial commitments are required, several members of top management. Under these circumstances there is a distressingly high probability that some member of the chain will be what C. Northcote Parkinson has called "an abominable no-man," objecting decisively to ideas that are untried or that stray too far from accepted ways of doing business.

One consequence of this syndrome, which has been noted time and again in case histories and treatises on research management, is a bias against really imaginative innovations in the laboratories of large firms. Inability to get ideas approved by higher management drives creative individuals out of large corporate



R&D organizations to go it alone with their own ventures. Thousands of research-based new enterprises have been founded by frustrated expatriates from the laboratories of such U.S. giants as IBM, Sperry-Rand (now Unisys), Western Electric, Hughes Aircraft, and Texas Instruments.[103]

Jewkes, Sawers, and Stillerman compiled case histories of seventy important twentieth-century "inventions" and learned that only twenty-four had their origin in industrial research laboratories.[104] More than half were pioneered by individuals working either completely independent of any formal research organization or in an academic environment. However, one must be wary of carrying this insight too far. Further analysis by Jewkes et al. revealed that sizable corporations often shouldered the burden of *developing* independent inventors' ideas for commercial utilization. And for complex innovations like high-performance aircraft, high-definition television equipment, and nuclear reactor systems requiring R&D expenditures of tens or even hundreds of millions of dollars before commercialization can commence, only very large firms can undertake the tasks of technical development with something approaching equanimity.[105]

It is nevertheless well-established that new entrants without a commitment to accepted technologies have been responsible for a substantial share of the really revolutionary new industrial products and processes. The illustrations are legion: arc lighting (Brush), the incandescent lamp (Edison), radio telegraphy (Marconi), radio telephony (Fessenden and de Forest), FM radio (Armstrong), the photoflash lamp (Wabash), the dial telephone (Automatic Electric), the turbojet engine (Whittle in England, Heinkel and Junkers in Germany), sound motion pictures (Western Electric and Warner Brothers), catalytic cracking of petroleum (Houdry), the electric typewriter (IBM), the ball-point pen (Reynolds), self-developing photography (Polaroid), electrostatic copying (Haloid), supine dentistry (Den-Tal-Ez), the microwave oven (Raytheon), the microprocessor chip (Intel),[106] the microcomputer (Altair and Apple), polytetrafluoroethylene arterial grafts (IMPRA), and (unsuccessfully) laser-actuated hydrogen fusion (KMS Industries), to name only a few. In several of these cases, well-established firms flatly rejected invitations to collaborate with the inventor of a concept that later revolutionized

101. This large firm advantage is minimized when small firms have ready access to outside specialists such as university science and engineering faculty. It is not clear, however, whether outside expertise is tapped as willingly as internal expertise.

102. See Albert N. Link, "Firm Size and Efficient Entrepreneurial Activity: A Reformulation of the Schumpeter Hypothesis," *Journal of Political Economy*, vol. 88 (August 1980), pp. 771–782, who found that productivity growth returns to R&D expenditures rose with company size within the broadly defined chemicals industry.

103. For an early study of the phenomenon, see Edward B. Roberts, "Entrepreneurship and Technology," *Research Management*, vol. 11 (July 1968), pp. 249–266. On the efforts of large corporations to stem the tide, see Gifford Pinchot III, *Intrapreneuring: Why You Don't Have To Leave the Corporation To Become an Entrepreneur* (New York: Harper and Row, 1986).

104. *The Sources of Invention*, supra note 17, pp. 65–78. See also Dan Hamberg, "Invention in the Industrial Research Laboratory," *Journal of Political Economy*, vol. 71 (April 1963), pp. 95–115; and Willard F. Mueller, "The Origins of the Basic Inventions Underlying du Pont's Major Product and Process Innovations, 1920 to 1950," in the National Bureau of Economic Research conference report, *The Rate and Direction of Inventive Activity* (Princeton: Princeton University Press, 1962), pp. 323–346.

105. Surprisingly, the fraction of business lines' patent portfolios devoted to (relatively complex) systems and subsystems inventions was observed to increase only weakly with unit size, for example, from 51.0 percent with sales of $100 million to 53.8 percent with sales of $1 billion. Scherer, "The Propensity To Patent," supra note 36, pp. 124–125.

