

COPYRIGHT AND PRODUCT DIFFERENTIATION

CHRISTOPHER S. YOO*

Existing economic analyses generally frame copyright as presenting a conflict between promoting efficient levels of access to creative works on the one hand and providing sufficient incentives to support their creation on the other. The supposed irreconcilability of the access-incentives tradeoff has led most scholars to regard copyright as a necessary evil and to advocate limiting copyright protection to the lowest level still sufficient to support creation of the work. In this Article, Professor Christopher Yoo breaks with the conventional wisdom and proposes a new approach to copyright law based on the economics of product differentiation. This differentiated products approach provides an explanation for market features that appear to be internal contradictions under the traditional approach. It also surpasses prior work by providing a basis for formalizing the incentives side of the tradeoff. In so doing, it underscores the importance of an alternative means for promoting access that has largely been ignored in the current literature: facilitating entry by close substitutes for existing works and allowing the ensuing competition to increase access by lowering prices. Focusing on this alternative means for promoting access further demonstrates that the access-incentives tradeoff may not be as intractable as generally believed, since facilitating entry can promote both considerations simultaneously. The differentiated products approach also assigns the government responsibilities that are better suited to its institutional capabilities than does the traditional approach. Equally importantly, it isolates the impact of three different dimensions of copyright protection, rather than depicting all aspects of copyright protection with a single variable in the manner of previous analyses. The more nuanced analysis made possible by the differentiated products approach suggests that economic welfare would best be promoted if copyright were strengthened along two of these dimensions and weakened along the third.

INTRODUCTION	214
I. THE TRADITIONAL APPROACH TO COPYRIGHT	226
A. <i>The Basic Economics of the Traditional Approach</i> ..	226

* Copyright © 2004 by Christopher S. Yoo. Associate Professor of Law, Vanderbilt University. A.B., Harvard College, 1986; M.B.A., The Anderson School at UCLA, 1991; J.D., Northwestern University School of Law, 1995. This Article has benefited from workshops conducted at the 2003 Annual Congress of the Society for Economic Research on Copyright Issues, the Thirty-First Annual Meeting of the American Law and Economics Association, the 2003 Works-in-Progress Intellectual Property Colloquium at Tulane University, Columbia Law School, and the Vanderbilt Law School. I also would like to thank Michael Abramowicz, Mark Brandon, Tim Brennan, Andy Daughety, Paul Edelman, Shubha Ghosh, John Goldberg, Wendy Gordon, Paul Heald, Larry Helfer, Steven Hetcher, Jack Hirshleifer, Doug Lichtman, Stan Liebowitz, Glynn Lunney, David McGowan, Michael Meurer, Tom Nachbar, Mark Nadel, Richard Nagareda, Neil Netanel, Erin O'Hara, Giovanni Ramello, Bob Rasmussen, Jennifer Reinganum, Pam Samuelson, and Art Snow for helpful comments on earlier drafts of this Article. Financial support from the Vanderbilt Dean's Fund is gratefully acknowledged.

INTRODUCTION

To date, the commentary on the economics of copyright¹ has largely revolved around the feature generally thought to distinguish markets for intellectual property from markets for other types of goods: nonrivalry in consumption.² Economists classify goods as nonrival when consumption by one person does not reduce the supply available for consumption by others. In the context of copyright, nonrivalry is generally taken to imply that once the first copy of a creative work has been produced, additional copies can be made virtually costlessly.³

Nonrivalry in consumption gives rise to a well-known economic conundrum.⁴ If authors are to break even, the per-copy price they

¹ This Article focuses solely on the economics of copyright. It therefore omits any discussion of noneconomic approaches to copyright. For a sampling of the alternatives, see, e.g., Wendy J. Gordon, *A Property Right in Self-Expression: Equality and Individualism in the Natural Law of Intellectual Property*, 102 *YALE L.J.* 1533 (1993) (exploring theories of copyright based in Lockean natural law); Justin Hughes, *The Philosophy of Intellectual Property*, 77 *GEO. L.J.* 287, 330–50 (1988) (advocating use of personality theory to shape copyright); Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 *YALE L.J.* 283 (1996) (analyzing copyright in terms of promoting particular vision of democracy). See generally William Fisher, *Theories of Intellectual Property*, in *NEW ESSAYS IN THE LEGAL AND POLITICAL THEORY OF PROPERTY* 168 (Stephen R. Munzer ed., 2001) (surveying noneconomic copyright literature). It should go without saying that in restricting the scope of my analysis in this manner, I do not mean to suggest that I regard noneconomic approaches as unimportant.

² Copyrighted works are also sometimes said to exhibit another feature associated with the theory of pure public goods: nonexcludability. Recent theoretical work on public good economics and the Coase Theorem suggests that nonexcludability does not necessarily induce market failure to the extent previously believed. See Christopher S. Yoo, *Rethinking the Commitment to Free, Local Television*, 52 *EMORY L.J.* 1579, 1594–96 (2003). Technological advances have also mitigated this concern by rendering copyrighted works much more excludable than in the past.

³ For the seminal statement of this principle, see Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, in *THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS* 609, 614–17 (1962). For other representative statements appearing in the literature, see Yochai Benkler, *An Unhurried View of Private Ordering in Information Transactions*, 53 *VAND. L. REV.* 2063, 2065–66 (2000); James Boyle, *Cruel, Mean, or Lavish? Economic Analysis, Price Discrimination and Digital Intellectual Property*, 53 *VAND. L. REV.* 2007, 2012 (2000); David J. Brennan, *Fair Price and Public Goods: A Theory of Value Applied to Retransmission*, 22 *INT'L REV. L. & ECON.* 347, 349–50 (2002) [hereinafter D. Brennan, *Fair Price*]; Timothy J. Brennan, *Copyright, Property, and the Right to Deny*, 68 *CHI.-KENT L. REV.* 675, 698 (1993) [hereinafter T. Brennan, *Right to Deny*]; William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 *HARV. L. REV.* 1659, 1700 (1988); William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 *J. LEGAL STUD.* 325, 326–27 (1989); Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 *TEX. L. REV.* 989, 994–95 (1997). Cf. William R. Johnson, *The Economics of Copying*, 93 *J. POL. ECON.* 158, 161 (1985) (incorporating nonrivalry into model of copying).

⁴ This Article focuses on copyrighted works solely as end products. Extended analysis of copyrighted works as inputs for the creation of other works thus exceeds the scope of

charge for a work must cover both a portion of the fixed costs needed to produce the work in the first place (often called “first-copy costs”) as well as the incremental cost of making the particular copy sold (which economists call “marginal cost”). Allowing third parties to copy freely would allow those third parties to underprice original authors, because the prices charged by those third parties would need only to cover the costs of producing an additional copy without having to include any surplus to defray the first-copy costs incurred by the authors. This would deprive authors of any reasonable prospect of recovering their fixed-cost investments and would thus leave rational authors with no economic incentive to invest in the production of creative works.

The need to enable authors to recover their first-copy costs is generally regarded as the primary justification for giving authors exclusive control over their creations.⁵ Copyright scholars have consistently raised the concern, however, that the exclusivity granted by copyright gives rise to the familiar welfare losses associated with monopoly pricing.⁶ Other commentators have expressed misgivings

this Article. I offer my preliminary thoughts *infra* note 204, in which I suggest that including considerations of cumulative innovation does not significantly change the analysis. For explorations of the effect of cumulative innovation in the context of copyright, see Landes & Posner, *supra* note 3, at 332–35; Lemley, *supra* note 3.

⁵ It should be noted that the earliest work on the economics of copyright suggested that transaction costs and strategic considerations could give authors sufficient lead time over copiers to recoup their fixed costs notwithstanding the problems associated with the lack of rivalry. See Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 HARV. L. REV. 281, 293–306 (1970); Robert M. Hurt & Robert M. Schuchman, *The Economic Rationale of Copyright*, 56 AM. ECON. REV. 421, 427–29 (1966); Arnold Plant, *The Economic Aspects of Copyright in Books*, 1 ECONOMICA 167, 171–75 (1934); see also Mark S. Nadel, *Why Copyright Law May Actually Have a Net Negative Effect on New Creations: The Overlooked Impact of Marketing*, 19 BERKELEY TECH. L.J. (forthcoming 2004). Those arguments were regarded as unpersuasive at the time they were offered. See Barry W. Tyerman, *The Economic Rationale for Copyright Protection for Published Books: A Reply to Professor Breyer*, 18 UCLA L. REV. 1100 (1971) (offering critique of Breyer’s arguments); Fisher, *supra* note 3, at 1708 n.231 (calling Tyerman’s refutation of Breyer “convincing”); Edmund W. Kitch, *Elementary and Persistent Errors in the Economic Analysis of Intellectual Property*, 53 VAND. L. REV. 1727, 1728 & nn.1–2 (2000) (noting Breyer’s acknowledgment of limitations of his arguments). The emergence of digital copying and transmission has made them even less so today.

⁶ For representative statements appearing in the economic literature, see, e.g., S.J. Liebowitz, *Copyright Law, Photocopying, and Price Discrimination*, 8 RES. L. & ECON. 181, 184 (1986); Ian E. Novos & Michael Waldman, *The Effects of Increased Copyright Protection: An Analytic Approach*, 92 J. POL. ECON. 236, 236–38 (1984). For representative statements appearing in the legal literature, see, e.g., Fisher, *supra* note 3, at 1700–03. See also Boyle, *supra* note 3, at 2012–13, 2028–29, 2032–33, 2037–38 (analyzing whether monopoly or perfect competition models should be applied to copyright); Kitch, *supra* note 5, at 1729–30, 1732–35 (same).

about the distributional implications of increasing authors' ability to control the terms of access to their works.⁷

Together, these premises have led most scholars to view copyright as mediating between the benefits flowing from the widespread dissemination of copyrighted works (often called the "access" side of the tradeoff) and the need to provide authors with sufficient compensation to support the creation of their works (typically called the "incentives" side of the tradeoff).⁸ Some degree of copyright protection may be necessary to provide sufficient incentives to support production of creative works, but such protection unavoidably reduces access. In the extreme case in which producing an additional copy of a creative work is essentially costless, allowing authors to charge anything for their works necessarily excludes some consumers even though the benefits they would derive from obtaining access to the work would exceed the costs of allowing them to do so. Conversely, pricing such works at efficient levels would lead them to be priced at zero, in which case they would generate no revenue whatsoever and authors would be unable to cover their first-copy costs.

The resulting tension between access and incentives has led most scholars to regard copyright as a necessary evil.⁹ These scholars conclude that copyright protection must exist, but should be calibrated to

⁷ See, e.g., Boyle, *supra* note 3, at 2024–25; Julie E. Cohen, *Copyright and the Jurisprudence of Self-Help*, 13 BERKELEY TECH. L.J. 1089, 1127 (1998); Fisher, *supra* note 3, at 1701–02; Raymond Shih Ray Ku, *The Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, 69 U. CHI. L. REV. 263, 319–20 (2002); Glynn S. Lunney, Jr., *Reexamining Copyright's Incentives-Access Paradigm*, 49 VAND. L. REV. 483, 497 (1996); Netanel, *supra* note 1, at 295.

⁸ See *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984) (describing copyright as requiring "a difficult balance between the interests of authors and inventors in the control and exploitation of their writings and discoveries on the one hand, and society's competing interest in the free flow of ideas, information, and commerce on the other hand"). William Landes and Richard Posner similarly observed in their landmark article on the economics of copyright: "Copyright protection . . . trades off the costs of limiting access to a work against the benefits of providing incentives to create the work in the first place. Striking the correct balance between access and incentives is the central problem in copyright law." Landes & Posner, *supra* note 3, at 326. For other representative statements appearing in the economic literature, see JACK HIRSHLEIFER & JOHN G. RILEY, *THE ANALYTICS OF UNCERTAINTY AND INFORMATION* 259 (1992); Liebowitz, *supra* note 6, at 184; Novos & Waldman, *supra* note 6, at 237. Cf. Johnson, *supra* note 3, at 159 (noting similar balancing in patent context). For representative statements appearing in the legal commentary, see Boyle, *supra* note 3, at 2013–14; Breyer, *supra* note 5, at 281–82; Julie E. Cohen, *Copyright and the Perfect Curve*, 53 VAND. L. REV. 1799, 1801 (2000); Fisher, *supra* note 3, at 1700–03; Lunney, *supra* note 7, at 485–86, 492–98.

⁹ Thomas Macaulay offered the classic exposition of this principle:

It is good that authors should be remunerated; and the least exceptionable way of remunerating them is by a monopoly. Yet monopoly is an evil. For the sake of the good we must submit to the evil; but the evil ought not to last a day longer than is necessary for the purpose of securing the good.

the lowest level that still provides sufficient return to support creation of a work.¹⁰ The argument for restraining copyright is thought to bear particular weight when a work is unique or has few substitutes, because it is under those circumstances that market power wielded by copyright holders and consumers is the greatest.¹¹

This mode of analysis, however, suffers from a well-recognized theoretical defect. Although courts and commentators routinely speak of copyright “monopolies,”¹² the exclusivity granted by intellectual property protection creates monopoly power only if substitutes are unavailable and entry barriers prevent the emergence of any such

THOMAS B. MACAULAY, *A Speech: Delivered in the House of Commons, February 5, 1841*, in *SPEECHES ON POLITICS & LITERATURE BY LORD MACAULAY* 176, 180 (New York, E.P. Dutton & Co. n.d.). The hostility to copyright embodied in Macaulay’s statement continues to influence modern analyses. See, e.g., James Boyle, *Fencing Off Ideas: Enclosure & the Disappearance of the Public Domain*, *DAEDALUS*, Spring 2002, at 13, 16; T. Brennan, *Right to Deny*, *supra* note 3, at 686–88; Niva Elkin-Koren, *Copyrights in Cyberspace—Rights Without Laws?*, 73 *CHI.-KENT L. REV.* 1155, 1171 (1998); Wendy J. Gordon, *An Inquiry into the Merits of Copyright: The Challenges of Consistency, Consent, and Encouragement Theory*, 41 *STAN. L. REV.* 1343, 1344 (1989); Neil Weinstock Netanel, *Asserting Copyright’s Democratic Principles in the Global Arena*, 51 *VAND. L. REV.* 217, 248–49 (1998).

¹⁰ See, e.g., Dan L. Burk, *Muddy Rules for Cyberspace*, 21 *CARDOZO L. REV.* 121, 133 (1999) (contending that “the incentive to create the work is purchased at the expense of restricted availability,” which “represents a loss of social welfare” and “is acceptable up to the point required to induce creation of the work, but not beyond”); William W. Fisher III, *Property and Contract on the Internet*, 73 *CHI.-KENT L. REV.* 1203, 1249 (1998) (arguing that goal of copyright law “is to give creators enough entitlements to induce them to produce the works from which we all benefit but no more”); Stewart E. Sterk, *Rhetoric and Reality in Copyright Law*, 94 *MICH. L. REV.* 1197, 1205 (1996) (“Any copyright protection beyond that necessary to compensate the author for lost opportunities would generate no additional incentive to create and would discourage production of additional copies even when the cost of producing those copies was less than the price consumers would be willing to pay.”); see also Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 *HARV. L. REV.* 501, 527 (1999) (describing “the concept of a restricted copyright—one that protects a copyrighted work to the extent necessary to induce creation, but no more”).

¹¹ See Cohen, *supra* note 8, at 1811; Michael J. Meurer, *Copyright Law and Price Discrimination*, 23 *CARDOZO L. REV.* 55, 82–83 (2001).

¹² The Supreme Court has long referred to copyrights as granting a “monopoly.” See, e.g., *Am. Tobacco Co. v. Werckmeister*, 207 U.S. 284, 293 (1907) (“[T]he purpose of copyright law is . . . to secure a monopoly”); *Holmes v. Hurst*, 174 U.S. 82, 85 (1899) (“[T]he right of an author to a monopoly of his publications is measured and determined by the copyright act”); *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53, 56 (1884) (“The monopoly which is granted to [authors] is called a copyright”). The Court’s antitrust tying jurisprudence has also long assumed that copyrights confer monopoly power. See, e.g., *U.S. Steel Corp. v. Fortner Enters.*, 429 U.S. 610, 619 (1977); *United States v. Loew’s, Inc.*, 371 U.S. 38, 45, 48 (1962); *Times-Picayune Publ’g Co. v. United States*, 345 U.S. 594, 611–13 (1953); *United States v. Paramount Pictures, Inc.*, 334 U.S. 131, 143–44 (1948). References to a copyright “monopoly” also appear in the Court’s most recent copyright decisions. See, e.g., *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23, 29–30 (2003); *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003). For examples of scholarly works treating copyright as conferring a monopoly, see *supra* note 6.

substitutes in the foreseeable future.¹³ Neither of these restrictive conditions is likely to be met with respect to copyright. Although some works exist for which there are few alternatives, substitutes are readily available for most works.¹⁴ More importantly, the “idea-expression dichotomy” limits copyright protection to the form of expression without offering any protection for the underlying ideas expressed in the work.¹⁵ This basic principle effectively guarantees that any competitor willing to undertake the same fixed-cost investment as the original author remains free to create alternative works with the same functional characteristics as any existing work.¹⁶ By preserving the possibility of entry into any market for copyrighted

¹³ See SENATE SUBCOMM. ON PATENTS, TRADEMARKS, AND COPYRIGHTS OF THE SENATE COMM. ON THE JUDICIARY, 85TH CONG., AN ECONOMIC REVIEW OF THE PATENT SYSTEM 53–54 & n.238 (Comm. Print 1958) (prepared by Fritz Machlup) (arguing that property and monopoly have different meanings in economic theory and should be distinguished); Edmund W. Kitch, *Patents: Monopolies or Property Rights?*, 8 RES. L. & ECON. 31 (1986) (offering classic statement of this principle in patent law context). To offer a common example outside the realm of intellectual property, the exclusive right to sell the land upon which my house is located does not give me the power to charge supracompetitive prices unless no other land with the same functional attributes is presently available or expected to be available in the near future.

¹⁴ Fisher, *supra* note 3, at 1702–03; see also Benkler, *supra* note 3, at 2069 (recognizing that particular information goods may or may not have close substitutes).

¹⁵ See 17 U.S.C. § 102(b) (2000) (providing that copyright does not protect “any idea, procedure, process, system, method of operation, concept, principle, or discovery” contained in work). For leading judicial statements articulating the idea-expression dichotomy, see *Eldred*, 537 U.S. at 218–21; *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 348–51 (1991); *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 556–60 (1985); *N.Y. Times Co. v. United States*, 403 U.S. 713, 726 n.* (1971) (Brennan, J., concurring); *Mazer v. Stein*, 347 U.S. 201, 217 (1954); *Int'l News Serv. v. Associated Press*, 248 U.S. 215, 234 (1918); *Baker v. Selden*, 101 U.S. 99, 103–04 (1879).

¹⁶ There are no barriers preventing another author from putting pen to paper and attempting to create a substitute for any written work. In other words, although copyright prohibits others from copying the specific words penned by J.K. Rowling without her permission, it does nothing to prevent any other person from writing stories about a school where children learn to perform magic. The inputs needed to create substitutes for more complex media are generally freely available. See Christopher S. Yoo, *Vertical Integration and Media Regulation in the New Economy*, 19 YALE J. ON REG. 171, 230 (2002) (noting that talent and hardware needed to produce television programming are “readily available in markets already highly organized to supply these same inputs to other industries” (citing BRUCE M. OWEN & STEVEN S. WILDMAN, VIDEO ECONOMICS 222 (1992))). There are, however, some factor markets in which this is not the case. For example, cable sports channels often enjoy local monopolies that are often alleged to inhibit the ability of others to offer competing services. See *Applications for Consent to the Transfer of Control of Licenses from Comcast Corp. and AT&T Corp., Transferors, to AT&T Comcast Corp., Transferee, Memorandum Opinion and Order*, 17 F.C.C.R. 23246, 23284 ¶ 96 (2002) (describing allegations that Comcast was using its control of Philadelphia’s professional sports teams to reduce competition from direct broadcast satellite (DBS) systems and rival cable operator RCN). The more straightforward way to remedy this problem is to redress the imperfections in the factor markets directly rather than using copyright law to compensate for them indirectly.

works, the idea-expression dichotomy dissipates authors' monopoly power.¹⁷ As a result, it would be a mistake to presume that copyright by itself is sufficient to confer monopoly power.¹⁸

Prior attempts to reflect the likely existence of substitute products have not proven wholly satisfactory. Some commentators have acknowledged that not all copyrighted works enjoy monopoly power, while nonetheless continuing to adhere to the intuitions associated with monopoly pricing.¹⁹ Others have treated the issue as a choice between the polar outcomes under perfect competition and monopoly, with the degree of substitutability determining the endpoint towards which copyright law should steer.²⁰ One commentator has turned to oligopoly models.²¹ Unfortunately, the variant of Cournot competition employed in this model does not allow for the possibility of entry.²² Still others have employed "dominant firm"

To say that copyright does not limit entry by competitive *content* is not to say that copyright might not serve to constrain entry of alternative media platforms. Although individuals always remain free to generate alternative music, text, or pictures, copyright has long constrained the emergence of new media, such as cable television, video cassette recorders (VCRs), DBS systems, music downloads, and streaming media over the Internet, just to name a few. For an insightful analysis of these issues, see Randal C. Picker, *Copyright as Entry Policy: The Case of Digital Distribution*, 47 ANTITRUST BULL. 423 (2002).

¹⁷ See Kitch, *supra* note 5, at 1730, 1734; Lemley, *supra* note 3, at 1041; Meurer, *supra* note 11, at 60; Douglas A. Smith, *Collective Administration of Copyright: An Economic Analysis*, 8 RES. L. & ECON. 137, 139 (1986); see also Landes & Posner, *supra* note 3, at 361 (noting that limitations on copyright duration reduce market power).

