

REGULATION AND THE MARGINALIST REVOLUTION

*Herbert Hovenkamp**

Abstract

The marginalist revolution in economics became the foundation for the modern regulatory State with its “mixed” economy. For the classical political economists, value was a function of past averages. Marginalism substituted forward looking theories based on expectations about firm and market performance. Marginalism swept through university economics, and by 1920 or so virtually every academic economist was a marginalist.

This Article considers the historical influence of marginalism on regulatory policy in the United States. My view is at odds with those who argue that marginalism saved capitalism by rationalizing it as a more defensible buttress against incipient socialism. While marginalism did permit economists and policy makers to strike a middle ground between laissez faire and socialism, the “middle ground” tilted very strongly toward public control. Ironically, regulation plus private ownership was able to go much further in the United States than socialism ever could because it preserved the rhetoric of capital as privately owned, even as it deprived firms of many of the most important indicia of ownership.

Marginalism upended many of the classical conceptions about the market, including assumptions about their robustness, as well as the need for government intervention and the optimal type. For regulatory policy the most important issues were: (1) The fixed-cost controversy and the scope of natural monopoly; (2) cost classification, incentives, and ratemaking; (3) the changing domain of market failure, and regulation and Pigouvian taxes as correctives; (4) market diversity and the rise of sector regulation; (5) deregulation; (6) concerns about the distribution of wealth; (7) the development of cost-benefit analysis; and (8) the assessment of risk. The final section examines risk management under marginalism by looking at two diverse but important areas: negligence and products liability in tort law, and administrative review of patents by the Patent Trial and Appeal Board.

* James G. Dinan University Professor, University of Pennsylvania School of Law and The Wharton School. Thanks to audiences at the conference on New Perspectives on U.S. Regulatory History, Harvard Business School; and a faculty seminar, Northwestern University Law School.

INTRODUCTION: CAPITALISM, SOCIALISM, AND REGULATION	456
I. FIXED COSTS, EQUILIBRIUM, AND NATURAL MONOPOLY	464
II. RETURN-BASED RATEMAKING	471
III. THE CHANGING DOMAIN OF MARKET FAILURE: TRADITIONAL REGULATION, PUBLIC CHOICE, AND PIGOUVIAN TAXES	484
IV. MARKET DIVERSITY AND AGENCY SECTOR REGULATION.....	492
V. DEREGULATION	494
VI. WEALTH DISTRIBUTION	497
VII. COST-BENEFIT ANALYSIS.....	500
VIII. ACTUARIAL CONCEPTIONS OF VALUE AND MANAGEMENT OF RISK	503
A. <i>Holmes’s Marginalism: Torts as Risk Management</i>	506
B. <i>Innovation Risk and PTAB Patent Review</i>	508
CONCLUSION.....	513

INTRODUCTION: CAPITALISM, SOCIALISM, AND REGULATION

The neoclassical, or marginalist, revolution in economics had a profound and lasting impact on the theory and policy of regulation. It became the foundation for the modern regulatory State with its “mixed” economy. Marginalism, which best identifies the dividing line between classical political economy and neoclassical economics, originated during the 1860s and 1870s in the work of William Stanley Jevons in England, Carl Menger in Austria, Leon Walras in Switzerland, and a little later John Bates Clark in the United States.¹ The classical political

1. See MARK BLAUG, ECONOMIC THEORY IN RETROSPECT 294–325 (4th ed. 1985); HERBERT HOVENKAMP, THE OPENING OF AMERICAN LAW: NEOCLASSICAL THOUGHT, 1870–1970, at 27–32 (2015) (expanding on marginalism). See generally JOHN BATES CLARK, THE DISTRIBUTION OF WEALTH: A THEORY OF WAGES, INTERESTS AND PROFITS (1899) (emphasizing natural law); WILLIAM STANLEY JEVONS, THEORY OF POLITICAL ECONOMY (1871) (analyzing political economy as a science); CARL MENGER, PRINCIPLES OF ECONOMICS (James Dingwall & Bert F. Hoselitz trans., Ludwig von Mises Institute 2007) (1871) (advocating value based on marginal utility and risk rather than labor); LÉON WALRAS, ELEMENTS OF PURE ECONOMICS

economists, including Adam Smith, David Ricardo, and John Stuart Mill in his earlier years,² had believed that value was determined mainly by the amount of socially necessary labor that went into something in the past. A capital investment reflected value because the production of capital required labor.

In sharp contrast, marginalists developed a subjective and behavioral conception that value is a function of marginal willingness to pay—that is, how much would someone pay to obtain or avoid a certain good or other outcome? The marginal value of any good decreased as one had more of it and, at the margin, people maximize value by equating marginal utilities. This observation enabled marginalists to develop highly useful theories about the optimal mixture of any person's goods and services. Computing these values, however, called for increasing use of mathematics at a scale that was unfamiliar to most classical political economists.

While classical conceptions of value were objective and backward looking, marginalism was both forward looking and based on subjective valuation, although in many cases the relevant value was that of a presumed “rational” actor.³ The move toward valuation based on preference also brought interest group theory into economic analysis in a way that it had never existed before. Depending on their characteristics, different interest groups value outcomes in different ways. The search for a “social” value required that these differences be accounted for in some meaningful way.

When coupled with the concept of declining marginal utility, marginalism facilitated greatly increased use of mathematics to explain how markets work. Formulating marginalism into what became modern price theory and industrial organization became the task of a second generation of marginalists—beginning with Alfred Marshall of Cambridge, whose *Principles of Economics* was published in 1890 and went through numerous editions.⁴

Marginalism completely upended classical political economy's theory of value, which had largely been drawn from averages of past experience. For example, under classicism the optimal rate of wages was dictated by the wage-fund doctrine, which held that the fund available for the

(William Jaffe trans., Augustus M. Kelley Publishers 1977) (1874) (establishing Walras's theory of general equilibrium).

2. On Mill's late career rejection of the wage-fund doctrine, see generally Arthur C. Pigou, *Mill and the Wages Fund*, 59 *ECON. J.* 171 (1949).

3. Herbert Hovenkamp, *The Marginalist Revolution in Legal Thought*, 46 *VAND. L. REV.* 305, 316 (1993).

4. See generally ALFRED MARSHALL, *PRINCIPLES OF ECONOMICS* (1890) (introducing marginalist price theory).

payment of wages was determined by the amount of capital retained from previous production. This then had to be divided among the number of workers.⁵ If a firm's wages exceeded this amount, it would end up borrowing against the future, which was a road to insolvency. An analytically similar classical doctrine was the theory of corporate finance: that the value of a corporation equaled the amount of capital that had been paid into it.⁶ The marginalist critique of these doctrines observed that they had little to do with actual market behavior. The wage that an employer is willing to pay is not based on the amount in some historical fund, but on the employer's expectations about how much value the employee will contribute to the firm.⁷ Equally, the value of a corporation is not a function of the amount of capital that has been paid in, but rather the firm's prospects for the future. As one early twentieth century treatise on corporate finance observed concerning the change, "there is very seldom even a close correspondence between the original investment of capital and the value of a corporation's assets."⁸

In the United States the dispute over marginalism led to an upheaval in the American Economic Association, which had been founded in 1885.⁹ The neoclassicists won out and marginalist approaches swept the field. Today every mainstream economist is some kind of marginalist.

Marginalism's most disquieting feature was its threat to the traditional dividing line between private property and public control. Private ownership had generally assumed private control, limited mainly by the common law of nuisance and regulations that fell within a triumvirate of exceptions to the Gilded Age Supreme Court's anti-regulatory bias—namely, health, safety, and morals.¹⁰ Beginning in the late nineteenth century the Supreme Court began permitting some rate regulation of railroads, public utilities, or other bottleneck industries.¹¹

By contrast, marginalism's critique of the laissez faire state cut very broadly, undermining much of classical political economy. My own

5. HERBERT HOVENKAMP, *ENTERPRISE AND AMERICAN LAW 1836–1937*, at 193–98 (1991).

6. *See generally* HOVENKAMP, *supra* note 1, at 159–71 (discussing the revolution in corporate finance).

7. *E.g.*, CLARK, *supra* note 1, at 397 (developing marginal productivity theory of wages).

8. WILLIAM H. LOUGH & FREDERICK W. FIELD, *CORPORATION FINANCE: AN EXPOSITION OF THE PRINCIPLES AND METHODS GOVERNING THE PROMOTION, ORGANIZATION AND MANAGEMENT OF MODERN CORPORATIONS* 125 (8th ed. 1916) (1909); *see also id.* at 127–28, 130–31 (discussing the correspondence between original investment and value of assets).

9. HOVENKAMP, *supra* note 1, at 75–76.

10. *See* HOVENKAMP, *supra* note 1, at 243–62 for a discussion of nineteenth century regulation of health, safety, and morals. *See* WILLIAM J. NOVAK, *THE PEOPLE'S WELFARE: LAW AND REGULATION IN NINETEENTH-CENTURY AMERICA* 191–233 (1996) (discussing pre-Gilded Age regulation).

11. *See* discussion *infra* notes 104–12.

views of marginalism in the United States are somewhat at odds with those expressed by historians such as Dorothy Ross. She presented marginalism as capitalism's savior to the extent that it created a new, more defensible paradigm that "offered Americans assurance that the new industrial world could still operate like the old."¹² To be sure, marginalism permitted economists to strike a middle ground between laissez faire and socialism, but the "middle ground" was in fact very far removed from the antistatist vision of the classical political economists.

Regulation under marginalist principles did not lead to a large increase in public business ownership, as socialism threatened. To a significant extent, however, it deprived private firms of the power to make their own economic decisions about market entry, price, and other central features of private property. Regulation became public ownership in disguise. Every element of private firm behavior within an agency's jurisdiction became subject to government review under a public interest standard.

Ironically, regulation plus private ownership was able to go much further in the United States than socialism ever could because it preserved the rhetoric of capital as privately owned—but all the while depriving private firms of the traditional indicia of ownership. Ronald Coase gave us the idea that anything that can be specified by a firm can also be specified as a set of contracts.¹³ Similarly, regulation under marginalist tools developed the idea that public regulators could specify for the firm nearly anything that the firm could specify for itself.

Socialism and public ownership never came close to being a dominant model for management of the economy in the United States, even for the Progressives.¹⁴ Socialist political parties never attained anything like the representation they were able to achieve in many European countries.¹⁵ But regulation could preserve the image of private ownership while completely undermining its most central indicia. The real key to

12. DOROTHY ROSS, *THE ORIGINS OF AMERICAN SOCIAL SCIENCE* 179 (1991).

13. Ronald H. Coase, *The Nature of the Firm*, 4 *ECONOMICA* (n.s.) 386, 391 (1937); see also OLIVER E. WILLIAMSON, *MARKETS AND HIERARCHIES: ANALYSIS AND ANTITRUST IMPLICATIONS* 3–4 (1975) (discussing Coase's article); Michael C. Jensen & William H. Meckling, *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, 3 *J. FIN. ECON.* 305, 310 (1976) (discussing Coase in relation to the firm).

14. See generally Eric Foner, *Why Is There No Socialism in the United States?*, 17 *HIST. WORKSHOP J.* 57 (1984) (discussing the failures of socialism). The title was borrowed from German historicist WERNER SOMBART, *WHY IS THERE NO SOCIALISM IN THE UNITED STATES?* (1908).

15. See, e.g., SEYMOUR MARTIN LIPSET & GARY MARKS, *IT DIDN'T HAPPEN HERE: WHY SOCIALISM FAILED IN THE UNITED STATES* 85–88 (2000) (examining the failure of socialism in America). See generally ROBERT KAGAN, *OF PARADISE AND POWER: AMERICA VS. EUROPE IN THE NEW WORLD ORDER* (2003) (contrasting the different perspectives of American and European economics).

economic behavior is not “who owns,” but rather “who decides.” Depending on the nature and extent of regulation, that question was often answered in favor of the government.¹⁶ Coase himself acknowledged the point in his Nobel Prize lecture, emphasizing that in a world of zero transaction costs private bargaining would even establish the law itself.¹⁷ When transaction costs are positive, however, “the rights which individuals possess . . . will be to a large extent what the law determines.”¹⁸ As a result, “the legal system will have a profound effect on the working of the economic system and may in certain respect be said to control it.”¹⁹

Regulation also had one big advantage over socialism, which was its ability to vary the amount and nature of public control to different circumstances. Public ownership tends to be binary, in the sense that a firm is either publicly owned or it is not. By contrast, regulation permits the government to begin with a welfare standard and then select particular elements of firm activity for regulation or competition. This ability accounts for the economic success of most parts of the deregulation movement.²⁰ Most deregulation is “partial,” in the sense that it restores to the market those functions thought to be best left to competition while leaving others within the control of regulators.

At the margin, a regulated firm could be made to behave in precisely the same way as a publicly-owned enterprise. Of course, regulation rarely approached that margin and, with the rise of deregulation,²¹ began to step back from it considerably. But the general point should be clear: regulation permitted policy makers to experiment with varying degrees of public control by not crossing that important rhetorical line into public ownership.

Marginalism upended many of the classical conceptions about the market and, in the process, about the need for regulation and the optimal type. The initial impact of marginalism was rapid expansion of the justifications for regulation as well as ideas about its scope. Later on came a certain amount of pushback.²² The most important issues were: (1) The fixed-cost controversy and the scope of natural monopoly; (2) cost

16. For good perspective, see Mark J. Roe, *Political Preconditions to Separating Ownership from Corporate Control*, 53 STAN. L. REV. 539, 551 (2000).

17. See Ronald H. Coase, *Lecture to the Memory of Alfred Nobel*, NOBEL PRIZE (Dec. 9, 1991), <https://www.nobelprize.org/prizes/economic-sciences/1991/coase/lecture/> [<https://perma.cc/MU2Y-W58C>].

18. Ronald H. Coase, *The Institutional Structure of Production*, in OCCASIONAL PAPERS 1, 9–10 (Univ. of Chi. Law, Occasional Paper No. 28, 1992).

19. *Id.* at 10.

20. See discussion *infra* Part V.

21. See discussion *infra* Part V.

22. See Hovenkamp, *supra* note 3, at 320.

classification, incentives, and ratemaking; (3) the changing domain of market failure and regulation or Pigouvian taxation as correctives; (4) market diversity and the rise of sector regulation; (5) deregulation; (6) concerns about the distribution of wealth; (7) the development of cost-benefit analysis; and (8) the assessment of risk.

The marginalist revolution made modern, sector-specific agency regulation inevitable, at least within the framework that marginalist economics chose.²³ To be sure, the development of regulatory agencies is historically contingent. One can also imagine a world in which we have fewer of them or in which their jurisdictional boundaries over subject matter or geography differ from the arrangements that we have. Clearly, however, the complexity of issues created by marginalism, including accounting for risk and distributional concerns, appeared to require greater expertise in determining the need for and the type of government intervention. It is one thing to roll back the clock on regulation; it is quite another to roll it back on marginalism itself.

Marginalism imposed a significant measure of subjective choice into theories of market behavior, and thus into policy making. The conception of markets in nineteenth century political economy was that they were an objective part of the laws of nature.²⁴ Classical political economists spoke of the laws of the economy as a part of natural law.²⁵ A popular nineteenth century American text initially written by Brown University's Francis Wayland and later edited by Aaron Chapin, President of Beloit College, declared that:

Political Economy is that branch of Social Science which treats the production and application of wealth to the well-being of men in society. It is a branch of true science.

By Science, as the word is here used, we mean a Systematic arrangement of the laws which God has established, so far as they have been discovered, of any department of human knowledge.²⁶

Marginalism and the regulatory policy that resulted were stunningly different from this conception of economic science. Nearly every decision to regulate leads to a menu of policy choices, and it is not always clear that one works better than another. Further, different approaches may work better in different environments. Regulation often invites experimentation and, necessarily, criticism about the amount or the

23. See discussion *infra* Part IV.

24. See Hovenkamp, *supra* note 3, at 307–09.

25. E.g., HENRY WOOD, *THE POLITICAL ECONOMY OF NATURAL LAW* 18–19 (1894).

26. FRANCIS WAYLAND, *THE ELEMENTS OF POLITICAL ECONOMY* 4 (Aaron L. Chapin ed., 1886) (1837) (emphasis omitted).

particular direction regulation has taken. Just to give one example, the perceived need for regulation of the business corporation and its shareholders or creditors began to loom large during the Gilded Age, particularly in the “watered stock” scandals. The regulatory experiments that resulted included state “blue sky” laws,²⁷ and in the federal government the U.S. Industrial Commission (1898–1902), the Federal Bureau of Corporations and Federal Trade Commission (1903–1915), and eventually the Securities Act and the founding of the Securities Exchange Commission (1933–1934). Each of these experiments took more control of the corporation from its shareholders and placed it under an agency.

Also inherent in marginalism’s emphasis on expectations was greatly increased use of mathematics. As a result, the initial division between traditionalists and marginalists was generational. Older economists lacked the training in mathematics that the new economics required, and they naturally resisted it.²⁸ The models that marginalist economists developed took expected value and risk into account. Because pricing decisions are made at the margin, neoclassical economics provided a technical basis for dividing the behavior of markets or firms into shorter or longer “runs,” or time periods. It produced technical cost classifications into long-run and short-run costs, including the development of the marginal cost curve, perhaps the most distinctive mathematical feature of marginalism.²⁹ The marginal revenue curve came a generation later.³⁰

Closely related was increased attention to the theory of competition, and increasing awareness that “perfect” competition is the exception rather than the rule. Rather, competition exists on a continuum, depending mainly on the number of firms, the types of costs that each firm faces, and the nature of its products.³¹ The full implications had to

27. See, e.g., William W. Cook, “Watered Stock”—Commissions—“Blue Sky Laws”—*Stock Without Par Value*, 19 MICH. L. REV. 583, 584, 590 (1921) (exemplifying prominent Gilded Age corporate law treatise author advocating state blue sky statutes); see also LOUIS LOSS & EDWARD M. COWETT, BLUE SKY LAWS 3–10 (1958) (providing an in-depth discussion of the blue sky laws); Paul G. Mahoney, *The Origins of the Blue-Sky Laws: A Test of Competing Hypotheses*, 46 J.L. & ECON. 229, 230–32 (2003) (further analyzing blue sky laws).

28. HOVENKAMP, *supra* note 1, at 75–76.

29. See Hovenkamp, *supra* note 3, at 313.

30. The construction of the marginal revenue curve is generally attributed to Joan Robinson, occasionally with Edward H. Chamberlin as an independent discoverer. See JOAN ROBINSON, THE ECONOMICS OF IMPERFECT COMPETITION 182–96 (1933) (illustrating the marginal revenue curve); see also EDWARD CHAMBERLIN, THE THEORY OF MONOPOLISTIC COMPETITION 178–93 (1933); Alfred S. Eichner, *Joan Robinson’s Legacy*, 27 CHALLENGE 42 (1984) (stressing the importance of Robinson’s contributions).

31. George J. Stigler, *Perfect Competition, Historically Contemplated*, 65 J. POL. ECON. 1, 1–3 (1957).

await the early 1930s and the work of Joan Robinson and Edward Chamberlin on imperfect and monopolistic competition.³² Joseph Schumpeter had the prescience to see this in the mid-thirties, in a review of Joan Robinson's *Economics of Imperfect Competition*.³³ Prior to that time economists had treated competition and monopoly as the only market structures worth systematic examination, and everything in between as "an academic curiosity."³⁴ However, if this intermediate range should emerge as dominant, then the traditional laissez faire assumption that markets should be left alone could cease to hold true. Instead, the circumstances under which governmental action could increase welfare "becomes so extended as to make these cases the rule rather than more or less curious exceptions."³⁵

Closely related was the concept of uncertainty—and relatedly, risk—which developed naturally out of theories that identified value with future expectations rather than past averages.³⁶ The revolution in corporate finance provides a good example.³⁷ Under the classical theory any judge who knew a little arithmetic and basic concepts of property valuation could decide whether stock was "watered," which meant that its stated par value exceeded the actual value historically paid in.³⁸ Answering that question required a judge to determine the amount of paid in cash and noncash capital and divide by the number of shares. The biggest problem was overstatement of the value of noncash property.³⁹ But the marginalist theory that a firm's value is a function of anticipated profitability was far more complex, requiring considerable information about the firm itself and its products and management, the value and duration of its intellectual property portfolio and other productive assets, the market in which the firm operated, and the valuation of uncertain future events.