106. For a detailed analysis of the innovative role of small new firms in computers and semiconductors, see Nancy Dorfman, *Innovation and Market Structure* (Cambridge: Ballinger, 1987).

their industry. Many other cases can be found in which the threat of entry through innovation by a newcomer stimulated existing members to pursue well-known technical possibilities more aggressively. Examples include General Electric's handling of the fluorescent lamp; AT&T's development of microwave radio relay systems, cordless telephones, and electronic office switchboards; IBM's response to the electronic computer innovations of Sperry Rand, Control Data, Digital Equipment, and Compaq; and the sudden awakening of old-line aircraft makers' interest in basic research and systems engineering when the U.S.Air Force chose the infant Ramo-Wooldridge Corporation to oversee its Atlas ICBM development program.

The qualitative evidence supports a preliminary conclusion that no single firm size is uniquely conducive to technological progress. There is a place for firms of all sizes. Technical progress thrives best in an environment that nurtures a diversity of sizes and, perhaps especially, that keeps barriers to entry by technologically innovative newcomers low.

## A Quantitative Perspective

With this qualitative generalization in mind, we move to the evidence on such quantitative indicators of innovative performance as research and development expenditures, patenting, and the origination of significant product and process innovations.

In 1982, there were approximately 294,000 manufacturing enterprises in the United States. National Science Foundation surveys reveal that only about 12,000 had expenditures on formally organized research and development programs. The fraction of companies conducting formal R&D rises with firm size. Thus, R&D programs were sustained in 1982 by 293 of the 300 manufacturing companies with 10,000 or more employees, by 184 of the 223 companies with from 5,000 to 9,999 employees, and by 693 of the 1,352 firms with 1,000 to 4,999 employees.[107] Enterprises with 10,000 or more employees performed 81.3 percent of all company-financed R&D in that year while employing 45 percent of the persons engaged in manufacturing. Companies with fewer than 1,000 employees had an employment share of 35 percent and a company-financed R&D share of 5 percent. Thus, formally organized R&D is much more the forte of relatively large than small firms. As a qualification to this conclusion, it must be recognized that inventive and innovative activities are also pursued outside the context of formal R&D programs, particularly in smaller corporations.[108]

Among the manufacturing corporations that conduct sufficient amounts of company-financed R&D to report the sums spent to their stockholders, there is no tendency for the very largest firms to contribute disproportionately. *Business Week's* tabulation of R&D expenditure disclosures for 1987 included 915 corporations.[109] When the companies were ranked on the basis of 1987 sales, the cumulative fractions of sales and R&D outlays for various size cohorts were as follows:

| Number of Firms, Ranked by Sales | Percentage of All 915 Firms' | |
|---|---|---|
| | **Sales** | **R&D Outlays** |
| Top 10 | 32.5 | 32.3 |
| Top 25 | 47.1 | 44.4 |
| Top 50 | 60.9 | 59.7 |
| Top 100 | 76.2 | 74.4 |
| Top 200 | 89.1 | 86.7 |
| All 915 | 100.0 | 100.0 |



There is a hint that the ten largest corporations had an R&D share closer to their sales share than companies of less gigantic size, but the differences are sufficiently small throughout the size distribution that we cannot confidently reject an inference that R&D outlays were proportional to sales.

This rough proportionality seems to be a phenomenon of long standing. David Mowery traced the spread of industrial R&D activity in the United States from 1921, when only 35 percent of the 200 largest manufacturers reported formal programs, to 1946, when 84 percent of the top 200 had programs.[110] Except in the chemicals industry, R&D employment increased less than proportionately with company size (measured by asset values). For chemical manufacturers, R&D employment rose more than proportionately with size in 1921 and 1933 but roughly proportionately in 1946.