¹⁸ See, e.g., Paul Goldstein, *Copyright, LAW & CONTEMP. PROBS.*, Spring 1992, at 79, 84 (arguing that because "one author's expression will always be substitutable for another's," copyright will generally not create monopoly power); Kitch, *supra* note 5, at 1730, 1734 (arguing that "copyrights do not prevent competitors from creating works with the same functional characteristics" and that "almost all copyrights" are not monopolies); Smith, *supra* note 17, at 139 ("The potential monopoly power for individual holders of copyright whose works must compete with each other is in most instances not likely to be substantial."). A parallel critique emerged in the antitrust literature on tying. See, e.g., *Jefferson Parish Hosp. Dist. No. 2 v. Hyde*, 466 U.S. 2, 37 n.7 (1984) (O'Connor, J., concurring in judgment) ("A common misconception has been that a patent or copyright . . . suffices to demonstrate market power."); U.S. Dep't of Justice & Fed. Trade Comm'n, *Antitrust Guidelines for the Licensing of Intellectual Property* §§ 2.0, 2.2 (1995), *reprinted in* 4 Trade Reg. Rep. (CCH) ¶ 13,132, at 20,734-35 ("The Agencies will not presume that a patent, copyright, or trade secret necessarily confers market power upon its owner."); PHILIP E. AREEDA ET AL., 2A ANTITRUST LAW § 518, at 136 (2d ed. 2002).

¹⁹ See, e.g., James Boyle, *Foreword: The Opposite of Property?*, LAW & CONTEMP. PROBS., Winter/Spring 2003, at 1, 8 n.23; Cohen, *supra* note 8, at 1801-05, 1811; Fisher, *supra* note 3, at 1700-03; Wendy J. Gordon, *Intellectual Property as Price Discrimination: Implications for Contract*, 73 CHI.-KENT L. REV. 1367, 1386-89 & n.76 (1998); Lemley, *supra* note 3, at 1041-42, 1072-83; Meurer, *supra* note 11, at 82-83, 90-100; Sterk, *supra* note 10, at 1205 & n.45.

²⁰ See Boyle, *supra* note 3, at 2025-27; Kitch, *supra* note 5, at 1734-35.

²¹ See RICHARD WATT, COPYRIGHT AND ECONOMIC THEORY 39-50, 177-90 (2000) (offering Cournot model of copyright).

²² See Yoo, *supra* note 16, at 203 & n.135.

market structures, in which a single large producer faces competition from fringe producers.²³ Although more promising in terms of capturing the dynamics of entry, these dominant firm models still regard an inherent tension between access and incentives.²⁴ In addition, for reasons that I will explain later, they do not capture the full range of dynamics associated with entry by substitute works.²⁵

I believe that the economic analysis of copyright would benefit from employing a different approach to imperfect competition that I believe better captures the key characteristics of markets for copyrighted works: the economic models associated with product differentiation.²⁶ Product differentiation occurs when competing goods act as imperfect rather than perfect substitutes for one another. An oft-cited example is breakfast cereals, which vary widely in flavor, sweetness, and crunchiness.²⁷ To use an example based in copyright, torts casebooks vary considerably in terms of length, topics covered, depth of treatment, and perspective, as well as price. Although some exploration of product differentiation has appeared in the patent literature,²⁸ to date copyright scholars have accorded it relatively little attention.²⁹

²³ See Landes & Posner, *supra* note 3, at 334 & n.17; Christian Koboldt, *Intellectual Property and Optimal Copyright Protection*, 19 J. CULTURAL ECON. 131, 138 (1995).

²⁴ See Landes & Posner, *supra* note 3, at 326.

²⁵ See *infra* note 85.

²⁶ For surveys of the economic literature on product differentiation, see generally JOHN BEATH & YANNIS KATSOUALACOS, *THE ECONOMIC THEORY OF PRODUCT DIFFERENTIATION* (1991); B. Curtis Eaton & Richard G. Lipsey, *Product Differentiation*, in 1 HANDBOOK OF INDUSTRIAL ORGANIZATION 723 (Richard Schmalensee & Robert D. Willig eds., 1989). Overviews of product differentiation theory also appear in all leading industrial organization textbooks. See, e.g., DENNIS W. CARLTON & JEFFREY M. PERLOFF, *MODERN INDUSTRIAL ORGANIZATION* 194–235 (3d ed. 2000); JEFFREY CHURCH & ROGER WARE, *INDUSTRIAL ORGANIZATION* 367–421 (2000); F.M. SCHERER & DAVID ROSS, *INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE* 600–10 (3d ed. 1990); JEAN TIROLE, *THE THEORY OF INDUSTRIAL ORGANIZATION* 277–303 (1988).

²⁷ See, e.g., CHURCH & WARE, *supra* note 26, at 379.

²⁸ See, e.g., Paul Klemperer, *How Broad Should the Scope of Patent Protection Be?*, 21 RAND J. ECON. 113 (1990); Manfredi La Manna et al., *The Case for Permissive Patents*, 33 EUR. ECON. REV. 1427 (1989); Stephen M. Maurer & Suzanne Scotchmer, *The Independent Invention Defence in Intellectual Property*, 69 ECONOMICCA 535 (2002); Ted O'Donoghue et al., *Patent Breadth, Patent Life, and the Pace of Technological Progress*, 7 J. ECON. & MGMT. STRATEGY 1 (1998); Michael Waterson, *The Economics of Product Patents*, 80 AM. ECON. REV. 860 (1990).

²⁹ The most complete discussion in the published literature of the connection between product differentiation and copyright is Lunney, *supra* note 7, at 582–89 (applying monopolistic competition to copyright). See also MICHAEL ABRAMOWICZ, *COPYRIGHT REDUNDANCY* (George Mason Law & Economics Research Paper No. 03-03) (Jan. 31, 2003) (applying spatial competition to copyright), available at <http://ssrn.com/abstract=374580>.

Some articles analyze product differentiation in the limited context of direct copying, in which the copies are treated as imperfect substitutes for the original. See Stanley M. Besen & Sheila Nataraj Kirby, *Private Copying, Appropriability, and Optimal Copying*

Shifting to a differentiated products approach provides insights that until now have largely been overlooked.³⁰ The differentiated products approach provides a theoretical explanation for features of markets for copyrighted works that appeared to be internal contradictions under previous theories. It also suggests that nonrivalry may not play as central a role in the analysis of copyright as generally believed. Product differentiation theory indicates that the fundamental tension between access and incentives traditionally attributed to nonrivalry would still exist even if consumption of copyrighted works were completely rival.

In addition, the differentiated products approach corrects one of the central normative shortcomings of the traditional perspective on the economics of copyright by providing a basis for formalizing the incentive side of the access-incentives tradeoff.

Even more importantly, it opens up the policy space by underscoring the existence of an alternative avenue for promoting the access side of the tradeoff that has largely been obscured by the assumptions underlying the traditional approach to copyright. In short, access to creative works can be increased by facilitating entry by new works and allowing the ensuing competition to narrow the spread between price and marginal cost, rather than by using regulation to reduce the price of access to extant works.

At the same time, the possibility of entry serves to ensure that authors obtain enough—but only enough—revenue to support creation of their works. The presence of any supracompetitive profits will attract entry by competitors until all such profits have been dissipated. At that point, the revenue captured by each entrant should be just

Royalties, 32 J.L. & ECON. 255 (1989); Johnson, *supra* note 3, at 161–63, 170–71; Koboldt, *supra* note 23, at 138; Liebowitz, *supra* note 6, at 184. Another working paper uses monopolistic competition to evaluate competition between originals and derivative works. See Shubha Ghosh, Rights of First Entry in “Derivative Markets”: Exploring Market Definition in Copyright (June 19, 2003) (unpublished manuscript, presented at Second Annual Congress of the Society for Economic Research on Copyright Issues), available at <http://www.serco.org/2003/ghosh.pdf>. These articles make no attempt to model competition arising from substitute works. The only other references have simply noted in passing the possibility of analyzing copyright as competition among differentiated products without engaging in any significant exploration of its implications. See Stanley M. Besen & Leo J. Raskind, *An Introduction to the Law and Economics of Intellectual Property*, J. ECON. PERSP., Winter 1991, at 3, 5 n.2; Timothy J. Brennan, *Harper & Row v. The Nation, Inc.: Copyrightability and Fair Use*, 33 J. COPYRIGHT SOC'Y U.S.A. 368, 375–76 & n.19 (1985–86); Fisher, *supra* note 3, at 1702–03; Johnson, *supra* note 3, at 159; Landes & Posner, *supra* note 3, at 327–28 & n.4; Robert P. Merges, *Intellectual Property Rights and the New Institutional Economics*, 53 VAND. L. REV. 1857, 1859 (2000); Meurer, *supra* note 11, at 60–61, 97.

³⁰ This Article draws on and extends my prior work applying the economics of product differentiation to the regulation of television programming. See Yoo, *supra* note 2.

sufficient to offset the author's cost of production. By bringing revenues in line with the cost of production, entry largely alleviates concerns about overstimulation of creative activity or sustained supracompetitive profits generally raised in the patent literature.³¹ Over the long run, any short-run profits will accrue back to consumers in the form of increased product variety, which in turn increases welfare by allowing individuals to consume works that better accord with their ideal preferences.³²

The differentiated products approach further suggests that the tension between access and incentives, commonly regarded as the central problem of copyright policy, may not be as intractable as generally believed. Because facilitating entry by substitute works typically involves strengthening certain aspects of copyright protection, promoting access in this manner can have the added benefit of simultaneously promoting the incentive side of the tradeoff as well. In this manner, the differentiated products approach also contradicts the conventional wisdom by demonstrating how strengthening certain aspects of copyright protection can actually cause economic welfare to increase.³³

Finally, the existing scholarship on copyright makes almost no attempt to draw any distinctions among the various dimensions along which copyright protection can be increased or decreased. Instead, previous analyses have represented the overall strength of copyright protection with a single variable.³⁴ The importance of differentiating among the various aspects of intellectual property protection has received some attention in the patent literature analyzing the distinction between length and breadth.³⁵ My analysis modifies these two

³¹ See Mark F. Grady & Jay I. Alexander, *Patent Law and Rent Dissipation*, 78 VA. L. REV. 305, 316, 321 (1992); Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265, 278-79 (1977).

³² The zero-profit result is subject to several important qualifications, which are discussed *infra* at notes 91, 102, 206-09 and accompanying text.

³³ This conclusion is subject to an important qualification: The economic literature indicates that when fixed costs are large as compared to marginal costs, it is possible that strong property rights may induce levels of entry that are inefficiently excessive. See *infra* note 160 and accompanying text.

³⁴ See, e.g., Koboldt, *supra* note 23, at 137; Landes & Posner, *supra* note 3, at 333-34; Novos & Waldman, *supra* note 6, at 239.

³⁵ See Vincenzo Denicolò, *Patent Races and Optimal Patent Breadth and Length*, 44 J. INDUS. ECON. 249, 252 (1996); Nancy T. Gallini, *Patent Policy and Costly Imitation*, 23 RAND J. ECON. 52, 60 (1992); Richard Gilbert & Carl Shapiro, *Optimal Patent Length and Breadth*, 21 RAND J. ECON. 106, 106 (1990); Klemperer, *supra* note 28, at 113; Maurer & Scotchmer, *supra* note 28, at 542-43; D.G. McFetridge & M. Rafiquzzaman, *The Scope and Duration of the Patent Right and the Nature of Research Rivalry*, 8 RES. L. & ECON. 91, 92-93 (1986); O'Donoghue et al., *supra* note 28, at 2-5; Pankaj Tandon, *Optimal Patents with Compulsory Licensing*, 90 J. POL. ECON. 470, 470-71 (1982).

categories to provide a better fit with markets for copyrighted works.³⁶ It also adds an additional consideration that the current literature does not take into account: the extent to which authors are able to appropriate the surplus created by their works.³⁷

Interestingly, the differentiated products approach indicates that the relationship between these three different facets of copyright protection is more complex than generally recognized. In contrast to current analyses, which presume that strengthening any aspect of copyright protection will have the same effect on economic welfare, the differentiated products approach suggests that entry would best be promoted by strengthening copyright protection along certain dimensions while weakening it along others.

The ability to distinguish among different aspects of copyright protection prevents the differentiated products approach from collapsing into a simple defense of stronger property rights.³⁸ On the contrary, the differentiated products approach provides a basis for determining when copyright protection is too strong. At the same time, it somewhat paradoxically contradicts the conventional wisdom by suggesting that the case for expanding copyright protection is at its strongest when markets are the most concentrated.

Together these insights suggest that the study of the economics of copyright would benefit from shifting towards a differentiated products approach. By providing a more complete picture of the dynamics of free entry, my proposal offers a basis for moving beyond the relatively static, monopoly-oriented vision implicit in the traditional approach. In so doing, it compensates for the blind spot that has led most scholarship on the economics of copyright to regard weakening copyright protection as the only available means for promoting the

³⁶ For example, the first consideration raised by the patent literature—length—frames the issue solely in terms of the duration of the patent term. See, e.g., Denicolò, *supra* note 35, at 249–51. My analysis *infra* Part IV recognizes that duration represents only one way to expand the number of surplus-generating activities encompassed by the right. Those authors that have considered aspects of length other than duration have done so under the rubric of breadth. See Denicolò, *supra* note 35, at 252 (citing C. Matutes et al., *Optimal Patent Design and the Diffusion of Innovations*, 27 RAND J. ECON. 60 (1996)) (considering number of applications of innovation which are reserved for patentee as element of breadth). This Article acknowledges that policymakers can increase the number of surplus-generating activities that fall within the right in other ways as well. See *infra* Part V.A.1.

³⁷ The only analysis of which I am aware that even touches upon this subject is the work of Terry Fisher, who argues that increases in appropriability made possible by facilitating price discrimination might justify weakening other aspects of copyright protection. See Fisher, *supra* note 10, at 1250–52. Fisher's analysis deviates from the one I advance in that he seeks to promote social values that are not primarily economic in character.

³⁸ See, e.g., Harold Demsetz, *Toward a Theory of Property Rights*, AM. ECON. REV. (papers and procs.), May 1967, at 347, 354–59.

access side of the tradeoff. Instead, it emphasizes how the same goal can be achieved by stimulating entry and allowing the decrease in price resulting from the increase in competition to facilitate more widescale dissemination of the work.

In presenting this argument, I acknowledge that markets for differentiated products always involve some degree of efficiency loss. As a result, the differentiated products approach does not offer a first-best solution, and the choice of regulatory regime ultimately turns on a comparison of second-best outcomes.³⁹

Several considerations suggest, however, that the market equilibrium under the differentiated products approach will likely prove more efficient than the outcome under the traditional approach. As I will subsequently discuss in some detail, the efficiency losses associated with the differentiated products approach are not likely to be significant so long as the fixed costs of entry are small when compared to the size of the market segment in which a particular work competes.

In addition, the differentiated products approach assigns the government responsibilities that are better suited to its institutional capabilities than does the traditional approach. Specifically, the traditional approach requires the government to calibrate the level of copyright protection so that authors earn the minimum return necessary to support creation of their works. The informational demands needed to implement such a system border on the prohibitive, and the academic commentary has grown increasingly critical of whether the government is capable of striking anything approaching the proper balance.⁴⁰ The differentiated products approach, in contrast, makes far more modest demands of the government. It requires only that the govern-

³⁹ See Harold Demsetz, *Information and Efficiency: Another Viewpoint*, 12 J.L. & ECON. 1, 1-4 (1969) (criticizing "nirvana approach," which compares real alternatives with ideal norm, as opposed to "comparative institution approach," which assesses best of real alternatives).

⁴⁰ See, e.g., Frank H. Easterbrook, *Cyberspace Versus Property Law?*, 4 TEX. REV. L. & POL. 103, 107-13 (1999); Fisher, *supra* note 10, at 1249; Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CAL. L. REV. 1293, 1308-16 (1996); Merges, *supra* note 29, at 1867-73; Steven Shavell & Tanguy Van Ypersele, *Rewards Versus Intellectual Property Rights*, 44 J.L. & ECON. 525, 541-44 (2001). In particular, copyright legislation has been singled out for being overly responsive to industry concerns. See, e.g., Boyle, *supra* note 3, at 2035 ("[I]t would be hard to find a more perfect example of rent-seeking than intellectual property legislation."); Jessica D. Litman, *Copyright, Compromise, and Legislative History*, 72 CORNELL L. REV. 857, 870-79 (1987) (describing legislative process leading up to enactment of 1976 Copyright Act); cf. Merges, *supra* note 29, at 1875 (noting "broad consensus that industry groups have unusually broad input into the drafting of . . . legislation" relating to intellectual property rights).

ment facilitate entry, depending instead upon market forces to bring revenues into balance with fixed costs.

What follows is a discursive and graphical presentation designed to capture the basic intuitions underlying the differentiated products approach, with a more formal specification of the theory postponed until later work.⁴¹ Part I offers an overview of the traditional, monopoly-oriented approach to copyright. In the process, it identifies several features of markets for copyrighted works that appear to be logical inconsistencies under the traditional approach and highlights the traditional approach's normative deficiencies.

Part II introduces the economics of product differentiation. Although these models are well known to most economists, they are less familiar to many legal academics. So, at the risk of belaboring points that are already familiar to some readers, I lay out the fundamentals of the two principal approaches to modeling product differentiation at some length. Section A describes the approach based on the theory of monopolistic competition first elucidated by Edward Chamberlin.⁴² Section B discusses the spatial competition models pioneered by Harold Hotelling.⁴³ In both cases, I emphasize what has become known as the "symmetric preferences branch" of each approach, which presumes that all available products are in equal competition with one another.⁴⁴

Part III offers an analysis of the differentiated products approach's positive implications, concluding that it provides a basis for resolving the seeming contradictions identified in Part I.

Part IV, which is the heart of the analysis, explores the differentiated products approach's normative implications. Most significantly, it breaks with the traditional approach by demonstrating how entry can promote both the access and incentive sides of the tradeoff simultaneously. It also suggests that strengthening copyright protection will not have the distributional impacts that many scholars suggest.

⁴¹ I also hope to apply a similar analysis to patent law. Patent law differs from copyright in that it does give patentees proprietary rights in particular ideas. In addition, the doctrine of equivalents gives patentees control over a penumbra of functionally similar technologies. As a result, in contrast with copyright, the issuance of a patent does place limits on others' ability to create products with the same functional characteristics as existing products. Consequently, the free entry assumption that drives much of the analysis of this Article may carry less validity with respect to patent law.

⁴² See EDWARD HASTINGS CHAMBERLIN, *THE THEORY OF MONOPOLISTIC COMPETITION: A RE-ORIENTATION OF THE THEORY OF VALUE* (8th ed. 1962).

⁴³ See Harold Hotelling, *Stability in Competition*, 39 *ECON. J.* 41 (1929).

⁴⁴ See BEATH & KATSOUACOS, *supra* note 26, at 5–6, 42; *infra* notes 78, 80 and accompanying text.

Part V demonstrates how the differentiated products approach makes it possible to separate the impact of three different aspects of copyright protection. In so doing, it offers a significant advance over the traditional approach, which, as noted earlier, tends to represent all aspects of copyright protection with a single variable.⁴⁵ This analysis reveals that economic efficiency would best be promoted by strengthening copyright protection along certain dimensions while simultaneously weakening it along others.

I

THE TRADITIONAL APPROACH TO COPYRIGHT

This Part describes and critiques the traditional economic approach that has come to dominate modern copyright theory. Section A describes the basic elements of the traditional analysis. Section B reviews its positive shortcomings. In short, real-world markets for copyrighted works exhibit a number of descriptive features that classic monopoly theory cannot explain. Section C analyzes the traditional approach from a normative perspective, concluding that it fails to provide a basis for formally assessing the incentives side of the access-incentives tradeoff.

A. *The Basic Economics of the Traditional Approach*

Traditional economic analyses of copyright either explicitly or implicitly are based on the intuitions underlying the graphical depiction in Figure 1, which has become a staple of copyright scholarship.⁴⁶ Two key aspects of this graph bear emphasizing. The first is the relative positions of the average cost curve (AC) and marginal cost curve (MC).⁴⁷ When consumption is rival, the average and marginal cost curves are normally depicted as “U” shaped.⁴⁸ This is because production of most goods requires the incurrence of both fixed and variable costs. At relatively low quantities, the impact of amortizing fixed

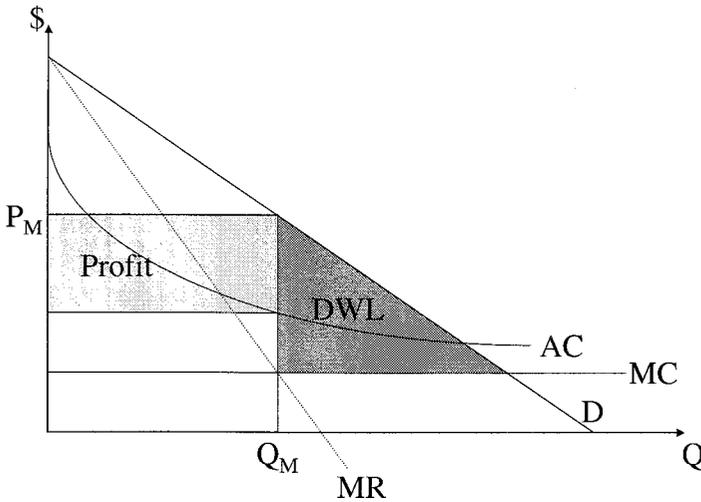
⁴⁵ See *supra* note 34 and accompanying text.

⁴⁶ See, e.g., Cohen, *supra* note 8, at 1802 fig.A; Fisher, *supra* note 10, at 1236 fig.2; Fisher, *supra* note 3, at 1701 n.201 fig.1, 1708 n.232 fig.2; Liebowitz, *supra* note 6, at 185 fig.1; see also Kitch, *supra* note 5, at 1732 fig.1 (employing similar graphical representation); Merges, *supra* note 29, at 1858 fig. (same); Sterk, *supra* note 10, at 1206 fig.1 (same). Figure 1 differs from similar diagrams appearing in prior scholarship that portray the entire producer surplus as profit. Such depictions overstate the degree of profit by ignoring the role of fixed costs. Because of fixed costs, only the portion of the producer surplus lying above the average cost curve properly can be regarded as profit.

⁴⁷ The ensuing discussion on the shape and relative position of the average and marginal cost curves for rival and nonrival goods is adapted from Yoo, *supra* note 2, at 1596–600.

⁴⁸ For an example, see *infra* p. 247 fig.5.

FIGURE 1: LONG-RUN EQUILIBRIUM UNDER THE TRADITIONAL APPROACH TO COPYRIGHT



costs across larger volumes causes average costs to decline. Variable costs initially decline as well, as the increase in volume allows the realization of the available economies of scale. Together these forces cause both average cost and marginal cost to decline, with the average cost curve lying above the marginal cost curve.