The initial effect of marginalism was to weaken the classical view that nearly all markets work well, and opinions about the robustness of

32. See generally ROBINSON, *supra* note 30 (discussing distribution, allocation, and exploitation); CHAMBERLIN, *supra* note 30 (discussing competition between different types of firms).

33. Joseph A. Schumpeter & A.J. Nichol, *Robinson's Economics of Imperfect Competition*, 42 J. POL. ECON. 249, 250–51 (1934).

34. HOVENKAMP, *supra* note 1, at 207–08

35. Schumpeter & Nichol, *supra* note 33, at 251.

36. Among the most prophetic was FRANK H. KNIGHT, *RISK, UNCERTAINTY, AND PROFIT* (1921).

37. See *supra* text accompanying notes 5–7.

38. *E.g.*, *Lehigh Ave. Ry.*, 18 A. 498, 500 (Pa. 1889) (stating stock certificate "stands in the hands of the subscriber for so much as, and no more than, the amount actually paid upon it").

39. *E.g.*, *Gillett v. Chi. Title & Trust Co.*, 82 N.E. 891, 904–05 (Ill. 1907) (holding a promoter's contribution of an unwritten play and an unpatented invention not worth the \$2 million evaluation placed on them); *cf.* *Carr v. Le Fevre*, 27 Pa. 413, 416–17 (1856) (holding the contribution of coal mining property at appraised value was permitted, given that there was no evidence suggesting that the appraisal was fraudulent).

markets differed widely, depending on assumptions about such things as the distribution of fixed and variable costs, the flow of information, or the degree and nature of risk or uncertainty that the firm faced. This in turn led increasingly to the view that markets differ from one another, and that many would work better if the state applied a corrective. In other words, marginalism provided a rationale for both a vastly increased amount of government intervention in the economy and the idea that the type of intervention should vary from one market to another. As with many revisionist ideas, its early history reflected positions that were later thought to be extreme, once marginalism became more normalized in economic and regulatory theory.

I. FIXED COSTS, EQUILIBRIUM, AND NATURAL MONOPOLY

One of the most important issues confronting neoclassical regulatory theory was determining the range of industries and firms to which price regulation should be applied. Policy makers needed to know whether an industry equilibrium with satisfactory results could be achieved without state intervention.

Among Alfred Marshall's most significant contributions to neoclassical price theory was his development of partial equilibrium analysis, which examined a portion of the economy limited to a single "commodity" or finite time period. Marshall realized that in an economy everything affects everything else. Nevertheless, he argued for the importance of studying a specific product over a limited time period, assuming that changes in demand and supply for that product had no effect on other goods. In defending this approach he wrote:

The forces to be dealt with are however so numerous, that it is best to take a few at a time; and to work out a number of partial solutions as auxiliaries to our main study. Thus we begin by isolating the primary relations of supply, demand and price in regard to a particular commodity. We reduce to inaction all other forces by the phrase "other things being equal": we do not suppose that they are inert, but for the time we ignore their activity. This scientific device is a great deal older than science: it is the method by which, consciously or unconsciously, sensible men have dealt from time immemorial with every difficult problem of ordinary life.⁴⁰

Marshall was unable to get an equilibrium in a competitive market, however, if the firms in it experienced significant fixed costs.⁴¹

40. ALFRED MARSHALL, *PRINCIPLES OF ECONOMICS* xiv (8th ed. 1949).

41. See, e.g., Renee Prendergast, *Increasing Returns and Competitive Equilibrium—the Content and Development of Marshall's Theory*, 16 *CAMBRIDGE J. ECON.* 447, 447 (1992).

Competition would drive prices to marginal cost, which covers only variable costs.⁴² The firms in such a market would be forced to charge prices so low that they could not stay in business.

One result of Marshall's dilemma was that economists became embroiled in a significant "fixed cost" controversy, in which economists and lawyers debated both the meaning and the policy implications of high fixed costs.⁴³ In antitrust for example, firms argued that in industries with high fixed costs competition would be "ruinous," forcing firms out of business until only a monopolist survived.⁴⁴ Otherwise they would be forced to collude or merge.⁴⁵ The economics literature produced many studies involving long- vs. short-period sales, the availability of price discrimination, or other theories attempting to explain how fixed costs could be consistent with stable competition.⁴⁶ By and large, these studies assumed that the products in question were commodities, and thus the only variable for consumer choice was price.

The policy responses were less extreme than the literature. First, the Supreme Court consistently rejected "ruinous competition" as a defense in antitrust cases, all of which involved fungible products or services.⁴⁷ The principal industry that came under price regulation as a result of the controversy was the railroads, where fixed costs were so high that they were thought to create natural monopolies across a wide range of situations.⁴⁸ The issue was also raised in the cast-iron pipe price fixing case, but both the Sixth Circuit in Judge Taft's well known opinion and the Supreme Court rejected it there as well.⁴⁹ There is reason to believe that the Supreme Court's hostility toward cartels in high fixed cost industries either led to or exacerbated the first great American merger

42. *Id.*

43. For an account of the debate in the United States and its effect on developing antitrust policy, see HOVENKAMP, *supra* note 5, at 308–22.

44. *Id.* at 313–22.

45. See, e.g., Eliot Jones, *Is Competition in Industry Ruinous?*, 34 Q.J. ECON. 473, 473–74 (1920).

46. One of the most prescient was the largely ignored JOHN M. CLARK, *STUDIES IN THE ECONOMICS OF OVERHEAD COSTS* 416–17 (1923). See also Spurgeon Bell, *Fixed Costs and Market Price*, 32 Q.J. ECON. 507, 509–22 (1918) (explaining the effects of fixed costs on market price in certain industries); Frank H. Knight, *Cost of Production and Price over Long and Short Periods*, 29 J. POL. ECON. 304, 304–10 (1921) (analyzing short-time and long-time economic problems).

47. See *United States v. Trans-Mo. Freight Ass'n*, 166 U.S. 290, 329–30 (1897) (rejecting the defense); accord *United States v. Joint Traffic Ass'n*, 171 U.S. 505, 519–23, 547–48, 569 (1898).

48. See HOVENKAMP, *supra* note 5 at 315–16.

49. *United States v. Addyston Pipe & Steel Co.*, 85 F. 271, 291 (6th Cir. 1898) (rejecting "ruinous competition" defense), *modified and aff'd*, 175 U.S. 211, 235–36 (1899).

movement, in which many firms with high fixed costs were forced to merge because antitrust law left them unable to collude.⁵⁰

Alfred Marshall's own solution to the fixed cost problem was widely regarded as unacceptable. Borrowing from biology, he reasoned that firms were like trees in a forest. The "representative firm" went through a finite life cycle, growing in the earlier part, maturing, and eventually withering away.⁵¹ As a result there would not be durable monopoly but rather ongoing cycling of firms.⁵² Marshall wrote in his Eighth edition:

[T]he very conditions of an industry which enable a new firm to attain quickly command over new economies of production, render that firm liable to be supplanted quickly by still younger firms with yet newer methods. Especially where the powerful economies of production on a large scale are associated with the use of new appliances and new methods, a firm which has lost the exceptional energy which enabled it to rise, is likely ere long quickly to decay; and the full life of a large firm seldom lasts very long.⁵³

This passage indicates that Marshall was aware of the role of innovation ("newer methods") in upsetting equilibrium and perhaps even making the search for equilibrium less important. But Marshall himself did not go very far down this road.⁵⁴

One limitation on his biological theory of the life of a representative firm, Marshall acknowledged, might be joint stock companies, or corporations, which did not necessarily go through this ageing cycle. However, Marshall persisted in the view that such companies were inherently inferior mechanisms of production, largely because of problems relating to the separation of ownership and control.⁵⁵

50. *E.g.*, NAOMI R. LAMOREAUX, *THE GREAT MERGER MOVEMENT IN AMERICAN BUSINESS, 1895-1904*, at 27-33 (1985); *see also* LESTER G. TELSER, *ECONOMIC THEORY AND THE CORE* 387 (1978) (providing a mathematic perspective). *See generally* George Bittlingmayer, *Decreasing Average Cost and Competition: A New Look at the Addyston Pipe Case*, 25 J.L. & ECON. 201 (1982) (introducing the question of what role *Addyston Pipe* played in the merger wave); George Bittlingmayer, *Did Antitrust Policy Cause the Great Merger Wave?*, 28 J.L. & ECON. 77 (1985) (positing that mergers were legal while cartels were not).

51. *See* G. F. Shove, *The Representative Firm and Increasing Returns*, 40 ECON. J. 94, 114 (1930).

52. For a good discussion, see Neil Hart, *Marshall's Theory of Value: The Role of External Economies*, 20 CAMBRIDGE J. ECON. 353, 360-62 (1996).

53. MARSHALL, *supra* note 40, at 287.

54. That largely awaited Schumpeter. *See* JOSEPH A. SCHUMPETER, *CAPITALISM, SOCIALISM, AND DEMOCRACY* 81-86 (1942).

55. MARSHALL, *supra* note 40, at 303 (stating shareholders are generally unable "to exercise an effective and wise control over the general management of the business"); *id.* at 303-04

Most of Marshall's intellectual descendants did not follow him into this biological forest. Rather, they looked for solutions to the fixed cost problem in two quite different places. Both were driven by the mathematics of marginalism rather than biological analogies. First was the idea that the long run average cost curve might be U-shaped rather than continuously downward sloping. The second was the availability of product differentiation and, eventually, monopolistic competition.

The proposition of U-shaped long run average cost is simply that scale economies do not produce increasing returns indefinitely. Rather, at some point the curve bottoms out and either remains flat or begins to rise. In that case the equilibrium number of firms in a market with fixed costs can be more than one, depending on the size of the market. For example, if average fixed costs decline only to an output level of 1000 and market demand at that level is 10,000 units, then this market has room for up to 10 technically efficient firms. This idea developed gradually in writings by Marshall's successor Arthur C. Pigou, and Francis Edgeworth, Piero Sraffa, and Jacob Viner.⁵⁶ Pigou in particular showed that when the supply price of the industry is higher than the marginal cost of a firm, that firm would expand until these costs were equalized. When a firm was in equilibrium, its marginal costs would equal the industry's supply curve. Firms could be expected to jockey for growth by competing to reduce costs.⁵⁷

Of course, the fact that a U-shaped cost curve is possible does not mean that it explains every situation. If the market is sufficiently small in relation to the availability of scale economies, there still might be room for only one firm. The other major development that addressed the equilibrium problem was the idea that firms differ from one another in geographic⁵⁸ and product space. The development of the idea of product

(discussing Marshall's highly suspicious account of corporations, and particularly the rise of the "trust" in the United States).

56. For the development of this concept, see generally ARTHUR C. PIGOU, *THE ECONOMICS OF WELFARE* (3d ed. 1929); Piero Sraffa, *On the Relations Between Cost and Quantity Produced* (translated from French "*Sulle Relazioni fra Costo e Quantità Prodotta*"), 11 *ANNALI DI ECONOMIA* 277 (1925); Piero Sraffa, *The Laws of Returns Under Competitive Conditions*, 36 *ECON. J.* 535 (1926); Jacob Viner, *Cost Curves and Supply Curves*, 2 *J. ECON.* 23 (1932). For a discussion of this development, see generally Francis Y. Edgeworth, *Contribution to the Theory of Railway Rates – IV*, 23 *ECON. J.* 206 (1913) (depicting the U-shape); and Jan Horst Keppler & Jerome Lallemand, *The Origins of the U-Shaped Average Cost Curve: Understanding the Complexities of the Modern Theory of the Firm*, 38 *HIST. POL. ECON.* 733 (2006) (identifying the earliest graphic depiction in Francis Y. Edgeworth).

57. Arthur C. Pigou, *An Analysis of Supply*, 38 *ECON. J.* 238, 252 (1928); see Herbert Hovenkamp, *Coase, Institutionalism, and the Origins of Law and Economics*, 86 *IND. L.J.* 499, 506 (2011).

58. See generally Harold Hotelling, *Stability in Competition*, 39 *ECON. J.* 41 (1929) (discussing differentiation in geographic space).

differentiation, and most particularly of monopolistic competition theory in the 1930s, very largely ended the fixed cost controversy.⁵⁹ Under monopolistic competition firms compete not only in pricing but also by differentiating their product in order to appeal to specific consumer tastes. Edward Chamberlin was able to show that even with easy entry and high fixed costs it was possible to have an equilibrium if products were differentiated.⁶⁰

The theory of monopolistic competition accomplished two things for regulatory policy. First was its assurance that at least in the world of manufactured products rather than commodities, ruinous competition and collusion or merger to monopoly were not inevitable. Rather, this became a problem of degree. Second, however, was the realization that monopolistic competition was not perfect either. Prices were higher than marginal cost. Further, firms under monopolistic competition dedicated considerable resources toward product differentiation, leading to the view that product differentiation was “excessive.” This led to discussions of such topics as whether annual style changes in the automobile industry should be regarded as an unfair method of competition.⁶¹ Relatedly, the intellectual property laws were thought to be harmful to the extent that they facilitated such differentiation. The decades following the New Deal saw rapid expansion of antitrust policy accompanied by considerable hostility toward intellectual property rights, most notably patents and trademarks.⁶² Chamberlin himself advocated widespread trademark infringement as a device for permitting firms to compete more effectively by making it more difficult for producers to differentiate their brands.⁶³

59. See BLAUG, *supra* note 1, 391–96.

60. E.g., John R. Hicks, *Survey of Economic Theory: The Theory of Monopoly*, 3 *ECONOMETRICA* 1, 11 (1935); Arthur Smithies, *Equilibrium in Monopolistic Competition*, 55 *Q.J. ECON.* 95, 95 (1940).

61. E.g., FED. TRADE COMM’N, REPORT ON MOTOR VEHICLE INDUSTRY, H.R. DOC. NO. 76-468, at 365–418 (1939). The issues are summarized in Note, *Annual Style Change in the Automobile Industry as an Unfair Method of Competition*, 80 *YALE L.J.* 567 (1971). In its main brief to the Supreme Court in the *Brown Shoe Merger Case*, the government argued that rapid style changes in the shoe industry made it very difficult for smaller shoe manufacturers to compete. Brief for the United States at 15, *Brown Shoe Co. v. United States*, 370 U.S. 294 (1962) (No. 4).

62. See HOVENKAMP, *supra* note 1, at 198–200, 227–28, 288–89.

63. EDWARD HASTINGS CHAMBERLIN, *THE THEORY OF MONOPOLISTIC COMPETITION: A RE-ORIENTATION OF THE THEORY OF VALUE* app. E (5th ed. 1946); see also Edward H. Chamberlin, *Product Heterogeneity and Public Policy*, 40 *AM. ECON. REV.* 85, 90–92 (1950) (concluding that the consequences of product heterogeneity have been ignored or misunderstood). For a modern perspective, see generally Christopher S. Yoo, *Intellectual Property and the Economics of Product Differentiation*, in 1 *RESEARCH HANDBOOK ON THE ECONOMICS OF INTELLECTUAL PROPERTY LAW* (Ben Depoorter & Peter Menell eds., forthcoming 2019), https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=3017&context=faculty_scholarship.

In any event, the ruinous competition problem itself became largely relegated to common carriers such as the railroads, public utilities, and other transport firms. It has reappeared from time to time in antitrust price fixing case law, only to be rejected. The most recent is the *United States v. Apple, Inc.*⁶⁴ eBooks decision, which rejected the publishers' argument that ruinous competition in electronic books justified their collective agreement to raise prices and impose these on Amazon.⁶⁵ That market, just as the railroads, involved a technology with very high fixed costs and variable (distribution) costs near zero, except for royalties.⁶⁶ One important dissimilarity, however, was product fungibility. Book titles are differentiated to the extent that most do not compete with each other on the consumer end. As a result, as long as they are under copyright, successful titles will sell at profitable prices even under high fixed costs. Once copyright has expired, the equilibrium price is close to zero.⁶⁷

One important but overhyped idea that originated in the 1960s and 1970s was that even firms with very high fixed costs, including utilities, could operate competitively if competition were seen as being "for" rather than "in" the market. Even if a market has room for only one seller at a time, competition to be that seller might keep prices at the competitive level without regulation. The basic theory had already appeared in "potential competition" antitrust cases.⁶⁸ It was also implicit in the common law of requirements contracts. For example, a store owner might prefer to have one person plow snow on an as needed basis over the course of a season, rather than taking bids for each snowstorm. Courts have approved such contracts since the Gilded Age.⁶⁹ Even though there is only one snow plower over the year, the price is competitive because prospective sellers must bid and rebid for that contract, or franchise.

64. 791 F.3d 290 (2d. Cir. 2015).

65. *Id.* at 333, 334; see Herbert Hovenkamp, *Antitrust and Information Technologies*, 68 FLA. L. REV. 419, 437–38 (2016) [hereinafter Hovenkamp, *Information*]; Herbert Hovenkamp, *Antitrust and the Design of Production*, 103 CORNELL L. REV. 1155, 1202, 1204 (2018) [hereinafter Hovenkamp, *Design*].

66. Hovenkamp, *Design*, *supra* note 65, at 1203.

67. See Hovenkamp, *Information*, *supra* note 65, at 426–27 (noting the example of scholarly journals, which command high prices even though they are distributed electronically and the authors obtain no royalties).

68. See, e.g., *United States v. El Paso Nat. Gas Co.*, 376 U.S. 651, 657–58 (1964) (concluding that two firms bidding on wholesale natural gas to Los Angeles should be regarded as competitors for merger purposes, even though only one of them had made any sales there).

69. E.g., *Fontaine v. Baxley*, 17 S.E. 1015, 1018 (Ga. 1892) (upholding requirements contract requiring plaintiffs to sell its full output of railroad ties to the defendant); *J. Winslow Jones & Co. v. Binford*, 74 Me. 439, 442, 443 (1883) (upholding requirements contract requiring defendant to sell its full output of sweet corn to the plaintiff); see also HOVENKAMP, *supra* note 1, at 126–29.

The late Harold Demsetz famously queried, why could not the same principle apply to public utilities?⁷⁰ Of course, the snow plower's truck can readily be moved from one site to another, permitting the market for plowing snow on a particular parking lot to be very competitive, provided that the prospective snow plowers do not collude. As Oliver Williamson countered, however, as soon we consider utilities that have durable and nonmoveable infrastructures, such as power, cable, or gas lines, the problem becomes much more complex.⁷¹ The winner of the first round would have a significant bidding advantage over any prospective entrant who would have to purchase or re-install the infrastructure. Alternatively, some mechanism might be established for transferring the infrastructure from the incumbent to the new winner, or perhaps a government such as a state or municipality could own the system, with successive firms simply bidding to be its operators for successive periods.⁷²

William Baumol, one of the most enthusiastic of "contestable market" proponents, proclaimed this idea of competition for the market to be "An Uprising in the Theory of Industry Structure."⁷³ To be sure, the theory of contestable markets was an important contribution to regulatory theory. However, as a policy matter it never lived up to expectations. Much of the technical theory turned into a discussion of cost classification—mainly, which costs were variable, which were fixed, and which were "sunk."⁷⁴ The later type was crucial. Even a high fixed cost need not interfere with contestability if the asset in question, such as the snow plowing truck, is readily and costlessly transferrable from one market to another. For obvious reasons, much of the theory focused on the airline industry. Although aircrafts are costly and durable, if they can easily be transferred from one market to another in response to diverging prices the result should be competition. But this theory of "hit and run" entry largely overlooked details such as the very substantial cost of airports, gate space leases, and other items that were not so readily transferable.⁷⁵ Institutions,

70. See Harold Demsetz, *Why Regulate Utilities?*, 11 J.L. & ECON. 55, 55 (1968).

71. See Oliver E. Williamson, *Franchise Bidding for Natural Monopolies—in General and with Respect to CATV*, 7 BELL J. ECON. 73, 85 (1976).