R&D spending is a measure of inputs into the process of advancing industrial technology. At least as vital to assessing the role firm size plays is how effectively those inputs are transformed into outputs.[111] Patent counts are the most comprehensive quantitative indicator of industrial technology outputs. The Federal Trade Commission's Line of Business survey for 1974 included the 250 largest U.S. manufacturing corporations in that year, ranked by domestic sales, plus 193 additional (uniformly sizable) producers. The 443 surveyed corporations obtained 61 percent of the invention patents issued to U.S. industrial corporations in

**107.** U.S. National Science Foundation, *Research and Development in Industry, 1982*, NSF 84–325 (Washington: microfiche, 1984), pp. 14 and 19; and U.S. Bureau of the Census, *1982 Enterprise Statistics*, vol. 1, "General Report on Industrial Organization (Washington: USGPO, October 1986), Table 3.

**108.** See, for example, Jacob Schmookler, "Bigness, Fewness, and Research," *Journal of Political Economy*, vol. 67 (December 1959), p. 630, who found that for every eight inventions stemming from full-time R&D employees, companies obtained five inventions from employees engaged only part-time in inventive activity. On the undercounting of R&D in the Netherlands, see Alfred Kleinknecht, "Measuring R&D in Small Firms: How Much Are We Missing?" *Journal of Industrial Economics*, vol. 36 (December 1987), pp. 253–256. For a case study, see Samuel Hollander, *The Sources of Increased Efficiency* (Cambridge: MIT Press, 1965), Chapters 7 and 8.

**109.** "A Perilous Cutback in Research Spending," *Business Week*, June 20, 1988, pp. 139–160. Companies classified to the service and financial industries are excluded. The *Business Week* listing is confined to companies with sales of $35 million or more and R&D expenses of at least $1 million or 1 percent of sales. This selection criterion (and the corporate reporting practices that underlie it) biases the list in favor of relatively research-oriented companies, excluding many others, small and large. Among the excluded large corporations were Philip Morris, Shell Oil, RJR Nabisco, Tenneco, and BP America — all on *Fortune's* list of the 25 largest U.S. industrial corporations in 1987. For a similar analysis of 1975 data, see the second edition of this text, p. 420.

**110.** David C. Mowery, "Industrial Research and Firm Size, Survival, and Growth in American Manufacturing, 1921–1946," *Journal of Economic History*, vol. 43 (December 1983), pp. 953–979.

**111.** On the importance of having output data and the theory underlying the interpretation of input–output relationships, see Franklin M. Fisher and Peter Temin, "Returns to Scale in Research and Development: What Does the Schumpeterian Hypothesis Imply?" *Journal of Political Economy*, vol. 81 (January-February 1973), pp. 56–70; Meir Kohn and John T. Scott, "Scale Economies in Research and Development: The Schumpeterian Hypothesis," *Journal of Industrial Economics*, vol. 30 (March 1982), pp. 239–249; and John Lunn, "Research and Development and the Schumpeterian Hypothesis: Alternate Approach," *Southern Economic Journal*, vol. 49 (July 1982), pp. 209–217. On Australian input-output relationships, see Ian W. McLean and David K. Round, "Research and Product Innovation in Australian Manufacturing Industries," *Journal of Industrial Economics*, vol. 27 (September 1978), pp. 1–12.

the lagged time period when patents from 1974 R&D were granted.[112] They accounted for 73 percent of company-financed research and development expenditures, 68 percent of the value of plant and equipment in manufacturing, and 52.4 percent of U.S. manufacturers' sales in 1974. Thus, the largest firms received patents more than proportionate to their sales, but less than proportionate to their capitalization and R&D outlays.[113] The same 443 FTC sample members won 55 percent of the "most significant technical advance of the year" citations bestowed upon U.S. corporations by the magazine *Research & Development* for the years 1976 through 1980.[114] Evidently, the largest manufacturers derived fewer patents and significant technical advances from their R&D money than smaller firms.

Another view of large versus small firm innovative output propensities emerges from a survey that sought to compile a comprehensive list of technical innovations introduced by U.S. manufacturing firms in 1982. Of the 4,531 innovations so identified, 42 percent came from small firms, defined as those with fewer than 500 employees. Those small firms averaged 322 innovations per million employees, while companies with 500 or more employees averaged 225 innovations per million employees.[115] Again, the implication is that relatively small enterprises make a disproportionate contribution to innovative output, especially in view of their modest formal R&D expenditures share.