As production increases, the impact of the amortization of fixed costs on average cost decays exponentially until it becomes essentially immaterial, and variable costs become the primary determinant of average cost. With respect to variable costs, the increase in volume eventually exhausts all of the available economies of scale. At some point, congestion in production facilities and the increase in demand for scarce factors of production cause the variable costs needed to produce additional quantities to rise, at which point the marginal cost curve begins to curve upwards. Once the increase in marginal costs becomes substantial enough, marginal cost will exceed average cost and the average cost curve will begin to curve upwards as well.

Basic principles of welfare maximization⁴⁹ require that works be priced at marginal cost, because it is at that point that the social benefits of producing an incremental unit no longer exceed the social costs. In addition, authors must be allowed to price their works so as to cover their average costs if those works are to exist. When goods are

⁴⁹ For purposes of this Article, I will follow the standard convention and use total surplus as the relevant measure of welfare. Although total surplus is not the only available welfare metric, if utility is linear in the composite commodity, maximization of welfare and maximization of total surplus are identical. See Eaton & Lipsey, *supra* note 26, at 729.

rival, there are volumes at which marginal cost lies at or above the average cost curve. As a result, prices exist that satisfy both the preconditions for economic efficiency and the profitability constraint needed for a stable equilibrium to exist.

A different situation obtains when consumption of a good is nonrival. Because consumption by one person does not reduce the supply available for consumption by others, one would not expect increases in production to cause variable cost to increase.⁵⁰ As a result, the average cost curve is constantly declining and lies above the marginal cost curve across all volumes. The fact that average cost is always greater than marginal cost gives rise to the distinctive pricing problem underlying much of traditional copyright analysis. When goods are nonrival, there is no volume at which marginal cost equals or exceeds average cost. As a result, no level of production exists that provides optimal access to the work while simultaneously providing authors with the expectation of sufficient revenue to induce the work's creation. This suggests that absent some other institutional solution,⁵¹ some degree of deadweight loss is endemic to markets for copyrighted works. Note that the nonexistence of any price that simultaneously promotes efficiency and allows authors to break even derives from the lack of rivalry rather than from any perceived monopoly power. No matter how much competition authors face, they will find it impossible to set a price that both equals marginal cost and equals or exceeds average cost.

The second feature that bears emphasis is that the traditional approach presumes that authors engage in monopoly pricing. Authors that face competition cannot raise price without losing all of their sales, since any attempt to increase price would simply induce all of their customers to transfer their sales to their competitors.⁵² A dif-

⁵⁰ Note that the basic conclusions of this analysis do not depend upon the strong assumption that marginal costs are zero. Marginal costs can be positive without changing the analysis so long as they are nonincreasing. See WILLIAM J. BAUMOL ET AL., *CONTESTABLE MARKETS AND THE THEORY OF INDUSTRY STRUCTURE* 301-02 (rev. ed. 1988).

⁵¹ It has, for example, long been recognized that government subsidies can make up for the shortfall in production. See, e.g., Paul A. Samuelson, *Aspects of Public Expenditure Theories*, 40 *REV. ECON. & STAT.* 332, 335-36 (1958). In the context of copyright, this can take the form of public libraries. See Cohen, *supra* note 8, at 1806; Gordon, *supra* note 19, at 1388. Alternatively, the government could implement a prize system, in which the government, rather than consumers, rewards authors for their creative efforts. See generally Michael Abramowicz, *Perfecting Patent Prizes*, 56 *VAND. L. REV.* 115, 127-211 (2003) (reviewing literature in patent context). Although these alternative forms theoretically might solve the pricing problem, I am skeptical as to whether governmental institutions are well situated to determine the value of copyrighted works. See *supra* note 40 and accompanying text.

⁵² For a more detailed description of the economics of perfect competition, see Yoo, *supra* note 2, at 1587-94.

ferent situation obtains if an author is assumed to be a monopolist. Monopolists by definition have no competitors and can therefore raise price without worrying about losing their entire sales to others. Instead, increasing price would simply reduce the number of consumers who actually purchase the product. As a result, it is common to model monopolists as facing a downward-sloping demand curve (D).

More importantly for our purposes, the assumption that the author is a monopolist also causes the marginal revenue curve (MR) for authors to lie somewhere below the demand curve. The intuition underlying this result is quite simple. Authors that already are charging the maximum possible for a given level of production can only attract new customers by dropping their price. If authors could offer the lower price only to new customers without also having to offer that price to their existing customers, their marginal revenue curve would coincide with the demand curve over the relevant range. However, most authors cannot reduce prices for new customers without also reducing prices for their existing customers. Any revenue gained by moving farther down the demand curve must thus at least partially be offset by the revenue lost by giving the lower price to existing customers. The marginal revenue curve must then be represented by the demand curve minus the revenue lost by offering the price reduction to the existing customers.⁵³ This causes the marginal revenue curve to lie below the demand curve.⁵⁴

Like any producer seeking to maximize its profits, a monopolist will increase its production until the marginal revenue it obtains from selling an additional unit no longer exceeds the marginal cost of producing the additional unit (represented in Figure 1 as Q_M , where the marginal revenue curve intersects the marginal cost curve). Having identified the profit-maximizing level of production, the firm will charge the maximum price that it can receive for that quantity (represented in Figure 1 by P_M , where Q_M falls along the demand curve).

Monopoly pricing has two well-recognized economic consequences.⁵⁵ First, because monopoly prices exceed marginal cost, they inefficiently exclude some consumers who would be willing to purchase the work at prices between P_M and marginal cost. The

⁵³ This analysis changes somewhat if the firm is able to price discriminate. *See infra* notes 56–60 and accompanying text.

⁵⁴ In fact, when the demand function for a particular product is linear, the marginal revenue curve is given by a straight line with twice the slope of the demand curve intersecting the vertical axis at the same point as the demand curve. *See* SCHERER & ROSS, *supra* note 26, at 21 n.13 (offering simple proof of this relationship).

⁵⁵ *See supra* note 7 and accompanying text.

reduction in economic welfare caused by this underproduction is called deadweight loss (DWL) and is represented by the darkly shaded triangular area in Figure 1. Second, if producer surplus exceeds fixed costs, authors can earn supracompetitive profits, represented by the lightly shaded rectangular area of Figure 1. Not only do these profits raise distributional concerns, they also threaten to overstimulate creative activity by overcompensating authors for their efforts. Some scholars have argued that this tension could be resolved by making it easier for authors to price discriminate.⁵⁶ Perfect price discrimination (i.e., if authors were able to charge each consumer the maximum amount she would be willing to pay) would eliminate deadweight loss while simultaneously optimizing authors' incentives to create.⁵⁷ The problem is that perfect price discrimination is never possible, and while it is more likely than not that imperfect price discrimination will cause total output to increase and deadweight loss to fall,⁵⁸ the opposite may well hold true.⁵⁹ In addition, the increase in profits made possible by price discrimination threatens to exacerbate the distributive and overstimulation problems.⁶⁰

The traditional approach thus provides ample reason for concern that markets will not provide efficient levels of access. Despite the well-recognized shortcomings of relying on the government to allocate

⁵⁶ The seminal discussion of price discrimination is JOAN ROBINSON, *THE ECONOMICS OF IMPERFECT COMPETITION* 188–95 (8th ed. 1948). For recent overviews in the context of copyrighted works, see Meurer, *supra* note 11; Yoo, *supra* note 2, at 1618–27, 1706–12. For other discussions in the context of nonrival consumption, see James M. Buchanan, *Public Goods in Theory and Practice: A Note on the Minasian-Samuelson Discussion*, 10 J.L. & ECON. 193, 193–96 (1967); Harold Demsetz, *The Private Production of Public Goods*, 13 J.L. & ECON. 293, 301–04 (1970); Earl A. Thompson, *The Perfectly Competitive Production of Collective Goods*, 50 REV. ECON. & STAT. 1, 3–5 (1968).

⁵⁷ See Tom W. Bell, *Fair Use vs. Fared Use: The Impact of Automated Rights Management on Copyright's Fair Use Doctrine*, 76 N.C. L. REV. 557, 588 & n.142 (1998); Fisher, *supra* note 10, at 1237–40; Fisher, *supra* note 3, at 1709–10; David Friedman, *In Defense of Private Orderings: Comments on Julie Cohen's "Copyright and the Jurisprudence of Self-Help"*, 13 BERKELEY TECH. L.J. 1151, 1168–71 (1998). For a related argument in the context of television programming, see Yoo, *supra* note 2, at 1618–27, 1706–12.

⁵⁸ See, e.g., RICHARD G. LIPSEY ET AL., *ECONOMICS* 241 (8th ed. 1987); SCHERER & ROSS, *supra* note 26, at 495.

⁵⁹ See, e.g., ROBINSON, *supra* note 56, at 190–94; Michael L. Katz, *Non-Uniform Pricing, Output and Welfare Under Monopoly*, 50 REV. ECON. STUD. 37, 51 (1983); Richard Schmalensee, *Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination*, 71 AM. ECON. REV. 242 (1981); Jun-ji Shih et al., *A General Analysis of the Output Effect Under Third-Degree Price Discrimination*, 98 ECON. J. 149, 152–54 (1988).

⁶⁰ See Boyle, *supra* note 3, at 2025–26, 2029; Cohen, *supra* note 8, at 1806. *But see* Fisher, *supra* note 10, at 1239 (noting that price discrimination provides more consumers with access to creative works); *id.* at 1240 & n.89, 1249–50 (arguing that facilitating price discrimination would make possible offsetting adjustments in other aspects of copyright designed to promote distributional equity).

goods in an efficient manner,⁶¹ adherents of the monopoly model rely upon policymakers to calibrate the level of copyright protection so as to permit the greatest possible degree of access while still providing sufficient incentives for the work to be created in the first place.⁶²

B. *The Positive Deficiencies of the Traditional Approach*

One measure of an economic theory is how well its predictions match up with real-world features of markets for copyrighted works. I have already noted the traditional approach's inability to account for the relative ease with which competitors can enter the market with near substitutes.⁶³ Upon closer inspection, it becomes clear that the traditional approach also fails to explain a number of other important market features.

1. *Nonrivalry and Zero Marginal Costs*

As noted earlier, traditional economic theories assume that copyrighted works can be copied costlessly.⁶⁴ It is this feature that transforms copyright into an intractable tradeoff between access and incentives.

There is reason to question whether the marginal cost curves will assume the shape and position posited by the traditional approach for all copyrighted works. On the contrary, it is more likely that copyrighted works will occupy a spectrum. At one extreme are works that can be copied an unlimited number of times at virtually no expense. Indeed, the pervasive shift towards digitization of media is causing the number of works at this end of the spectrum to increase dramatically. At the other extreme are three-dimensional analog works, such as stone sculptures, textured paintings, and other pieces of physical art, that are rather hard to copy. Although digital images can inexpensively capture some aspects of such works, copies that capture the full range of the aesthetic experience remain extremely expensive.⁶⁵ The extent to which particular works will exhibit some degree of rivalry in

⁶¹ See *supra* note 40 and accompanying text.

⁶² See *supra* note 10 and accompanying text.

⁶³ See *supra* notes 14–17 and accompanying text.

⁶⁴ See *supra* notes 3, 46–51 and accompanying text.

⁶⁵ See Lunney, *supra* note 7, at 493 n.24; John Shepard Wiley Jr., *Copyright at the School of Patent*, 58 U. CHI. L. REV. 119, 150 (1991); cf. Richard A. Epstein, *Transaction Costs and Property Rights: Or Do Good Fences Make Good Neighbors?*, in CHICAGO LECTURES IN LAW AND ECONOMICS 175, 177 (Eric A. Posner ed., 2000) (noting that “copying” farm produce is impossible).

consumption depends on the extent to which the copyrighted material must be combined with physical inputs.⁶⁶

Transmission costs could also be a source of nonzero marginal costs. Transmission costs may be substantial if every copy of the creative work must be fixed into a physical form. Conventional wisdom holds that the cost of transmitting digital content over the Internet is trivial, but it is possible that distributors of electronic media will incur more substantial transmission costs.

Consider, for example, the costs associated with transmitting Internet content to an incremental household via a cable modem system. If the network has not yet been extended to that household, the provider may well have to incur additional costs in order to provide the service. Even more to the point, many components of the system of content servers, backbones, local servers, and neighborhood nodes necessary to provide Internet service are subject to congestion.⁶⁷ Although it initially seems possible to add an incremental user at no additional cost, doing so in fact imposes costs by decreasing the quality of service provided to other users and by hastening the day at which additional capital investments will be necessary to expand network capacity. The lumpiness of these investments thus justifies conceiving of transmission costs as a step function rather than as constant and flat. Viewed from a distance, systems that appear to exhibit zero marginal costs begin to manifest some degree of rivalry.

It is thus not clear that all real-world markets for copyrighted works will exhibit the declining average cost structure that is central to the traditional approach. The implication is either that the tension between access and incentives identified by monopoly theory does not always exist, in which case one would expect the scope of copyright protection doctrine to vary across different types of works, or, as I will subsequently suggest, that the relationship between average and marginal cost usually attributed to nonrivalry in consumption may in fact stem from some other source.⁶⁸

2. *The Absence of Natural Monopoly*

One of the standard results in the economic literature is that markets for nonrival goods will exhibit a strong tendency towards natural

⁶⁶ See I. Trotter Hardy, *Not So Different: Tangible, Intangible, Digital, and Analog Works and Their Comparison for Copyright Purposes*, 26 U. DAYTON L. REV. 211, 233 (2001); Kitch, *supra* note 5, at 1737 & n.21; Merges, *supra* note 29, at 1859. For examples of this effect in the context of television programming, see Demsetz, *supra* note 56, at 296–303; Yoo, *supra* note 2, at 1631–32.

⁶⁷ See Yoo, *supra* note 16, at 263.

⁶⁸ See *infra* Part III.A.

monopoly.⁶⁹ This is because the presence of zero marginal costs causes average cost to decline across all volumes. The inexhaustibility of the economies of scale gives authors with the largest volumes a decisive cost advantage that should allow them to drive all of their competitors out of the market.

This effect can be illustrated by the example of torts casebooks. Assume that two authors split a market of one thousand students. Each must charge a price that covers both the marginal cost of producing each copy and an allocation of the fixed costs needed to produce the first copy. Authors that sell more textbooks can spread their fixed costs over larger volumes, permitting them to undercut the unit price charged by other authors. If the works are completely fungible, every buyer will base her decision solely on price and choose the casebook with the larger volume. In this manner, the cost advantages allow the author with the higher volume to capture the entire market.

Of course, a casual perusal of available creative works reveals that most, if not all, face sustainable competition from multiple producers. I do not mean to imply that anyone will find it surprising that not all copyrighted works devolve into natural monopolies. My point is that the inability of the traditional approach to provide an explanation of why this is so signals a basic problem with the theory underlying it. The presence of competition despite the presence of unexhausted scale economies strongly suggests the presence of another economic consideration that is not adequately being taken into account.

⁶⁹ See Brennan, *supra* note 29, at 375 (noting that lack of rivalry “implies that for each particular piece of intellectual property, one firm could produce and distribute it less expensively than a number of firms”); see also RICHARD CORNES & TODD SANDLER, *THE THEORY OF EXTERNALITIES, PUBLIC GOODS AND CLUB GOODS* 348–49 (2d ed. 1996) (noting that for pure public goods, “[t]he entire population is in a single provision association”); WILLIAM W. SHARKEY, *THE THEORY OF NATURAL MONOPOLY* 47 (1982) (noting that increasing returns to scale associated with natural monopoly “are closely related to the concept of public goods”); STEPHEN SHMANSKE, *PUBLIC GOODS, MIXED GOODS, AND MONOPOLISTIC COMPETITION* 27 (1991) (noting that high overhead costs and zero marginal costs associated with nonrivalry “certainly evokes the standard natural monopoly model”); Dagobert L. Brito & William H. Oakland, *On the Monopolistic Provision of Excludable Public Goods*, 70 *AM. ECON. REV.* 691, 691 (1980) (noting that if left to private sector, scale economies would cause many public goods to be “characterized by monopoly elements”). The tendency towards natural monopoly would also give markets a winner-take-all quality that promotes the type of destructive races to get to the market that have commanded so much attention in patent law. See Meurer, *supra* note 11, at 97. See generally Jennifer F. Reinganum, *The Timing of Innovation: Research, Development, and Diffusion*, in 1 *HANDBOOK OF INDUSTRIAL ORGANIZATION*, *supra* note 26, at 849, 853 (reviewing literature on patent races). For the classic statement of how intellectual property can allow a firm to dominate an entire market, see JOSEPH A. SCHUMPETER, *CAPITALISM, SOCIALISM, AND DEMOCRACY* 81–87 (2d ed. 1947).

3. *The Relationship Between Market Power, Free Entry, and Profit*

As noted earlier, the idea-expression dichotomy ensures that entry into markets for copyrighted works should generally be quite easy.⁷⁰ This suggests that entry by substitutes should force prices towards marginal cost and dissipate any monopoly power that may exist. At the same time, if authors are to produce creative works, they must possess sufficient market power to charge prices that exceed marginal cost.

The traditional approach cannot adequately explain how these conditions of free entry and market power can coexist, since theoretically the influx of substitutes should dissipate authors' ability to exert any power over price. The absence of widescale monopoly profits presents a similar puzzle. Monopoly theory suggests that such profits should be endemic. Free entry, on the other hand, suggests that no profits should exist.

This apparent paradox has confounded economically sophisticated commentators on both sides of the copyright debate. For example, in his critique of price discrimination, James Boyle asserts that the coexistence of free entry and market power represents an irreconcilable internal contradiction that cannot be resolved in a principled manner.⁷¹ Edmund Kitch sounds a similar theme when he criticizes Landes and Posner for asserting that copyrighted works confront downward-sloping demand curves while acknowledging at the same time that copyrights generally do not confer monopoly power.⁷² Despite the significant differences in their views about the proper scope of copyright protection, Boyle and Kitch are united in pointing out that the simultaneous coexistence of these features represents a paradox that the traditional approach is unable to resolve.

Again, I do not mean to suggest that anyone would be surprised that copyrighted works face competition from substitutes, possess a degree of power over price, and do not typically garner supracompetitive profits. My point is that the traditional approach's inability to provide an explanation for how these features can occur at the same time suggests the possibility of a fundamental deficiency in the existing theories.

⁷⁰ See *supra* notes 14–17 and accompanying text.

⁷¹ Boyle, *supra* note 3, at 2028–29, 2032, 2037–38.

⁷² Kitch, *supra* note 5, at 1734–35 (citing Landes & Posner, *supra* note 3, at 327 & n.4, 361).

C. *The Normative Shortcomings of the Traditional Approach*

In addition to the positive shortcomings identified above, the traditional approach to copyright suffers from significant normative deficiencies. The economic analysis of the problems of nonrivalry has long been criticized for focusing solely on the allocation of already-created works without providing any basis for determining which and how many works should be created in the first place.⁷³ In other words, the traditional approach adopts an ex post perspective that focuses solely on the access to copyrighted works. In so doing, it fails to provide any means for assessing the incentives to produce creative works, which arise on an ex ante basis before authors undertake the sunk-cost investments needed to create the first copy of the work.⁷⁴

The inability of the traditional approach to provide a basis for formalizing the incentives side of the tradeoff renders any balance struck between access and incentives somewhat arbitrary. In the absence of such a basis, scholars have adopted less formal approaches. For example, some scholars have asserted a preference for the incentives side of the tradeoff on the ground that, when compounded over time, the long-run benefits associated with creating new works will exceed any short-run static inefficiency losses.⁷⁵ Other scholars have developed rough metrics to indicate whether creation of a particular work would generate net economic benefits.⁷⁶ Scholars employing these metrics acknowledge that they do not measure welfare directly and thus only can provide a vague sense of how much creative activity would be optimal.

* * *

The traditional approach to copyright therefore suffers from critical positive and normative inadequacies. Not only does it fail to provide a theoretical explanation for a number of real-world features for markets of copyrighted works; it also fails to provide a basis for evaluating the incentives side of the copyright tradeoff. I now turn to product differentiation theory, which can help remedy these descriptive and analytical shortcomings.

⁷³ See, e.g., Jora R. Minasian, *Television Pricing and the Theory of Public Goods*, 7 J.L. & ECON. 71, 73, 79 (1964).

⁷⁴ For an interesting comparison of the ex post and ex ante perspectives on nonrivalry, see D. Brennan, *Fair Price*, *supra* note 3, at 351–55.

⁷⁵ See *id.* at 355; Janusz Ordover & William Baumol, *Antitrust Policy and High-Technology Industries*, 4 OXFORD REV. ECON. POL'Y 13, 32 (1988).

⁷⁶ See Fisher, *supra* note 3, at 1707–17 (using ratio of producer surplus to deadweight loss as rough measure of tradeoff between access and incentives). Fisher adapted the basic approach from Louis Kaplow, *The Patent-Antitrust Intersection: A Reappraisal*, 97 HARV. L. REV. 1813, 1829–34 & n.54 (1984) (developing metric for assessing patent duration).

II

A BASIC ECONOMICS OF THE DIFFERENTIATED PRODUCTS APPROACH

This Part lays out the basic elements of the differentiated products approach to copyright. Section A introduces the theory of monopolistic competition, paying particular attention to the extensions of Chamberlin's work offered during the 1970s and 1980s by Avinash Dixit and Joseph Stiglitz, Oliver Hart, and Michael Spence.⁷⁷ Section B describes the basic characteristics of spatial competition.⁷⁸

A. *Monopolistic Competition*

The theory of monopolistic competition retains most of the assumptions underlying perfect competition, including free entry and the presence of a substantial number of sellers.⁷⁹ This ensures that every product group is "large" enough to justify ignoring strategic interactions and assuming instead that each producer will set prices without anticipating that its competitors will react to its pricing decisions. The key difference is that monopolistic competition relaxes the assumption that competing works are homogeneous and serve as perfect substitutes for one another. Homogeneous works compete solely on the basis of price, and authors cannot charge above the equilibrium price without losing all of their customers to their competitors. In contrast, when works are differentiated, the structure of demand for any particular work allows authors to raise their prices without losing all of their sales, because they will be able to retain those customers who most value the particular version they offer. Product differentiation thus provides each author with a degree of power over price sufficient to justify modeling each work as facing a downward-sloping demand curve.