72. *Id.* at 90, 102 (arguing that "franchise bidding for incomplete long-term contracts is a much more dubious undertaking than Demsetz' discussion suggests," but suggesting that markets such as "[l]ocal service airlines" and "postal delivery" might work).

73. William J. Baumol, *Contestable Markets: An Uprising in the Theory of Industry Structure*, 72 AM. ECON. REV. 1, 1 (1982). The theory was expanded and later published in WILLIAM J. BAUMOL ET AL., *CONTESTABLE MARKETS AND THE THEORY OF INDUSTRY STRUCTURE* (1982).

74. See, e.g., Don Coursey et al., *Market Contestability in the Presence of Sunk (Entry) Costs*, 15 RAND J. ECON. 69, 70–71 (1984).

75. See U.S. GEN. ACCOUNTING OFFICE, *AIRLINE COMPETITION: EFFECTS OF AIRLINE MARKET CONCENTRATION AND BARRIERS TO ENTRY ON AIRFARES* 13 (1991),

it seems, always get in the way. The effect of deregulation of airlines has not been contestability, but rather imperfect competition or oligopoly among competitors, with frequent claims of price fixing. The literature generally shows an inverse correlation between price and the number of carriers on a route—which is inconsistent with contestability but quite consistent with oligopoly or collusion.⁷⁶ From a consumer welfare perspective, price competition among competing carriers is very likely much superior to the previous regime of price regulation by the now defunct Civil Aeronautics Board, but it is a far cry from contestability.

II. RETURN-BASED RATEMAKING

The historical rationale for rate regulation had been that certain firms were “affected with a publick interest,” in the words of Lord Justice Matthew Hale, because of their location as gateways to commerce.⁷⁷ He made that argument in reference to strategically located English seaports which held effective monopolies in their service area.⁷⁸ The United States Supreme Court quoted Hale in upholding state price regulation of grain elevators strategically located along railroad tracks.⁷⁹ Blackstone’s less

<https://www.gao.gov/assets/220/214117.pdf> [<https://perma.cc/WVB3-2SQC>] (explaining that under deregulation, airline markets still exhibit higher prices as they become more concentrated, indicating lack of contestability).

76. See, e.g., Federico Ciliberto et al., *Collusive Pricing Patterns in the US Airline Industry*, INT’L J. INDUS. ORG. (forthcoming 2019) (manuscript at 15), <https://www.sciencedirect.com/science/article/pii/S0167718717304125> [<https://perma.cc/3DXM-66ZU>]; Gaurab Aryal et al., *Public Communication and Collusion in the Airline Industry* 27 (Becker Friedman Inst. for Research in Econ., Working Paper No. 2018-11, 2018), https://bfi.uchicago.edu/sites/default/files/research/WP_No.2018-11.pdf [<https://perma.cc/6JKS-GXSS>].

77. LORD CHIEF-JUSTICE HALE, *De Portibus Maris*, in A TREATISE, IN THREE PARTS (c. 1670), reprinted in 1 A COLLECTION OF TRACTS RELATIVE TO THE LAW OF ENGLAND 45, 77–78 (Francis Hargrave ed., 1787).

78. *Id.*

79. *Munn v. Illinois*, 94 U.S. 113, 127, 129, 154 (1876). The Court quoted *De Portibus Maris* as follows:

A man, for his own private advantage, may, in a port or town, set up a wharf or crane, and may take what rates he and his customers can agree for crantage, wharfage, housellage, pesage; for he doth no more than is lawful for any man to do, viz., makes the most of his own. . . . If the king or subject have a public wharf, unto which all persons that come to that port must come and unlade or lade their goods as for the purpose, because they are the wharfs only licensed by the king, . . . or because there is no other wharf in that port, as it may fall out where a port is newly erected; in that case there cannot be taken arbitrary and excessive duties for crantage, wharfage, pesage, &c., neither can they be enhanced to an immoderate rate; but the duties must be reasonable and moderate, though settled by the king’s license or charter. For now the wharf and crane and other conveniences are affected with a public interest, and they cease to be *juris*

sophisticated justification for regulation used the term “king’s prerogative” as opposed to “common callings.”⁸⁰ Anyone could engage in a common calling, but only the Crown could authorize someone to exercise one of its prerogatives, which became the subject of exclusive grants and, in many cases, prices set by the government.⁸¹

What distinguished these special firms was both that they were essential to commerce and that competition was not believed to be possible because the market terrain, whether physical or economic, permitted only one firm. Under marginalism, this rationale shifted to the one developed in the previous section—namely, that the firms had very high fixed costs tending toward natural monopoly. As a result, any market with multiple suppliers would be threatened with ruinous competition. More technically, the long run average cost curve sloped downward continuously for such a distance that the optimal number of sustainable firms in a market was one. That fact provided a rationale for both monopoly status and price regulation, leaving the question of how to select the proper price.

This neoclassical economic rationale for monopoly rate regulation was significantly narrower than the common law classification scheme that Blackstone embraced. He supported legally enforced monopoly status for things that were almost certainly not natural monopolies, but merely sellers regarded as needing protection from “excessive” competition. For example, according to Blackstone, at common law a retail market authorized by the local Lord was entitled to protection from competition for a seven-mile radius from its location, which was one-third of a day’s journey.⁸² As late as 1827, the Court of King’s Bench held it unlawful for the defendant to sell fish from his house within the specified distance from an authorized local market.⁸³ In his 1812 *Livingston v. Van Ingen*⁸⁴ decision involving a state issued steamboat patent, New York’s Chancellor Kent opined that the government prerogative to grant exclusive charters for “beneficial public purposes” was clear.⁸⁵ Further, “[a]ll our bank charters, turnpike, canal and bridge

privati only

Id. at 127 (first and second alterations in original) (quoting HALE, *supra* note 77, at 77–78).

80. 1 WILLIAM BLACKSTONE, COMMENTARIES 79 (1765–1769).

81. See Barry E. Hawk, *English Competition Law Before 1900*, 63 ANTITRUST BULL. 350, 368 (2018); William L. Letwin, *The English Common Law Concerning Monopolies*, 21 U. CHI. L. REV. 355, 368 (1954).

82. 3 BLACKSTONE, *supra* note 80, at 218–19.

83. *Mosley v. Walker* (1827) 108 Eng. Rep. 640, 642.

84. 9 Johns. 507 (N.Y. 1812), *overruled by* *N. River Steamboat Co. v. Livingston*, 3 Cow. 182 (N.Y. 1825).

85. *Id.* at 573.

companies, ferries, markets, &c. are grants of exclusive privileges”⁸⁶ He also acknowledged that the extent of these grants might be “inexpedient or unwise,” but that did not undermine the state’s power to grant them.⁸⁷ His list included some things that could have been natural monopolies (canals and bridges) but others (banks and public markets) that almost certainly were not.⁸⁸ Although *Livingston* was a patent decision, most of these examples also did not come close to satisfying any “novelty” requirement that patent law assesses today. A few, however, were consistent with the British view, which a few American colonies and states had adopted, that patents should be awarded not merely to inventors, but also to those that introduced existing technology or products into a new area.⁸⁹

The American theory of price regulated monopoly went through three distinct stages between the early national period and World War II, each one strongly influenced by the economic theory of the day. Prior to the Jacksonian revolution the prevailing theory among Federalists and Whigs was that monopoly grants, often coupled with price regulation, were essential for encouraging economic development. This was Blackstone’s position and also that of the American Federalists as reflected in Chancellor Kent’s *Livingston* decision and Justice Story’s dissent in the *Charles River Bridge v. Warren Bridge*⁹⁰ case.⁹¹ The Jacksonian revolution, which got underway in the 1830s, introduced a classical economic theory that was hostile toward monopoly of all kinds and believed that the market itself was sufficient to incentivize any worthwhile project.⁹² After the turn of the twentieth century, neoclassicism substituted a more nuanced theory driven by the nature of a firm’s costs and its optimal size in relation to its market. The neoclassical theory, which largely survives to this day, allows considerably more room for regulation than Jacksonian classicism had, but less than early national Federalist ideology embraced.⁹³

Through all three periods, exclusive grants to public service companies were usually accompanied by price regulation. Prior to the

86. *Id.*

87. *Id.*

88. *Id.*

89. See Herbert Hovenkamp, *The Emergence of Classical American Patent Law*, 58 ARIZ. L. REV. 263, 276, 283 (2016). When federal patent law began to require novelty as a condition of patentability roughly a third of challenged patents were declared invalid on that ground. *Id.* at 276.

90. 36 U.S. (11 Pet.) 420 (1837).

91. See Herbert Hovenkamp, *Inventing the Classical Constitution*, 101 IOWA L. REV. 1, 21 (2015); *supra* notes 82–86 and accompanying text.

92. Hovenkamp, *supra* note 91, at 4–5.

93. *Id.* at 7.

Gilded Age, however, the prices were typically stipulated in a particular grantee's corporate charter rather than applied by legislation to the market as a whole.⁹⁴ In the 1837 *Charles River Bridge* case the plaintiff's charter authorizing a toll bridge stipulated the tolls that it could charge.⁹⁵ The charter did not expressly guarantee a monopoly right, however. Erection of the bridge had been costly and was regarded as a substantial feat of engineering.⁹⁶ The plaintiff had argued that government-mandated spatial separation of toll bridges was necessary in order to prevent ruinous competition between them.⁹⁷ Justice McLean accepted that argument in his dissent, observing that "great risk and expense were incurred" in its construction, and that "[t]he unrestricted profits contemplated, were necessary to induce or justify the undertaking."⁹⁸

94. See HOVENKAMP, *supra* note 5, at 125–30.

95. STANLEY KUTLER, *PRIVILEGE AND CREATIVE DESTRUCTION: THE CHARLES RIVER BRIDGE CASE 9–10, 32–45* (1971). The charter stipulated:

Each foot passenger (or one person passing), two-thirds of a penny; one person and horse, two pence two-thirds of a penny; single horse cart or sled, or sley, four pence; wheelbarrows, hand-carts, and other vehicles capable of carrying like weight, one penny, one-third of a penny; single horse and chaise, or sulkey, eight pence; coaches, chariots, phaetons and curricles, one shilling each; all other wheel carriages or sleds drawn by more than one beast, six pence; meat cattle and horses passing the said bridge, exclusive of those rode or in carriages or teams, one penny, one-third of a penny; swine and sheep, four pence for each dozen, and at the same rate for a greater or less number; and in all cases the same toll shall be paid for all carriages and vehicles passing the said bridge, whether the same be loaded or not loaded; and to each team one man and no more shall be allowed as a driver to pass free from payment of toll, and in all cases double toll shall be paid on the Lord's day; and at all times when the toll gatherer shall not attend his duty the gate or gates shall be left open.

Id. at 10. The 1868 monopoly charter at issue in the *Slaughter-House Cases* also stipulated the rates in the charter itself. Section 7 of the Charter provided:

That all persons slaughtering or causing to be slaughtered, cattle or other animals in said slaughter-houses, shall pay to the said company or corporation the following rates or perquisites, viz.: For all beeves, \$1 each; for all hogs and calves, 50 cents each; for all sheep, goats, and lambs, 30 cents each

Slaughter-House Cases, 83 U.S. 36, 42 (1872). The entire charter is reproduced at the beginning of the Court's opinion. *Id.* at 36–38.

96. See *Charles River Bridge v. Warren Bridge*, 36 U.S. (11 Pet.) 420, 562 (1837) (McLean, J., dissenting) ("When the Charles River bridge was built, it was considered a work of great magnitude. It was, perhaps, the first experiment made to throw a bridge of such length over an arm of the sea . . .").

97. *Id.* at 442 (majority opinion).

98. *Id.* at 562 (McLean, J., dissenting).

However, the closely divided Court held that the Constitution's Contract Clause did not require that an exclusive right should be implied.⁹⁹ Speaking for the majority, Chief Justice Taney famously wrote that "in grants by the public, nothing passes by implication."¹⁰⁰ The Court's split reflected a sharp economic division by Federalists such as Story and the new and more classical Jacksonians over the use of exclusive rights to encourage the development of infrastructure. Justice Story complained in his dissent that he could "conceive of no surer plan to arrest all public improvements, founded on private capital and enterprise, than to make the outlay of that capital uncertain and questionable."¹⁰¹

The two central features of the modern public utility—a monopoly grant and regulated rates—were thus in place long prior to the marginalist revolution.¹⁰² The *Charles River Bridge* approach of stipulating the rates in the corporate charter created some serious problems, however. Because the rates were specific to each firm they could vary from one situation to the next. Further, because the rates were stipulated in the charter they could not be changed without legislative revision. That might have been acceptable for a wooden toll bridge in a zero inflation environment, but it never would have worked for more modern utilities that had changing costs. So price regulation migrated from corporate charters to statutes that accommodated more frequent changes.¹⁰³ This occurred during the emergent Substantive Due Process era, which was generally hostile toward state imposed regulation of prices. It also meant, however, that the modern theory of rate-of-return price regulation occurred simultaneously with the early development of marginalism.

The Supreme Court approved *statutory* price regulation in principle in *Munn v. Illinois*,¹⁰⁴ although it did not consider whether the Constitution required any limitations.¹⁰⁵ In *Munn* there was no charter in which rates could be specified, because *Munn & Scott* was a common law

99. *Id.* at 466 (majority opinion).

100. *Id.* at 546.

101. *Id.* at 608 (Story, J., dissenting); see also Herbert Hovenkamp, *Technology, Politics, and Regulated Monopoly: An American Historical Perspective*, 62 TEX. L. REV. 1263, 1268–92 (1984).

102. On the legal evolution see HOVENKAMP, *supra* note 5, at 125–30.

103. *Id.*

104. 94 U.S. 113 (1877).

105. *Id.* at 136.

partnership.¹⁰⁶ In the *Spring Valley Water Works v. Schottler*¹⁰⁷ case in 1884, the Supreme Court again approved statutory rate regulation, in this case of water rates by a privately operated, incorporated water works whose charter did not specify rates.¹⁰⁸ The statute required that water rates be “reasonable,” and created a commission to determine reasonableness.¹⁰⁹ In approving this procedure, however, the court warned that it was reserving judgment on the possibility that such a commission might set a “manifestly unreasonable” rate.¹¹⁰ Justice Field was the lone dissenter, complaining that the plaintiff was incorporated, and imposing rates on it after the fact violated the Contract Clause.¹¹¹ Finally, in 1898 the Supreme Court held that the Fourteenth Amendment prohibited a state from imposing a railroad rate that generated less than “fair value” on the railroad’s property.¹¹²

While the Interstate Commerce Commission had been created in 1887, it did not have full authority to set rates until 1920.¹¹³ Prior to that, most rate regulation befell the states. The statutory or administrative setting of rates in price regulated industries had always involved some kind of notion that rates must be “above cost,” sufficient to give the regulated firm a reasonable rate of return.¹¹⁴ A major contribution of marginalist price theory was to provide a cost classification system that aided regulators in establishing theoretically workable rates of return. It did not do so well, however, in establishing the Constitution’s minimum standard of reasonableness.

Aside from sporadic attempts to regulate commodity prices, price regulation prior to the Civil War was largely ad hoc, negotiated on an individual firm basis. By contrast, neoclassical price theory sorted industries by cost structure. In 1887, the same year that the Interstate Commerce Act was passed, Henry Carter Adams wrote an important article classifying industries into three groups: those with decreasing

106. See *Wabash, St. Louis & Pac. Ry. v. Illinois*, 118 U.S. 557, 569 (1886) (“[T]he case of *Munn v. Illinois* was selected by the [C]ourt as the most appropriate one in which to give its opinions on that subject, because that case presented the question of a private citizen, or unincorporated partnership . . .”).

107. 110 U.S. 347 (1884).

108. *Id.* at 356.

109. *Id.* at 353.

110. *Id.* at 354.

111. *Id.* at 369 (Field, J., dissenting).

112. *Smyth v. Ames*, 171 U.S. 361 (1898); see also Robert L. Hale, *Does the Ghost of Smyth v. Ames Still Walk?*, 55 HARV. L. REV. 1116 (1942). For a contemporary defense of agency rate regulation, see generally Adelbert Moot, *Railway Rate Regulation*, 19 HARV. L. REV. 487 (1906).

113. Transportation Act of 1920, Pub. L. No. 66-152, 41 Stat. 456 (codified in scattered sections of 49 U.S.C.).

114. See Herbert Hovenkamp, *Progressive Legal Thought*, 72 WASH. & LEE L. REV. 653, 662 (2015).

returns to scale, those with constant returns to scale, and those with increasing returns to scale.¹¹⁵ For industries in the third classification, he argued, competition could not be relied on to make firms perform well.¹¹⁶ Firms with higher output would have lower costs than smaller ones, leading to a monopoly unless they were constrained.¹¹⁷ Adams's principal example of such industries was the railroads.¹¹⁸

Much of the early neoclassical theory of cost-based price regulation developed in debates over the proper treatment of railroads, an industry that was tailor-made for study of the marginalist economics of cost.¹¹⁹ The railroads had very high fixed costs for land, track networks, and equipment, but also significant operating (variable) costs, including fuel and labor.¹²⁰ Left to itself, the market appeared not to do a very good job of providing the conditions for competitive equilibrium.¹²¹ If a line was too isolated its owners would earn monopoly returns.¹²² State officials often responded by chartering additional railroads, resulting in significant overbuilding and complaints about "ruinous competition" already in the 1890s.¹²³ Because most shipping services were fungible, excess capacity tended to drive rates to marginal cost, without enough remaining to amortize fixed costs. In this atmosphere it is no wonder that several economists and legal writers advocated "pooling," which was essentially cartelization of competing railroads.¹²⁴ Indeed, both the Interstate Commerce Commission and the Eighth Circuit had approved the very rate-setting pool that the Supreme Court condemned in its first antitrust decision on the merits, the *United States v. Trans-Missouri Freight Ass'n*¹²⁵ case in 1897.¹²⁶

115. For this article, see generally Henry C. Adams, *Relation of the State to Industrial Action*, 1 PUB. AM. ECON. ASS'N, Jan. 1887, at 1, 55.

116. *Id.*

117. *Id.* at 60.

118. *Id.* at 61.

119. Herbert Hovenkamp, *Regulatory Conflict in the Gilded Age: Federalism and the Railroad Problem*, 97 YALE L.J. 1017, 1035–36 (1988).

120. *Id.*

121. *Id.* at 1035–37.

122. *See id.* at 1031–32, 1035–36.

123. *See discussion supra* notes 47–48.

124. *E.g.*, Arthur T. Hadley, *The Prohibition of Railway Pools*, 4 Q.J. ECON. 158, 158–61, 164 (1890); Edwin R.A. Seligman, *Railway Tariffs and the Interstate Commerce Law II*, 2 POL. SCI. Q. 369, 373, 378–80 (1887). For further discussion see generally Hovenkamp, *supra* note 119.

125. 58 F. 58 (1893), *rev'd*, 166 U.S. 290 (1897).