The superior innovative performance of smaller firms does not persist across all industries. In some fields, firms with fewer than 500 employees contributed a share of innovations lower than their employment share; in others, as the overall averages require, they surpassed their employment share. Analyzing the relative innovative performance of large as compared to small companies, Acs and Audretsch discovered that small firms had higher innovation-per-employee ratios than their larger counterparts in four-digit industries with high innovation rates, an employment mix rich in professional workers, and modest small-firm employment shares.[116] The first two variables reinforce the prediction from theory that firms with small market shares do well when technical progress is rapid. Companies with 500 or more employees "out-innovated" their smaller compatriots in industries where consumer goods advertising was important, production processes were capital intensive, and (for the most innovative industries only) four-firm concentration ratios were relatively high.

Evidence from an international survey of major innovations introduced between 1953 and 1973 suggests that the superior innovative record of small enterprises in the United States is not always mirrored elsewhere. From that survey, the percentages of innovations credited to companies with sales at the time of less than $50 million for various nations were as follows:[117]

| | |
|---|---|
| United States | 50% |
| France | 57 |
| West Germany | 37 |
| United Kingdom | 33 |
| Japan | 20 |

The low fraction for Japan is consistent with well-known facts about that nation's culture: the most able technical graduates favor lifetime jobs in large companies, venture capital sources are meager, and those who leave a large corporation to start their own high-technology ventures suffer ostracism from former colleagues.[118] The United Kingdom's experience is particularly interesting. For it, a comprehensive tally of more than 4,000 significant industrial innovations commercialized between 1945 and 1983 exists. Between 1956 and 1970, companies with fewer than 500 employees contributed a lower share of innovations than their share of industrial employment. But since then, firms in the 100 to 499 employment range have innovated more than proportionately to their employment shares, and indeed, by growing margins.[119] How much this change owes to the emergence of a high-technology venture capital market in the UK is unknown.[120] In contrast to patterns observed for the United States, British companies with 25,000 or more employees have consistently originated a much larger share of innovations than their employment share, while companies in the 2,000 to 9,999 employee range are notable technological laggards. Evidently, the relative roles of small and large businesses depend more upon variations in business culture and the supply of entrepreneurial talent than upon hard-and-fast technological imperatives.



### R&D Activity within Large, Established Firms

A more detailed picture of R&D performance within large, well-established companies can be gleaned by tapping once again the data assembled under the Federal Trade Commission's Line of Business surveys. In those surveys, to reiterate, U.S. manufacturing companies disaggregated information on their operations to individual *lines of business,* defined homogeneously at the three- or four-digit Standard Industrial Classification industry level. For each of the 196 industries in which there were five or more reporting units with nonzero company-financed R&D expenditures in 1974, nonlinear regression equations were computed with the form:

$$R_i = a + b_1 S_i + b_2 S_i^2 + e_i, \tag{17.1}$$

where $R_i$ is the level of R&D expenditures for the $i^{th}$ company's line, $S_i$ is that line's sales, and $e_i$ is an error term.[121] Interpretation of the squared sales variable's coefficient (and in a few cases, the value of the *a* term) permitted the industries to be