⁷⁷ See generally Avinash K. Dixit & Joseph E. Stiglitz, *Monopolistic Competition and Optimum Product Diversity*, 67 AM. ECON. REV. 297 (1977); Oliver D. Hart, *Monopolistic Competition in the Spirit of Chamberlin: A General Model*, 52 REV. ECON. STUD. 529 (1985) [hereinafter Hart, *General Model*]; Oliver D. Hart, *Monopolistic Competition in the Spirit of Chamberlin: Special Results*, 95 ECON. J. 889 (1985) [hereinafter Hart, *Special Results*]; Michael Spence, *Product Differentiation and Welfare*, 66 AM. ECON. REV. 407 (1976) [hereinafter Spence, *Product Differentiation*]; Michael Spence, *Product Selection, Fixed Costs, and Monopolistic Competition*, 43 REV. ECON. STUD. 217 (1976) [hereinafter Spence, *Product Selection*].

⁷⁸ See *supra* note 44 and accompanying text.

⁷⁹ For formal statements of the assumptions of monopolistic competition, see BEATH & KATSOUACOS, *supra* note 26, at 45; Hart, *General Model*, *supra* note 77, at 529; Nicholas Kaldor, *Market Imperfection and Excess Capacity*, 2 ECONOMICA 33, 34-37 (1935).

Monopolistic competition theory also assumes that consumer preferences are symmetric with respect to each work in the group.⁸⁰ The primary effect of this assumption is to place each work in equal competition with all other works in the group rather than in localized competition with a smaller set of near neighbors.⁸¹

Because monopolistic competition portrays market interactions in a classic price-quantity space, it is quite easily integrated into a conventional welfare analysis.⁸² However, because each author is assumed to make independent price and quantity decisions, monopolistic competition theories model competition at the producer level rather than at the industry level. This alters the welfare analysis in several ways. When markets are modeled at the industry level, welfare implications can be determined simply by looking at whether total surplus on the relevant graph increases or decreases. When the market is analyzed at the producer level, however, total surplus depends not only on the amount of surplus generated in any particular graph, but also on the total number of graphs. In other words, economic welfare depends on the total number of works created as much as it does on the surplus generated by any particular work.⁸³ It is thus possible for total welfare to increase even though the total output of a particular producer falls, because any reduction in the surplus generated by a particular work may be offset by an increase in the total number of works produced. In addition, as I discuss in greater detail below,⁸⁴ product differentiation introduces a dimension other than price into the welfare calculus. When products are differentiated, authors can also compete by offering works that are closer to consumers' ideal preferences.

⁸⁰ See *supra* notes 44, 78 and accompanying text. Stated more formally, this assumption posits that the cross-price elasticities of demand are equal as between all works. See Kaldor, *supra* note 79, at 35 n.2. I will elaborate further on the symmetry assumption *infra* notes 81, 99–112 and accompanying text. I discuss the implications of relaxing the symmetry assumption *infra* notes 206–09 and accompanying text.

⁸¹ Chamberlin's original formulation also made a number of other symmetry assumptions, none of which turn out to be central to the analysis. For example, Chamberlin posited that each producer faced identical cost and demand curves. This allowed him to employ a single graph portraying the price-quantity response of a representative firm to model the entire market. Allowing the cost and demand curves to vary across products simply causes equilibrium price and quantity to differ with respect to each firm, which is completely reasonable given the assumption that products are differentiated. Firm-to-firm variations in price and quantity do not, however, change any essential aspects of the equilibrium. See G.C. Archibald, *Chamberlin Versus Chicago*, 29 REV. ECON. STUD. 2, 6–7 (1961); Kaldor, *supra* note 79, at 43–45.

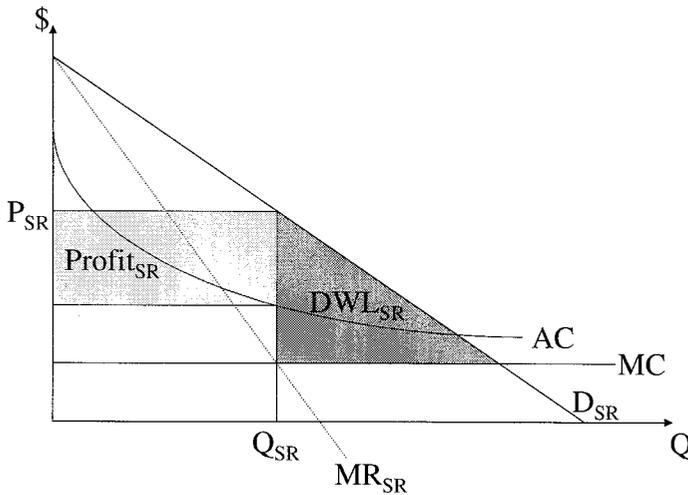
⁸² The primary drawback is that doing so requires indirect modeling of product variety.

⁸³ See Landes & Posner, *supra* note 3, at 341.

⁸⁴ See *infra* Part IV.A.1.

Profit-maximizing authors will produce at the level at which their marginal revenue and marginal cost curves intersect. Because monopolistically competitive authors face downward-sloping demand curves, in the short run they will set prices in the same manner as a monopolist, as depicted in Figure 2 (which essentially is identical to Figure 1). This results in deadweight loss. Should price exceed average cost, an author could also earn short-run supracompetitive profits.

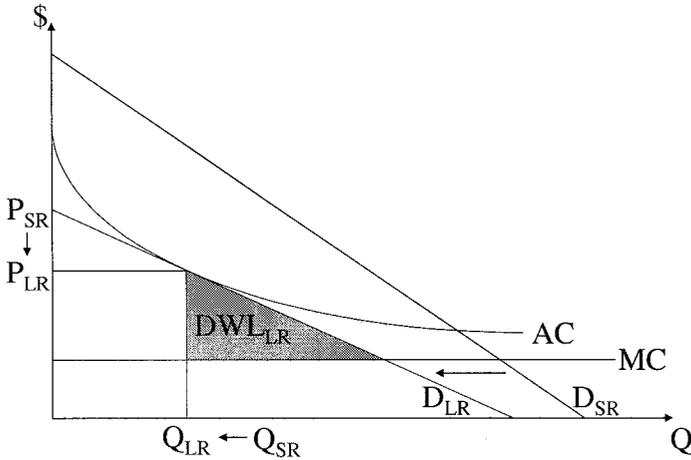
FIGURE 2: SHORT-RUN EQUILIBRIUM UNDER MONOPOLISTIC COMPETITION



Were entry impossible, the analysis would end here, just as it does under a monopoly analysis. The existence of barriers to entry would make this a stable equilibrium. Monopolistic competition, however, assumes that entry is always possible. When entry is free, the presence of supracompetitive profits attracts other authors selling close substitutes. Because all of the works in the market are in equal competition with one another, new entrants take business equally from each of the incumbents. This causes the demand curve confronting each incumbent to shift backwards, as customers substitute purchases of the new work for those of the incumbents. In addition, the demand curve confronting each author should flatten as the increase in the number of imperfect substitutes available causes demand to become

more elastic.⁸⁵ The resulting long-run equilibrium is depicted in Figure 3.⁸⁶

FIGURE 3: LONG-RUN EQUILIBRIUM UNDER MONOPOLISTIC COMPETITION



Entry continues to divide the available surplus into increasingly smaller fragments until no profits remain. Under Chamberlin’s original formulation, this occurs when the surplus appropriated by each author is just enough to cover the fixed costs of entry, a condition which exists when the demand curve is tangent to the average cost curve.⁸⁷ The equilibrium number of works can be determined by dividing the total surplus associated with the entire market by the fixed costs needed for entry by an additional work.⁸⁸ Indeed, as the

⁸⁵ This flattening of the demand curve distinguishes the differentiated products approach from dominant firm models. See *supra* note 23 and accompanying text. Because the residual demand in dominant firm models is simply the market demand minus the quantity supplied by the competitive fringe, entry by competitors simply causes the demand curve facing the dominant firm to shift backwards without any change in slope.

⁸⁶ See BEATH & KATSOUALACOS, *supra* note 26, at 136–37; LIPSEY ET AL., *supra* note 58, at 264; Oliver D. Hart, *Monopolistic Competition in a Large Economy with Differentiated Commodities*, 46 REV. ECON. STUD. 1, 15 (1979).

⁸⁷ See CHAMBERLIN, *supra* note 42, at 194–95. The importance of the tangency solution and some important qualifications to this conclusion are discussed *infra* note 139 and accompanying text.

⁸⁸ For example, if a particular market generated \$100 in surplus and the fixed costs needed to enter were \$5, entry would divide the surplus into increasingly small parts until it exactly equaled the fixed costs, at which point there would be twenty works in the market.

The existence of a maximum limit to the number of players caused Oliver Hart to observe that this type of model is more properly regarded as oligopolistic rather than monopolistically competitive. See Hart, *General Model*, *supra* note 77, at 529; Hart, *Special Results*, *supra* note 77, at 889–90. Hart then proceeded to model monopolistic competition in a way that ensured that every incumbent continued to face competition from

size of the market expands or the size of the fixed costs declines, the number of works asymptotically approaches infinity and the dead-weight loss approaches zero.⁸⁹ Whether an economy is “large” in this manner does not depend upon the size of the fixed costs relative to the size of the marginal costs, as suggested by the traditional approach; it depends on the magnitude of the fixed costs relative to the overall market.⁹⁰ There is, however, a well-known exception to Chamberlin’s zero-profit result.⁹¹ The lumpiness of fixed costs may create a situation in which n works would earn small profits while $n + 1$ works would run losses. This so-called “integer problem” allows for an equilibrium in which n works each earn sustainable profits. So long as the economy is sufficiently “large” (i.e., so long as n is relatively sizeable), such profits will be negligible.

The theory’s impact has been limited somewhat by the fact that large-group, monopolistically competitive markets are relatively rare.⁹² If fixed costs are sufficiently high relative to the size of the overall market, the number of sellers will be so small that it is no longer realistic to assume that competitors will not strategically react to other producers’ pricing decisions. When that is the case, differen-

potential entrants. Even if the total number of potential entrants is bounded, equilibrium prices will tend to approach competitive levels so long as the relevant market is significantly larger than the fixed costs of entry. See *infra* notes 89–90 and accompanying text.

⁸⁹ This is a general result (often called the “Folk Theorem”) that applies across markets without regard to their cost structure or the nature of competition. For a classic proof of this result in the context of Cournot competition with free entry and exit as well as U-shaped cost curves, see William Novshek, *Cournot Equilibrium with Free Entry*, 47 *REV. ECON. STUD.* 473 (1980).

⁹⁰ See BEATH & KATSOLACOS, *supra* note 26, at 137, 140; Hart, *supra* note 86, at 1–2; Hart, *Special Results*, *supra* note 77, at 899; Larry E. Jones, *The Efficiency of Monopolistically Competitive Equilibria in Large Economies: Commodity Differentiation with Gross Substitutes*, 41 *J. ECON. THEORY* 356, 356–59 (1987); N. Gregory Mankiw & Michael D. Whinston, *Free Entry and Social Inefficiency*, 17 *RAND J. ECON.* 48, 56–57 (1986).

⁹¹ This effect was first identified by Nicholas Kaldor. See Kaldor, *supra* note 79, at 42–43.

⁹² Other criticisms of monopolistic competition are less compelling. Chicago School advocates condemned it as an attack on the free market system. See Benjamin Klein, *Market Power in Antitrust: Economic Analysis After Kodak*, 3 *SUP. CT. ECON. REV.* 43, 75 & n.60 (1993). As this Article demonstrates, monopolistic competition is simply an analytical tool that may or may not have deregulatory implications. Other critics have decried monopolistic competition’s inability to generate simple and testable policy inferences. See, e.g., Archibald, *supra* note 81, at 9–15. As Paul Samuelson has noted, a theory should be measured by the insights that it provides, not for its elegance or for the simplicity of the policy inferences it is able to generate. See Paul A. Samuelson, *The Monopolistic Competition Revolution*, in *MONOPOLISTIC COMPETITION THEORY: STUDIES IN IMPACT* 105, 108 & n.5 (Robert E. Kuenne ed., 1967); see also EDWARD HASTINGS CHAMBERLIN, *TOWARDS A MORE GENERAL THEORY OF VALUE* 304–05 (1957) (noting that “there might also be simplifying assumptions other than those identified with perfect competition”).

tiated products markets are better modeled as oligopolies.⁹³ Nevertheless, given the relative ease of generating additional creative works, and the large market available for creative works, markets for many types of copyrighted works likely are sufficiently competitive to justify treating them as monopolistically competitive.⁹⁴ Indeed, Chamberlin himself regarded copyright as one of the paradigmatic examples of product differentiation under his theory.⁹⁵

B. Spatial Competition

Because monopolistic competition depicts market interactions in the classic price-quantity space of microeconomics, product differentiation cannot be portrayed directly. Spatial competition models, in contrast, follow the opposite tack, making product differentiation the primary variable without attempting to model variations in price and quantity directly. The original formulation assumed that, rather than compete on price, firms instead vie for business by choosing a location along a linear geographic space. One common example imagines hot dog vendors locating themselves along a beach, while another imagines stores positioning themselves along the main street of a city

⁹³ See LIPSEY ET AL., *supra* note 58, at 265; Hart, *General Model*, *supra* note 77, at 529; Hart, *Special Results*, *supra* note 77, at 889–90.

⁹⁴ In addition, several regulatory responses are available to markets that are too concentrated. Such responses will tend to mitigate this objection by rendering competition in those segments more robust. See *infra* Part V.A.

⁹⁵ CHAMBERLIN, *supra* note 42, at 58–59; see also Lunney, *supra* note 7, at 582–89 (applying monopolistic competition theory to copyright). Although Chamberlin notes that intellectual property rights might place some restrictions on entry, his primary concern appears to be with respect to patents and trademarks. See CHAMBERLIN, *supra* note 42, at 111–12. In any event, as noted earlier, to the extent that intellectual property rights create cost asymmetries among works without forestalling entry altogether, they simply alter the equilibrium price and quantity with respect to a particular work without causing any fundamental change in the nature of the equilibrium. See *supra* note 81.

Interestingly, Glynn Lunney expands the analysis of substitutability outside the realm of copyright by arguing that copyright protection, if too strong, can impose opportunity costs by diverting resources from more economically beneficial activities. See Lunney, *supra* note 7, at 488–89. In so doing, he builds on an insight first offered by Pigou and reiterated by Arnold Plant. See A.C. PIGOU, *THE ECONOMICS OF WELFARE* 169 n.1 (2d ed. 1924) (“[I]nventions may actually diminish aggregate economic welfare; for they may cause labour to be withdrawn from other forms of productive service to make a new variety of some article to supersede an old one”); Plant, *supra* note 5, at 183–84 (“Monopoly is, of course, a common enough device for securing in this way the diversion of scarce resources to particular uses. . . . What is generally overlooked by the more enthusiastic advocates of these schemes is the alternative output which the resources would have yielded in other employment.”). This point is well taken so long as the overall market is already in general equilibrium. If not, it is quite possible that the distortions caused by strong copyright protection might well be welfare enhancing. As a general matter, discerning which is the case requires the simultaneous solution of an intractably large number of equations.

or a railroad. Because of transportation costs, customers derive greater utility from purchasing from sellers that are closer to their locations. Utility declines as the distance from the store increases until the entire surplus is completely consumed by transportation costs, at which point the customer decides not to purchase from that vendor. Spatial differentiation thus gives sellers a degree of power over price, as they can increase price without losing those customers who are situated closest.⁹⁶

Economists quickly recognized that the same framework could be used to model competition among products distributed along a characteristics space rather than a geographic space. Manufacturers of apple cider, for example, could face a decision of where to produce along a spectrum of product characteristics running from sweet to sour;⁹⁷ or to use an example involving copyrighted works, popular music can be conceived as occupying a spectrum from "hard rock" to "easy listening." Under this approach, customers decide whether to purchase a particular work based on how closely that work matches their ideal preferences. The decline in utility represented by transportation costs in geographic location models is replaced by divergence from a consumer's preferred characteristics.

Examples of two such product spaces are depicted in Figure 4. The horizontal dimension depicts where along the continuum of characteristics a particular work is located. The vertical dimension in the graph represents the net surplus available from consumers occupying any particular location. Consumers' ideal preferences are assumed to be distributed uniformly across the characteristics space. Each product is produced by a different firm, and the surplus captured by each work is depicted by a triangle. The decline in utility resulting from the good's divergence from the consumer's ideal preferences is represented by the slope of the triangle's sides.⁹⁸ The slope of this line is determined by the structure of demand, as reflected in the relevant cross-price elasticities. If a work serves as a relatively good substitute for similar works, the slope will be relatively flat. If not, the slope will be relatively steep.

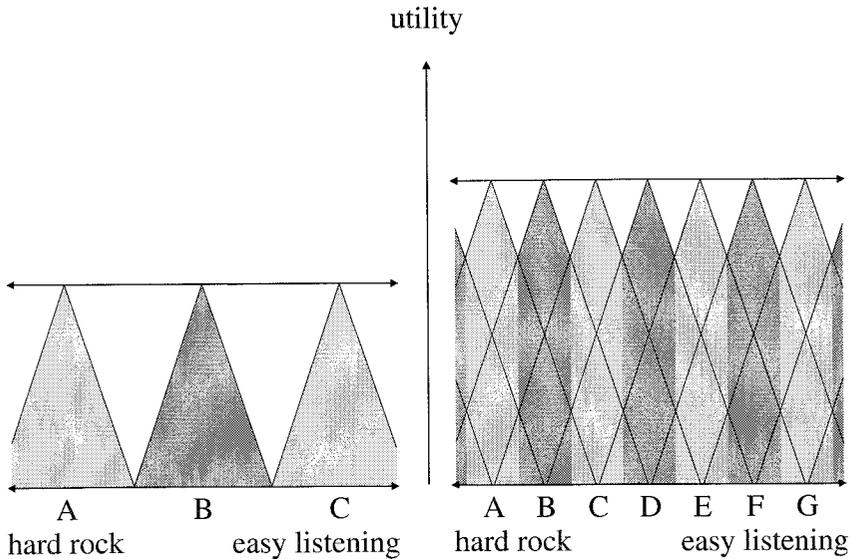
Consumers are assumed to purchase whichever work lies closest to their ideal preferences. As noted earlier, the symmetric preference

⁹⁶ Hotelling, *supra* note 43, at 44-45.

⁹⁷ *Id.* at 54; *see also* Kaldor, *supra* note 79, at 37.

⁹⁸ Although the assumption that utility falls linearly as distance from a work increases represents the standard approach in the literature, it is a somewhat strong assumption. One can obtain radically different results by varying the shape of this function. *See, e.g.*, C. d'Aspremont et al., *On Hotelling's "Stability in Competition,"* 47 *ECONOMETRICA* 1145 (1979) (modeling transportation costs as quadratic function).

FIGURE 4: IMPACT OF FIXED COSTS ON THE EQUILIBRIUM UNDER SPATIAL COMPETITION



assumption posits that all works within a group are in equal competition with one another and that entry by a new work takes business from all incumbents evenly.⁹⁹ This assumption is represented in spatial models either by positing that all works enter simultaneously or by assuming that incumbent works respond to product entry by shifting their position. The result is an equilibrium in which works are evenly distributed across the relevant product space.¹⁰⁰

⁹⁹ See *supra* notes 44, 78 and accompanying text.

¹⁰⁰ See Steven C. Salop, *Monopolistic Competition with Outside Goods*, 10 BELL J. ECON. 141, 143, 145 & n.3 (1979); cf. Samuelson, *supra* note 92, at 125–26 & n.19 (comparing results under sequential entry and optimal entry in which products are evenly distributed across characteristics space). This is tantamount to assuming that works do not have sunk costs in any particular location. Interestingly, the assumption that prices are constant and endogenously determined can give rise to a different equilibrium in which works, rather than being spread evenly across the characteristics space, divide the market by entering at the same point. See Hotelling, *supra* note 43, at 51–54, 57. The oddities resulting from refusing to allow for price competition are well recognized. For example, the assumption that firms do not compete on price necessarily implies that effective competition exists only with respect to consumers located equidistantly from two works. See SCHERER & ROSS, *supra* note 26, at 502–03. This effect disappears, however, if the model is broadened to allow for endogenous pricing in which price competition is possible. It also disappears if there are sunk costs in particular locations and works are unable to shift positions in response to entry. See CHURCH & WARE, *supra* note 26, at 384–88, 395–404. In any event, such minimum-differentiation solutions presumably are preempted in the context of copyright, in which multiple works are not allowed to occupy the exact same location.

As was the case under monopolistic competition,¹⁰¹ entry by additional works divides the available surplus into increasingly smaller fragments until the size of those fragments equals the fixed costs of entry, at which point no work earns supracompetitive profits.¹⁰² Again, the equilibrium number of works can be determined by dividing the total surplus by the fixed costs of entry.¹⁰³ If the total surplus is significantly greater than the fixed costs of entry, the economy will be relatively "large" and the competition among sellers will be relatively substantial. As the available surplus increases or the size of the fixed costs approaches zero, the number of works will approach infinity and all customers will purchase works exactly matched to their tastes.¹⁰⁴ If the total surplus is small relative to the fixed costs, the economy will be relatively "small" and works may well enjoy a degree of local monopoly or oligopoly power. Although entry by competitors remains possible under these conditions, the volume may be too low to support additional sellers, as demonstrated by the local monopolies enjoyed by gas stations located along thinly populated stretches of highway.

The two different depictions in Figure 4 illustrate how the relative size of the total available surplus affects the degree of competition. The available surplus in the right-hand graph is fifty percent larger than that in the left-hand graph. In both graphs, each individual work captures an identical surplus. As the graphs illustrate, increasing the size of the overall market yields a fairly substantial increase in the degree of competition. More sophisticated models that allow for price reactions among competitors further underscore the importance of the relative size of the economy.¹⁰⁵ These models demonstrate that so long as marginal costs are nonzero, increases in the size of the

¹⁰¹ See *supra* note 88 and accompanying text.

¹⁰² This result is subject to the integer problem resulting from fixed cost indivisibilities. See *supra* note 91 and accompanying text.

¹⁰³ As was the case under monopolistic competition, the existence of a maximum limit to the number of possible players arguably renders the ensuing competition oligopolistic. See Hart, *General Model*, *supra* note 77, at 529; Hart, *Special Results*, *supra* note 77, at 889-90 & n.1. The robustness of the competition depends upon the number of entrants that the relevant market is able to support.