126. *Id.* at 76–77 ("To make railroads of the greatest possible service to the country, contract relations would be essential, because there would need to be joint tariffs, joint running arrangements and interchange of cars, and a giving of credit to a large extent, some of which were obviously beyond the reach of compulsory legislation, and, even if they were not, could be best

While collusion would protect the railroads from ruinous competition, it would not protect customers, however. A railroad cartel would charge its profit-maximizing price just as a single firm monopolist would, not the minimum price needed to sustain investment in the industry. In the 1880s, Gilded Age scholars of railway regulation such as eventual Yale president Arthur Twining Hadley embarked on a serious economic and legal analysis of railroad costs and pricing.¹²⁷ Already by this time, railroad operators and economists understood that paying off fixed cost debt required keeping output high, and that this would be facilitated by permitting price discrimination.¹²⁸ Hadley showed how adding incremental freight at reduced rates served to lower overall costs, provided that the rate was greater than incremental operating costs.¹²⁹ Hadley was already closing in on a fundamental conception of mid-twentieth century rate regulation¹³⁰: Absent capacity constraints, the optimal rate should bring in every class of customers at the highest rate that class is willing to pay, thus maximizing output. But of course, that was price discrimination, which could be trusted to produce loud complaints from those required to pay the higher prices.

The need to price discriminate in order to sustain high output led to rate classification rules that seemed unnecessarily complicated and fundamentally unfair to outside observers,¹³¹ although their purpose was clear: Any freight rate sufficient to cover the variable costs of shipment would make a positive contribution to net revenue. By bringing in every such customer at the highest price they were willing to pay, the railroad maximized its revenue and could divide its fixed costs over a larger number of sales. Offsetting this, of course, would be the administrative cost of developing so many classifications, including disputing classification costs with unhappy customers or their rivals.¹³² These solutions, extended to other public utilities, produced much of the modern

settled, and all the incidents and qualifications fixed, by the voluntary action of the parties in control of the roads respectively.” (quoting 1 I.C.C. ANN. REP. 1, 33 (1887)).

127. See generally, e.g., ARTHUR TWINING HADLEY, *RAILROAD TRANSPORTATION: ITS HISTORY AND ITS LAW* (1885).

128. *Id.* at 114–18 (discussing this phenomenon in the context of local price discrimination).

129. *Id.* at 117; see also Frank W. Taussig, *A Contribution to the Theory of Railway Rates*, 5 Q.J. ECON. 438, 454–59 (1891) (showing more technically the efficiency of price discrimination).

130. See *infra* notes 143–44 and accompanying text (discussing Ramsey pricing).

131. See Sidney S. Alderman, *How Shall the Railroad Rate Structure be Regulated in the Public Interest?*, 12 L. & CONTEMP. PROBS. 579, 579 (1947) (considering certain cases and their position that rate rules were unfair to certain shippers).

132. See, e.g., *United States v. W. Pac. R.R.*, 352 U.S. 59, 60, 69 (1956) (involving a dispute over higher tariffs for napalm-containing bombs); *Great N. Ry. v. Merchants’ Elevator Co.*, 259 U.S. 285, 288 (1922) (involving a dispute over grain).

neoclassical economics of regulation, including a more sophisticated understanding of second- and third-degree price discrimination. For example, the Ramsey pricing solution discussed below contemplated very elaborate rate classifications, literally assessing a rate equal to each individual customer's willingness to pay.¹³³

Once marginalism led to classification of costs as fixed, variable, and marginal, the basic theory of rate regulation became simple enough. The devil was in the details. The regulated utility, transportation, or other firm was entitled to a "fair rate of return" on its fixed cost investment, plus "pass through" of its variable costs.¹³⁴ This has come to be called "rate of return" regulation or "cost of service" ratemaking.¹³⁵ It has been widely used for a century, but also widely criticized, mainly for providing insufficient incentives for firms to innovate or reduce costs.¹³⁶ It also produced a lengthy constitutional debate about the rate base—namely, whether the firm's return should be based on replacement cost of worn out plant and equipment, or whether it was enough that the firm receive a positive return on its actual historical investment, which was typically much lower.¹³⁷ It also depended on a clear delineation between fixed and variable costs that was sometimes difficult to maintain in practice.

During the Substantive Due Process era, the Supreme Court had insisted on "fair value" and this was generally interpreted by contemporaries as replacement cost or even more.¹³⁸ The Court's

133. See *infra* notes 143–44 and accompanying text.

134. Paul M. Sotkiewicz & Lynne Holt, *Public Utility Commission Regulation and Cost-Effectiveness of Title IV: Lessons for CAIR*, 18 *ELECTRICITY J.* 68, 74 (2005).

135. The standard treatment is attributed to ALFRED E. KAHN, *THE ECONOMICS OF REGULATION: PRINCIPLES AND INSTITUTIONS* 31, 48 (Mass. Inst. of Tech. 2d ed. 1988) (1970–1971). See also JAMES C. BONBRIGHT ET AL., *PRINCIPLES OF PUBLIC UTILITY RATES* 109 (2d ed. 1988) (describing cost of service ratemaking); CHARLES F. PHILLIPS, JR., *THE REGULATION OF PUBLIC UTILITIES: THEORY AND PRACTICE* 409–10 (2d ed. 1988) (describing the shift in regulatory proceedings towards cost of service rate regulation).

136. See, e.g., Harvey Averch & Leland L. Johnson, *Behavior of the Firm Under Regulatory Constraint*, 52 *AM. ECON. REV.* 1052, 1052 (1962).

137. The debate also became a focus of the early law and economics movement. See generally Robert L. Hale, *Rate Making and the Revision of the Property Concept*, 22 *COLUM. L. REV.* 209 (1922) (arguing for historical cost).

138. See, e.g., *Smyth v. Ames*, 169 U.S. 466, 541 (1898), *overruled by Fed. Power Comm'n v. Nat. Gas Pipeline Co. of Am.*, 315 U.S. 575 (1942). The Court indicated that even a rate that provided for full recovery of all costs, including interest and a dividend to shareholders, would be insufficient if less than fair market value:

It cannot, therefore, be admitted that a railroad corporation maintaining a highway under the authority of the State may fix its rates with a view solely to its own interests, and ignore the rights of the public. But the rights of the public would be ignored if rates for the transportation of persons or property on a railroad are exacted without reference to the fair value of the property used for

formulation of the requirements in *Smyth v. Ames*¹³⁹ was:

[I]n order to ascertain that [fair] value, the original cost of construction, the amount expended in permanent improvements, the amount and market value of its bonds and stock, the present as compared with the original cost of construction, the probable earning capacity of the property under particular rates prescribed by statute, and the sum required to meet operating expenses, are all matters for consideration, and are to be given such weight as may be just and right in each case. We do not say that there may not be other matters to be regarded in estimating the value of the property.¹⁴⁰

After the Roosevelt administration and the New Deal had relegated Substantive Due Process to history, however, the Supreme Court began to uphold rates based on historical cost as the constitutional minimum.¹⁴¹

Two of the most serious criticisms of rate of return regulation are, first, that the basic formulation leads to inefficient pricing to the extent that pass through of variable costs is not a sufficiently close approximation of marginal cost. Second, the administration of cost-of-service ratemaking takes away firms' incentives to innovate and reduce costs.¹⁴²

On the first, cost of service rate making is often not a close approximation to marginal cost pricing. A single price equal to marginal cost would be far too low for most public utilities because it would not compensate the utility for its fixed costs. But how should fixed costs be allocated? The Ramsey solution, which was intended to maximize output and thus reduce the impact of fixed costs, was to price to each customer at the inverse of that customer's elasticity of demand.¹⁴³ While Ramsey

the public or the fair value of the services rendered, but in order simply that the corporation may meet operating expenses, pay the interest on its obligations, and declare a dividend to stockholders.

Id. at 544.

139. 169 U.S. 466 (1898), *overruled by* Fed. Power Comm'n v. Nat. Gas Pipeline Co. of Am., 315 U.S. 575 (1942).

140. *Id.* at 546–47.

141. *See generally* Fed. Power Comm'n v. Hope Nat. Gas, 320 U.S. 591, 606–07 (1944) (approving rate based on historical cost less depreciation); Fed. Power Comm'n v. Nat. Gas Pipeline Co. of Am., 315 U.S. 575 (1942) (rejecting argument that the Constitution required rates based on anticipated replacement cost).

142. MARK A. JAMISON, PUB. UTIL. RESEARCH CTR., RATE OF RETURN: REGULATION 3 (2005), https://bear.warrington.ufl.edu/centers/purc/docs/papers/0528_Jamison_Rate_of_Return.pdf [<https://perma.cc/Q874-5UT3>].

143. *See, e.g.*, Burlington N. R.R. v. ICC, 985 F.2d 589, 596 (D.C. Cir. 1993) (“Under Ramsey pricing, the regulator allows firms to charge each user a premium over marginal cost in

formalized this solution in the 1920s and it bears his name, railroads had already been doing a crude version of this for decades, as Arthur Twining Hadley's 1885 book on railroad rates had elaborated.¹⁴⁴ Regulators understood that the key to profitability was to maximize output within the capacity of the existing fixed cost network.¹⁴⁵ Any shipper willing to pay more than running expenses was profitable in the short run.¹⁴⁶ At the same time, however, someone had to amortize fixed costs. As a result, the elaborate railroad rate classification schemes made rates for lower value goods such as coal or cement that were far below the rates charged for higher value finished goods.¹⁴⁷

Railroads also charged significantly higher prices per mile for short hauls than for long hauls.¹⁴⁸ The rationale for that phenomenon is readily apparent to anyone who looks at a map of the American railroad network. Often there was only one railroad between two fairly nearby towns. However, as the drop and pickup points became further apart the network provided more alternatives, making it possible for more railroads to compete for the same shipment. That is, the amount of competition in the market increased as the distance between the two shipping points increased. One effect of this difference is that the federal concern, limited to interstate shipments, was largely with rates that were too low.¹⁴⁹ By contrast, the state concern, limited mainly to intrastate shipments, was with rates that were too high.¹⁵⁰

One idea intended to make regulated pricing resemble market pricing more closely was Ronald Coase's suggestion of two-part pricing, which attempted to segregate the fixed and variable cost components, permitting marginal cost pricing of the latter.¹⁵¹ Coase argued that every customer

inverse proportion to the elasticity of the user's demand. Because the highest charges fall on the most inelastic demanders, the impact on total usage is minimized. Thus, the Commission believed, it would reconcile the railroad's need for revenue to cover total costs with the least possible distortion of demand (i.e., railroad usage would approximate as nearly as possible the level that would prevail under perfect competition)."). See generally Frank P. Ramsey, *A Contribution to the Theory of Taxation*, 37 *ECON. J.* 47 (1927) (explaining the Ramsey Solution).

144. See HADLEY, *supra* note 127, at 118. Hadley defends a railroad's practice of rate discrimination in oyster shipping rates as between two buyers with differential willingness to pay by arguing that "[a]t the higher rate the road cannot get sufficient volume of business. At the lower rate it cannot get sufficient profit. It must . . . get what it can at high rates, and fill up at the lower ones." *Id.*

145. See Hovenkamp, *supra* note 119, at 1036–37.

146. See *id.* at 1037.

147. HADLEY, *supra* note 127, at 121.

148. *Id.* at 117.

149. See Hovenkamp, *supra* note 119, at 1056–62.

150. *Id.* at 1056.

151. Ronald H. Coase, *The Marginal Cost Controversy*, 13 *ECONOMICA* (n.s.) 169, 173 (1946). For elaboration, see KAHN, *supra* note 135, at 95–100.

could be required to pay an access fee which was fixed, in the sense that it did not vary with the number of units that the customer used.¹⁵² This fee would be calculated so as to cover the fixed cost components of public utility costs, and Ramsey concerns about maximizing output could be included in the computation. Then each customer's actual use would be priced out at marginal cost.

Coase's article had actually been written as a response to Harold Hotelling's far more interventionist approach, which would require the government to pay the fixed cost proportion of transportations and utilities by providing the infrastructure, and then charging customers the marginal cost of operation.¹⁵³ Hotelling's argument was highly praised by prominent public utility scholars of the day, such as James Bonbright.¹⁵⁴ By contrast, Coase, who was always fierce in searching out private alternatives, sought one that would minimize government control. It is worth noting that this debate, which occurred in the 1930s and 1940s, reveals how central marginalist conceptions had become in economic thinking about regulated pricing—both among those who advocated for a high degree of public ownership as well as libertarians such as Coase.

The second development, which was to have a major impact on the deregulation movement, was Averch and Johnson's theory of gold-plating.¹⁵⁵ Economically, the "Averch-Johnson effect" operates as a severe qualification on Coase's *The Nature of the Firm*¹⁵⁶ when the firm is in a price-regulated environment. Coase actually conceived of his own article as a first attempt to apply marginalist economics to questions about the optimal size and shape of a business firm.¹⁵⁷ He opened his paper with praise for Marshall:

It is hoped to show . . . that a definition of a firm may be obtained which is not only realistic . . . but is tractable by two of the most powerful instruments of economic analysis developed by Marshall, the idea of the margin and that of substitution, together giving the idea of substitution at the margin.¹⁵⁸

152. See Coase, *supra* note 151.

153. Harold Hotelling, *The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates*, 6 *ECONOMETRICA* 242, 242, 258–60, 264 (1938).

154. James C. Bonbright, *Major Controversies as to the Criteria of Reasonable Public Utility Rates*, 30 *AM. ECON. REV.* 379, 385 (1941) (following Hotelling); see also Nancy Ruggles, *Recent Developments in the Theory of Marginal Cost Pricing*, 17 *REV. ECON. STUD.* 107, 107–10 (1949) (describing Hotelling's theory).

155. See Averch & Johnson, *supra* note 136, at 1068.

156. Coase, *supra* note 13.

157. *Id.* at 386.

158. *Id.* at 386–87.

Coase argued that the firm, driven entirely by the need to maximize its profits, relentlessly compares the marginal cost of doing something internally against the marginal cost of procurement from outside, choosing whichever produces the largest payoff.¹⁵⁹ As a result, the firm's structure is efficient to the extent that its procurement decisions are efficient.¹⁶⁰

But suppose the firm is guaranteed a profitable price, equal to or slightly above marginal cost, on all of its internal production. The firm will have an incentive to perform activities internally rather than through purchase on the market, even though the latter is the better choice in a competitive market. The firm would have a regulation-imposed incentive to integrate vertically or to expand into other markets. For example, to the extent that it gets its rate of return on its capital investment and is limited to pass through of costs for variable cost items such as labor,¹⁶¹ it has an incentive to invest relatively more in capital assets. As a result, regulated firms tend to be excessively capital intensive when compared with the unregulated market.¹⁶² Perhaps more ominously, the regulated firm would have an incentive to build unneeded infrastructure, knowing that it would be guaranteed a profitable rate of return.¹⁶³ By contrast, a competitive firm would enlarge its plant only if anticipated receipts, determined entirely by the market, exceeded anticipated costs.

The Averch-Johnson literature produced interesting collateral issues, such as the regulated utility's right to recover for its "stranded" costs, which are costs incurred in enlargement of infrastructure that later turned out to be unnecessary or unwise.¹⁶⁴ In some cases this happened because

159. *Id.* at 406.

160. For example, monopoly in a supply market might induce a firm to do something internally even though external procurement might be cheaper if that market were competitive.

161. Although, regulation will lead to higher wages to the extent that the firm is guaranteed pass through of higher wage costs. *See, e.g.,* Nancy L. Rose, *Labor Rent Sharing and Regulation: Evidence from the Trucking Industry*, 95 J. POL. ECON. 1146, 1146, 1173–75 (1987) (determining that deregulation in the trucking industry leads to lower wages).

162. For a simple economic analysis, together with a summary of the most important critiques see JEAN-JACQUES LAFFONT & JEAN TIROLE, *A THEORY OF INCENTIVES IN PROCUREMENT AND REGULATION* 91–93 (1993) and W. KIP VISCUSI ET AL., *ECONOMICS OF REGULATION AND ANTITRUST* 433–36 (4th ed. 2005). For deep skepticism see Stephen M. Law, *Assessing the Averch-Johnson-Wellisz Effect for Regulated Utilities*, 6 INT'L J. ECON. & FIN. 41, 51–52 (2014). For more technical treatment in the context of natural monopolies see generally Paul L. Joskow, *Regulation of Natural Monopolies*, in 2 HANDBOOK OF LAW AND ECONOMICS (A. Mitchell Polinsky & Steven Shavell eds., 2007).

163. *See* KAHN, *supra* note 135, at 35–36.

164. *See* Herbert Hovenkamp, *The Takings Clause and Improvident Regulatory Bargains*, 108 YALE L.J. 801, 831 (1999) (reviewing J. GREGORY SIDAK & DANIEL F. SPULBER, *DEREGULATORY TAKINGS AND THE REGULATORY CONTRACT: THE COMPETITIVE TRANSFORMATION OF NETWORK INDUSTRIES IN THE UNITED STATES* (1997)).

of policy changes that made previous investments improvident, such as coal-fired plants or hospitals' certificates of need.¹⁶⁵ Firms acting under regulation often propose capital investments to a regulator, and Averch-Johnson suggests this can happen even if those investments would have been inefficient in a competitive market. Then later, when the regulator views things differently or the increased demand that the regulated utility promised does not materialize, the regulator withdraws its authorization. This might leave a partially built or recently on line plant "stranded," in the sense that it is no longer needed or demand is insufficient. The firm may then claim that the regulatory approval and its subsequent retraction is an improper taking of its property without compensation.¹⁶⁶

No matter how one feels about Averch-Johnson, one thing that emerges clearly is that cost-of-service rate making is not the best way to determine price and output in any market where competition is available as a workable alternative. As a result, price regulated markets must be pared down so that price and output is regulated only in those portions of the market where competition seems not to work. This has been one of the principal consequences of the deregulation movement discussed below.¹⁶⁷

III. THE CHANGING DOMAIN OF MARKET FAILURE: TRADITIONAL REGULATION, PUBLIC CHOICE, AND PIGOUVIAN TAXES

The marginalist revolution led United States policy to the "mixed" economy that it has today.¹⁶⁸ Under marginalism, markets appear less robust than pre-marginalists believed, but they remain the dominant means by which resources move around. The most important economic policy function of government today is identifying markets that work well when left relatively untended and those that require intervention, and then determining the correct prescription for the latter group.

165. See generally Michael E. Granfield, *Resource Allocation Within Hospitals: An Unambiguous Analytical Test of the A-J Hypothesis*, 7 J. APPLIED ECON. 241 (1975) (explaining the application of the A-J hypothesis in a hospital setting).

166. See Emily Hammond & Jim Rossi, *Stranded Costs and Grid Decarbonization*, 82 BROOK. L. REV. 645, 650–59, 691 (2017) (discussing stranded costs in the context of attempts to migrate power generation away from carbon-based fuels). See generally J. GREGORY SIDAK & DANIEL F. SPULBUR, *DEREGULATORY TAKINGS AND THE REGULATORY CONTRACT: THE COMPETITIVE TRANSFORMATION OF NETWORK INDUSTRIES IN THE UNITED STATES* xiii (1997) (addressing "deregulatory policies that threaten to reduce or destroy the value of private property," or "deregulatory takings").

167. See discussion *infra* Part V.

168. For a robust defense, see generally JACOB S. HACKER & PAUL PIERSON, *AMERICAN AMNESIA: HOW THE WAR ON GOVERNMENT LED US TO FORGET WHAT MADE AMERICA PROSPER* (2016) (discussing the "mixed" economy and the prosperity it has created in America).