**112.** F. M. Scherer, "Technological Change and the Modern Corporation," in Betty Bock et al., ed., *The Impact of the Modern Corporation* (New York: Columbia University Press, 1984), pp. 284–284; and Federal Trade Commission, *Statistical Report: Annual Line of Business Report, 1974* (Washington: 1981), p. 47.
**113.** Within the FTC sample, patenting tended to rise most frequently in rough proportion to line of business R&D outlays, but in the nonconforming cases, patenting rose less than proportionately with R&D more often than disproportionately. See Scherer, "The Propensity To Patent," supra note 36, p. 115.
**114.** Scherer, "Technological Change," pp. 282–283. At the time, the journal was called *Industrial Research & Development.*
**115.** See Zoltan J. Acs and David B. Audretsch, "Innovation, Market Structure, and Firm Size," *Review of Economics and Statistics*, vol. 69 (November 1987), p. 568.
**116.** "Innovation," pp. 571–573.
**117.** Stephen Feinman and William Fuentevilla, *Indicators of International Trends in Technological Innovation*, Final Report to the National Science Foundation, NTIS document PB-263-738 (Jenkintown, PA: Gelman Research Associates, April 1976). The sample sizes for some nations were quite small.
**118.** See also Edwin Mansfield, "Industrial R&D in Japan and the United States," *American Economic Review*, vol. 78 (May 1988), p. 227, who observed that R&D expenditures on entirely new products and processes increased more than proportionately with firm size in Japan, whereas the opposite was true in the United States.
**119.** Keith Pavitt, Michael Robson, and Joe Townsend, "The Size Distribution of Innovating Firms in the UK: 1945–1983," *Journal of Industrial Economics*, vol. 35 (March 1987), pp. 302–304. The innovation tallies are the same as those used in Geroski's analyses, supra notes 93 and 97.
**120.** See "Venture Adventures," *The Economist*, December 19, 1987, p. 67. The U.K. venture capital market is sufficiently new that one might appropriately view it more as a response than an inducement to the change.
**121.** Lines of business with no R&D were retained in the analysis if the five non-zero line criterion was satisfied.

classified into one of three categories, as illustrated in Figure 17.9. When $b_2$ and $a$ were insignificantly different from zero, R&D performance increased proportionately with line of business sales, so the situation described in Figure 17.9(a) as *constant returns* prevailed. When $b_2 > 0$, one has the increasing returns case of Figure 17.9(b); that is, R&D performance increases more than proportionately with size. With $b_2 < 0$, the Figure 17.9(c) diminishing returns case holds.[122]

The results of this curve fitting for 196 industries yielded the following breakdown of cases:

| | Number of Industries |
|---|---|
| No significant departure from constant returns | 140 |
| Increasing returns | 40 |
| Diminishing returns | 16 |

The constant returns case was predominant, indicating that firms with relatively small market shares invested as intensively relative to their size as market leaders. Deviations from the constant returns pattern were biased on the side of increasing returns, implying that leading firms were more aggressive R&D supporters than counterparts with lower market shares.[123]

A similar analysis was carried out for 124 industries in which five or more firms had nonzero patenting. The distribution of cases was as follows:

| | Number of Industries |
|---|---|
| No significant departure from constant returns | 91 |
| Increasing returns | 14 |
| Diminishing returns | 19 |