¹⁰⁴ See Eaton & Lipsey, *supra* note 26, at 761; B. Curtis Eaton & Myrna Holtz Wooders, *Sophisticated Entry in a Model of Spatial Competition*, 16 RAND J. ECON. 282, 290-92 (1985).

¹⁰⁵ Most early analyses assumed that competitors would not change their prices in reaction to entry. See, e.g., B. Curtis Eaton, *Free Entry in One-Dimensional Models: Pure Profits and Multiple Equilibria*, 16 J. REGIONAL SCI. 21, 22 (1976); B. Curtis Eaton & Richard G. Lipsey, *Freedom of Entry and the Existence of Pure Profit*, 88 ECON. J. 455, 461 (1978); W.J. Lane, *Product Differentiation in a Market with Endogenous Sequential Entry*, 11 BELL J. ECON. 237, 239 (1980); William Novshek, *Equilibrium in Simple Spatial (or Differentiated Product) Models*, 22 J. ECON. THEORY 313, 315 (1980). For a discussion of

economy cause prices to approach marginal cost and reduce profits by bringing the revenue captured by each author more into line with fixed costs.¹⁰⁶

Spatial competition models provide for a more intuitive presentation of the impact of product diversity, although the fact that these models do not employ the price-quantity space of neoclassical economics means that price competition must be modeled separately and makes it more difficult to integrate spatial models into conventional analyses of economic welfare. The foregoing exposition only presents the basic intuitions underlying spatial competition models and is subject to a number of qualifications and refinements. Not all works can be organized into a simple spectrum of characteristics. Genres such as jazz and classical music, for example, may demand a more multidimensional approach to product differentiation.¹⁰⁷

Furthermore, the analysis becomes significantly more complicated when one relaxes the rather restrictive assumptions that typify the basic model of spatial competition described above. More refined models allow for the possibility of sunk costs in location and sequential entry.¹⁰⁸ Other models relax the assumption that prices are fixed and instead allow prices to be determined endogenously.¹⁰⁹ Still other models allow for the possibility that a single firm might produce multiple works occupying multiple locations¹¹⁰ or that consumers base their purchases on multiple dimensions of characteristics.¹¹¹ Finally,

the complications associated with assuming away price undercutting, see SCHERER & ROSS, *supra* note 26, at 502–08; Eaton & Lipsey, *supra* note 26, at 748–49.

¹⁰⁶ See Eaton & Wooders, *supra* note 104, at 291–95.

¹⁰⁷ As I subsequently discuss, multidimensionality can have profound effects on the welfare characteristics of the differentiated products equilibrium. See *infra* notes 111, 210–11 and accompanying text.

¹⁰⁸ See William J. Baumol, *Calculation of Optimal Product and Retailer Characteristics: The Abstract Product Approach*, 75 J. POL. ECON. 674, 679 n.4 (1967); Giacomo Bonanno, *Location Choice, Product Proliferation and Entry Deterrence*, 54 REV. ECON. STUD. 37 (1987); B. Curtis Eaton & Richard G. Lipsey, *Exit Barriers Are Entry Barriers: The Durability of Capital as a Barrier to Entry*, 11 BELL J. ECON. 721 (1980); D.A. Hay, *Sequential Entry and Entry-Detering Strategies in Spatial Competition*, 28 OXFORD ECON. PAPERS 240 (1976); Lane, *supra* note 105; Damien J. Neven, *Endogenous Sequential Entry in a Spatial Model*, 5 INT'L J. INDUS. ORG. 419 (1987); Edward C. Prescott & Michael Visscher, *Sequential Location Among Firms with Foresight*, 8 BELL J. ECON. 378 (1977).

¹⁰⁹ See Eaton & Wooders, *supra* note 104, at 282–83; Salop, *supra* note 100, at 143–45.

¹¹⁰ See James A. Brander & Jonathan Eaton, *Product Line Rivalry*, 74 AM. ECON. REV. 323 (1984); B. Curtis Eaton & Richard G. Lipsey, *The Theory of Market Pre-emption: The Persistence of Excess Capacity and Monopoly in Growing Spatial Markets*, 46 ECONOMICA 149 (1979); Kenneth L. Judd, *Credible Spatial Preemption*, 16 RAND J. ECON. 153 (1985); Richard Schmalensee, *Entry Deterrence in the Ready-To-Eat Breakfast Cereal Industry*, 9 BELL J. ECON. 305 (1978).

¹¹¹ See G.C. Archibald & G. Rosenbluth, *The "New" Theory of Consumer Demand and Monopolistic Competition*, 89 Q.J. ECON. 569 (1975); Robert C. Feenstra & James A.

the results change significantly when one allows for the possibility that consumer preferences are not distributed equally across the characteristics space.¹¹² The basic model is nonetheless sufficient to capture the key intuitions about how differentiated products compete and to provide useful insights into markets for copyrighted works.

III

THE DESCRIPTIVE ADVANTAGES OF THE DIFFERENTIATED PRODUCTS APPROACH

Compared with the traditional approach, the predictions of the differentiated products approach to copyright fit better with features of real-world markets for copyrighted works. The equilibrium under monopolistic competition is particularly helpful in providing an explanation for each of the key features identified in Part I.B for which the traditional approach is unable to account.

A. *The Noncentrality of Nonrivalry*

One of the most interesting aspects of the differentiated products approach is that it provides an explanation for the tension between access and incentives that is completely independent of the assumption that consumption of copyrighted works is nonrival. According to the conventional wisdom, the fundamental pricing problem arises from the assumption that marginal costs are zero and that average cost therefore exceeds marginal cost across all volumes.¹¹³

The differentiated products approach offers a different explanation for the relationship between average cost and marginal cost. Monopolistically competitive industries reach long-run equilibrium at a point where the demand curve is tangent to the average cost curve.¹¹⁴ The fact that demand curves are downward sloping implies that this equilibrium necessarily occurs where the average cost curve is downward sloping as well, which in turn implies that the average cost curve must be above the marginal cost curve.

This suggests that as long as products are differentiated, the conflict between access and incentives identified in the current literature will arise whether consumption is rival or nonrival. Indeed, as illus-

Levinsohn, *Estimating Markups and Market Conduct with Multidimensional Product Attributes*, 62 REV. ECON. STUD. 19 (1995).

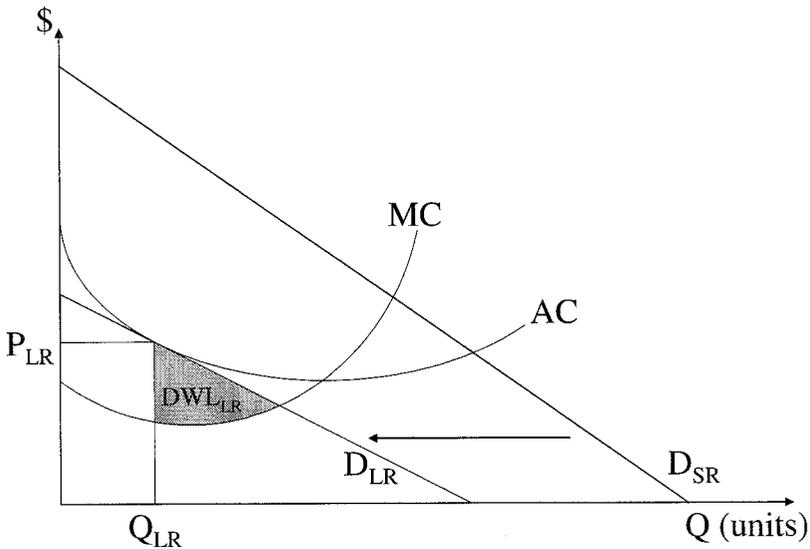
¹¹² The seminal analysis of the impact of preference asymmetries was offered by Kaldor. See Kaldor, *supra* note 79, at 37–40. For more recent embellishments on this insight, see B. Curtis Eaton & Richard G. Lipsey, *The Non-Uniqueness of Equilibrium in the Löschian Location Model*, 66 AM. ECON. REV. 77 (1976); Waterson, *supra* note 28.

¹¹³ See *supra* notes 46–57 and accompanying text.

¹¹⁴ See *supra* note 87 and accompanying text.

trated in Figure 5, the equilibrium positions of the average and marginal cost curves under monopolistic competition are remarkably similar to those customarily used to model the economics of copyright. It is thus no accident that the leading monopolistic competition analyses model the marginal cost curve as flat and lying below the average cost curve in precisely the same manner as occurs with nonrival goods.¹¹⁵

FIGURE 5: LONG-RUN EQUILIBRIUM UNDER MONOPOLISTIC COMPETITION WHEN CONSUMPTION IS RIVAL



Thus, nonrivalry in consumption may not play as central a role in the analysis of copyright as generally believed. This explains why copyrighted works pose difficult pricing problems even when copying is expensive or when the copyrighted material must be combined with rival inputs before the final product can be sold to consumers. My point is not that nonrivalry has no effect on the resulting equilibrium. On the contrary, nonrivalry will determine the shape and position of the average cost curves, which in turn play a critical role in determining the magnitude of the deadweight loss. My point is simply that the equilibrium will exhibit the same tension between access and

¹¹⁵ See, e.g., BEATH & KATSOUACOS, *supra* note 26, at 42–46, 52–55; Dixit & Stiglitz, *supra* note 77, at 298–99; Eaton & Lipsey, *supra* note 26, at 728–29; Spence, *Product Differentiation*, *supra* note 77, at 409 fig.1, 411. The discussions of monopolistic competition in many leading industrial organization textbooks also make the same assumption. See, e.g., CARLTON & PERLOFF, *supra* note 26, at 202–07 & figs.7.1–7.2; CHURCH & WARE, *supra* note 26, at 377–78 & figs.11.3–11.4.

incentives regardless of whether consumption is rival or nonrival. Although nonrivalry clearly exacerbates matters, it is not a necessary condition for that tension to arise.

B. *The Absence of Natural Monopoly*

Product differentiation also explains why the presence of unexhausted (and inexhaustible) economies of scale does not cause markets for copyrighted works to devolve into natural monopolies. As noted earlier, when competition involves homogeneous products, constantly declining average costs give rise to a Schumpeterian, winner-take-all competition in which intellectual property allows one author to drive its competitors from the market.¹¹⁶

A different situation obtains when products are differentiated. The fact that works serve as imperfect substitutes for one another allows producers to coexist even when faced with increasing returns to scale, as exhibited by the fact that monopolistically competitive markets reach an equilibrium in which multiple producers operate on the declining portion of their average cost curves. Variations in consumer preferences allow producers to survive by targeting those segments of the overall customer base that place the highest value on their particular version of the good.¹¹⁷

Stated in terms of the torts casebook example discussed above,¹¹⁸ product differentiation explains why one casebook does not necessarily monopolize the market even though its larger volume would almost certainly allow it to underprice the competition. Authors of other casebooks remain free to offer differentiated products—by covering different topics, emphasizing different perspectives, drawing upon different materials, or posing different questions—and market them to those who most highly value each particular set of features.¹¹⁹

Product differentiation thus helps prevent the presence of unexhausted economies of scale from causing markets for copyrighted

¹¹⁶ See *supra* note 69 and accompanying text.

¹¹⁷ See Yoo, *supra* note 2, at 1603, 1657–59. This is the supply-side analog to the manner in which customer heterogeneity can mitigate the demand-side economies of scale resulting from network economic effects. See Daniel F. Spulber & Christopher S. Yoo, *Access to Networks: Economic and Constitutional Connections*, 88 CORNELL L. REV. 885, 928 (2003); Yoo, *supra* note 16, at 272, 280–81.

¹¹⁸ See *supra* Part I.B.2.

¹¹⁹ This explains how HBO has been able to generate one-half the revenue of CBS despite having an audience base that is five times smaller, see Yoo, *supra* note 2, at 1679–80, and the key role that allowing broadband providers to customize their networks for different types of applications might play in preventing any one carrier from monopolizing broadband services, see Christopher S. Yoo, *Fighting Traffic on the Disinformation Superhighway*, THE TENNESSEAN (Nashville), July 8, 2003, at 7.

products to collapse into natural monopolies.¹²⁰ The possibility that a work may compete on factors other than cost makes it possible for a subsequent entrant to survive despite the presence of cost disadvantages that would be insuperable if the works were homogeneous.¹²¹

C. *The Relationship Between Market Power, Free Entry, and Profit*

The differentiated products approach also provides an explanation for how authors are able to retain sufficient market power to charge prices that exceed marginal cost despite facing competition from new entrants.¹²² This is because under the differentiated products approach, the requisite power over price comes from imperfections in substitutability (represented in economic terms by negative, nonunitary cross-price elasticities of demand). These imperfections in substitutability cannot be completely dissipated no matter how much entry occurs. As a result, every author retains a degree of market power despite the fact that other authors remain free to create works with the same functional characteristics. Were it otherwise, competition would eliminate authors' ability to generate sufficient revenues to cover the fixed costs incurred in producing their works in the first

¹²⁰ The manner in which product differentiation can prevent markets from devolving into natural monopolies also provides an explanation of how copyright can avoid the destructive races to capture the entire market identified by patent scholarship. See *supra* note 69.

¹²¹ See Maurer & Scotchmer, *supra* note 28, at 540–41 (offering similar observation with respect to patents).

¹²² The most complete exploration of the persistence of market power in the face of free entry occurs in the literature analyzing price discrimination in competitive markets. The seminal contribution on competitive price discrimination is Daniel F. Spulber, *Non-Cooperative Equilibrium with Price Discriminating Firms*, 4 ECON. LETTERS 221 (1979). For subsequent work, see DANIEL F. SPULBER, REGULATION AND MARKETS 544–48 (1989); Mark Armstrong & John Vickers, *Competitive Price Discrimination*, 32 RAND J. ECON. 579 (2001); Severin Borenstein, *Price Discrimination in Free-Entry Markets*, 16 RAND J. ECON. 380 (1985); Kenneth S. Corts, *Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment*, 29 RAND J. ECON. 306 (1998); Peter C. Coyte & C. Robin Lindsey, *Spatial Monopoly and Spatial Monopolistic Competition with Two-Part Pricing*, 55 ECONOMICA 461 (1988); Thomas J. Holmes, *The Effects of Third-Degree Price Discrimination in Oligopoly*, 79 AM. ECON. REV. 244 (1989); Michael L. Katz, *Price Discrimination and Monopolistic Competition*, 52 ECONOMETRICA 1453 (1984); Klein, *supra* note 92, at 74–78 & n.59; Daniel F. Spulber, *Competition and Multiplant Monopoly with Spatial Nonlinear Pricing*, 25 INT'L ECON. REV. 425 (1984); Daniel F. Spulber, *Spatial Nonlinear Pricing*, 71 AM. ECON. REV. 923 (1981); Lars A. Stole, *Nonlinear Pricing and Oligopoly*, 4 J. ECON. & MGMT. STRATEGY 529 (1995). For recent surveys, see Einer Elhauge, *Why Above-Cost Price Cuts to Drive Out Entrants Are Not Predatory—and the Implications for Defining Costs and Market Power*, 112 YALE L.J. 681, 732–43 (2003); Lars A. Stole, *Price Discrimination and Imperfect Competition* (Dec. 22, 2003) (unpublished manuscript), available at <http://gsblas.uchicago.edu/papers/hio.html>. For discussions in the antitrust law context, see Symposium, *Competitive Price Discrimination*, 70 ANTITRUST L.J. 593 (2003).

place. Market power is thus more properly regarded as coming from the product differentiation inherent in the underlying market than from the exclusivity granted by copyright.

Although free entry may not be sufficient to dissipate market power, it may be sufficient to dissipate any short-run profits. As demonstrated earlier, the existence of short-run profits attracts long-run entry by close substitutes until no such profits remain.¹²³ Any supracompetitive returns earned in the short run will ultimately accrue to the benefit of consumers in the form of increased product variety. At the same time, the differentiated products equilibrium provides a self-regulating mechanism for ensuring that authors do not face such excessive entry that they are no longer able to capture sufficient surplus to cover their first-copy costs. There is thus no need to calibrate the scope of copyright protection to protect against the danger that entry will proceed to the point where the surplus captured will no longer support creation of the work.¹²⁴ The free entry made possible by the idea-expression dichotomy should ensure that copyrighted works capture sufficient revenue to cover their fixed costs and no more.

The differentiated products approach thus resolves another paradox under the traditional approach: how market power can exist without allowing most works to earn supracompetitive returns. This revelation highlights the importance of analyzing markets in terms other than just the opposing poles of perfect competition and pure monopoly. Simply put, models of imperfect competition sever the apparent tie between market power and supracompetitive profits that seems inevitable when viewed solely through the lens of those polar models.

Finally, the differentiated products approach to copyright largely renders moot the objection that strengthening copyright protection and facilitating price discrimination raise distributional concerns.¹²⁵ Instead, the possibility of free entry breaks the link between market power, supracompetitive profits, and large deadweight losses posited by monopoly theory. Although fixed-cost indivisibilities may create an "integer problem" that allows some degree of sustainable profits,

¹²³ See *supra* notes 87, 102 and accompanying text.

¹²⁴ See Bell, *supra* note 57, at 588 n.142 (raising concern that "[s]ubstitutability across copyrighted works presents would-be monopolists with frustrating flat demand curves, forcing access prices back down towards marginal costs"); cf. Grady & Alexander, *supra* note 31 (arguing that patent doctrine should be tailored to prevent excessive "rent dissipation").

¹²⁵ See *supra* notes 7, 60 and accompanying text.

any such profits will be insignificant so long as the overall market is large relative to the fixed costs of entry.¹²⁶

IV

THE NORMATIVE IMPLICATIONS OF THE DIFFERENTIATED PRODUCTS APPROACH

In addition to offering a better description of many observed features of markets for creative works, the differentiated products approach also provides new insights into the welfare analysis of copyright. Most critically, the differentiated products approach models the dynamics of entry by substitute works, an aspect that the traditional approach is poorly situated to take into account. It demonstrates that widescale access to copyrighted works may be promoted through means other than reducing the level of copyright protection. Specifically, it reveals that access instead may be promoted by stimulating entry, which in turn requires the strengthening of copyright protection.

The differentiated products approach also rectifies the central analytical deficiency of the traditional approach by providing a basis for formalizing the incentive side of the access-incentives tradeoff. In this respect, the differentiated products approach is reminiscent of the key insight from classic property theory that efficient levels of investment depend upon the existence of well-defined property rights.

Together these insights falsify the claim that simultaneous promotion of access and incentives is impossible and that copyright necessarily devolves into a tradeoff between the two. The supposed tension between access and incentives turns out to be nothing more than an artifact of the traditional approach's reliance on monopoly and oligopoly models that fail to account for entry. The differentiated products approach reveals that encouraging entry can promote both types of efficiency simultaneously.

This conclusion is subject to an important qualification discussed below in Section B.2. When fixed costs are large relative to marginal costs, property rights may be so strong as to attract entry even when such entry is inefficient. This analysis suggests, contrary to classic property theory, that it is theoretically possible for intellectual property rights to be too strong.

¹²⁶ See *supra* note 91 and accompanying text.

A. *A New Perspective on Access*

The shift to a differentiated products approach creates two significant changes in the analysis of the extent to which markets allow optimal access. First, it adds an additional dimension to the welfare calculus—product variety—thereby going beyond the total surplus represented in a price-quantity space. Second, it underscores how static efficiency can be promoted not by adjusting price directly, but rather by adjusting it indirectly through the facilitation of entry by substitutes. So long as entry is easy, the resulting static efficiency losses should be relatively small.

1. *The Multidimensionality of Welfare Under Product Differentiation*

The fact that markets for differentiated products necessarily reach long-run equilibrium at prices that exceed marginal cost suggests that some degree of static inefficiency is endemic to markets for differentiated products. In addition, the fact that monopolistically competitive equilibrium does not minimize average cost led Chamberlin to conclude that these markets necessarily operate with “excess capacity.”¹²⁷

Later theorists pointed out that these initial conclusions were too simplistic and failed to take into account the full implications of product differentiation.¹²⁸ When products are homogeneous, authors can compete only on a single dimension—price—which greatly simplifies the welfare analysis by limiting it to the cumulative spread between reservation prices and actual prices (i.e., total surplus). The welfare calculus changes dramatically when the competing works are differentiated. In such markets, works compete not only by offering cheaper prices, but also by incorporating attributes that come closer to particular customers’ ideal preferences. The multidimensionality of this competition makes simple price-cost comparisons an incomplete way to determine social welfare.

As a result, product differentiation raises the possibility that any deadweight losses caused by nonmarginal cost pricing might be offset

¹²⁷ See CHAMBERLIN, *supra* note 42, at 104–09. It is this fact that led Chamberlin to regard market failure as endemic. *Id.* at 109 (pointing to systematic excess capacity as explanation for what was often called “wastes of competition”).

¹²⁸ See BEATH & KATSOUACOS, *supra* note 26, at 62–63; Robert L. Bishop, *Monopolistic Competition and Welfare Economics*, in *MONOPOLISTIC COMPETITION THEORY*, *supra* note 92, at 251, 261; Harold Demsetz, *The Nature of Equilibrium in Monopolistic Competition*, 67 J. POL. ECON. 21, 22 (1959); Dixit & Stiglitz, *supra* note 77, at 301–02; Mankiw & Whinston, *supra* note 90, at 49, 54–55; Spence, *Product Differentiation*, *supra* note 77, at 411.

in part by welfare gains resulting from product variety. Although entry appears to be a waste of resources when products are homogeneous,¹²⁹ when products are differentiated it may actually reflect the market's attempt to satisfy differences in consumers' preferences.¹³⁰

2. *Entry as a Means for Reducing Deadweight Loss*

Even more importantly, the differentiated products approach highlights the existence of a different way to reduce deadweight loss that the traditional approach largely ignores. Specifically, it illustrates how access can be promoted not by lowering the degree of protection provided by copyright, but rather by facilitating entry by similar works. Entry by near substitutes causes the demand curve facing each work to shift inwards and flatten,¹³¹ reducing the spread between price and marginal cost that is the cause of deadweight loss (as revealed by a comparison of Figures 2 and 3 above).¹³² Although imperfections in substitutability necessarily prevent price from completely converging to marginal cost, the ready availability of substitutes should cause any remaining deadweight losses to be relatively small. The larger the degree of entry, the smaller the deadweight losses will be.¹³³

In this manner, the differentiated products approach opens up the policy space by identifying how entry can promote access, an insight that the traditional approach is ill equipped to consider. When nonrival goods are homogeneous, entry is unnecessarily duplicative

¹²⁹ See *infra* note 134.