This makes the concept of “market failure” central to modern public decision making. In the political economy of regulation, the term has more than a single meaning. The dominant definition within neoclassical economics relates to the inability of a market to reach an equilibrium on its own,¹⁶⁹ or otherwise to reach only equilibria that exhibit unsatisfactory output and prices.¹⁷⁰ As an example of the latter, when entry is impossible, monopoly may be a stable equilibrium, but one in which price and output are suboptimal. Another example is externalities that cause harm in secondary markets. Internally, the primary market might attain an equilibrium, and perhaps even at competitive prices. However, that market might impose unnecessarily heavy costs elsewhere, indicating that these prices are not covering the full social cost of the activity. The debate over so-called Pigouvian taxes, which has waxed and waned and waxed again over nearly a century, principally concerns mechanisms for reducing or eliminating these externalities by taxing the harmful conduct, and in some cases using the funds to correct harms elsewhere.¹⁷¹

Under the strictest definition, static market failure is any durable deviation from marginal cost pricing. Of course, dynamic factors such as innovation may simultaneously increase the innovator’s price-cost margins while also increasing output or welfare, at least in the long run.¹⁷² For example, one of the longest-standing responses to market failure is the intellectual property system, which involves government creation of exclusive rights to facilitate limited periods of high price-cost margins as an inducement to innovation. In any event, market failure is often defined in ways that do not use Pareto efficiency or perfect competition as a baseline—for example, one can speak of it as “the failure of the market to bring about results that are in the best interests of society as a

169. See generally LESTER TELSER, *ECONOMIC THEORY AND THE CORE* (1978) (outlining the core theory in economics and the effect on market equilibrium); see generally, e.g., Stephen Craig Pirrong, *An Application of Core Theory to the Analysis of Ocean Shipping Markets*, 35 J.L. & ECON. 89 (1992) (applying core theory to ocean shipping markets); Timothy K. Smith & Ronald B. Lieber, *Why Air Travel Doesn’t Work*, FORTUNE, Apr. 3, 1995, at 42 (discussing core theory in the context of airlines).

170. The standard discussion is Francis M. Bator, *The Anatomy of Market Failure*, 72 Q.J. ECON. 351 (1958). See also Paul A. Samuelson, *The Pure Theory of Public Expenditure*, 36 REV. ECON. & STAT. 387, 387–89 (1954) (attempting to assign a strict definition to market failure). For even further commentary, see generally PUBLIC GOODS AND MARKET FAILURES: A CRITICAL EXAMINATION (Tyler Cowen ed., 2017).

171. See *infra* text accompanying notes 208–18.

172. See, e.g., Richard R. Nelson, *Thinking About Technology Policy: “Market Failures” versus “Innovation Systems”* 8 (UCL IIPP, Working Paper No. 2017-02, 2017), https://www.ucl.ac.uk/bartlett/public-purpose/sites/public-purpose/files/thinking_about_technology_policy_-_market_failures_versus_innovation_systems.pdf [<https://perma.cc/JNV8-R8NN>].

whole,”¹⁷³ or as the failure of market institutions “to sustain ‘desirable’ activities or to estop ‘undesirable’ activities.”¹⁷⁴

Technical neoclassical definitions of market failure generally exclude purely distributive factors. Of course, sometimes income inequality can lead to market failure, particularly when inequality makes it more costly for people without resources to move into more productive occupations,¹⁷⁵ or when lack of competitive pressure on certain groups induces them to be less productive.¹⁷⁶ To the extent maldistribution of wealth impairs productivity, and thus output, it can be a market failure even under a strictly neoclassical definition.

The classical political economists had strong faith in markets. To be sure, there were important qualifications, such as Thomas Malthus’ population argument that in the long run the market would force the population to subsistence levels,¹⁷⁷ Ricardo’s concerns with monopoly in land rents,¹⁷⁸ or John Stuart Mill’s discussion of natural monopoly and the British postal system or lines for gas lighting.¹⁷⁹ But for the most part the classical political economists regarded these as rare exceptions to the general theory of markets. That was even truer of classical political economy in the United States, where an abundance of undeveloped land was widely seen as making the concerns expressed by both Malthus and Ricardo relatively unimportant.¹⁸⁰ Increasingly after the Jackson era, American legislatures and courts trusted markets to allocate resources

173. Alain Marciano & Steven G. Medema, *Market Failure in Context: Introduction*, 47 HIST. POL. ECON. 1, 1 (2015).

174. Bator, *supra* note 170, at 351. Vaguely, Bator adds that the meaning of “desirable” is determined by the solution to “some explicit or implied maximum-welfare problem.” *Id.*

175. *See generally, e.g.*, Kim A. Weeden & David B. Grusky, *Inequality and Market Failure*, 38 AM. BEHAV. SCIENTIST 473 (2014) (discussing market failure as a result of income inequality and occupational, educational, managerial, and capital “rents”).

176. For example, if it could be shown that a tax and transfer system that levelled income also increased productivity, then this would be a correction of a market failure in the neoclassical sense.

177. *See generally* THOMAS MALTHUS, AN ESSAY ON THE PRINCIPLE OF POPULATION (1798) (discussing population’s effect on markets).

178. *See* DAVID RICARDO, ON THE PRINCIPLES OF POLITICAL ECONOMY AND TAXATION 49–76 (1817) (describing the nature of land rents and the interplay between rent and commodities).

179. *See generally* JOHN STUART MILL, PRINCIPLES OF POLITICAL ECONOMY WITH SOME OF THEIR APPLICATION TO SOCIAL PHILOSOPHY (1848) (discussing the concept of natural monopoly).

180. *See* Herbert Hovenkamp, *The Political Economy of Substantive Due Process*, 40 STAN. L. REV. 379, 421–31 (1988) (summarizing how nineteenth century American political economists believed that the vast amounts of undeveloped land in the United States established that the threats offered by Malthus and Ricardo would occur in the distant future, if at all).

properly. Some states even invalidated monopoly rights for rather clear natural monopoly industries, such as gas lighting.¹⁸¹

A common characteristic of the bridges, toll roads, canals, and railroads that received monopoly charters prior to the Civil War is that they required a significant fixed cost investment, and in most cases a single installation could handle all of the traffic.¹⁸² As a result, they were true natural monopolies. Competition would drive prices to marginal cost without enough remaining to cover fixed costs. Indeed, in the *Charles River Bridge* case the plaintiffs raised the “ruinous competition” defense sixty years before it came into use in railroad cases.¹⁸³ Tolls decreased by half to two-thirds upon the opening of the competing Warren Bridge.¹⁸⁴ By the time the case reached the Supreme Court the Charles River Bridge had already closed.¹⁸⁵ It would reopen in 1841, only when the Commonwealth turned it into a free bridge.¹⁸⁶

At least for a time, the marginalist revolution very largely brought the classicists’ robust faith in markets to an end—a view undoubtedly aided by World War I and the Depression. The concerns about fixed costs and failure of equilibrium discussed previously¹⁸⁷ were one significant manifestation, but there were others. The dramatic rise of antitrust was driven by the belief that even markets that are not natural monopolies can fail. Indeed, the extent to which antitrust was brought to bear in *nonmonopolized* markets is striking. After the New Deal, Antitrust became a vehicle for controlling manufacturer-created distribution systems in competitively structured markets,¹⁸⁸ as well as controlling

181. See, e.g., *Norwich GasLight Co. v. Norwich City Gas Co.*, 25 Conn. 19, 37–38 (1856) (invalidating monopoly grants for a gas light utility granted by both the state and the local government and permitting second firm to install competing lines).

182. See HOVENKAMP, *supra* note 5, 105–68, 199–206 (discussing fixed costs and market failure history).

183. *Charles River Bridge v. Warren Bridge*, 36 U.S. (11 Pet.) 420, 436 (1837) (“No other ferry or bridge could be erected . . . without being *so near*, in the language of Blackstone, as to draw away the custom of the elder ferry or bridge; or without producing, in the language of Chancellor Kent, *ruinous* competition.”). On the railroads and ruinous competition, see discussion *supra* notes 112–26.

184. See *Charles River Bridge v. Warren Bridge*, 24 Mass. (7 Pick.) 349, 538 (1829).

185. HOVENKAMP, *supra* note 5, at 112.

186. See KUTLER, *supra* note 95, at 32–45.

187. See discussion *supra* Part I.

188. See HOVENKAMP, *supra* note 1, 206–39. See generally LAURA PHILLIPS SAWYER, *AMERICAN FAIR TRADE: PROPRIETARY CAPITALISM, CORPORATISM, AND THE “NEW COMPETITION,” 1890–1940* (2018) (exploring the development of antimonopoly law and economics, including antitrust law, in the wake of the New Deal).

price differences among competing dealers,¹⁸⁹ and for limiting the effects of product differentiation. Underlying this was a fear of vertical integration that, at least as a matter of economics, can only be described as hysterical.¹⁹⁰

Increasingly after the mid-twentieth century, regulatory theory embraced a renewed neoclassicism that once again saw markets as robust, although perhaps not as robust as they were perceived by the classicists. Coasean thinking strongly emphasized market solutions, even in markets that did not meet the requirements for perfect competition. Indeed, the Coase Theorem itself focused on bargaining in bilateral monopolies.¹⁹¹ Entitlements such as exclusive rights in communications spectrum¹⁹² or the right to pollute became tradeable and the scope of market failure declined to near non-existence. In the Coasean vision, even the lighthouse, frequently given as an example of resistance to market supply, became freed from government control.¹⁹³

Worse yet, with the rise of modern public choice theory in the 1960s, interest group capture, rather than market failure, became the dominant positive rationale for regulation.¹⁹⁴ Writers like Buchanan and Tullock began with perfect competition as the baseline, and then sought to explain how democratic political voting produced harmful deviations.¹⁹⁵ Mancur Olson's much more popular but equally devastating *Logic of Collective Action* also began with competitive markets as a baseline and then tried to explain regulatory deviations as political oligopolies or cartels that

189. See HOVENKAMP, *supra* note 1, 220–39 (discussing the effects of the passage of the Robinson-Patman Act of 1936, Pub. L. No. 74-692, 49 Stat. 1526 (codified at 15 U.S.C. § 13 (2012))).

190. See *id.* at 206–19.

191. See Herbert Hovenkamp, *The Coase Theorem and Arthur Cecil Pigou*, 51 ARIZ. L. REV. 633, 640–41 (2009). See generally Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960) (explaining that in a world without transaction costs, people would negotiate with each other to produce the most mutually beneficial outcomes, focusing heavily on bilateral monopolies).

192. See, e.g., Ronald H. Coase, *The Federal Communications Commission*, 2 J.L. & ECON. 1, 25 (1959).

193. See Ronald H. Coase, *The Lighthouse in Economics*, 17 J.L. & ECON. 357, 376 (1974) (disputing economists' use of the lighthouse as an example of a service that the government must provide rather than the private market).

194. See generally NAT'L BUREAU OF ECON. RESEARCH, *THE REGULATED ECONOMY: A HISTORICAL APPROACH TO POLITICAL ECONOMY* (Claudia Goldin & Gary D. Libecap eds., 1994) (exploring the history of market regulation and focusing on the influence of interest groups on the development of governmental policies).

195. See generally JAMES M. BUCHANAN & GORDON TULLOCK, *THE CALCULUS OF CONSENT: LOGICAL FOUNDATIONS OF CONSTITUTIONAL DEMOCRACY* (1965) (discussing the behavior of an individual as he participates in the democratic process, analogizing constitutional choice with the theory of private choice in economics).

enriched their organizers at the expense of society.¹⁹⁶ Ironically, Olson borrowed his regulatory oligopoly theory straight from left leaning Edward Chamberlin, who had been his dissertation director.¹⁹⁷ Under the theory, small relatively homogeneous groups of individuals or firms with a common cause could wield disproportionate political power over larger but less unified or focused interest groups that represented the larger part of society.¹⁹⁸ Under the Stigler-Posner-Peltzman model of regulation, its roots lay in political bargaining among special interests rather than neoclassical price theory or industrial organization economics.¹⁹⁹

Theories of interest group capture actually entered American political-economic discourse from the left, through writers such as Charles Beard and Gabriel Kolko. Beard argued during the Progressive Era that the Constitution was largely a product of capture by urban owners of personal property over much more populous but widely dispersed agrarian interests.²⁰⁰ Kolko, a New Left disciple of Herbert Marcuse, argued in the 1960s that Gilded Age railroad regulation really occurred at the behest of the railroads themselves, who wanted regulation in order to shelter themselves from competition.²⁰¹ After that, regulatory capture theory moved rightward.²⁰²

Today we seem to have reached a balance in regulatory theory in which a combination of traditional price and organization economics and public choice theory are called upon as the best positive explainers of regulatory policy. In a democratic society, to see regulation as nothing more than the consequence of market structure or cost characteristics is naïve, but so is seeing it as nothing more than the outcome of fights among competing interest groups. For example, today in the great majority of states groceries, clothing, and automobiles are sold in markets that are more-or-less competitive, with prices set predominantly through

196. See MANCUR OLSON, JR., *THE LOGIC OF COLLECTIVE ACTION: PUBLIC GOODS AND THE THEORY OF GROUPS* 133–67 (1965).

197. See HOVENKAMP, *supra* note 1, at 313 n.72.

198. *Id.* at 313–14.

199. Sam Peltzman, *An Economic Interpretation of the History of Congressional Voting in the Twentieth Century*, 75 AM. ECON. REV. 656, 656, 674 (1985); Sam Peltzman, *The Economic Theory of Regulation After a Decade of Deregulation*, 1989 BROOKINGS PAPERS ON ECON. ACTIVITY: MICROECONOMICS 1, 1; Sam Peltzman, *Toward a More General Theory of Regulation*, 19 J.L. & ECON. 211, 212 (1976); Richard A. Posner, *Taxation by Regulation*, 2 BELL J. ECON. & MGMT. SCI. 22, 31 (1971); Richard A. Posner, *The Social Costs of Monopoly and Regulation*, 83 J. POL. ECON. 807, 819 (1975); George J. Stigler, *The Theory of Economic Regulation*, 2 BELL J. ECON. & MGMT. SCI. 3, 3, 7 (1971).

200. CHARLES A. BEARD, *AN ECONOMIC INTERPRETATION OF THE CONSTITUTION OF THE UNITED STATES* 73–151 (1913).

201. See generally GABRIEL KOLKO, *RAILROADS AND REGULATION, 1877–1916* (1965) (discussing railroad regulation).

202. See HOVENKAMP, *supra* note 1, at 308–14.

voluntary transactions. By contrast, retail electricity and natural gas are sold at retail through regulated monopolies. Few would seriously argue that this is because electric power lobbyists have more clout than those who work in the food and agricultural markets.

To be sure, interest group theory may explain many of the details of regulatory policy. As noted below, the deregulation movement has done a great service by distinguishing those attributes of the economy that are best left to the market from those that require more active state intervention.²⁰³ In the process, however, the deregulation movement faces abundant criticism from those who think it has gone too far on the one side, and those who think that it has not gone far enough on the other.

The history of “Pigouvian” taxes as a corrective for market failure long precedes both Pigou and the marginalist revolution. In 1776, Adam Smith’s *Wealth of Nations* proposed a tax on carriages to offset the damage they do to roads. He suggested that the tax be proportional to weight because “each carriage is supposed to pay exactly for the wear and tear which that carriage occasions of the roads.”²⁰⁴ In the United States, taxes rather than prohibitions were frequently used by the mid-nineteenth century to control liquor consumption, lotteries, gambling and other vices.²⁰⁵ Reflecting on these, in 1876 Thomas M. Cooley included a chapter in his *Treatise on Taxation* on “Taxes under the Power of Police.”²⁰⁶ Such a tax does not have “for its object the raising of revenue,” but rather “the conservation of order . . . , to the encouragement of industry, and the discouragement of pernicious employments.”²⁰⁷

A half century later Arthur C. Pigou, successor to Marshall’s professorship at Cambridge, devoted a chapter of his *Study in Public Finance* to “Taxes and Bounties to Correct Maladjustments.”²⁰⁸ These maladjustments were situations in which “the value of the marginal social net product falls short of the value of the marginal private net product when resources yield, besides the commodity which is sold and paid for, a dis-commodity for which those on whom it is inflicted are unable to

203. See discussion *infra* Part V.

204. 2 ADAM SMITH, AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS 338 (2d ed. 1778) (1776).

205. See, e.g., *Washington v. State*, 13 Ark. 752, 752 (1853) (tax on billiard parlors); *Wendover v. City of Lexington*, 15 B. Mon. 258, 262, 264 (Ky. App. 1854) (lottery tax); *Youngblood v. Sexton*, 32 Mich. 406, 413 (1875) (liquor); *City of Hannibal v. Guyott*, 18 Mo. 515, 516 (1853) (same); *J.F. Baker & Co. v. Panola Cty.*, 30 Tex. 86, 87 (1867) (same). For a fuller discussion, see HOVENKAMP, *supra* note 1, at 104–05.

206. THOMAS M. COOLEY, A TREATISE ON THE LAW OF TAXATION, INCLUDING THE LAW OF LOCAL ASSESSMENTS 396–415 (1876).

207. *Id.* at 396.

208. ARTHUR C. PIGOU, A STUDY IN PUBLIC FINANCE, Ch. 8 (2d Ed. 1929) (1928).

exact compensation.”²⁰⁹ The theory behind Pigouvian taxes was that the tax would impose a cost that would correct harmful externalities by making the harm causing activity less profitable, and thus reducing it to a safer point.²¹⁰ Meanwhile, the tax monies could be used to correct the harms.²¹¹ The trick is to increase the marginal cost of the harmful activity to the point that its marginal cost equals its marginal social benefit.²¹²

Pigou’s views were aggressively criticized by Ronald Coase, who argued in *The Problem of Social Cost* that the problem of divergences between “private” and “social” net product is really one of identifying who is hurting whom.²¹³ Further, Coase argued, private bargaining will address such situations, provided that transaction costs are sufficiently low.²¹⁴ The burdened and benefitted sides would bargain their way to an efficient result that would eliminate non-cost-justified harms. On the other hand, if a harm was cost justified the person imposing it would pay for it rather than stop.²¹⁵ *The Problem of Social Cost* was in large part an assault on Pigou.²¹⁶

Today Pigouvian taxation has experienced a revival in the academic literature, if not yet in policy.²¹⁷ The principal targets are carbon fuels and other emissions whose victims are not in a position to bargain with those who injure them.²¹⁸ That is to say, it is not so much that Coase was wrong

209. *Id.* at 119. He observed: “Thus, incidental uncharged disservices are rendered to third parties when the owner of a site in the residential quarter of a city builds a factory there and so destroys a great part of the amenities of neighbouring sites” *Id.*

210. See Jonathan S. Masur & Eric A. Posner, *Toward a Pigouvian State*, 164 U. PA. L. REV. 93, 100 (2015).

211. *Id.*

212. *Id.* at 100–01.

213. Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 44 (1960).

214. *Id.* at 15–18.

215. *Id.*

216. *Id.* at 28–29. Coase’s critique was in turn severely criticized in A.W. Brian Simpson, *Coase v. Pigou Reexamined*, 25 J. LEGAL STUD. 53, 53–97 (1996).

217. Masur & Posner, *supra* note 210, at 100–01; see also Louis Kaplow & Steven Shavell, *On the Superiority of Corrective Taxes to Quantity Regulation*, 4 AM. L. & ECON. REV. 1, 3–4 (2002) (discussing Pigouvian taxation); N. Gregory Mankiw, *Smart Taxes: An Open Invitation to Join the Pigou Club*, 35 E. ECON. J. 14, 14–15 (2009) (discussing the disagreement among economists on Pigouvian taxation); Steven Shavell, *Corrective Taxation Versus Liability as a Solution to the Problem of Harmful Externalities*, 54 J.L. & ECON. S249, S249–50 (2011) (discussing the limited application of Pigouvian taxation). One optimistic evaluation is CASS R. SUNSTEIN, *RISK AND REASON: SAFETY, LAW AND THE ENVIRONMENT* ix (2002). For a more pessimistic assessment of policy results so far, see Robert W. Hahn & Robert A. Ritz, *Does the Social Cost of Carbon Matter? Evidence from U.S. Policy*, 44 J. LEGAL STUD. 229, 232 (2015).

218. E.g., Evan N. Turgeon, *Triple-Dividends: Toward Pigovian Gasoline Taxation*, 30 J. LAND RESOURCES & ENVTL. L. 145, 146–49 (2010). On the metrics, see generally Shi-Ling Hsu, *A Complete Analysis of Carbon Taxation: Considering the Revenue Side*, 65 BUFF. L. REV. 857 (2017).

about the proposition that private bargaining could be a corrective if transaction costs were sufficiently low. Rather, it is that the presence of high transaction costs is in fact ubiquitous, and makes his observations irrelevant over a wide range of circumstances.

