Copyright and Public Good Economics: A Misunderstood Relation

Christopher S. Yoo

ABSTRACT

The conventional approach to analyzing the economics of copyright is based on the premise that copyrightable works constitute pure public goods, which is generally modeled by assuming that such works are nonexcludable and that the marginal cost of making additional copies of them is essentially zero. These assumptions in turn imply that markets systematically produce too few copyrightable works and underutilize those that are produced. In this Article, Professor Christopher Yoo argues that the conventional approach is based on a fundamental misunderstanding. A close examination of the foundational literature on public good economics reveals that the defining characteristic of public goods is the need to satisfy an optimality criterion known as the "Samuelson condition," which suggests that the systematic bias toward underproduction is the result of the inability to induce consumers to reveal their preferences rather than nonexcludability and zero marginal cost. Reframing the analysis in terms of the Samuelson condition also expands the number of ways in which the assumptions underlying pure public goods can be relaxed. In so doing, it suggests that markets for copyrighted works are more properly analyzed as impure public goods. Unlike markets for pure public goods, markets for impure public goods exhibit no systematic bias toward underproduction and are not bounded away from providing efficient levels of utilization. The insights of impure public goods theory thus have broad implications for a wide range of copyright-related issues, including fair use, duration, compulsory licenses, database protection, digital rights management, and derivative works.

Copyright and Public Good Economics: A Misunderstood Relation

Christopher S. Yoo[†]

INTRODUCTION

Scholarship on the economics of copyright has been dominated by the assumption that copyrightable works are pure public goods.[1] The most frequently cited definition of pure public goods focuses on two characteristics. First, pure public goods are *nonexcludable*, in that producers cannot provide their benefits to one consumer without simultaneously providing the benefits to other consumers. Second, pure public goods are *nonrival*, in that the consumption of

---

[†] Professor of Law and Director of the Technology and Entertainment Law Program, Vanderbilt University Law School. I would like to thank participants at workshops conducted at the University of Michigan Law School and the Vanderbilt University Law School, as well as John Conley, Andy Daughety, Paul Edelman, Gerry Faulhaber, James Gibson, Paul Heald, Bob Inman, Mark Lemley, Glynn Lunney, Kristen Madison, Ted Parsons, Adam Pritchard, Bob Rasmussen, Jennifer Reinganum, Chris Sanchirico, Polk Wagner, and Joel Waldfogel, for comments on earlier drafts of this Article, and Catherine Sloan and Daniel Burks-Goodman for their research assistance. Financial support from the Vanderbilt Dean's Fund is gratefully acknowledged. All errors are my responsibility.

[1] *See* Ben Depoorter & Francesco Parisi, *Fair Use and Copyright Protection: A Price Theory Explanation*, 21 INT'L REV. L. & ECON. 453, 465 n.4 (2002) (calling the assumption that copyright is a pure public good "part of the collective wisdom of mainstream economic analysis"). For the seminal statement tying intellectual property to the theory of pure public goods, 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, 614-16 (Nat'l Bureau of Econ. Research ed., 1962). For leading examples within the copyright literature, see Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 HARV. L. REV. 281, 281 (1970); William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659, 1700-05 (1988); 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, 1610-11 (1982); William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J. LEGAL STUD. 325, 326 (1989); Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989, 994-99 (1997). For overviews of the economics of pure public goods, see RICHARD CORNES & TODD SANDLER, THE THEORY OF EXTERNALITIES, PUBLIC GOODS, AND CLUB GOODS 143-239 (2d ed. 1996); William H. Oakland, *Theory of Public Goods*, *in* 2 HANDBOOK OF PUBLIC ECONOMICS 485, 486-99, 502-22 (Alan J. Auerbach & Martin Feldstein eds., 1987).

Of course, a wide range of noneconomic justifications for copyright also exist. *See generally* William Fisher, *Theories of Intellectual Property*, *in* NEW ESSAYS IN THE LEGAL AND POLITICAL THEORY OF PROPERTY 168, 170-73, 184-94 (Stephen R. Munzer ed., 2001). I leave extended discussion of copyright's noneconomic aspects to other work. *See* Christopher S. Yoo, *Copyright and Democracy: A Cautionary Note*, 53 VAND. L. REV. 1933, 1953-63 (2000) (critiquing democratic theories of copyright).

the good by one consumer does not reduce the supply available for consumption by others.[2] Nonrivalry is generally modeled by assuming that the marginal cost of making an additional copy of a copyrightable work is zero.[3] These assumptions imply that markets provide insufficient incentives to produce copyrightable works and provide insufficient access to those works that are produced. They also imply that any attempt to alleviate the problems of underproduction necessarily worsens the problems of underutilization and vice versa. The conventional approach thus frames copyright as a tradeoff between access and incentives that is necessarily second best in both dimensions.[4]

In my prior work, I have critiqued the conventional approach, focusing on how product differentiation can mitigate these economic problems.[5] In this Article, I extend my critique by returning to the fundamental economic characteristics of pure public goods first identified by Paul Samuelson. Interestingly, Samuelson did not regard either nonexcludability or zero marginal cost as the distinctive characteristic of pure public goods.[6] Instead, Samuelson focused

---

[2] *See* R.A. Musgrave, *Provision for Social Goods*, *in* PUBLIC ECONOMICS: AN ANALYSIS OF PUBLIC PRODUCTION AND CONSUMPTION AND THEIR RELATIONS TO THE PRIVATE SECTORS 124, 126-29 (Julius Margolis & Henri Guitton eds., 1969).