¹³⁰ See KELVIN LANCASTER, VARIETY, EQUITY, AND EFFICIENCY 14 (1979) (concluding that it is better to have increased product variety from producing more goods at outputs below minimum average cost level than to produce goods at minimum average cost level); J.N. ROSSE, MONOPOLISTIC COMPETITION, EXCESS CAPACITY, AND INEFFICIENCY (Stanford Univ. Dept. of Econ. Studies in Indus. Econ., Working Paper No. 74, n.d.); Dixit & Stiglitz, *supra* note 77, at 301–02; see also CHAMBERLIN, *supra* note 42, at 94 (calling equilibrium under monopolistic competition “a sort of ideal”); BEATH & KATSOULACOS, *supra* note 26, at 61 (confirming this conclusion).

¹³¹ See *supra* note 85 and accompanying text.

¹³² Those unaccustomed to viewing markets through the lens of product differentiation may find aspects of this new equilibrium somewhat counterintuitive. For example, the fact that demand is shifting backwards suggests that the output and total surplus generated by any particular work will fall. The key to unlocking this puzzle is to recall that modeling market interactions at the producer rather than the industry level makes total welfare a function of the number of works created as well as the surplus generated by any particular work. The welfare gains from new entry should more than compensate for the reductions in surplus generated by any particular product. See *supra* note 83 and accompanying text.

¹³³ See BEATH & KATSOULACOS, *supra* note 26, at 52–54, 140–47; ROBERT S. PINDYCK & DANIEL L. RUBINFELD, MICROECONOMICS 424–29 (5th ed. 2001); William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 481 (2003).

and simply wastes resources.¹³⁴ Moreover, the tendency towards natural monopoly created by declining cost structures strongly suggests that no such entry would be viable. Both of these problems disappear when viewed through the lens of product differentiation.

How closely prices approach marginal cost is largely determined by the equilibrium number of firms in the relevant market segment. This suggests that with respect to market segments that are particularly large (i.e., where the total size of the relevant market is significantly greater than the fixed costs of entry), the size of the deadweight loss is likely to be rather small. It is true that even if modest, the presence of fixed costs and finite limitations to the overall market will render some degree of deadweight loss endemic to markets for differentiated products.¹³⁵ The differentiated products approach's inability to generate a first-best outcome, however, is not by itself sufficient grounds to justify rejecting it. The theory of the comparative second-best also requires comparing its shortcomings to the inefficiencies posed by other approaches to copyright. Given the acknowledged shortcomings of the traditional approach, it is quite likely that the differentiated products approach will prove superior with respect to certain segments of the market. Indeed, it provides a potential basis for allowing the degree of copyright protection to vary with respect to different categories of copyrighted works.

Even with respect to market segments that are too concentrated, the differentiated products approach identifies a fundamentally different policy tool to improve the level of access to copyrighted works. Specifically, it suggests that instead of promoting access directly by lowering the level of copyright protection, access should be promoted indirectly by stimulating entry and allowing the ensuing competition to drive price closer towards marginal cost. Since it is the presence of short-run profits that induces authors to enter the market with similar works in the first place, the differentiated products approach suggests that under some circumstances access may be better fostered by

¹³⁴ See C. Edwin Baker, *Giving the Audience What It Wants*, 58 OHIO ST. L.J. 311, 339–40 (1997); Steven T. Berry & Joel Waldfoegel, *Free Entry and Social Inefficiency in Radio Broadcasting*, 30 RAND J. ECON. 397, 397 (1999). This is analogous to concerns about “overbuilding” in natural monopoly industries, such as cable television. For a critique of such arguments, see Thomas W. Hazlett, *Private Monopoly and the Public Interest: An Economic Analysis of the Cable Television Franchise*, 134 U. PA. L. REV. 1335 (1986).

¹³⁵ Economic welfare will also depend on the time it takes the market to reach long-run equilibrium. Since entry will necessarily be non-instantaneous, the long-run benefits of the differentiated product equilibrium must be offset by the short-run deadweight losses incurred while waiting for entry to occur. The magnitude of these losses will be determined by contextual factors, such as the shape of the demand curve and the magnitude of the relevant discount rate.

increasing certain aspects of copyright protection (although as we shall see, the analysis suggests that efficiency would best be promoted by weakening other aspects of copyright protection). This stands in stark contrast to the only means for promoting access identified by the traditional approach: lowering the level of copyright protection. Indeed, failure to strengthen copyright only will serve to deter entry by reducing the number of works that can enter successfully. Consequently, as I will discuss in greater detail later, it may have the perverse effect of cementing an excessively concentrated market structure in place.¹³⁶

Entry also mitigates any distributional concerns raised by increasing the amount of surplus captured by authors. As noted earlier, product differentiation theory predicts that entry will continue until all supracompetitive profits are dissipated.¹³⁷ In other words, any short-run transfer of surplus from consumers to producers eventually will accrue to consumers' benefit over the long run in the form of entry by works that lie closer to their ideal preferences. At the same time, the product differentiation approach addresses the concern, often raised with respect to patent law, that entry will dissipate surplus to the point at which authors no longer have sufficient incentive to invest.¹³⁸ The equilibrium under the differentiated products approach effectively should bring the amount of surplus into alignment with the fixed costs needed to create the first copy of the work.

The conclusion that entry will narrow the spread between price and marginal cost depends on certain key aspects of the Chamberlinian equilibrium. Specifically, it follows from the determination that monopolistically competitive markets will reach equilibrium where the demand curve is tangent to the average cost curve. If this is the case, the fact that entry causes the demand curve to flatten necessarily implies that the market will reach equilibrium at a point

¹³⁶ See *infra* notes 148–49 and accompanying text. Some might think that the differentiated products approach's emphasis on the role of short-run profits in stimulating entry bears some resemblance to Schumpeterian competition. See SCHUMPETER, *supra* note 69, at 81–86. A moment's reflection reveals the inaptness of such a comparison. Schumpeterian competition involves vertical competition “for” the market, in which the industry is dominated by a succession of monopolists. The differentiated products approach envisions horizontal competition “within” the market, in which multiple producers vie for the same customers. The fact that the differentiated products approach relies on short-run profits to stimulate entry does not make it Schumpeterian. In this respect, the differentiated products approach is no different from the textbook microeconomic model of perfect competition, which also relies upon short-run supracompetitive returns to stimulate the entry needed to bring the market into long-run equilibrium. See Yoo, *supra* note 2, at 1591.

¹³⁷ See *supra* notes 87–88, 102 and accompanying text.

¹³⁸ See Grady & Alexander, *supra* note 31.

where the average cost curve is flatter, at which point the marginal cost curve will lie closer to the average cost curve and the spread between price and marginal cost (and the accompanying deadweight loss) will be smaller.

It bears noting that whether the market will reach long-run equilibrium on a flatter portion of the demand curve depends on the assumption that the relevant demand curve is linear. If the demand function is curved, it could be tangent to the average cost curve at any one of a number of points. In that case, it is no longer inevitable that the market will reach equilibrium at a place where the spread between price and marginal cost is narrower. In addition, the tangency solution also presupposes that the author is charging the same price to all consumers. Allowing for price discrimination raises the possibility that a firm might still earn supracompetitive profits even at the point of tangency, meaning that further entry would occur. Moreover, as noted earlier, although perfect price discrimination unambiguously would enhance welfare, perfect price discrimination is impossible, and imperfect price discrimination theoretically may cause deadweight losses to increase.¹³⁹

I thus do not mean to argue that entry by close substitutes will necessarily reduce deadweight loss in all cases. For the purposes of this Article, it is sufficient to demonstrate how relying on entry may, under certain circumstances, represent a better second-best solution than the solution suggested by the traditional approach. The differentiated products approach also reveals how access can be promoted indirectly by stimulating entry rather than directly by lowering the level of copyright protection. Although that may not be the inevitable result in all cases, it suggests that under certain circumstances, economic welfare might be better promoted by following precisely the opposite of the policies prescribed by the traditional approach.

B. The Formalization of Efficient Incentives

1. Appropriability as a Determinant of Efficient Incentives

In addition to suggesting a radically different way to promote access, the differentiated products approach also surpasses the traditional approach by offering a basis for determining the optimal level of incentives. Product differentiation theory suggests that a work should be produced if and only if the surplus it would create exceeds the costs needed to produce it.¹⁴⁰ When reproduction of additional

¹³⁹ See *supra* notes 58–59 and accompanying text.

¹⁴⁰ See Dixit & Stiglitz, *supra* note 77, at 297; Roger W. Koenker & Martin K. Perry, *Product Differentiation, Monopolistic Competition, and Public Policy*, 12 BELL J. ECON.

copies of a work is costless, this condition is met whenever the total surplus generated by the work exceeds the fixed costs needed to produce the first copy.

This criterion for determining efficient levels of entry underscores the key role played by authors' ability to appropriate surplus. It suggests that unless authors are able to capture the entire surplus created by their works, they may find it unprofitable to enter even when doing so would cause total welfare to increase.¹⁴¹ The larger the slippage in authors' ability to appropriate surplus, the fewer welfare-enhancing works will be created.

Consider the following illustration. Suppose that a pair of works would each create \$10 million in surplus if produced. One requires the incurrence of \$4 million in fixed costs, while the other requires \$6 million. Because the surplus created by each work exceeds the fixed costs needed to create them, the basic welfare metric identified above indicates that both works should be created. Suppose further that authors are only able to capture fifty percent of the available surplus (as is the case with a monopolist facing a linear demand curve), thereby generating \$5 million in revenue. If so, only the first work would be able to cover its costs despite the fact that both works would be welfare enhancing. Both works would be created if the copyright holders were allowed to capture in excess of sixty percent of the surplus, but one can easily imagine similar works costing \$7 million and \$8 million that still would not be created. Indeed, this thought experiment can be generalized into a range of works requiring fixed costs of anywhere up to \$10 million that each would cause total surplus to increase if produced. Any slippage in the author's ability to appropriate surplus would cause some of these welfare-enhancing works to be lost.

At this point, the differentiated products approach would appear to suggest that economic welfare would best be promoted by maximizing authors' ability to appropriate surplus. Complete appropriation of surplus, however, requires perfect price discrimination, which is a practical impossibility.¹⁴² Were this the only consideration, one

217, 226–27 (1981); Spence, *Product Differentiation*, *supra* note 77, at 407–08; Spence, *Product Selection*, *supra* note 77, at 218–20, 224, 230; Michael Spence & Bruce Owen, *Television Programming, Monopolistic Competition, and Welfare*, 91 Q.J. ECON. 103, 110–12 (1977).

¹⁴¹ For a related argument, see R. Polk Wagner, *Information Wants to Be Free: Intellectual Property and the Mythologies of Control*, 103 COLUM. L. REV. 995, 1017–24 (2003) (arguing that increasing appropriability of information goods leads to overall increase in production of distinct works). This conclusion is subject to the qualification regarding demand diversion discussed *infra* Part IV.B.2.

¹⁴² See *supra* notes 58–59 and accompanying text.

would conclude that markets would exhibit a systematic tendency towards underproduction of copyrighted works that could not be rectified no matter how much copyright law were structured to enhance appropriability. The burden of incomplete appropriability would fall particularly heavily on works with high fixed costs relative to their surplus, since authors must be able to appropriate a higher proportion of the available surplus to create those works.¹⁴³

This represents a fairly sharp departure from the view of appropriability taken by the traditional approach. The most sanguine perspective regards the transfer of surplus from consumers to producers that results from an increase in appropriability as welfare-neutral.¹⁴⁴ More cautious commentators raise concerns that enhancing appropriability would create excessive incentives to create and would raise distributional concerns.¹⁴⁵

The differentiated products approach casts the appropriability of surplus in a far different light. It reveals that the transfer of surplus from consumers to producers, far from being economically irrelevant, is instead a necessary condition for ensuring optimal incentives for the creation of copyrighted works.¹⁴⁶ It also suggests that concerns about overstimulation of creative activity are misplaced, since entry by competitive works will help ensure that no work garners excessive returns. Increasing appropriability even lacks distributional consequences in the long run, since free entry will dissipate any profits that initially accrue to the author and cause them instead to add to consumer welfare in the form of increased product variety.

The differentiated products approach also appears to assign the government responsibilities for which it is better suited than does the traditional approach. The traditional approach demands that the government calibrate the level of copyright protection to the lowest level possible that still supports the production of creative works. Under the best of circumstances, the informational demands needed to implement such a system threaten to exceed the government's institu-

¹⁴³ See BEATH & KATSOULACOS, *supra* note 26, at 59; Dixit & Stiglitz, *supra* note 77, at 307–08; Spence, *Product Differentiation*, *supra* note 77, at 408, 413; Spence, *Product Selection*, *supra* note 77, at 224, 234; Spence & Owen, *supra* note 140, at 112.

¹⁴⁴ See, e.g., Fisher, *supra* note 3, at 1700–02; Lunney, *supra* note 7, at 497–98. It should be noted that wealth transfers have no effect on welfare only if one adopts the standard Marshallian device of ignoring income effects. Even when income effects are present, the percentage errors involved in taking areas under Marshallian demand curves are likely to be relatively small. See Spence & Owen, *supra* note 140, at 104 n.4. For an interesting discussion of how income effects can give wealth transfers greater economic significance, see HAL R. VARIAN, *MICROECONOMIC ANALYSIS* 160–71 (3d ed. 1992).

¹⁴⁵ See *supra* note 60 and accompanying text.

¹⁴⁶ See Dixit & Stiglitz, *supra* note 77, at 297; Spence, *Product Selection*, *supra* note 77, at 218–20; Spence & Owen, *supra* note 140, at 121–22.

tional capability.¹⁴⁷ The differentiated products approach, in contrast, requires only that the government facilitate entry by enhancing authors' ability to appropriate surplus, relying instead on the market to calibrate prices at the levels that ensure that authors do not earn supracompetitive profits.

Indeed, the foregoing analysis demonstrates how the traditional approach may have the perverse effect of making matters worse by entrenching uncompetitive market structures.¹⁴⁸ The existence of copyright licensing fees that are prohibitively high is a sign that a market is too concentrated. Basic competition policy's typical response when faced with such problems is to deconcentrate the relevant market by inducing competition from new entrants.¹⁴⁹ The traditional approach to copyright, in contrast, attempts to force authors to charge lower prices by lowering the level of copyright protection that their works enjoy. Reducing the ability of such works to generate revenue, however, has the perverse effect of making it less likely that new competitors will enter the market. Lowering the level of copyright protection thus can become the cause of market failure rather than its remedy, cementing the existing, overly concentrated market structure into place. The better long-run solution might be to increase authors' ability to appropriate surplus by raising the level of copyright protection and to depend upon the ensuing competition to drive prices down towards marginal cost.¹⁵⁰

¹⁴⁷ Such intervention is also subject to a host of government failures and rent-seeking. Public choice theory, for example, has underscored how governmental processes tend to overrepresent certain types of private interests. See *supra* note 40 and accompanying text.

¹⁴⁸ For a similar exposition in the context of media regulation, see Yoo, *supra* note 16, at 246–47, 269, 293–95.

¹⁴⁹ This insight is related to the frequent observation in antitrust law that the true source of most, if not all, problems of vertical market power is horizontal market power. See, e.g., Timothy J. Brennan, *Understanding "Raising Rivals' Costs,"* 33 ANTITRUST BULL. 95 (1988). In other words, the problem stems from the existence of too few copyrighted works, not from too little access to existing works. The most effective solution is thus not to increase access to extant works, but rather to eliminate the core problem by using entry to reduce the horizontal concentration that represents the true cause of the problem.

¹⁵⁰ Interestingly, enhancing producers' ability to appropriate surplus tends to reduce consumer surplus. At the same time, producer surplus is driven down by the assumption that entry will occur until the surplus captured by authors is exactly offset by the fixed costs needed to create the first copy of the work. It would be a mistake, however, to conclude that this decline in total surplus necessarily implies a decline in economic welfare. The reduction in total surplus is counterbalanced by the welfare benefits of enabling individual consumers to purchase works that provide a better fit with their ideal preferences. See Christopher S. Yoo, *The Role of Politics and Policy in Television Regulation*, 53 EMORY L.J. (forthcoming 2004).

2. Demand Diversion as a Countervailing Consideration

The analysis advanced thus far suggests that economic welfare would best be promoted by maximizing authors' ability to appropriate surplus. The practical impossibility of perfect price discrimination, however, would appear to make underproduction an endemic feature of markets for copyrighted works. Fortunately from the standpoint of economic welfare, an additional consideration exists that tends to offset imperfections in appropriability: the extent to which a new entrant's sales represent "demand diversion" rather than "demand creation."¹⁵¹

To understand the distinction between demand creation and demand diversion, one must recognize that the surplus captured by a new entrant may come from two different sources. On the one hand, some of the sales captured by a new entrant may represent incremental sales to new customers who previously were not purchasing any competing works (i.e., demand creation). The surplus generated by these incremental sales represents an increase in economic welfare made possible by the creation of the new work. As a result, to the extent that sales result solely from demand creation, the profitability constraint has the natural effect of maximizing welfare, since a comparison of the revenue and costs provides an accurate reflection of the net economic benefits.¹⁵²

On the other hand, some of the new entrant's sales may be cannibalized from works already in the market (i.e., demand diversion). Because these are sales that would have occurred even without entry of a new work, most of the surplus attributable to demand diversion represents a transfer of surplus from one author to another rather than the creation of incremental surplus. Demand diversion does provide some increase in total surplus by enabling some consumers to purchase works that are closer to their ideal preferences. Nonethe-

¹⁵¹ The terminology used in this discussion is taken from Borenstein, *supra* note 122, at 388–89. For similar analyses using other terminology, see BEATH & KATSOUACOS, *supra* note 26, at 57 (“cannibalisation”); Mankiw & Whinston, *supra* note 90, at 49, 52, 54–55 (“business stealing effect”). See also Koenker & Perry, *supra* note 140, at 226–27 (describing effect without employing distinctive moniker); Spence, *Product Differentiation*, *supra* note 77, at 410 (same); Spence, *Product Selection*, *supra* note 77, at 230–31 (same). For other discussions of this effect appearing in the copyright literature, see Meurer, *supra* note 11, at 96–97; ABRAMOWICZ, *supra* note 29, at 57–76.

¹⁵² Spatial competition models take the existing distribution of potential customers as fixed. As a result, demand creation represents movement along the demand curve rather than an exogenous shift in demand. It is possible, however, that in addition to causing movement along the demand curve, copyright law can affect shifts in the demand curve by facilitating or limiting the emergence of alternative media platforms. Such a possibility, while real, exceeds the scope of this Article. For an insightful discussion, see Picker, *supra* note 16.

less, the total revenue attributable to demand diversion systematically tends to overstate its social benefits.

As a result, in determining a new work's net contribution to total welfare, the surplus realized by consumers purchasing the new work must be offset by the surplus that would have been realized even if the new entrant had not appeared.¹⁵³ A new entrant, however, will be concerned solely with profitability, not welfare. It will enter whenever gross surplus (as opposed to net surplus) exceeds cost. Its indifference to whether its revenue comes from demand creation or demand diversion may lead it to find entry profitable even when the incremental benefits provided by the new work fall short of the costs of entry.¹⁵⁴

Whether the free entry equilibrium results in efficient levels of entry thus depends upon a balance of two offsetting forces. On the one hand, authors' inability to appropriate all of the surplus created by their works tends to depress entry below efficient levels.¹⁵⁵ On the other hand, the presence of demand diversion tends to provide excess incentives to enter. Whether entry levels will exceed or fall short of the optimum depends upon which effect dominates.¹⁵⁶

The spatial models depicted in Figure 6 provide an intuitive illustration of these effects. The left-hand graph depicts entry by a new work that is a relatively poor substitute for other works already

¹⁵³ As Jack Hirshleifer and John Riley point out, demand diversion is analogous to forces that tend to induce overfishing of common pools. That is because each angler will not take into account that her catch includes fish that would have been caught anyway by other anglers. This causes the private marginal costs to understate the social marginal costs. Because anglers can be expected to increase their activity until their marginal benefit no longer exceeds their personal marginal costs, one would expect the number of fish caught to exceed the level that would maximize social welfare. See J. Hirshleifer & John G. Riley, *The Analytics of Uncertainty and Information—An Expository Survey*, 17 J. ECON. LIT. 1375, 1404 (1979).

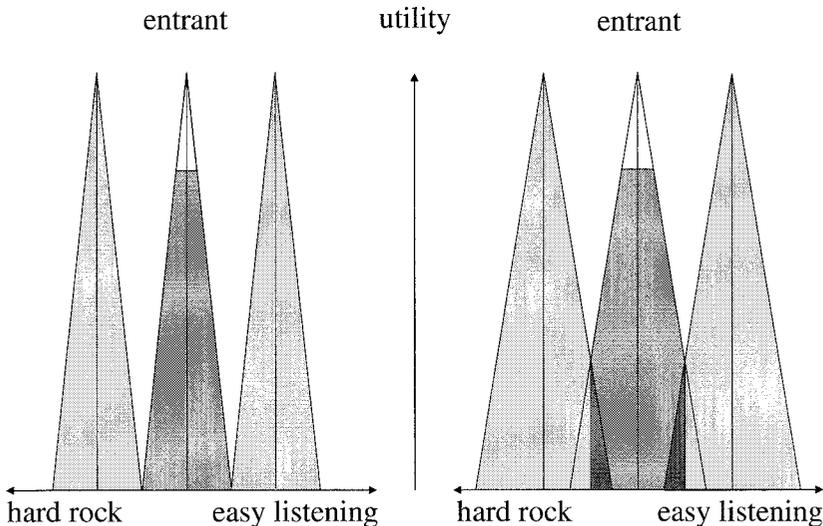
¹⁵⁴ See *infra* notes 160–61 and accompanying text. As discussed therein, policy tools exist to mitigate this distortion.

¹⁵⁵ Although the analysis advanced in this Article focuses on structural features, such as appropriability and demand diversion, I recognize that frictional considerations also can lead to systematic underinvestment. For an interesting discussion, see HIRSHLEIFER & RILEY, *supra* note 8, at 262–63 (describing how informational asymmetries can induce what they term “the speculative effect”).