IV. MARKET DIVERSITY AND AGENCY SECTOR REGULATION

Market failure and market diversity go hand in hand. Markets are like Tolstoy's families. Happy families are all pretty much alike, but each unhappy family is unhappy in its own way.²¹⁹ When markets work well they require relatively little government attention other than the rules of property and contract, supervised by courts of general jurisdiction. But markets fail for reasons that are typically quite specific to each one. Even natural monopolies, while exhibiting a common set of characteristics, have unique features and information requirements that operate on the regulator.

An important consequence of the marginalist revolution was this idea that markets differ from one another, particularly when they fail. For a variety of reasons, they were thought to require different regulatory fixes. First, the cost classifications that occupied marginalists, including the emergent theory of regulated monopoly, not only made price and output analysis more complex, they also revealed that while firms within the same market and using the same technology might be similar, the firms in one market could have cost structures that were quite different from those in another market. Further, the relationship between fixed costs and market size varied enormously. A relatively small railroad operating within a single state might require price regulation, while giant petroleum or steel refiners might be made to operate competitively. Geographically large markets tend to require larger regulators—most notably the federal government if they spilled over more than a single state. Small markets were best controlled by smaller regulators—namely, states and municipalities. The diversity was not merely geographic, however. It was also production or technology specific. For example, the cost classification problems that might show up in electric power generation differed from those for natural gas production or telecommunications. Agencies at all levels of government with unique and exclusive jurisdictions were in large part a response to this perception of market diversity, replacing a nineteenth century conception in which most “regulating” was carried on by courts of general jurisdiction.

George Stigler once considered how economists from Adam Smith through Frank Knight, working over a century and a half, attempted to

219. See LEO TOLSTOY, *ANNA KARENINA* 1 (Joel Carmichael trans., Bantam Books 1960) (1877) (“Happy families are all alike; every unhappy family is unhappy in its own way.”). Thanks to Suzanna Sherry for this.

identify the prerequisites for perfect competition. He concluded such a list is impossible because the complete theory is “open ended”:

[I]t is always possible that a new range of problems will be posed in this framework, and then, no matter how well developed the theory was with respect to the earlier range of problems, it may require extensive elaboration in respects which previously it glossed over or ignored.²²⁰

What Stigler was able to harvest from these writers was three families of characteristics, each of which exhibited wide variations. The first is structural, pertaining to the number of actors in a market and the nature of their costs. Second is the ability of individual actors to make maximizing choices without excessive restraint. Third is the free flow of reliable, actionable information. Oddly, Stigler never mentioned externalities, or the inability of bargaining parties to internalize all the effects of their transactions. This is peculiar, because contemporaries such as Pigou and Coase were virtually obsessed with them. Paul Bator’s contemporary essay on market failure also commented on them at length,²²¹ and they have become a central concern in such as areas as environmental regulation. Perhaps for this reason, Stigler’s essay makes no mention of patents or other intellectual property rights. If anything, Stigler understated the varieties of market failure.

Market diversity and complexity invariably led to market-specific expertise and the rise of sector-specific regulatory agencies. While accountability in adjudication and rule making are big issues, the most fundamental conflict over the legitimacy of agency regulation concerns the extent and diversity of market failure. The Constitution was drafted long prior to the marginalist revolution and its increasing appreciation of market diversity. Within the framework acknowledged by the Constitution the principal regulators of the economy were Congress and the state and federal Courts. Even the IP Clause, which authorized the Patent Act,²²² did not expressly authorize the creation of an agency similar to the USPTO. Rather, it empowered *Congress* to “secur[e] for limited Times . . . the exclusive Right.”²²³ That language is more consistent with direct Congressional issuance of patents, just as state issued patents prior to the Constitution were typically issued directly by

220. George J. Stigler, *Perfect Competition, Historically Contemplated*, 65 J. POL. ECON. 1, 14 (1957).

221. Bator, *supra* note 170, at 371–75.

222. Ch. 7, 1 Stat. 109 (1790).

223. U.S. CONST. art. I, § 8 & cl. 8 (“The Congress shall have power . . . to promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”).

state legislatures.²²⁴ At the same time, however, the Framers had the foresight not to *prohibit* the creation and use of an intermediary such as an agency.

As discussed later, another problem leading to agency regulation was risk management.²²⁵ Once again, however, the nature and technical features of the relevant risks could vary widely from one technology or product to another. Classical valuations that depended on the past did not take risk into account because relevant risks were already encountered and included in the calculus. But risk became a significant, even overwhelming, part of valuation and optimization within the more forward looking marginalist framework. This is particularly true of activities that produce numerous situations where the market itself or the common law cannot force people to internalize the social harm their activities cause.

V. DEREGULATION

The deregulation movement forced reexamination of many issues in the debate over the economic rationales for government intervention. The movement began in government policy during the waning years of the Carter administration and accelerated during the 1980s.²²⁶ The initial impact of deregulation was substantial. By one measure, industries characterized by “full” regulation of price and entry produced about 17% of the GNP in 1977, but about 6.6% in 1988.²²⁷ The deregulation movement was significantly a reaction to excesses in the granting of monopoly status (entry restrictions) and needless price regulation. Then-Professor Stephen Breyer’s 1984 book, *Regulation and its Reform*, pointed out how legal entry restrictions and price regulation were imposed in industries such as trucking that were competitively structured or at least capable of being so.²²⁸ Breyer termed these instances of “regulatory mismatch.”²²⁹

224. See P.J. Federico, *State Patents*, 13 J. PAT. OFF. SOC’Y 166, 167–69 (1931); Hovenkamp, *supra* note 89, at 267–68.

225. See discussion *infra* Part VIII.

226. See, e.g., Motor Carrier Act of 1980, Pub. L. No. 96–296, 94 Stat. 793 (codified as amended in scattered sections of 49 U.S.C.) (removing entry restrictions and most rate regulation, except for the obligation to file a rate). A later regulation removed most obligations to file rates. See 49 U.S.C. § 13710 (2012). Rail rates were also largely deregulated. See, e.g., *id.* § 10502(a).

227. See W. KIP VISCUSI ET AL., *ECONOMICS OF REGULATION AND ANTITRUST* 305 (3d ed. 2000).

228. See STEPHEN BREYER, *REGULATION AND ITS REFORM* 222–39 (1982).

229. See *id.* at 247–48.

In addition, deregulation was a response to several market developments. An important one was changes in technology.²³⁰ For example, thanks to the development of wireless transmission and advanced switching capabilities, the technical justifications for a single natural monopoly telephone system disappeared. The real instigators were wireless firms such as MCI and Sprint, who convinced courts and the FCC to upset AT&T's strenuously defended resistance to interconnection.²³¹

More generally, it became clear that regulation could be a very costly enterprise in relation to any benefits that it produced. One well known early example of this critique is the evolution of James Landis, Chairman of the SEC during the New Deal. Landis initially lauded regulation as the savior of the economy,²³² but a generation later lamented that it had not come close to satisfying expectations.²³³

However, most of the low hanging fruit has already been picked. Further deregulation may threaten the environmental, health, or safety by amounts considerably greater than the cost savings themselves, depending on the industry.²³⁴ That is more likely to be the case as a greater amount of regulation is made subject to mandatory cost-benefit analysis, and as the methodologies of CBA have improved.²³⁵ If a regulation was fully and accurately evaluated and approved under CBA, then removing it by implication will cause more harm than good.

Contestability theory, which was widely touted as opening the path toward broad deregulation,²³⁶ has actually had much less to do with implementing deregulation than have more fundamental concerns about the values of incumbent competition and the economic inefficiency of regulation, particularly those expressed in the Averch-Johnson literature.²³⁷ When deregulation has moved markets from regulated monopoly to some sort of competition, the competition that emerged has for the most part been among incumbent firms, not monopoly incumbents

230. See IA PHILLIP E. AREEDA & HERBERT HOVENKAMP, *ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION* ¶ 241b1 (4th ed. 2013).

231. See generally, e.g., *MCI Commc'ns Corp. v. AT&T Co.*, 708 F.2d 1081 (7th Cir. 1981) (applying antitrust's essential facility doctrine to force AT&T to permit wireless carrier to connect into its system), *cert. denied*, 464 U.S. 891 (1983); *In re Carterfone*, 13 F.C.C.2d 420 (1968).

232. See JAMES M. LANDIS, *THE ADMINISTRATIVE PROCESS* (1938).

233. See JAMES M. LANDIS, *REPORT ON REGULATORY AGENCIES TO THE PRESIDENT-ELECT* 83–86 (1960) (proposing extensive reorganization plans); HOVENKAMP, *supra* note 1, at 308–09.

234. See Ed Dolan, *Is Overregulation Really Holding Back the U.S. Economy*, *HARV. BUS. REV.* (Jan. 8, 2018), <https://hbr.org/2018/01/is-overregulation-really-holding-back-the-u-s-economy> [<https://perma.cc/7D4B-222R>].

235. See discussion *infra* text accompanying notes 264–81.

236. See discussion *supra* text accompanying notes 226–34.

237. See discussion *supra* text accompanying notes 164–67.

who vied with one another for the right to occupy an exclusive position. In other words, the Williamson analysis of contestability, emphasizing the difficulties in the presence of costly and durable infrastructure, has been much more influential on actual policy than the Demsetz analysis.²³⁸

The greatest focus of deregulation has been on situations where competition among actual incumbents is possible and markets or political institutions can be designed in a way that will accommodate them. The earliest targets for deregulation, which included trucking and passenger airlines, certainly fell into that category. More networked industries eventually followed. For example, the dramatic increases in the amount of telecommunications competition since the 1982 antitrust consent decree breaking up the phone company²³⁹ and the Telecommunications Act of 1996²⁴⁰ have largely been increases in competition among multiple incumbent firms. Contestability theory has not had a significant role. AT&T's divestiture of Western Electric²⁴¹ led to a fiercely competitive market for devices among incumbent rivals, which exists to this day. For a lengthy period local telephony operated as a price-regulated monopoly, while long distance provision was competitively structured.²⁴² Today even local service is competitive, although the competition is typically between the resident ILEC,²⁴³ wireless, and one or more cable or other internet service providers that sell VOIP²⁴⁴ or perhaps a different technology. Looking at the economy as a whole, true contestability, where a single firm is the current seller in a market with significant infrastructure but must periodically re-bid for that right, is a rarity.

The other thing that has accompanied deregulation is a counterbalancing increase in the use of antitrust law.²⁴⁵ Under the comprehensive regulation formerly thought to govern industries such as the airlines, market wide antitrust immunity was often the rule.²⁴⁶ That is hardly the case today. Substantially, this is the result of a paradigm shift

238. See *supra* text accompanying notes 64–76. For the Demsetz analysis, see generally Demsetz, *supra* note 70. For the Williamson analysis, see generally Williamson, *supra* note 71.

239. See *United States v. AT&T Co.*, 552 F. Supp. 131, 222–25 (D.D.C. 1982), *aff'd sub nom. Maryland v. United States*, 460 U.S. 1001 (1983).

240. Pub. L. No. 104-104, 110 Stat. 56 (1996) (codified as amended in scattered sections of 47 U.S.C.).

241. See *United States v. W. Elec. Co.*, 604 F. Supp. 256, 261 n.23 (D.D.C. 1984), *cert. denied*, 480 U.S. 922 (1987).

242. See *AT&T Corp. v. Iowa Utils. Bd.*, 525 U.S. 366, 414–16 (1999).

243. Incumbent Local Exchange Carrier, or the firm designated as the incumbent carrier in a given region.

244. Voice Over Internet Protocol, which permits real time voice transmission on the internet.

245. See AREEDA & HOVENKAMP, *supra* note 230, ¶ 241.

246. *E.g.*, *Hughes Tool Co. v. Trans World Airlines, Inc.*, 409 U.S. 363, 380 (1973); *Pan Am. World Airways, Inc. v. United States*, 371 U.S. 296, 305, 309 (1963).

in the theory of regulation, away from the control of markets and toward the control of practices. In most cases of deregulation, regulators do not simply exit from a market. Rather, they apply regulatory tools more selectively.

During the 1960s we tended to think of federal agency regulation as “pervasive,” in the sense that it left little discretion over pricing and exclusionary practices to the individual firm, and when these occurred they were usually placed within the jurisdiction of the regulatory agency.²⁴⁷ Today, however, we view regulation as more “transactional,” in that it applies to a particular practice rather than an entire industry. When a specific transaction is within the jurisdiction of the agency and the agency is actually overseeing it, then antitrust has no role to play. But antitrust can be brought to bear in those areas that regulators do not tend.²⁴⁸ This hybrid approach is particularly descriptive of telecommunications, the airlines, healthcare, and to a lesser extent energy.²⁴⁹

VI. WEALTH DISTRIBUTION

The issues swirling around regulation and wealth distribution tend to focus on two concerns. The first is whether government control should be used to redistribute wealth for its own sake, simply because we regard certain distributions as unfair or we think social welfare is greater by some measure as wealth is more evenly distributed. The other is whether maldistribution of wealth affects productivity, and thus redistribution may spur economic growth. The latter one falls within modern neoclassical concerns about regulation. Under most conceptions of economic welfare, the former one does not.

For the first two generations of marginalists, concerns about the distribution of wealth in the economy were paramount. Further, since most did not expect that maldistribution of wealth would right itself

247. See AREEDA & HOVENKAMP, *supra* note 230, ¶ 240c1.

248. See, e.g., *Nat'l Gerimedical Hosp. & Gerontology Ctr. v. Blue Cross*, 452 U.S. 378, 389 (1981).

[W]here Congress did intend to repeal the antitrust laws, that intent governs, but this intent must be clear. Even when an industry is regulated substantially, this does not necessarily evidence an intent to repeal the antitrust laws with respect to every action taken within the industry. Intent to repeal the antitrust laws is much clearer when a regulatory agency has been empowered to authorize or require the type of conduct under antitrust challenge.

Id. (citations omitted).

249. See, e.g., *Oneok, Inc. v. Learjet, Inc.*, 135 S. Ct. 1591, 1594 (2015) (holding Natural Gas Act did not preempt application of state antitrust law to claim of anticompetitive price manipulation).

through ordinary market forces, some kind of government intervention was thought necessary.

The early marginalists believed in declining marginal utility, but also in the policy value of interpersonal utility comparisons. All things equal, a more even distribution of wealth produced greater welfare. Individuals were thought to experience declining marginal utility for any good, including wealth, as they have more. The leap the early marginalists were willing to make was that this attribute applied across persons as well as to a single individual. For example, Marshall wrote in the third edition of *Principles* that “a pound’s worth of satisfaction to an ordinary poor man is a much greater thing than a pound’s worth of satisfaction to an ordinary rich man.”²⁵⁰ Marshall’s successor at Cambridge, Arthur C. Pigou, agreed: a “transference of income from a relatively rich man to a relatively poor man of similar temperament, since it enables more intense wants to be satisfied at the expense of less intense wants, must increase the aggregate sum of satisfaction.”²⁵¹ Many early American marginalists also agreed. Chicago economist Jacob Viner embraced the most significant consequence of interpersonal utility comparisons in 1925—namely, that “[c]hanges in the relative distribution of income as between different classes will bring about changes in the amount of welfare, even though the aggregate real income of the community remains the same.”²⁵² Within this framework, a social welfare scheme that transferred wealth away from the wealthy and toward the poor increased aggregate welfare to the extent that the wealthy experienced less welfare from a given sum of money than the poor did.

In the mid-thirties Lord Lionel Robbins exploded this argument in his *Essay on the Nature and Significance of Economic Science*. He argued that the proposition that one person experiences the same utility as another from a given amount of wealth was scientifically meaningless because it could not be tested.²⁵³ In a widely quoted passage Robbins reasoned:

[S]uppose that we differed about the satisfaction derived by A from an income of £1,000, and the satisfaction derived by B from an income of twice that magnitude. Asking them

250. ALFRED MARSHALL, *PRINCIPLES OF ECONOMICS* 206 (3d ed. 1895).

251. ARTHUR C. PIGOU, *THE ECONOMICS OF WELFARE* 89 (4th ed. 1938).

252. Jacob Viner, *The Utility Concept in Value Theory and Its Critics*, 33 J. POL. ECON. 368, 644 (1925); see also FRANK WILLIAM TAUSSIG, *PRINCIPLES OF ECONOMICS* 132 (3d ed. 1921) (“[I]nequality of incomes brings a less sum of human well-being than equality of incomes . . .”); John Bates Clark, *The Ultimate Standard of Value*, 1 YALE REV. 258, 258 (1893) (discussing the social nature of value); Simon N. Patten, *The Scope of Political Economy*, 2 YALE REV. 264, 265 (1894) (similar).

253. See LIONEL ROBBINS, *AN ESSAY ON THE NATURE AND SIGNIFICANCE OF ECONOMIC SCIENCE* 136–43 (2d ed. 1935); HOVENKAMP, *supra* note 1, at 111–12.

would provide no solution. Supposing they differed. A might urge that he had more satisfaction than B at the margin. While B might urge that, on the contrary, he had more satisfaction than A. We do not need to be slavish behaviourists to realise that here is no scientific evidence. There is no means of testing the magnitude of A's satisfaction as compared with B's. If we tested the state of their blood-streams, that would be a test of blood, not satisfaction. Introspection does not enable A to discover what is going on in B's mind, nor B to discover what is going on in A's. There is no way of comparing the satisfactions of different people.²⁵⁴

Given his assumptions, everything that Robbins said in this passage is very likely true. Nevertheless, it did not end the debate over interpersonal utility comparisons. First, Robbins was writing in a positivistic tradition obsessed with limitations on scientific inquiry into state of mind.²⁵⁵ He simply assumed that the only "scientific" economic inquiry into welfare accordingly concerned mental state.²⁵⁶ For behaviorists and other social scientists less prone to recognize a hard mind-body dichotomy, these arguments were not persuasive.²⁵⁷ If welfare for policy purposes is estimated objectively, by looking at such external factors as material wealth, health, education, nutrition, shelter, productivity, or numerous other external indicia, then interpersonal comparability is readily possible. As a matter of government policy, objective welfare judgments may be superior in any event because they more readily permit generalization about large numbers.²⁵⁸

Second, Robbins's analysis disregarded the relationship between wealth and human *production* functions.²⁵⁹ If a transfer from a rich person to a poor person also increases productivity by taking wealth from a person for whom it has little marginal impact and giving it to someone who becomes much more productive, then the transfer may increase welfare by increasing economic growth. This makes the transfer welfare positive even on traditional neoclassical assumptions that avoid interpersonal utility comparisons.

One effect of the welfarism debate was that for some time it created a sharp divide between neoclassical welfare economics and more applied

254. ROBBINS, *supra* note 253, at 139–40 (emphasis omitted).

255. See HOVENKAMP, *supra* note 1, at 111–12.

256. *Id.* at 112.

257. See Robert Cooter & Peter Rappoport, *Were the Ordinalists Wrong About Welfare Economics?*, 22 J. ECON. LITERATURE 507, 523 (1984); Hovenkamp, *supra* note 57, at 521–25.

258. For a more-or-less contemporary critic on this point, see generally I.M.D. LITTLE, *A CRITIQUE OF WELFARE ECONOMICS* (1950).

259. See HOVENKAMP, *supra* note 1, at 112.

branches of economics, including state economic policy making. The same issue also divided neoclassical welfare economics from most of the other social sciences.²⁶⁰ In neoclassical economics textbooks, authors talked about “mere” wealth transfers as if economics should be indifferent to them. The real concern of economics was said to be allocative efficiency, typically measured by Paretian criteria, and the deadweight loss caused by monopoly. By contrast, economic policy during the New Deal and after took the distribution of wealth very seriously and developed important governmental programs designed to give effect to these concerns.²⁶¹ Even in public utility law, the goal of universal service for such things as electricity and telephone might be viewed as distributive, at least for those customers who cannot even pay the incremental (variable) costs of their service.²⁶²

Today, it seems safe to say, concerns about distribution have dramatically re-entered the picture, although the concerns tend to focus on productivity rather than on Robbins’s states of mind.²⁶³

VII. COST-BENEFIT ANALYSIS

The fundamental idea of cost-benefit analysis (CBA) is as old as Anglo-American thought’s first marginalists. In the late eighteenth century Jeremy Bentham proposed his “principle of utility,” which would evaluate government policy by querying whether its “tendency . . . to augment the happiness of the community is greater than any it has to diminish it.”²⁶⁴ To be sure, this was not a modern theory of CBA. Utilitarianism would not make the jump from political philosophy to

260. See Herbert Hovenkamp, *Knowledge About Welfare: Legal Realism and the Separation of Law and Economics*, 84 MINN. L. REV. 805, 810 (2000).