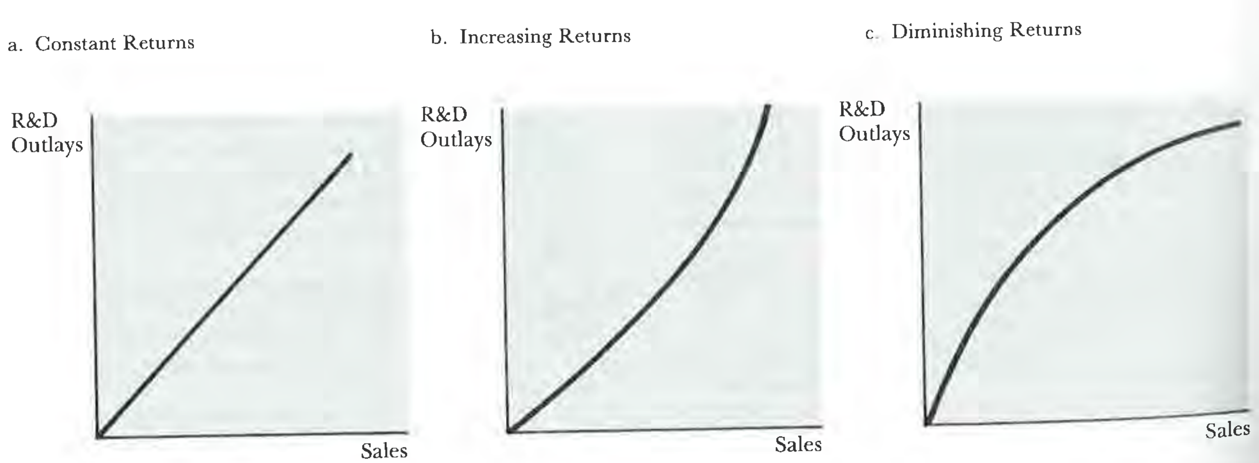



**Figure 17.9**
Three Firm Size–Innovative Activity Cases



Here the story is somewhat different. The tendency toward constant returns was slightly stronger—73 percent of the cases, as contrasted to 71 percent for R&D expenditures. And for the cases that departed from constant returns, the bias was on the side of diminishing rather than increasing returns. To the extent that patent counts are better measures of innovative output than R&D input data, the indicated conclusion is that leading firms were not on average more vigorous innovators. But again, the overriding implication from both the R&D and patent analyses is that large and small units contributed roughly proportionately to the advance of technology.[124]

These results are for activity within lines of business, which for diversified corporations were often much smaller than the corporate whole. The question remains, Does size beyond the line of business level, that is, diversified size, affect the vigor with which innovative activity is pursued? Diversification might contribute affirmatively if firms with many lines were better able to exploit unexpected results emerging from their research laboratories, achieve richer cross-fertilization among specialized technical talent, secure lower-cost R&D financing, and/or utilize common channels of distribution to market new products. On the other hand, it could hold back progress if it meant greater bureaucratization of decision-making processes.

The most powerful tests of the diversification hypothesis have been conducted using FTC Line of Business data. In one such analysis, a Herfindahl numbers-equivalent index of diversification[125] was added to the analysis of 1974 R&D-sales and patent-sales regression equations described above. The diversification index was found to have minute and statistically insignificant effects on the intensity of R&D and patenting.[126] Cohen et al. focused on R&D/sales ratios by line of business averaged over the years 1975 to 1977.[127] Their diversification measure was the amount of companywide domestic sales *not* originating from the line of business being observed. They found diversification to have a slight but statistically

122. For a more complete discussion of the classifications when $a \neq 0$, see Scherer, "Technological Change," supra note 112, pp. 287–289, from which the results here are drawn.
123. For similar results from analyses of R&D/sales ratios as a function of line of business sales, see Wesley M. Cohen, Richard C. Levin, and David C. Mowery, "Firm Size and R&D Intensity: A Re-Examination," *Journal of Industrial Economics*, vol. 35 (June 1987), pp. 543–565; and Ravenscraft and Scherer, *Mergers*, supra note 91, pp. 120–121. In all these analyses, inter-industry differences are found to have a much more powerful effect on R&D expenditures than intra-industry size differences.
124. These results, it must be reemphasized, relate innovative activity to size at the line of business level, not the whole-company level. But they are generally similar to the results of earlier, less well-controlled studies using observations measured at the whole-company level. See, for example, Scherer, *Innovation and Growth*, supra note 9, Chapter 9 and pp. 213–215; Albert N. Link, Terry G. Seaks, and Sabrina R. Woodbery, "Firm Size and R&D Spending: Testing for Functional Form," *Southern Economic Journal*, vol. 54 (April 1988), pp. 1027–1038; J. D. Howe and D. G. McFetridge, "The Determinants of R&D Expenditures," *Canadian Journal of Economics*, vol. 9 (February 1976), pp. 57–71; W. J. Adams, "Firm Size and Research Activity: France and the United States," *Quarterly Journal of Economics*, vol. 84 (August 1970), pp. 386–409; Jörg Tabbert, *Unternehmensgrösse, Marktstruktur und technischer Fortschritt* (Göttingen: Vandenhoeck & Ruprecht, 1975), pp. 56–108; Louis Phlips, *Effects of Industrial Concentration: A Cross-Section Analysis for the Common Market* (Amsterdam: North-Holland, 1971), pp. 121–132; D. J. Smyth, J. M. Samuels, and J. Tzoanos, "Patents, Profitability, Liquidity and Firm Size," *Applied Economics*, vol. 4 (June 1972), pp. 77–86; Bengt Johanisson and Christian Lindström, "Firm Size and Inventive Activity," *Swedish Journal of Economics*, vol. 73 (December 1971), pp. 427–442; Noriyuki Doi, "Diversification and R&D Activity in Japanese Manufacturing Firms," *Managerial and Decision Economics*, vol. 6 (September 1985), pp. 147–152; and K. Gannicott, "The Determinants of Industrial R&D in Australia," *Economic Record*, vol. 60 (September 1984), pp. 231–235.
125. See p. 92 supra.
126. Scherer, "Technological Change," supra note 112, pp. 292–293. But see Scott, "Diversification versus Cooperation," supra note 81, who found that R&D spending increased with *purposive* diversification, identified as diversification across apparently complementary fields.
127. "Firm Size and R&D Intensity," supra note 123.