¹⁵⁶ Although demand diversion allows a new product to capture the same *number* of buyers as would result under complete appropriability, it does not result in the capture of the *same* buyers. Instead, demand diversion substitutes buyers who already were purchasing other products for new buyers whose purchases represent incremental sales. Thus, although the total number of sales may reach optimal levels, the total surplus generated by those sales is likely to fall somewhat short of welfare-maximizing levels because the buyers who actually purchase the product are not necessarily those who place the highest value on the good. Some consumers may purchase goods that provide a better fit with their ideal preferences, while others may purchase goods in which the fit is worse. As a result, the equilibrium amounts to a close approximation of a first-best outcome that falls somewhat short of maximizing welfare.

offered.¹⁵⁷ The surplus actually captured by the entrant is represented by the dark gray area of the triangle, with the white triangle at the top of each representing the amount lost due to incomplete appropriability.¹⁵⁸ Because of the relative lack of substitutability, the entire surplus captured by the entering work represents demand creation (i.e., incremental surplus generated by sales to new customers). The inevitable slippage in the author's ability to appropriate surplus causes the revenue generated to understate the work's social benefits, which in turn renders the incentives for creation too weak. Thus, the inability to capture the entire surplus created by the work inevitably prevents some works that would make a marginal contribution to welfare from being created.

FIGURE 6: SPATIAL REPRESENTATION OF ENTRY WITH AND WITHOUT DEMAND DIVERSION



A different situation obtains, however, when demand diversion is present, as portrayed in the right-hand graph. In this case, the work offered by the new entrant is a somewhat better substitute for its near neighbors, with the adjacent works splitting the areas that overlap. Part of the surplus captured by the new entrant comes from demand

¹⁵⁷ Note that the graphs depict situations in which utility decays linearly with distance from the consumer's ideal preference, with the only distinction being the slope of the line. This is for illustrative purposes only; the particular equilibrium is sensitive to the rate at which the utility function decays. See *supra* note 98 and accompanying text.

¹⁵⁸ The figure depicts the area of nonappropriability as residing at the top of the triangle for illustrative purposes only. The nonappropriable surplus may occur at any point or may even be evenly distributed throughout the dark gray region.

creation (depicted by the darkly shaded, hexagonal area in the middle of the graph), which again represents the new work's incremental contribution to total welfare. Part of the surplus captured by the new entrant consists of demand diversion (represented by the two small, triangular, cross-hatched areas). When demand diversion is present, the fact that a new entrant is unable to appropriate the entire surplus created by its work does not necessarily prevent the work from being created. Because the author responds to appropriable *gross* demand, the surplus attributable to demand diversion can compensate for an author's inability to appropriate the entire surplus created by the work.

In fact, demand diversion creates the possibility of excess entry, in which authors produce new works even when the costs of doing so exceed the societal benefits. As noted earlier, efficient entry requires that authors produce new works only when the surplus attributable to demand creation exceeds the fixed costs needed to produce the work.¹⁵⁹ The problem is that a profit-maximizing author will enter whenever the total surplus captured exceeds the fixed costs of entry regardless of whether the surplus captured results from demand creation or demand diversion. Such an author could finance the fixed costs with surplus cannibalized from other producers already in the market rather than incremental surplus generated from new consumers.¹⁶⁰ Under these circumstances, the profitability constraint does not necessarily prevent the waste of resources. This danger is greatest when the works in question serve as relatively good substitutes for one another and is least problematic when the degree of substitutability is the lowest.

The traditional remedies to excess entry stimulated by demand diversion include direct regulation of entry, price, and output, or ideally a combination of all three.¹⁶¹ When such problems arise in the context of copyright, policymakers have another regulatory device at their disposal. It is possible to redress excessive entry caused by demand diversion by adjusting the degree of similarity needed to constitute copyright infringement. By doing so, adjacent works can be

¹⁵⁹ See *supra* notes 87–88 and accompanying text.

¹⁶⁰ Consider, for example, a situation in which a new entrant offers a product that is identical to that of an incumbent. The entrant may find it profitable to split the market. The fact that the products are identical, however, means that there are no compensating welfare benefits to justify the incurrence of additional fixed costs. See, e.g., Berry & Waldfogel, *supra* note 134, at 397; Peter O. Steiner, *Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting*, 66 Q.J. ECON. 194, 200 (1952).

¹⁶¹ For discussions of such regulatory approaches, see CHURCH & WARE, *supra* note 26, at 374; Hart, *Special Results*, *supra* note 77, at 900–03; Koenker & Perry, *supra* note 140, at 227–28. See also CARLTON & PERLOFF, *supra* note 26, at 208.

forced to occupy positions along the characteristics space more distant from one another than they otherwise would in equilibrium. This would reduce the extent of surplus created by demand diversion, which in turn would reduce the tendency towards excess entry. Such a solution provides a number of institutional and practical benefits as well. It allows policymakers to redress any excess entry problems without having to assume the considerable administrative burdens associated with directly regulating entry, prices, and output.

V

THE REMEDIAL IMPLICATIONS OF THE DIFFERENTIATED PRODUCTS APPROACH

The differentiated products approach thus has the potential to revolutionize the study of copyright. It exposes the traditional approach's inability to provide a theoretical explanation for many salient features of real-world markets for copyrighted works. In addition, it suggests that the key economic features of those markets would still exist even if copyrighted works did not exhibit the lack of rivalry that is generally regarded as their most distinctive economic feature. Most importantly, it breaks with the standard approach by identifying remedies that can promote access and incentives simultaneously. In so doing, it reveals the supposed tension between those two considerations to be something of a false conflict.

The differentiated products approach also surpasses the traditional approach by providing a basis for distinguishing among the available policy instruments. As noted earlier, previous models tended to represent the overall level of copyright protection with a single variable.¹⁶² Although copyright scholars have made some attempts to analyze the interaction among various aspects of copyright protection,¹⁶³ it is the patent literature that has given this issue its fullest articulation.¹⁶⁴ In particular, the patent literature focuses on the tradeoff between a patent's "length," typically determined by the duration of the patent term, and its "breadth," which is most usefully described for our purposes as how close a competing product may

¹⁶² See *supra* note 34 and accompanying text.

¹⁶³ These analyses tend to emphasize the tradeoff between the remaining length of the copyright term and other aspects of copyright protection. Terry Fisher has suggested that the most efficient copyright legislation would include an infinite term but few other entitlements. Fisher, *supra* note 3, at 1719 n.265. Joseph Liu and Justin Hughes propose broadening the scope of fair use as the copyright term approaches expiration. See Justin Hughes, *Fair Use Across Time*, 50 UCLA L. REV. 775 (2003); Joseph P. Liu, *Copyright and Time: A Proposal*, 101 MICH. L. REV. 409 (2002).

¹⁶⁴ See *supra* note 35 and accompanying text.

come in the characteristics space to a patented product without constituting infringement.¹⁶⁵ There is significant debate in the literature as to whether patents should be long and narrow¹⁶⁶ or short and broad.¹⁶⁷

The differentiated products approach to copyright suggests a similar analysis, but with two important modifications. First, it goes beyond the patent literature's emphasis on length by acknowledging that duration represents only one way to expand the number of surplus-generating activities encompassed by the right. Second, it shows how a third consideration—appropriability—also influences economic efficiency.¹⁶⁸

Expanding the analysis in this manner highlights the complex interactions among these factors. Interestingly, the policy implications do not all point in the same direction. Specifically, the differentiated products approach indicates that the best way to render a market more competitive is by strengthening the first and third aspects of copyright protection (i.e., increasing the number of welfare-generating activities encompassed by the right and the authors' ability to appropriate surplus), while weakening it along the second dimension (i.e., narrowing copyright breadth). In other words, markets for copyrighted works would be most competitive if the right were large (in that it encompassed a wide sweep of surplus-generating activity within its scope) and intense (in that authors were able to appropriate a high proportion of the surplus determined by the first criterion to be within the right), but narrow (in that new works could exhibit a high degree of similarity without constituting infringement). The differentiated products approach is therefore not an unqualified endorsement for strengthening copyright protection. Instead, by providing a basis for distinguishing among different aspects of copyright protection, it allows for a more nuanced approach to copyright policy.

¹⁶⁵ See, e.g., Klemperer, *supra* note 28, at 114. For a more complete enumeration of the various conceptions of patent breadth, see Denicold, *supra* note 35, at 251–53.

¹⁶⁶ See, e.g., Gilbert & Shapiro, *supra* note 35 (examining infinite-term patents); Klemperer, *supra* note 28, at 120–21 (describing conditions under which long, narrow patent is most efficient); Tandon, *supra* note 35 (advocating infinite life for patents, subject to compulsory licensing).

¹⁶⁷ See, e.g., Gallini, *supra* note 35 (arguing for broad patents, with length determined accordingly); Klemperer, *supra* note 28, at 121–23 (describing conditions under which short, broad patent is most efficient).

¹⁶⁸ See Koenker & Perry, *supra* note 140, at 226–27 (describing relationships among fixed costs, substitutability, and appropriability); Spence, *Product Selection*, *supra* note 77, at 230–31, 234 (same); Spence, *Product Differentiation*, *supra* note 77, at 410, 413 (same).

A. *Determinants of the Differentiated Products Equilibrium*

1. *The Size of the Relevant Market Relative to the Fixed Cost of Entry*

The most important determinant of the overall competitiveness of markets for copyrighted works is the magnitude of the relevant market relative to the fixed costs of entry. As noted earlier, this ratio determines the number of works that will exist in equilibrium.¹⁶⁹ As such, it has a direct impact on the extent to which entry will ameliorate deadweight loss. If the relevant market segment is small relative to the fixed costs of entry, the number of works in the market will be relatively small, competition will be thin, and the spread between price and marginal cost will be relatively large. If the relevant market is large relative to fixed costs, the number of works will be relatively large, competition will be robust, and the spread between price and marginal cost will be relatively narrow.

Since copyright policy has relatively little impact on the absolute size of the fixed costs needed to produce the first copy of a creative work, the primary means for policymakers to increase this ratio is to expand the number of surplus-generating activities that fall within the scope of each copyrighted work. Several specific policy tools are available to change the “size” of copyright. One determinant of the size of the right that has received a great deal of attention in recent months is the length of the copyright term.¹⁷⁰ But other examples abound, such as the extent to which copyright allows authors to earn royalties for performances of their works.

The differentiated products approach suggests that increasing the number of surplus-generating activities contained within the scope of each copyrighted work can promote both access and incentives by increasing the equilibrium number of works. The suggestion that access would be promoted best by raising rather than lowering the level of copyright protection may seem counterintuitive. After all, it implies that the proper policy response to markets that are too concentrated is to increase the degree of copyright protection that authors enjoy. This apparent contradiction disappears when viewed in light of the traditional approach’s inability to capture the dynamics of entry. So long as entry is free, any strengthening in the level of copy-

¹⁶⁹ See *supra* notes 88–90, 102 and accompanying text.

¹⁷⁰ In 1998, Congress extended the term of copyright protection to the life of the author plus seventy years. See Sonny Bono Copyright Term Extension Act, Pub. L. No. 105-298, § 102(b) & (d), 112 Stat. 2827, 2827–28 (1998) (amending 17 U.S.C. §§ 302, 304 (2000)). This legislation recently withstood a constitutional challenge in *Eldred v. Ashcroft*, 537 U.S. 186 (2003).

right protection will not ultimately accrue to the benefit of the incumbents. Instead, it will only attract more entry, which will in turn reduce deadweight loss and bring the number of works closer to the optimum. Any short-term profits made possible by the expansion of the size of the right will eventually accrue back to consumers in the form of increased product variety.

2. *The Appropriability of Surplus*

Another consideration that determines the market's ability to promote economic welfare is the intensity of authors' rights, as measured by their ability to appropriate the surplus created by their works. As noted earlier, any slippage in appropriability tends to reduce the equilibrium number of works.¹⁷¹ Such a reduction can harm the incentives side of the tradeoff by causing the total number of works produced to drop below optimal levels. It also harms the access side of the tradeoff by limiting the extent to which entry by new works will flatten authors' demand curves and thereby reduce deadweight loss.

This suggests that increasing authors' ability to capture the surplus created by their works will tend to enhance welfare. Again, this argument may seem counterintuitive from the standpoint of the traditional approach to copyright. The solution lies in understanding that access may be promoted as much by increasing the number of works available for consumption as by mandating access to the limited number of works that already have been created. The power of this insight easily can be discerned when the differentiated products approach is used to analyze two of the most salient issues in copyright law: fair use and price discrimination.

(a) Fair Use

The fair use doctrine provides authors who incorporate limited portions of a copyrighted work into their own creative works with an affirmative defense against copyright liability. In determining whether a particular use of a copyrighted work constitutes fair use, the governing statute directs courts to consider four factors:

- (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit education purposes;
- (2) the nature of the copyrighted work;
- (3) the amount and substitutability of the portion used in relation to the copyrighted work as a whole; and

¹⁷¹ See *supra* Part IV.B.1.

- (4) the effect of the use upon the potential market or value of the copyrighted work.¹⁷²

Copyright scholars have offered two distinct economic justifications for fair use. Some have regarded fair use as one way that the law tips copyright towards the access side of the access-incentives tradeoff by ensuring that the total surplus appropriated by the copyright holder is not excessive.¹⁷³ The other economic justification regards fair use as a means of compensating for market failures induced by transaction costs.¹⁷⁴ In the tradition of Calabresi and Melamed,¹⁷⁵ this perspective posits that the government should mandate access to a work whenever friction in the bargaining process prevents a welfare-enhancing transaction from occurring.¹⁷⁶

Technological innovations such as digital rights management and the Internet have sharpened the conflict between these two theories and have touched off a debate over whether the reduction in transaction costs made possible by these developments should have any impact on the fair use doctrine's scope.¹⁷⁷ If fair use is intended to redress market failures caused by transaction costs, these technological developments should cause the fair use doctrine to contract. Any such narrowing of fair use would alarm those theorists who view the doctrine as a way to limit monopoly power.¹⁷⁸ Quite the contrary,

¹⁷² 17 U.S.C. § 107 (2000).

¹⁷³ See, e.g., 1 PAUL GOLDSTEIN, COPYRIGHT § 1.14.2.3.a, at 1:51–1:54 (2d ed. 1996 & Supp. 2004); Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1110, 1135–36 (1990); Sterk, *supra* note 10, at 1211–12.

¹⁷⁴ See, e.g., Wendy J. Gordon, *Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and Its Predecessors*, 82 COLUM. L. REV. 1600, 1627–30 (1982); Landes & Posner, *supra* note 3, at 357–58.

¹⁷⁵ Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089 (1972).

¹⁷⁶ Transaction cost theories of property can be seen as part of the larger debate about the optimal size and degree of standardization in property rights. See, e.g., Henry Hansmann & Reinier Kraakman, *Property, Contract, and Verification: The Numerus Clausus Problem and the Divisibility of Rights*, 31 J. LEGAL STUD. 373 (2002) (analyzing property rights regimes in light of conclusion that purpose of limitations on types of property rights is to facilitate verification of ownership); Michael A. Heller, *The Boundaries of Private Property*, 108 YALE L.J. 1163 (1999) (arguing that increasing fragmentation of property rights undermines usefulness of private property as economic institution and constitutional category); Thomas W. Merrill & Henry E. Smith, *What Happened to Property in Law and Economics?*, 111 YALE L.J. 357 (2001) (contrasting in rem conception of property with conception of property as bundle of in personam rights).

¹⁷⁷ See, e.g., PAUL GOLDSTEIN, COPYRIGHT'S HIGHWAY 224 (1994); Bell, *supra* note 57, at 579–600; Trotter Hardy, *Property (and Copyright) in Cyberspace*, 1996 U. CHI. LEGAL F. 217, 239–42; Edmund W. Kitch, *Can the Internet Shrink Fair Use?*, 78 NEB. L. REV. 880 (1999); Robert P. Merges, *The End of Friction? Property Rights and Contract in the "Newtonian" World of On-Line Commerce*, 12 BERKELEY TECH. L.J. 115, 130–35 (1997).

¹⁷⁸ Any contraction of fair use also would be a source of considerable concern to those who view the doctrine as a means to promote other, noneconomic goals. See, e.g., Fisher,

because the reduction in transaction costs allows authors to appropriate a larger amount of the available surplus, these technologies arguably would support a concomitant weakening in the strength of copyright protection by expanding the number of activities that constitute fair use.

The differentiated products approach offers a structural basis for the fair use doctrine capable of resolving the conflict between these two opposing points of view. In emphasizing the importance of maximizing authors' ability to appropriate surplus, the differentiated products approach supports keeping fair use as narrow as possible. Somewhat counterintuitively, it suggests that fair use should be broadest when competition is the most intense and narrowest when markets are the most highly concentrated and the chance of supracompetitive profits the greatest.

Consonant with the transaction cost approach, it would sanction a fair use doctrine designed to facilitate welfare-enhancing uses of copyrighted works that otherwise would be barred by frictional considerations. Indeed, the transaction cost approach does not itself provide a theory by which the optimal price, quantity, and mix of creative works can be determined. Instead, it implicitly presupposes the existence of an efficient outcome and simply examines when frictional considerations prevent markets from achieving it. By providing a basis for determining a baseline efficient outcome, the differentiated products approach can be regarded as a complementary analysis that provides the type of external foundation that the transaction cost approach needs, but cannot provide itself.

The differentiated products approach also accepts the fact that the scope of the fair use doctrine ought to narrow as technology causes transaction costs to decrease.¹⁷⁹ It also dismisses as fundamentally misplaced the concerns that doing so will upset the balance between access and incentives by putting too much weight on the

supra note 3, at 1744–94 (arguing that fair use should be designed to promote just social order); David Lange, *Reimagining the Public Domain*, LAW & CONTEMP. PROBS., Winter/Spring 2003, at 463, 478–82 (defending fair use as means for protecting individual creative expression); Lloyd L. Weinreb, *Fair's Fair: A Comment on the Fair Use Doctrine*, 103 HARV. L. REV. 1137, 1150–61 (1990) (advocating vision of doctrine based on normative conception of fairness).

¹⁷⁹ It remains possible that transaction costs might nonetheless remain high for uses such as parody and criticism, which necessarily require access to a specific work. In that case, the follow-on author would be locked into a bilateral monopoly with the original author in which no real substitutes would be available. Any fair use doctrine justified in this manner would necessarily be limited in scope to parody, criticism, and other forms of derivative use that raise these types of concerns. See Robert P. Merges, *Are You Making Fun of Me?: Notes on Market Failure and the Parody Defense in Copyright*, 21 AIPLA Q.J. 305 (1993).

latter. Although constricting the scope of the fair use doctrine will increase authors' ability to appropriate surplus, entry by substitute products simultaneously should increase access and dissipate any excess investment incentives.

(b) Price Discrimination

Another salient issue in current copyright scholarship is the extent to which copyright law should facilitate price discrimination.¹⁸⁰ It could do so by relaxing the first sale doctrine¹⁸¹ or by encouraging digital rights management¹⁸² and a host of other devices designed to permit increasingly fine degrees of customer segmentation.

As noted earlier, the traditional approach focuses primarily on whether price discrimination will reduce deadweight loss by increasing output of works that already have been created, as well as on the distributional implications of authors' enhanced ability to earn profits.¹⁸³ It should be noted from the outset that far from being an iniquitous business device, price discrimination is more properly regarded as an inevitable consequence of the fact that consumption of copyrighted works is often nonrival. For rival goods, the central problem is how to allocate output to individual customers. When goods are nonrival, however, allocation of output is unproblematic. Once the first copy has been created, the author can create an endless stream of copies without incurring any additional costs. Instead, the central economic problem for nonrival goods becomes allocating the first-copy costs rather than output.¹⁸⁴ The classic pricing schemes for the efficient

¹⁸⁰ See *supra* notes 56–57, 60 and accompanying text.

¹⁸¹ The first sale doctrine recognizes that copyright holders' right to control the terms under which their works can be transferred only applies to the first sale of the work. Once a person has purchased a copyrighted work, that person remains free to resell it at any price he or she sees fit. See 17 U.S.C. § 109(a) (2000); *Quality King Distribs., Inc. v. L'anza Research Int'l, Inc.*, 523 U.S. 135, 140–42 (1998). Because low-price customers remain free to resell their copies to those customers that the author would like to charge higher prices, the first sale doctrine sharply limits copyright holders' ability to engage in price discrimination.

¹⁸² Digital rights management (DRM) is the use of software and other technological devices to control access to creative works. By preventing resale to third parties and by increasing authors' ability to meter the intensity of usage, DRM substantially enhances the ability to engage in price discrimination. See Bell, *supra* note 57, at 588 n.142; Cohen, *supra* note 7, at 1121; Maureen A. O'Rourke, *Copyright Preemption After the ProCD Case: A Market-Based Approach*, 12 BERKELEY TECH. L.J. 53, 62–63, 85–86 (1997); R. Anthony Reese, *Will Merging Access Controls and Rights Controls Undermine the Structure of Anticircumvention Law?*, 18 BERKELEY TECH. L.J. 619, 659 (2003).

¹⁸³ See *supra* notes 58–60 and accompanying text.

¹⁸⁴ See SHARKEY, *supra* note 69, at 46.

allocation of fixed costs all depend upon the producer's ability to charge different prices to different customers.¹⁸⁵

The differentiated products approach reveals that the problems associated with deadweight loss also can be ameliorated by stimulating entry by new works, which in turn relies on competition to reduce the spread between price and marginal cost. In this manner, facilitating price discrimination and its accompanying increase in appropriability of the available surplus can cause welfare to increase regardless of whether the production of any particular work increases or decreases. The differentiated products approach also reveals that price discrimination is unlikely to give rise to sustainable supracompetitive profits. Any such profits should be dissipated by new entry, which inures to the benefit of consumers by allowing them to consume works that lie closer to their ideal preferences.

By underscoring the crucial role that the appropriability of surplus plays in maximizing total welfare, the differentiated products approach provides a strong justification for supporting authors' ability to engage in price discrimination. Somewhat counterintuitively, the case for doing so is the strongest when the markets are the most concentrated and the fewest substitutes exist, since it is under those circumstances that inducing entry by new works is the most critical.

3. *The Degree of Substitutability*

Finally, the total return captured by any particular work turns on the extent to which consumers regard competing works as substitutes. As noted earlier, works that serve as relatively good substitutes for other works will capture a larger amount of the surplus available in the adjacent areas. The degree of substitutability depends upon the structure of demand, as reflected in the cross-price elasticities. As a result, competition will be most robust when substitutability is the highest.