261. See HOVENKAMP, *supra* note 1, at 112–14.

262. On the development of the concept, see generally Herbert S. Dordick, *The Origins of Universal Service: History as a Determinant of Telecommunications Policy*, 14 TELECOMM. POL’Y 223 (1990). On telephone service at prices lower than variable costs, see Milton Mueller, *Universal Service in Telephone History: A Reconstruction*, 18 TELECOMM. POL’Y 352, 359 (1993).

263. For discussions on this, see generally, for example, JONATHAN D. OSTRY ET AL., IMF, REDISTRIBUTION, INEQUALITY AND GROWTH (2014); INEQUALITY AND GROWTH: PATTERNS AND POLICY (Kaushik Basu & Joseph E. Stiglitz eds., 2016); Sutirtha Bagchi & Jan Svejnar, *Does Wealth Inequality Matter for Growth? The Effect of Billionaire Wealth, Income Distribution, and Poverty*, 43 J. COMP. ECON. 505, 506 (2015); and Federico Cingano, *Trends in Income Inequality and Its Impact on Economic Growth* (Org. for Econ. Co-operation & Dev., Working Paper No. 163, 2014), <https://www.oecd-ilibrary.org/docserver/5jxrjncwvxv6j-en.pdf?expires=1526298689&id=id&acname=guest&checksum=237F380132F4FA78141554ECACA903D> [https://perma.cc/F3Y3-DKBY].

264. JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLE OF MORALS AND LEGISLATION 3 (1781) (“An action then may be said to be conformable to the principle of utility . . . when the tendency it has to augment the happiness of the community is greater than any it has to diminish it.”).

market economics for another century.²⁶⁵ Bentham was speaking of utility as a fairly abstract conception of “happiness,” or pain and pleasure, not as a tradeable standard of market exchange.

Bentham’s American contemporary, Benjamin Franklin, was also an important precursor. In a 1772 letter to Joseph Priestley which Franklin titled “Moral Algebra, or Method of Deciding doubtful Matters with oneself,” he offered this calculus of decision making:

[M]y way is, to divide half a sheet of paper by a line into two columns: writing over the one pro, and over the other con: then during three or four days consideration, I put down under the different heads, short hints of the different motives that at different time occur to me, for or against the measure. When I have thus got them altogether in one view, I endeavor to estimate their respective weights, and where I find two, (one on each side) that seem equal, I strike them both out. If I find a reason pro equal to some two reasons con, equal to some three reasons pro, I strike out the five; and thus proceeding, I find at length where the balance lies; and if after a day or two of further consideration, nothing new that is of importance occurs on either side, I come to a determination accordingly.²⁶⁶

The context suggests that Franklin was thinking not so much about social economic welfare as about personal decision making,²⁶⁷ but the methodology was similar nonetheless. Franklin also considered that under his methodology “the weight of reasons cannot be taken with the precision of algebraic quantities,” if nevertheless made him “less liable to make a rash step.”²⁶⁸

The modern theory of CBA grew more directly out of the early neoclassical debate over wealth distribution and interpersonal utility comparisons. In the wake of the Lionel Robbins’s critique noted above,²⁶⁹ Robbins’ younger colleague Nicholas Kaldor, and a little later John Hicks, proposed variations on the Pareto principle for assessing welfare without making interpersonal comparisons of utility or (in their view)

265. See HOVENKAMP, *supra* note 1, at 28–29.

266. Letter from Benjamin Franklin to Joseph Priestley (Sept. 19, 1772), in I THE PRIVATE CORRESPONDENCE OF BENJAMIN FRANKLIN, 19–20 (William Temple Franklin ed., 1817) [hereinafter Franklin Letter].

267. The letter was in Leeds in response to a query from Priestley about whether Priestley should give up his comfortable life in Leeds in order to become librarian to Lord Shelburne and tutor of his children, a lucrative but demanding position. Priestley ended up accepting the position. See I JOHN T. RUTT, LIFE AND CORRESPONDENCE OF JOSEPH PRIESTLEY 180–81 (1831).

268. Franklin Letter, *supra* note 266, at 20.

269. See *supra* text accompanying notes 254–63.

other judgments about fairness.²⁷⁰ What became the Kaldor-Hicks criterion of welfare was that a change is welfare superior if the marginal gains experienced by gainers are large enough that the gainers could compensate the marginal losers entirely out of their gains, thus leaving them indifferent.²⁷¹ Actual compensation is not required. This more-or-less became the question that CBA was designed to answer. No idea in the theory of regulation is more closely tied to marginalism than CBA, which queries whether the marginal social benefits of any proposed change in a rule or practice are greater or less than the marginal social costs.²⁷² The development of CBA was thus closely aligned with the analysis of social welfare functions in neoclassical welfare economics.²⁷³

The history of CBA as applied to regulatory problems is far too vast to be described here.²⁷⁴ Briefly, however, at least since President Reagan's 1981 Executive Order Number 12,291²⁷⁵ it has played a prominent role in both the creation and the legal analysis of economic

270. For the Hicks variation, see generally, for example, J. R. Hicks, *The Foundations of Welfare Economics*, 49 *ECON. J.* 696 (1939). For the Kaldor variation, see generally, for example, Nicholas Kaldor, *Welfare Propositions of Economics and Interpersonal Comparisons of Utility*, 49 *ECON. J.* 549 (1939).

271. Such a change would not be Pareto superior because it contemplates at least one loser, but it would be "potential" Pareto superior to the extent that the gainers would be able to compensate the losers sufficiently to make them indifferent.

272. See BENTHAM, *supra* note 264, at 3–4.

273. See Eric A. Posner & Cass R. Sunstein, *Moral Commitments in Cost-Benefit Analysis*, 103 *VA. L. REV.* 1809, 1819–20 (2017) (tracing out the history).

274. For a discussion of the CBA in this context, see generally MATTHEW D. ADLER & ERIC A. POSNER, *COST-BENEFIT ANALYSIS: LEGAL, ECONOMIC, AND PHILOSOPHICAL PERSPECTIVES* (2001); ANTHONY E. BOARDMAN ET AL., *COST-BENEFIT ANALYSIS: CONCEPTS AND PRACTICE* (5th ed. 2018); CASS R. SUNSTEIN, *THE COST-BENEFIT REVOLUTION* (2018); Don B. Hardin, Jr., *Why Cost-Benefit Analysis? A Question (And Some Answers) About the Legal Academy*, 59 *ALA. L. REV.* 1135, 1140–48 (2008); Cass R. Sunstein, *The Real World of Cost-Benefit Analysis: Thirty-Six Questions (and Almost as Many Answers)*, 114 *COLUM. L. REV.* 167, 168–70 (2014); and Richard O. Zerbe, *The Legal Foundation of Cost-Benefit Analysis*, 2 *CHARLESTON L. REV.* 93, 100–11 (2007).

275. Exec. Order No. 12,291, 46 *Fed. Reg.* 13,193 (Feb. 17, 1981) (ordering that a regulation not be implemented "unless the potential benefits to society . . . outweigh the potential costs"); see also Exec. Order No. 13,563, 3 *C.F.R.* § 215 (2012), *reprinted in* 5 *U.S.C.* § 601 (2012) (ordering a listing of relevant costs and benefits for proposed regulations, even when the statute in question contemplates only the assessment of benefits). President Trump's only relevant executive order as of this writing is inconsistent with CBA. It requires an agency to eliminate two regulations for every one that it adopts. See Exec. Order No. 13,771, 82 *Fed. Reg.* 9339 (Jan. 30, 2017) ("Unless prohibited by law, whenever an executive department or agency (agency) publicly proposes for notice and comment or otherwise promulgates a new regulation, it shall identify at least two existing regulations to be repealed."). To the extent that order is indifferent to the scope or economic weight of a regulation or to its beneficial or harmful effects, it appears to have nothing to do with the economic or social impact of regulations.

regulation.²⁷⁶ At the same time, it remains subject to very fundamental disputes about such things as whether neoclassical rational actor assumptions should control all of CBA, or whether more behavioral analysis or other frameworks that contemplate evaluations outside of neoclassical norms should be considered.²⁷⁷ Another dispute is whether willingness-to-accept estimates of consumer value should be used in addition to or instead of willingness-to-pay evaluations.²⁷⁸ Another is how finely tuned CBA must be when used in actual agency rule making, or indeed, whether costs should be considered at all when certain benefits are technologically feasible or a statute requires consideration only of technical feasibility.²⁷⁹ In addition are more technical issues pertaining to risk and discount rates, spillovers,²⁸⁰ and quantification of things that are not readily subject to observable market trades, such as human life or dignity.²⁸¹

VIII. ACTUARIAL CONCEPTIONS OF VALUE AND MANAGEMENT OF RISK

The marginalist revolution in economics led to two fundamental shifts in thinking about economic value. One was the relationship between schedules of demand and the various types of costs that determine output and price. This shift accounts for previously discussed issues of regulatory policy such as natural monopoly and cost-of-service ratemaking.²⁸²

The other shift was in forward rather than backward looking conceptions of value—a change that accounted for the development of modern cost-benefit analysis, which is largely focused on expected value.²⁸³ That shift brought much needed realism to the commercial and legal system, although its eventual impact reached much further. For example, the classical notion that the value of a corporation is determined

276. See David Pearce, *Cost-Benefit Analysis and Environmental Policy*, 14 OXFORD REV. ECON. POL'Y 84, 87–88 (1998) (explaining the history of the use of CBA by U.S. regulators).

277. See, e.g., Brian F. Mannix, *Benefit-Cost Analysis as a Check on Administrative Discretion*, 24 SUP. CT. ECON. REV. 155, 164 (2016).

278. E.g., John Bronsteen et al., *Well-Being Analysis vs. Cost-Benefit Analysis*, 62 DUKE L.J. 1603, 1645–46 (2013); Cass R. Sunstein, *Cost-Benefit Default Principles*, 99 MICH. L. REV. 1651, 1721 (2001).

279. E.g., *Entergy Corp. v. Riverkeeper, Inc.*, 556 U.S. 208, 235 (2009) (suggesting that avoidance of “extreme disparities” is sufficient).

280. E.g., Sarah E. Light, *Valuing National Security: Climate Change, the Military, and Society*, 61 UCLA L. REV. 1772, 1809 (2014).

281. See, e.g., Richard L. Revesz, *Environmental Regulation, Cost-Benefit Analysis, and the Discounting of Human Lives*, 99 COLUM. L. REV. 941, 943 (1999) (noting that often a critical component in CBA of environmental regulation is the value of affected human lives).

282. See *supra* text accompanying notes 77–167.

283. See *supra* text accompanying notes 28–29.

by the amount of historically contributed capital bore no useful relationship to the value of a firm in the eyes of prospective shareholders or creditors.²⁸⁴ The only date on which a firm could meaningfully be said to be worth its paid in value was the day it commenced operations. After that its value could go either up or down, and often in a relatively short period of time. Some firms that started out with little capital became enormous and extravagantly wealthy. Others, which were much better financed to begin with, soon went broke. It all depended on luck, managerial success, the market success of new products or services, and the comparative success or failures of rivals. As firms became more complex and took up a greater portion of the economy, the backward system of valuation led to a crisis in the law of corporate finance: legal valuation and economic valuation, including the determinants of stock prices, had virtually nothing to do with each other.²⁸⁵

While the marginalist theory of corporate valuation was a significant improvement, reflecting market realities, it enormously complicated decisions about valuation. Under classical corporate finance theory, a judge could determine whether stock was “watered”—i.e., whether its capital was overstated—by doing some arithmetic. Contributions of noncash property complicated the calculations, but even these were within the experience of a commercial judge accustomed to settling estates or determining lost profits. But basing the value of a firm on its market prospects was much more difficult, because it included risk and uncertainty about a large variety of factors.²⁸⁶ The value of a firm became a composite answer to questions about the future value of a firm’s products, the quality of its management, its capacity to manufacture, various anticipated headwinds in sources of supply and labor relations, the number and anticipated health of present and future rivals, the value and remaining life of intellectual property portfolios, the expected health of the economy overall, and expectations about the national or perhaps world physical and regulatory environment. Thus the modern 10-K and reports to shareholders.

In addition, actuarial conceptions of risk, which depend entirely on marginalist tools, facilitated significant changes in the theory of firm risk

284. See *infra* text accompanying notes 285–87.

285. See HOVENKAMP, *supra* note 1, at 159–62.

286. Under Frank Knight’s terminology a “risk” occurs when the occurrence of a future event is unknown but can be calculated as a probability, such as the future toss of dice. By contrast, “uncertainty” deals with the occurrence of future events whose probability cannot be calculated. See FRANK H. KNIGHT, RISK, UNCERTAINTY, AND PROFIT 231–32 (1921); Milton Friedman & L.J. Savage, *The Utility Analysis of Choices Involving Risk*, 56 J. POL. ECON. 279, 279 (1948).

bearing as well as insurance theory and law.²⁸⁷ The common law itself began to adopt more actuarial, or risk driven, conceptions of legal duties that governed things such as long-term (relational) contracting and the implementation of negligence and products liability rules in tort.²⁸⁸

These developments began early in the history of marginalist thought and have had staggering implications for regulation. The science of risk management has essentially developed into economically sophisticated and technical private and public branches. Private risk management refers mainly to how firms make investment or management decisions in the presence of risk. In *The Nature of the Firm* (1937), Ronald Coase tied risk assessment into his theory of firm size and structure. For example, minimizing risk might be accomplished by contracting for supply from knowledgeable outsiders rather than attempting to make something for oneself.²⁸⁹ The public branch of risk management is devoted mainly to problems of externalities or spillovers where government intervention is thought necessary.²⁹⁰ These are most likely to occur when one person's or firm's activities cause harm to someone else that the market does not internalize.

The balance of this Article briefly examines two areas in which risk management has played an important role in the design of regulatory policy, but there are many, many others.

287. See generally ALLAN H. WILLET, *ECONOMIC THEORY OF RISK AND INSURANCE* (1901) (discussing risk and insurance). The Casualty Actuarial Society was created in 1914 as a professional association of insurance actuaries. Then-Professor, later Justice, William O. Douglas became one of the first legal scholars to write about the administration of risk. See, e.g., William O. Douglas, *Vicarious Liability and Administration of Risk I*, 38 *YALE L.J.* 584 (1929); William O. Douglas, *Vicarious Liability and Administration of Risk II*, 38 *YALE L.J.* 720 (1929).

288. See HOVENKAMP, *supra* note 1, at 123–55.

289. See Coase, *supra* note 13, at 398–401, much of which is taken from KNIGHT, *supra* note 286. See also Coase, *supra* note 13 (theorizing that the attitude toward risk may explain preference between short and long term contracting). On private risk management and tort exposure, see Robert E. Keeton, *Conditional Fault in the Law of Torts*, 72 *HARV. L. REV.* 401 (1959)

290. On management of public risks, see generally 1 *THE PRESIDENTIAL/CONG. COMM'N ON RISK ASSESSMENT & RISK MGMT., FRAMEWORK FOR ENVIRONMENTAL HEALTH RISK MANAGEMENT* (1997), http://oaspub.epa.gov/eims/eimscomm.getfile?p_download_id=36372 [<https://perma.cc/7LG4-3PP9>]. The Federal Aviation Administration, which is in the U.S. Department of Transportation, also maintains a handbook that describes the development and implementation of safety risk management processes. See *FED. AVIATION ADMIN., SYSTEM SAFETY HANDBOOK* (2000), https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/risk_management/ss_handbook/ [<https://perma.cc/WQ63-JAEE>]. On private, or entrepreneurial risk management, see generally JAMES LAM, *ENTERPRISE RISK MANAGEMENT: FROM INCENTIVES TO CONTROLS* (2d ed. 2014) and ROBERT JARROW, *THE ECONOMIC FOUNDATIONS OF RISK MANAGEMENT: THEORY, PRACTICE, AND APPLICATIONS* (2016).

A. Holmes's Marginalism: Torts as Risk Management

The first marginalist legal scholar in the United States was Oliver Wendell Holmes, Jr., who devoted much of his academic writing to risk management, particularly within the common law legal system. Much has been written about the influence of Darwin on Holmes's thought, but much less about Holmes's marginalism, the other Victorian science that certainly had a more significant impact on his thinking.²⁹¹ This was already reflected in Holmes's lectures published as *The Common Law* (1881).²⁹² In his chapter on tort law he wrote that "the safest way to secure care is to throw the risk upon the person who decides what precautions shall be taken."²⁹³ Further, the risk must generally be assigned to the person in control of outcomes.²⁹⁴ He also defended aggressive rules for highly dangerous conduct that "throw the risk upon the party pursuing it."²⁹⁵ In contract law he argued that contracts should be viewed as "the taking of a risk" and a set of bets about the future. Consequential damages were not appropriate "unless the assumption of that risk is to be taken as having fairly entered into the contract."²⁹⁶ While Holmes was not an economically technical marginalist, many of these ideas eventually worked their way into the law and economics movement.

Relatedly, Holmes's external standard became a way of conforming individual activity to social norms so as to minimize risk. The legal system cannot assess states of mind; it can only control behavior.²⁹⁷ For that, Holmes posited the "average" person in "temperament, intellect and education"²⁹⁸ as the common law norm. Conduct that fell below that norm could rightfully be condemned as negligent. For example, "when the

291. On Holmes's marginalism, see HOVENKAMP, *supra* note 1, at 38–41.

292. See generally O. W. HOLMES, JR., *THE COMMON LAW* (1881) (showing the influence of marginalism on Holmes).

293. *Id.* at 117.

294. *Id.* at 149.

295. *Id.*

296. *Id.* at 300–05.