WCK1130
Wockhardt Bio AG v. Janssen Oncology, Inc.
IPR2016-01582

significant positive effect on R&D intensity in one version of their regression model, but it faded to insignificance when seven extreme-valued observations (out of 1,797) were removed or when variables controlling for technological opportunity and appropriability were included. These ambiguities led them to conclude that no significant relationship existed between company size and R&D intensity. In still another study using Line of Business data for 1977, Ravenscraft and Scherer found that R&D/sales ratios were slightly lower in lines with a history of diversification mergers, although the result fell short of statistical significance by conventional standards.[128]

It seems clear that firm size increases associated with greater diversification do not in general have a favorable effect on the vigor of research and development efforts. To this finding, one noteworthy qualification must be added. There is statistical evidence that the fraction of total industrial R&D outlays devoted to basic research rises with overall corporate size and greater diversification.[129] Vigorous support of basic research in turn appears to be positively correlated with higher innovative output across individual firms and higher productivity growth across broadly defined industry sectors, although the underlying chain of causation remains poorly understood.[130]

## *Conclusion*

Viewed in their entirety, the theory and evidence suggest a threshold concept of the most favorable climate for rapid technological change. A bit of monopoly power in the form of structural concentration is conducive to innovation, particularly when advances in the relevant knowledge base occur slowly. But very high concentration has a positive effect only in rare cases, and more often it is apt to retard progress by restricting the number of independent sources of initiative and by dampening firms' incentive to gain market position through accelerated R&D. Likewise, given the important role that technically audacious newcomers play in making radical innovations, it seems important that barriers to new entry be kept at modest levels. Schumpeter was right in asserting that perfect competition has no title to being established as the model of dynamic efficiency. But his less cautious followers were wrong when they implied that powerful monopolies and tightly knit cartels had any stronger claim to that title. What is needed for rapid technical progress is a subtle blend of competition and monopoly, with more emphasis in general on the former than the latter, and with the role of monopolistic elements diminishing when rich technological opportunities exist.

**128.** *Mergers*, supra note 91, pp. 120–121.

**129.** See Albert N. Link and James E. Long, "The Simple Economics of Basic Scientific Research: A Test of Nelson's Diversification Hypothesis," *Journal of Industrial Economics*, vol. 30 (September 1981), pp. 105–109; Link, "The Changing Composition of R&D," *Managerial and Decision Economics*, vol. 6 (June 1985), pp. 125–128; Mansfield, "Composition of R&D Expenditures," supra note 87, pp. 612–613; and Richard R. Nelson, "The Simple Economics of Basic Scientific Research," *Journal of Political Economy*, vol. 67 (June 1959), pp. 297–306.

**130.** Edwin Mansfield, "Basic Research and Productivity Increase in Manufacturing," *American Economic Review*, vol. 70 (December 1980), pp. 863–873; and Zvi Griliches, "Productivity, R&D, and Basic Research at the Firm Level in the 1970s," *American Economic Review*, vol. 76 (March 1986), pp. 141–154.

SCHERER/ROSS

# INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE

THIRD EDITION

# Industrial Market Structure and Economic Performance

## Third Edition

**F.M. Scherer**
Harvard University

**David Ross**
Williams College

**Houghton Mifflin Company Boston**
Dallas Geneva, Illinois Palo Alto Princeton, NJ