The degree of substitutability plays a natural role in determining how many firms will enter at equilibrium and how robust the competition among those firms will be. The "breadth" of copyright, or the

¹⁸⁵ One of the classic solutions to the pricing of nonrival goods is known as Lindahl pricing, in which different consumers are charged different prices (and thus bear a different proportion of the fixed costs) based on their marginal valuations of the total quantity. See Erik Lindahl, *Just Taxation—A Positive Solution*, in CLASSICS IN THE THEORY OF PUBLIC FINANCE 168 (Elizabeth Henderson trans., Richard A. Musgrave & Alan T. Peacock eds., 1958). Lindahl pricing bears some similarities to a pricing scheme popular in the regulated industries literature known as Ramsey pricing, in which fixed costs are allocated among buyers based on the elasticity of their demand, with those buyers with the least elastic demand bearing the heaviest burden. See F.P. Ramsey, *A Contribution to the Theory of Taxation*, 37 ECON. J. 47, 58–59 (1927).

scope of infringement, is a legal constraint on substitution because infringing works cannot compete absent permission from the copyright holder. This analysis thus initially suggests that copyright should be kept as narrow as possible so as not to add any exogenous constraints that would prevent the market from becoming as competitive as possible.

There is, however, one consideration that cuts the other way. Substitutability also determines the extent to which the surplus captured by any particular work derives from demand diversion. The higher the degree of substitutability between the works, the greater the proportion of the total surplus captured that demand diversion will represent. As noted earlier, demand diversion plays a critical role in determining how closely the total number of works produced will approximate the optimum.¹⁸⁶ Up to a point, demand diversion is beneficial, as it can replace the surplus that authors are unable to appropriate because of their imperfect ability to price discriminate. Beyond that point, demand diversion creates the possibility of excess entry. Thus, the danger of excess entry is greatest when goods are the most substitutable.

As discussed earlier, policymakers may reduce the impact of demand diversion by using the standard of copyright infringement to increase the distance between adjacent works. Doing so would require a delicate balance. Widening the breadth of copyright protection would tend to force the number of works below equilibrium levels. This decrease in the number of works would in turn harm the access side of the tradeoff by causing deadweight loss to increase. At the same time, increasing copyright breadth might simultaneously offer compensating benefits on the incentives side of the tradeoff, since increasing copyright breadth would potentially bring the total number of works closer to optimal levels. When demand diversion is most likely to cause excessive entry (i.e., when substitutability among works is the highest), policymakers may have to strike a difficult balance.

B. Interactions Among the Different Aspects of Copyright Protection

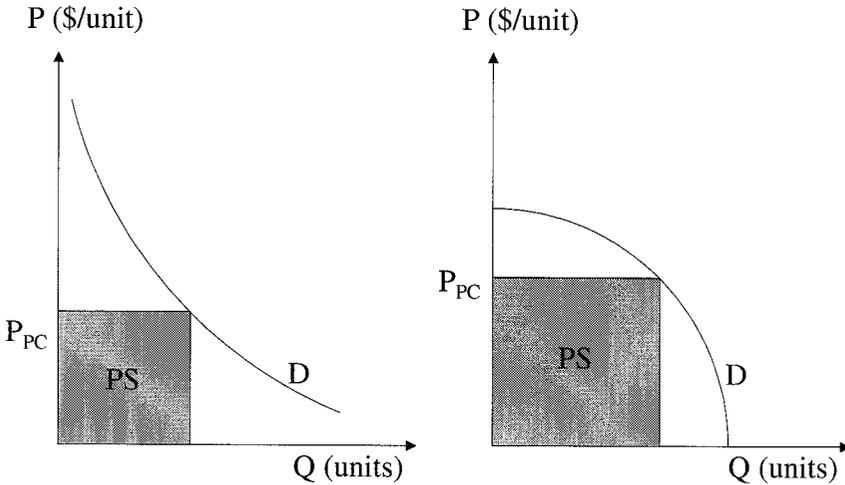
The differentiated products approach thus provides a framework that is able to distinguish among three different types of policy instruments. Although this degree of nuance enriches the power of the analysis, it also makes copyright policy considerably more difficult to implement. Indeed, two of the leading differentiated products theo-

¹⁸⁶ See *supra* Part IV.B.2.

rists candidly have conceded that “we believe that we would be quite unable to recognize an optimum if we saw one.”¹⁸⁷

One problem is that the available policy tools are not completely independent. As noted earlier, one of the key determinants of optimal entry is appropriability. This suggests that the market will exhibit a bias against works that are able to appropriate only a relatively small proportion of the surplus they create. Product differentiation theory suggests that a work’s ability to appropriate surplus is determined in no small part by the shape of its demand function. Specifically, the market will be biased against works whose demand curves are convex to the origin, such as the one represented in the left-hand graph in Figure 7.¹⁸⁸ It will be biased in favor of works whose demand curves are concave to the origin, such as the one represented in the right-hand graph.

FIGURE 7: APPROPRIABILITY AS A FUNCTION OF INVERSE DEMAND



Whether a demand curve is convex or concave to the origin depends on the steepness of the relevant inverse demand function.¹⁸⁹ Works with steep inverse demand functions tend to be products with low own-price elasticities of demand. The market thus will be especially biased against works that have the fewest substitutes, as it is

¹⁸⁷ Eaton & Lipsey, *supra* note 26, at 760.

¹⁸⁸ Figure 7 is adapted from Yoo, *supra* note 2, at 1613 fig.8. For similar representations, see BEATH & KATSOUALACOS, *supra* note 26, at 70 figs.4.7–4.8; OWEN & WILDMAN, *supra* note 16, at 112 figs.4.5–4.6; Spence, *Product Differentiation*, *supra* note 77, at 409 fig.1.

¹⁸⁹ See Spence, *Product Differentiation*, *supra* note 77, at 409–10; Spence & Owen, *supra* note 140, at 111–12. See generally Yoo, *supra* note 2, at 1612–13 (reviewing this literature).

these works that capture the lowest proportion of the available surplus. This implies that the case for increasing copyright protection and strengthening authors' ability to capture surplus is the strongest when the market is the most concentrated.

Derivative works provide another example of where two factors interact.¹⁹⁰ On the one hand, authorization of a derivative work expands the first factor by increasing the number of surplus-generating activities captured by the author. On the other hand, it also increases the number of substitutes with which the licensed work must compete. In other words, the decision to license simultaneously makes the triangle taller and narrower.¹⁹¹

The overlapping nature of these considerations further complicates isolating the impact of any particular aspect of copyright policy.¹⁹² It nevertheless may be possible to simplify the analysis with respect to particular industries or categories of copyrighted works. If one aspect of copyright protection can be taken as fixed with respect to certain types of works, the problem that must be solved becomes much simpler.

Empirical studies may provide additional insights into the balance of these problems. For example, ruling out as an empirical matter the dangers of excess entry induced by demand diversion would simplify the analysis considerably, as it would allow policymakers to focus on ensuring that copyright be as large, intense, and narrow as possible. Some initial illustrative calculations suggest that so long as the relevant economy is relatively large, any welfare losses resulting from excessive entry are likely to be quite small.¹⁹³

The empirical evidence on the magnitude of these welfare losses, however, is somewhat equivocal. A study of entry patterns in the radio industry conducted by Steven Berry and Joel Waldfogel estimates that the deadweight losses attributable to excess entry may be substantial.¹⁹⁴ They acknowledge, however, that the radio industry is somewhat unusual in that it serves two different groups of cus-

¹⁹⁰ As noted earlier, this Article captures only part of the debate surrounding derivative use as it sets to one side the impact of cumulative innovation and analyzes each work as an independent creation. See *supra* note 4. For my preliminary observations on cumulative innovation, see *infra* note 204.

¹⁹¹ I thank Louis Kaplow for providing me with this insight.

¹⁹² See T. Brennan, *Right to Deny*, *supra* note 3, at 705 n.101.

¹⁹³ See G.K. Yarrow, *Welfare Losses in Oligopoly and Monopolistic Competition*, 33 J. INDUS. ECON. 515, 520 (1985) (estimating welfare losses for large economies at 0.5% of total revenue). A very different result obtains in the case of small economies. See *id.* at 523 (concluding that welfare losses will be several times larger if market segment is oligopolistic); see also BEATH & KATSOULACOS, *supra* note 26, at 64–66 (reaching similar conclusion).

¹⁹⁴ Berry & Waldfogel, *supra* note 134, at 411–12.

tomers—advertisers and listeners—only one of which (advertisers) is able to make direct payments for programming. What appears to be excessive entry when measured solely in terms of benefits to advertisers may in fact be efficient when measured in terms of both advertisers and listeners.¹⁹⁵ In addition, the fact that their study assumed that the radio market is composed of homogeneous products¹⁹⁶ led them to overlook potential welfare benefits resulting from product differentiation. The existing theoretical literature suggests that this simplifying assumption can have a fairly dramatic effect on the welfare implications.¹⁹⁷

Another empirical study, by Ronald Goettler and Ron Shachar on spatial competition among the three major television networks, reached a somewhat different conclusion. They concluded that the networks' program offerings nearly fully achieved the optimum suggested by the underlying Nash equilibrium.¹⁹⁸ The shortfall was largely (but not completely) explained by the networks' adherence to the rules of thumb against airing sitcoms after 10:00 p.m. and against airing news magazines before 10:00 p.m.¹⁹⁹ Goettler and Shachar attribute the remainder of the shortfall to bounded rationality and the difficulty in identifying competing products when they have several attributes.²⁰⁰

These studies suggest the possibility of isolating the impact of each of the three factors discussed in this Part (i.e., the size of the

¹⁹⁵ *Id.* at 412–14. Some indirect data from Europe raise doubts as to whether this is plausible. *Id.*

¹⁹⁶ *Id.* at 398–99. They offer a preliminary discussion of how taking product differentiation into account might change the analysis, concluding that product heterogeneity explains part of the proliferation of stations, but that demand diversion continues to play a role. Unfortunately, the current state of the literature does not permit them to estimate the magnitude of the welfare loss or to determine how much of it was attributable to demand diversion. *Id.* at 414–17.

¹⁹⁷ The differences between homogenous and differentiated products are illustrated by Mankiw and Whinston's landmark analysis of demand diversion. They first consider a homogeneous product market in which entry creates a net reduction in demand for each incumbent (i.e., demand diversion dominates demand creation). When the integer problem is ignored, markets unambiguously tend to produce excess diversity. Mankiw & Whinston, *supra* note 90, at 51–52. Taking the integer problem into account creates the possibility of insufficient entry by no more than one firm, although the welfare losses from this shortfall may be substantial. *Id.* at 53. When products are differentiated, however, markets may create either excess or insufficient entry. *Id.* at 54–55. It is worth noting that all of these analyses assume that entry causes the output per firm to fall. *Id.* at 51 (assumption 2). It is theoretically possible that demand creation could so dominate demand diversion that entry might instead cause output per firm to rise. *Id.* at 52 n.7.

¹⁹⁸ Ronald L. Goettler & Ron Shachar, *Spatial Competition in the Network Television Industry*, 32 RAND J. ECON. 624, 647–48 (2001).

¹⁹⁹ *Id.* at 649–51.

²⁰⁰ *Id.* at 651–52.

right, the appropriability of surplus, and substitutability). Even if the empirical problem proves intractable, the differentiated products approach still should provide useful intuitions about the way these factors interact.²⁰¹ It suggests, for example, that excess entry is least likely to be a problem when a work has few substitutes. Thus, contrary to the conventional wisdom, it is when a work is the most unique that the case for strengthening the level of copyright protection in terms of the size of the right and appropriability is the strongest.

CONCLUSION

In the final analysis, product differentiation offers significant promise as a way to reconceptualize the economic analysis of copyright law. Not only does it offer a better theoretical explanation for a number of market features, it also provides a basis for formalizing both the access and incentives sides of the tradeoff in a way that yields insights into their structural interrelationship.

What emerges is an approach that demonstrates how stimulating entry can promote both access and incentives simultaneously. This stands in stark contrast to the position that dominates existing copyright scholarship, which views these two considerations as being in inexorable tension.

The differentiated products approach also suggests that the best policy response to a highly concentrated market might well be to strengthen the degree of copyright protection in ways that tend to stimulate new entry. The ensuing competition should dissipate supracompetitive returns and ensure that authors capture only enough revenue to cover their costs of production. Such a task is more consonant with the conventional understanding of government's institutional capabilities than is the task envisioned by the traditional approach.

It may seem counterintuitive that copyright protection should be the greatest when high fixed costs and low substitutability cause the market to become the most concentrated, but this apparent paradox is resolved once one understands the complex manner in which access and incentives interact with one another. In this sense, the differentiated products approach to copyright captures some of the insights of classic property theory, which emphasizes the importance of well-defined property rights in ensuring optimal investment and deployment. In so doing, it corrects for the blind spot that results when markets for copyrighted works are treated as monopolies and allows for

²⁰¹ Samuelson, *supra* note 92, at 108 n.5 (defending Chamberlin against similar criticism); *see also* CHAMBERLIN, *supra* note 92, at 304-05 (same).

serious consideration of the role of short-run profits in stimulating entry and promoting economic efficiency. At the same time, it moves beyond classic property theory by identifying ways in which a property right can be too strong.

It bears noting that the differentiated products approach cannot completely resolve the tension between access and incentives. The presence of a downward-sloping demand curve renders some degree of deadweight loss endemic. In addition, the fact that perfect price discrimination is impossible prevents authors from appropriating all of the surplus created by their works. As a result, markets may exhibit a systematic tendency towards having too little diversity. However, demand diversion makes it possible that the market will produce the optimal number of works. Any such solution to the incentives side of the tradeoff necessarily requires accepting a degree of inefficiency in terms of access. As the theory of the comparative second-best aptly points out, the differentiated products approach's inability to generate first-best outcomes is not by itself sufficient to justify rejecting it. At least with respect to some market segments, in which the fixed costs of entry are small relative to the overall size of the market segment, the well-documented flaws associated with governmental calibration of the level of copyright protection²⁰² make it quite plausible that the differentiated products approach may offer the more attractive policy alternative. At the same time, demand diversion raises the possibility that the equilibrium level of entry may be inefficiently excessive.

In addition, the differentiated products approach allows for a more nuanced analysis by making it possible for policymakers to distinguish among different aspects of copyright protection. This represents a substantial improvement over the traditional approach, which tends to represent all aspects of the strength of copyright protection with a single variable and fails to distinguish among copyright's different aspects. In so doing, it suggests that efficiency best might be served by making the right large (in terms of surplus-generating activities within its scope) and intense (in terms of the proportion of that surplus that authors are able to appropriate), but narrow (in terms of how close another work can come to an existing work without infringing the copyright). Thus, the differentiated products approach does not amount to a blanket endorsement for strengthening copyright protection. On the contrary, the resulting theory allows for a degree of subtlety that is impossible under other approaches.

Although my analysis has clear theoretical implications, considerable additional work remains to be done before it can be fully opera-

²⁰² See *supra* note 40 and accompanying text.

tionalized. As noted earlier, further work should incorporate elements of cumulative innovation that take into account the extent to which current works serve as inputs to subsequent works,²⁰³ although, for reasons I set forth in the margin, I suspect that such considerations ultimately will not prove problematic.²⁰⁴ Furthermore, the differentiated products approach should be broadened to account for endogenous pricing as well as the preemptive strategies available when entry is sequential and when firms can occupy more than one location.²⁰⁵

Perhaps most importantly, the models should consider the implications of relaxing the symmetrical preferences assumption, either by allowing for variations in the distribution of consumers across the characteristics space or by allowing the extent to which particular

²⁰³ See *supra* note 4 and accompanying text. The literature on sequential innovation in the context of patents is more extensive. See, e.g., Howard F. Chang, *Patent Scope, Anti-trust Policy, and Cumulative Innovation*, 26 RAND J. ECON. 34 (1995); Jerry R. Green & Suzanne Scotchmer, *On the Division of Profit in Sequential Innovation*, 26 RAND J. ECON. 20 (1995); Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839 (1990); Richard R. Nelson, *Intellectual Property Protection for Cumulative Systems Technologies*, 94 COLUM. L. REV. 2674 (1994); Ted O'Donoghue, *A Patentability Requirement for Sequential Innovation*, 29 RAND J. ECON. 654 (1998); Suzanne Scotchmer, *Protecting Early Innovators: Should Second-Generation Products Be Patentable?*, 27 RAND J. ECON. 322 (1996); Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, 5 J. ECON. PERSP. 29 (1991); Michael Waterson, *The Economics of Product Patents*, 80 AM. ECON. REV. 860 (1990).

²⁰⁴ Just as stimulating entry of close substitutes should promote access by readers through using increased competition to lower price, it should also promote access by follow-on authors who seek to build on prior work. Once the market for the work becomes sufficiently competitive, the problem of cumulative innovation—in which a copyrighted work is licensed simultaneously to both consumers and other authors who seek to use the work as an input in creating other works—becomes analogous to the classic problem of transfer pricing, in which a particular good simultaneously serves as an end product and as an input for another product. The transfer pricing literature indicates that welfare is maximized when the price of the good as an input is set equal to the price of the good as a final product, so long as the final product market is sufficiently competitive. See PAUL MILGROM & JOHN ROBERTS, *ECONOMICS, ORGANIZATION AND MANAGEMENT* 79–83 (1992).

Thus, so long as the total surplus in the market is sufficiently large relative to the fixed costs of entry, there is nothing inefficient about charging the market price to follow-on authors who seek access to a copyrighted work as an input in creating other works. Should the market for the works not be sufficiently competitive, the differentiated products approach suggests that the problem might be redressed by stimulating entry to make the market more competitive rather than by lowering the price paid by follow-on authors.

It is true that holdout behavior may lead particular authors to refuse to license their works. Competition policy, however, focuses on protecting competition, not particular competitors. Thus, unless such refusals create losses for more than just particular individuals, no intervention is warranted. The most compelling cases for protecting follow-on innovation are presented by parody and criticism, which necessarily require access to a specific work. As noted earlier, any intervention undertaken to remedy this problem would be designed solely for the benefit of parodists and would not take the form of a general revision of the scope of copyright protection. See *supra* note 179.

²⁰⁵ See *supra* notes 108–10 and accompanying text.

works serve as substitutes for other works to vary.²⁰⁶ Relaxing the symmetry assumption allows for the possibility that the impact of entry by a new work will no longer be spread evenly across all of the incumbents. Instead, it suggests that the entry will affect only some of the works. This localization of competition has the effect of dividing the relevant market into subsegments, with the overall competitiveness of each subsegment determined by the size of the total surplus of the subsegment relative to the fixed cost. The lack of robust competition within a subsegment may limit the extent to which entry can push price towards marginal cost. It can also allow the “integer problem” to arise simultaneously with respect to multiple portions of the overall market, as the single “large economy” is chopped into a series of “small economies” that each are capable of supporting sustainable profits.²⁰⁷ If these effects arise with respect to multiple subsegments, the combined adverse impact may be quite substantial,²⁰⁸ although the resulting policy prescription may be the same as when consumer preferences are assumed to be symmetric.²⁰⁹

Countervailing considerations exist as well. The discussion of spatial competition assumes that product characteristics vary along a single dimension, in which case works compete exclusively with their two adjacent neighbors. The localized nature of differentiated products competition can be substantially mitigated if spatial competition

²⁰⁶ See *supra* note 112 and accompanying text.

²⁰⁷ Interestingly, the market need not be divided into discrete subsegments in order for this effect to occur. Variations in the density of firms across the product space can balkanize the industry into a chain of “overlapping oligopolies,” each comprised of a small number of firms engaged in localized competition regardless of how many firms are operating in the overall market. This can give rise to the same problems even in the absence of actual gaps in the product continuum. See Eaton & Lipsey, *supra* note 26, at 750, 763–64; Nicholas Kaldor, *Mrs. Robinson’s “Economics of Imperfect Competition,”* 1 *ECONOMICA* 335, 339–40 (1934). This result provides an answer to the criticism that Chamberlinian product groups are nothing more than Marshallian industries. See MILTON FRIEDMAN, *The Methodology of Positive Economics*, in *ESSAYS IN POSITIVE ECONOMICS* 3, 38–39 (1953) (offering this criticism); GEORGE J. STIGLER, *Monopolistic Competition in Retrospect*, in *FIVE LECTURES ON ECONOMIC PROBLEMS* 12, 17 (1949) (same), *reprinted in* GEORGE J. STIGLER, *THE ORGANIZATION OF INDUSTRY* 309, 314 (1968). It suggests that products that do not compete directly with one another may nonetheless be linked together through a chain of competitive products.

²⁰⁸ See Eaton & Lipsey, *supra* note 26, at 750.

²⁰⁹ By underscoring the role that fixed costs and substitutability among works plays in the analysis, preference asymmetries simply may reinforce the same conclusions discussed above. In other words, to the extent that preference asymmetries cause competition within the subsegment to be insufficiently robust, the solution may be to thicken the market by increasing the size and the intensity of copyright protection.

occurs along more than one dimension.²¹⁰ Empirical studies have confirmed these insights.²¹¹

Finally, the policy instruments that follow from the differentiated products approach are by their nature extremely contextual and do not lend themselves to simple application. In addition, the interrelationships among the available policy instruments make calibrating them simultaneously an extremely difficult empirical exercise.²¹² The fact that the differentiated products approach is contextual and nuanced should not obscure its basic analytical power.²¹³ Indeed, the intuitions that the theory reveals about the relationship between access and efficiency, the manner in which the various aspects of copyright protection interrelate, and the true relationship between copyright and public goods theory are sufficient to justify further inquiry.

²¹⁰ One theoretical analysis concluded that if spatial competition takes place on three dimensions, each work may compete with as many as six adjacent neighbors. If competition expands to four dimensions, works may theoretically compete with as many as half the works operating in the product group. See Archibald & Rosenbluth, *supra* note 111, at 576–84.

²¹¹ See Feenstra & Levinsohn, *supra* note 111, at 36–41 (modeling competition among automobiles sold in 1987 model year as taking place along four product dimensions—weight, horsepower, reliability, and whether car was European—and finding that, on average, cars competed with 5.90 other models).

²¹² See *supra* note 192 and accompanying text.

²¹³ See *supra* note 201 and accompanying text.