297. *E.g., id.* at 49 (noting that even criminal law is indifferent to states of mind; rather its purpose is "to induce external conformity to rule"); *id.* at 62–63 ("The charge of malice aforethought in an indictment for murder has been shown not to mean a state of the defendant's mind, as is often thought, except in the sense that he knew circumstances which did in fact make his conduct dangerous."); *id.* at 136 ("[R]ecklessly' . . . does not mean actual personal indifference to the truth . . . [Rather,] [i]t means only that the data for the statement were so far insufficient that a prudent man could not have made it without leading to the inference that he was indifferent. . . . [I]f a man makes his statement on those data, he is liable, whatever was the state of his mind . . ."). On culpability, see *id.* at 146. "Foresight is a possible common denominator of wrongs at the two extremes of malice and negligence." *Id.* On murder, see *id.* at 53–54. If an act is of such a nature that it would likely cause death "the law will not inquire whether he [i.e., the accused] did actually foresee the consequences or not." *Id.*

298. *Id.* at 108.

question of the defendant's negligence is left to a jury, negligence does not mean the actual state of the defendant's mind, but a failure to act as a prudent man of average intelligence would have done."²⁹⁹

What Holmes did not fully articulate is that he was turning the "private" common law into a social control device. Holmes worked out the skeleton of a system in which deterrence at the margin became the goal of judge made legal policy, with his hypothetical "average man" as the determinant of the standard. Further, it was forward looking. Holmes repeatedly emphasized that the operative characteristic of his average person was reasonable foresight. "[I]f the intervening events are of such a kind that no foresight could have been expected to look out for them, the defendant is not to blame for having failed to do so," Holmes wrote about proximate cause in tort law.³⁰⁰

Judge Learned Hand's objective cost-benefit test for negligence was built entirely on these ideas, adding quantification to foresight.³⁰¹ That test—that an action is negligent if the marginal cost of the untaken precaution would have been less than the marginal cost of the accident it would have prevented (probability \times magnitude)—represented the triumph of pure marginalism in tort law.³⁰² As one pair of authors articulate the test: "The marginal Hand rule states that the injurer is negligent if the marginal cost of his or her precaution is less than the resulting marginal benefit. Thus, the injurer is liable under the Hand rule when further precaution is cost-justified."³⁰³

Today it has become conventional to regard the *United States v. Carroll Towing Co.*³⁰⁴ test for negligence as requiring a comparison of the marginal cost of precautions against the anticipated social cost of an accident. At the margin, an efficient actor who internalizes all these costs

299. *Id.* at 112. On criminal law, see *id.* at 54. "The test of foresight is not what this very criminal foresaw, but what a man of reasonable prudence would have foreseen." *Id.*

300. *Id.* at 92; see also *id.* at 147 ("Negligence is not foresight, but precisely the want of it; and if foresight were presumed, the ground of the presumption, and therefore the essential element, would be the knowledge of facts which made foresight possible."); *id.* at 130–31 (necessity to show foresight in order to establish criminal intent); *id.* at 57 ("A harmful act is only excused on the ground that the party neither did foresee, nor could with proper care have foreseen harm."). See generally HOVENKAMP, *supra* note 1, at 123–43 (tracing the development of tort law and the doctrine of proximate cause); Nicholas St. John Green, *Proximate and Remote Cause*, 4 AM. L. REV. 201 (1870) (discussing the definition and application of proximate cause).

301. See *United States v. Carroll Towing Co.*, 159 F.2d 169, 173 (2d Cir. 1947), which was largely adopted by Guido Calabresi in *THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS* 24–26 (1970), although couched in an argument that favored strict liability.

302. On the commonly accepted view that the relevant costs are marginal, see WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* 87 (1987).

303. ROBERT COOTER & THOMAS ULEN, *LAW AND ECONOMICS* 214–15 (6th ed. 2012).

304. 159 F.2d 169 (2d Cir. 1947).

will equate the two.³⁰⁵ That approach has also been substantially written into the test for dangerous and defective products given in the *Restatement (Third) of Products Liability* that “[a] product . . . is defective in design when the foreseeable risks of harm posed by the product could have been reduced or avoided by the adoption of a reasonable alternative design.”³⁰⁶ That approach creates a preference for negligence as a product design rule, although the Restatement (Third) retains a strict liability rule for manufacturing defects, where the costs of establishing negligence are much higher.³⁰⁷

This approach effectively turns product liability tort law into an exercise in risk management and cost-benefit analysis. As a common law rule, of course, it does not “regulate” in the command and control sense, but generally imposes liability after the fact. However, federal agencies such as the Consumer Product Safety Commission (CPSC) operate under the same standard for product safety requirements that are imposed *ex ante*.³⁰⁸ The more salient fact is that this rule, much like the risk-utility test used in tort cases, incorporates regulatory principles into the common law, using formulations that compare the marginal cost of precautions against marginal benefits. In the words of one scholar of products liability, that is “*the level at which the marginal cost of the investment equals the marginal cost of product-related accidents thereby avoided.*”³⁰⁹

B. Innovation Risk and PTAB Patent Review

Innovation often requires firms to take on a great deal of cost, risk, and uncertainty in contemplation of significant reward. A well designed patent system should manage these risks and rewards effectively. The

305. See, e.g., HOVENKAMP, *supra* note 1, at 147; LANDES & POSNER, *supra* note 302; Robert Cooter, *Unity in Tort, Contract, and Property: The Model of Precaution*, 73 CAL. L. REV. 1, 7, 7 n.7 (1985); Allan M. Feldman & Jeonghyun Kim, *The Hand Rule and United States v. Carroll Towing Co. Reconsidered*, 7 AM. L. ECON. REV. 523, 524 (2005).

306. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) (AM. LAW INST. 1998); see also Aaron D. Twerski & James A. Henderson, Jr., *Manufacturers’ Liability for Defective Product Designs: The Triumph of Risk-Utility*, 74 BROOK. L. REV. 1061, 1065 (2009).

307. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(a).

308. See 15 U.S.C. § 2058(f)(2) (2012) (prohibiting the CPSC from promulgating a product safety rule “unless it has prepared . . . a final regulatory analysis of the rule containing . . . [a] description of the potential benefits and potential costs of the rule”); see also *Zen Magnets, LLC v. Consumer Prod. Safety Comm’n*, 841 F.3d 1141, 1148 (10th Cir. 2016) (disapproving a CPSC rule where cost-benefit analysis was unsupported); cf. *United States v. Atl. Richfield Co.*, 429 F. Supp. 830, 832–33, 836 (E.D. Pa. 1977) (interpreting 33 U.S.C. § 1321 (Supp. 1976), a federal standard for determining liability for oil spills, as creating a marginal *Carroll Towing* test).

309. James A. Henderson, Jr., *Product Liability and the Passage of Time: The Imprisonment of Corporate Rationality*, 58 N.Y.U. L. REV. 765, 765 (1983).

consensus is strong that a patent system is necessary to facilitate innovation,³¹⁰ and of course the Constitution expressly authorizes its creation.³¹¹ The appropriate question is not whether to regulate, but rather how to regulate. If the marginal exclusion created by the patent system is too small, too little will be invested in innovation. On the other hand, if the patent system over-excludes, it acts as a clog on the flow of technology. The trick is to find the sweet spot between excessive and inadequate protection.³¹²

Considered in isolation, the initial system under which patent applications are evaluated and granted suggests significant deficiencies. First, it is largely *ex parte*, which means that the examiner hears almost exclusively from proponents of the patent.³¹³ Second, patent examiners do not have the time or resources to give each patent application adequate consideration. On average a patent receives 19 or fewer hours of examination.³¹⁴ The average number of claims in a patent hovers at around 16–18,³¹⁵ which suggests that examiners spend about an hour per claim evaluating patents, disregarding the time they need to spend on other portions of the patent application. As a result, too many invalid patents are granted. When patent validity is litigated in an adversarial setting, where considerably more resources are put into the analysis, as

310. Strong, but not unanimous. For one of the exceptions, see generally MICHELE BOLDRIN & DAVID K. LEVINE, *AGAINST INTELLECTUAL MONOPOLY* (2008) (arguing that patent and copyright be largely abolished).

311. U.S. CONST. art. I, § 8, cl. 8.

312. See generally CHRISTINA BOHANNAN & HERBERT HOVENKAMP, *CREATION WITHOUT RESTRAINT: PROMOTING LIBERTY AND RIVALRY IN INNOVATION 16–59* (2012) (discussing the costs and benefits associated with patent systems).

313. See, e.g., Lee Petherbridge et al., *The Federal Circuit and Inequitable Conduct: An Empirical Assessment*, 84 S. CAL. L. REV. 1293, 1296 (2011) (linking *ex parte* nature of patent prosecution to need for an inequitable conduct defense).

314. See John R. Allison & Mark A. Lemley, *The Growing Complexity of the United States Patent System*, 82 B.U. L. REV. 77, 135 (2002); Michael D. Frakes & Melissa F. Wasserman, *Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data*, 99 REV. ECON. & STAT. 550 app. at 2 tbl.A1 (2017); Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495, 1500 (2001); see also Roger Allan Ford, *The Patent Spiral*, 164 U. PA. L. REV. 827, 838 (2016); Janet Freilich, *Patent Clutter*, 103 IOWA L. REV. 925, 964 (2018) (“Examiners have a small amount of time allocated to examining each patent—an average of 18 hours per patent.”); John R. Thomas, *Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties*, 2001 U. ILL. L. REV. 305, 314 (“[T]he average time allocated for an examiner to address one application is understood to be between sixteen and seventeen hours.”); Brenda Sandburg, *Speed Over Substance?*, INTELL. PROP. MAG., Mar. 1999.

315. For recent data, see Dennis Crouch, *Standard Patent Size*, PATENTLYO (Oct. 22, 2017), <https://patentlyo.com/patent/2017/10/standard-patent-size.html> [<https://perma.cc/9VPH-NSFR>].

many as 60% or even more of these already issued patents are found completely or partially invalid.³¹⁶

Invalidity rates are only part of the problem of patent quality, however. In addition, a very large number of *legally* valid patents have no economic value. They may be perfectly valid as a matter of patent law, but they are worthless because there is no market for the idea that they offer. The technologies they invent may be too costly in relation to what they provide. There may be better methods for doing the same thing. They may be solutions in search of a problem. In any event, more than 90% of issued patents are never licensed, and of these a high percentage are never commercialized at all.³¹⁷ Only 1–3% of issued patents are ever litigated, and the rate has been declining, which is consistent with the proposition that the commercial value of issued patents (aside from validity questions) is declining as well.³¹⁸ Further, patent litigation is very expensive, ranging from \$600k where less than \$1m is at risk, to approximately \$5m when the amount at risk exceeds \$25m.³¹⁹

While patent examiners are trained to consider questions of legal validity, they do not generally consider commercial value or success, except in obvious cases of patents that do not work, claim physically

316. See Ronald J. Mann & Marian Underweiser, *A New Look at Patent Quality: Relating Patent Prosecution to Validity*, 9 J. EMPIRICAL LEGAL STUD. 1, 7 (2012) (looking at Federal Circuit decisions, 59.8% found invalid). Another study largely in accord is John R. Allison et al., *Understanding the Realities of Modern Patent Litigation*, 92 TEX. L. REV. 1769, 1777–79 (2014). That study updates John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AIPLA Q.J. 185, 188–93 (1998). See also Paul M. Janicke & Lilan Ren, *Who Wins Patent Infringement Cases?*, 34 AIPLA Q.J. 1, 16, 17, 19, 30 (2006) (breaking down data by technology, identity of inventors, and accused infringers and law firms); Mark A. Lemley, *The Surprising Resilience of the Patent System*, 95 TEX. L. REV. 1, 10 (2016) (explaining that the Supreme Court has made it easier to invalidate patents).

317. See Daniel Fisher, *The Real Patent Crisis is Stifling Innovation*, FORBES (June 18, 2014, 8:45 AM), <https://www.forbes.com/sites/danielfisher/2014/06/18/13633/#75d2ca2e6f1c> [<https://perma.cc/98AG-8EGR>]. For a more detailed breakdown of the data, see generally RON D. KATZNELSON, *A CENTURY OF PATENT LITIGATION IN PERSPECTIVE* (2014), <http://www.iplayership.org/articles/century-patent-litigation-perspective> [<https://perma.cc/PM4T-TFJ4>].

318. See KATZNELSON, *supra* note 317, at 11 fig.3; PWC, *2017 PATENT LITIGATION STUDY: CHANGE ON THE HORIZON?* 4 fig.1 (2017), http://www.ipwatchdog.com/wp-content/uploads/2017/05/2017-Patent-Litigation-Study_PwC.pdf [<https://perma.cc/JF3E-WGUG>] (showing the decline in patent litigation case filings); see also Hannibal Travis, *Counter-IP Conspiracies: Patent Alienability and the Sherman Antitrust Act*, 71 U. MIAMI L. REV. 758, 768 (2017) (discussing the decline in patent value).

319. See Samson Vermont, *AIPLA Survey of Costs of Patent Litigation and Inter Partes Review*, PATENTATTORNEY.COM (Jan. 30, 2017), <https://www.patentattorney.com/aipla-survey-of-costs-of-patent-litigation-and-inter-partes-review/> [<https://perma.cc/3695-F6JF>].

impossible things,³²⁰ or are useless on their face.³²¹ One can only imagine, but requiring examiners to test unissued patents for commercial value would increase examination costs enormously, even assuming they were capable of doing it. The best way to test commercial, as opposed to legal, value is through the market. A legally valid patent is commercially valuable if at least one firm wants to license it or produce technology or processes that infringe it.

A rational system for optimizing patent issuance must be responsive to concerns about both invalidity and economic value. Further, it must make these evaluations at reasonable cost. The patent examining process assesses legal validity, but economic value is best assessed by the relevant parties' willingness to put resources into validity determinations. If a patent is economically worthless it is unlikely that people will object to it or dispute its validity, because it does not interfere with their business.

In its 2018 *Oil States Energy Services, LLC v. Greene's Energy Group, LLC*³²² decision the Supreme Court approved a sensible regulatory approach to the system that assesses the legal validity of patents prior to litigation.³²³ The challenged administrative process, called *inter partes* review,³²⁴ permits any person to ask the United States Patent and Trademark Office (USPTO) to reconsider and, if necessary, cancel one or more claims in an issued patent.³²⁵ The procedure is administrative but adversarial. Both the patentee and the challenger have participation rights, including the right to present evidence, make arguments, and have a hearing before the Patent Trial and Appeal Board (PTAB).³²⁶ This makes the process significantly different and more demanding than the type of *ex parte* examination that goes on during initial patent prosecution. However, it is also much less costly than patent infringement litigation in a federal district court. Most particularly, PTAB review is limited to questions of validity under §§ 102 and 103 of the

320. See Daniel C. Rislove, *A Case Study of Inoperable Inventions: Why Is the USPTO Patenting Pseudoscience?*, 2006 WIS. L. REV. 1275, 1302, 1304 (2006) (discussing, *inter alia*, patents on perpetual motion or cold fusion machines, which are routinely rejected). For perpetual motion applications the USPTO does not reject them out of hand, but requires a working model. See MPEP § 608.03 (9th ed. Rev. 7, Nov. 2015).

321. For good perspective, see Michael Risch, *Reinventing Usefulness*, 2010 B.Y.U. L. REV. 1195, 1197 (2010) and Sean B. Seymore, *Making Patents Useful*, 98 MINN. L. REV. 1046, 1047 (2014).

322. 138 S. Ct. 1365 (2018).

323. *Id.* at 1369.

324. Created in the Leahy-Smith America Invents Act, Pub. L. No. 112-29, 125 Stat. 284 (2011) (codified as amended in scattered sections of 35 U.S.C.). See *Oil States*, 138 S. Ct. at 1371.

325. 35 U.S.C. §§ 311–19.

326. The process is briefly laid out in *Oil States*, 138 S. Ct. at 1371–72.

Patent Act; it does not consider infringement.³²⁷ The PTAB decision is fully reviewable on appeal to the Federal Circuit in which both the patentee and the challenger are entitled to appear, as well as PTAB itself.³²⁸

The PTAB proceeding permits the Agency to do a much more thorough and yet relatively low cost review of patent validity.³²⁹ The costs of administrative *inter partes* review are dramatically lower than the cost of litigation, running around \$275k through a PTAB hearing and \$350k through appeal.³³⁰ One Amicus in *Oil States* estimated the cost differential between litigation and *inter partes* review as exceeding ten to one.³³¹ Importantly, however, the costs are not zero, and as a result someone must have a financial incentive to challenge a patent. Most likely, this is someone producing technology that arguably infringes the patent in question, assuming it is valid. That is to say, patents that are challenged in a PTAB *inter partes* proceeding very likely do have positive *commercial* value, assuming they are valid, or no one would bother to challenge them.

Some critics of this administrative *inter partes* review system lament that it undermines a “strong patent system.”³³² That is an odd use of terminology, in which “strong patent system” becomes a synonym for protection of weak patents. It is as if a military general who, upon being instructed to raise a strong army, responded by abolishing or downgrading requirements for physical and mental fitness. That might produce a lot of soldiers, but not likely a stronger army. You don’t protect a strong patent system by approving lots of weak patents, but rather by coming up with a rational, cost-effective system for distinguishing those

327. See 35 U.S.C. § 311(b).

328. See *id.* § 319; *Oil States*, 138 S. Ct. at 1371–72 (describing the appeals procedure).

329. Accord Carl Shapiro, *Patent Reform: Aligning Reward and Contribution*, 8 INNOVATION POL’Y & ECON. 111, 136–37 (2007); see also Joseph Farrell & Carl Shapiro, *How Strong are Weak Patents?*, 98 AM. ECON. REV. 1347, 1347 (2008) (arguing that it would be efficient to create a process that determines patent validity prior to licensing).

330. See Vermont, *supra* note 319.

331. Brief for Apple Inc. as Amicus Curiae in Support of Respondents at 18, *Oil States*, 138 S. Ct. 1365 (No. 16-712).

332. E.g., Brief of Univ. of N.M. as Amicus Curiae in Support of Petitioner at 4, 16, *Oil States*, 138 S. Ct. 1365 (No. 16-712) (equating elimination of *inter partes* review with protection of a “strong patent system”); James Carmichael & Brad Close, *Despite Oil States, Inter Partes Review May Still be Held Unconstitutional*, IPWATCHDOG (Apr. 25, 2018), <https://www.ipwatchdog.com/2018/04/25/despite-oil-states-inter-partes-review-may-still-be-held-unconstitutional/id=96406/> [https://perma.cc/PD36-4CAT]; Gene Quinn, *Predicting Oil States in Advance of SCOTUS Oral Arguments*, IPWATCHDOG (Nov. 12, 2017), <http://www.ipwatchdog.com/2017/11/12/predicting-oil-states-advance-scotus-oral-arguments/id=89661/> [https://perma.cc/KN86-WXTK] (identifying abolition of *inter partes* review as protecting a “strong patent system”).

patents that make the requisite contribution to technological progress from those that do not.

The two-stage patent evaluation system that *inter partes* review permits is an efficient application of the economics of information, although there is some potential for abuse.³³³ In George Stigler's words, "the optimum amount of search will be such that the marginal cost of search equals the expected increase in receipts."³³⁴ At the initial grant stage relatively few resources are committed to establishing validity—a rational decision, given that so many patents are commercially worthless even if they are valid. While the error rate is high, the likelihood that anyone will be harmed is relatively low. Once their commercial value has been declared via a PTAB validity challenge, however, the stakes have gone up and a more costly examination is called for. One additional likely effect is that in the residual patent infringement cases in the district courts a relatively higher percentage of resources will go to questions of infringement, including claim construction and damages, rather than validity. Questions of infringement typically do not second guess the patent issuance process.³³⁵ Rather, they accept the patentee's claims as given and consider the scope of the accused device or process.

The patent system, including *inter partes* review before PTAB, is intended to manage both the very considerable social risk and cost attending innovation as well as its private costs. As with so many marginalist institutions, its job is to thread the needle between a policy that properly incentivizes useful and worthwhile technology while leaving the channels open for other development. An overly aggressive patent system is bad for the dissemination of technological progress just as much as an underdeterrent one.

CONCLUSION

In our current era in which agency review is under attack, it is worth considering exactly what we would have to do to roll the clock back. While this Article does not prove the point, it does suggest that returning

333. One troublesome example is PTAB challenges intended to do nothing more than reduce the value of the firm that owns the patent, thus making short-selling of that firm's stock a profitable activity. See Feng Ye, *Trading on the Outcomes of Patent Challenges: Short-Selling Petitioners and Possible Modifications to the Inter Partes Review Process*, 98 J. PAT. & TRADEMARK OFF. SOC'Y 557, 561 (2016). An empirical study supporting the strategy is Rabea Kruppert, *Dispute the Patent, Short the Stock: Empirical Analysis of a New Hedge Fund Strategy*, 50 INT'L REV. L. & ECON. 25, 31–34 (2017). Another problematic practice is the equivalent of "pay for delay" settlements of PTAB proceedings. See Erik Hovenkamp & Jorge Lemus, *Delayed Entry Settlements at the Patent Office*, 54 INT'L REV. L. & ECON. 30, 30 (2018).

334. George J. Stigler, *The Economics of Information*, 69 J. POL. ECON. 213, 216 (1961).

335. One exception is *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722 (2002) and its progeny, which govern the narrowing of claims during the patent prosecution process.

to a world in which generalist judges are responsible for devising and executing regulatory policy will require nothing less than reversing the marginalist revolution itself.