[3] For illustrative examples, see Yochai Benkler, *An Unhurried View of Private Ordering in Information Transactions*, 53 VAND. L. REV. 2063, 2066, 2070, 2078 (2000); James Boyle, *Cruel, Mean, or Lavish? Economic Analysis, Price Discrimination and Digital Intellectual Property*, 53 VAND. L. REV. 2007, 2013 (2000); Timothy J. Brennan, *Copyright, Property, and the Right To Deny*, 68 CHI.-KENT L. REV. 675, 698 (1993); Mark A. Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1053-54 (2005); Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 292 (1996). Other commentators assume that marginal cost is nonzero, but constant. *E.g.*, Landes & Posner, *supra* note 1, at 326-27, 333. Relaxing the strict assumption of zero marginal cost in this manner does not materially affect the analysis.

[4] *See, e.g.*, 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"); Landes & Posner, *supra* note 1, at 326 ("Striking the correct balance between access and incentives is the central problem in copyright law.").

[5] Christopher S. Yoo, *Copyright and Product Differentiation*, 79 N.Y.U. L. REV. 212 (2004).

[6] *See* Paul A. Samuelson, *The Pure Theory of Public Expenditure*, 36 REV. ECON. & STAT. 387 (1954) [hereinafter Samuelson, *Pure Theory*]. Indeed, Samuelson only noted the pricing problems posed by declining average cost as an afterthought. *See* Paul A. Samuelson, *Diagrammatic Exposition of a Theory of Public Expenditure*, 37 REV. ECON. & STAT. 350, 356 (1955) [hereinafter Samuelson, *Diagrammatic Exposition*] ("I

on another feature: the fact that the same quantity of production can appear as an argument in more than one person's consumption function.[7] Indeed, each person who purchases the public good simultaneously consumes the entire output of the public good.[8] This characteristic gives rise to an interesting inversion of the conditions for the efficient allocation of private goods. For private goods, consumers pay the *same price* and signal the different valuations that they place on the good by purchasing *different quantities*. For pure public goods, consumers consume the *same quantity* of production and signal the intensity of their preferences by their willingness to pay *different prices*.

This characteristic dictates that optimal production of public goods requires satisfying the "Samuelson condition," which is generally recognized as the key feature distinguishing public goods from private goods.[9] The Samuelson condition requires expanding the production of public goods so long as the aggregate marginal benefits derived by all consumers exceeds the marginal cost of increasing production of those goods. The problem is that when consumers express the intensity of their preferences through prices rather than quantities, there is no way to induce consumers to reveal their marginal valuations. On the contrary, the fact that the same quantity can appear as an argument in more than one person's consumption function gives

---

believe I did not go far enough in claiming for [my theoretical model] relevance to the vast area of decreasing costs that constitutes an important part of economic reality . . . . I must leave to future research discussion of these vital issues.").

[7] Paul A. Samuelson, *Aspects of Public Expenditure Theories*, 40 REV. ECON. & STAT. 332, 334 (1958) [hereinafter Samuelson, *Aspects*] (noting that public goods "*simultaneously* enter into many persons' indifference curves").

[8] Samuelson, *Diagrammatic Exposition*, *supra* note 6, at 350; Samuelson, *Pure Theory*, *supra* note 6, at 387.

[9] *See* CORNES & SANDLER, *supra* note 1, at 23-24 (describing how the Samuelson condition distinguishes public and private goods); Oakland, *supra* note 1, at 489 (calling the Samuelson condition "novel"). Indeed, Samuelson himself regarded the formulation of this condition as his primary contribution to the study of public goods. *See* Samuelson, *Pure Theory*, *supra* note 6, at 388 (defining the condition and calling it the "new element" that serves as the basis for his "pure theory of government expenditure on collective consumption goods").

consumers the incentive to understate the value they place on the public good in the hopes that other consumers will bear a larger proportion of the first-copy costs.

The absence of any reliable way to determine the aggregate marginal value that consumers place on a public good makes it all but impossible to determine the optimal level of production for any public good.[10] As Samuelson himself noted, this problem of incentive incompatibility would remain even if the problems associated with nonexcludability and nonmarginal cost pricing were somehow solved.[11] Although scholars have proposed a number of ingenious methods for inducing consumers to reveal their true demands,[12] all of these methods suffer from shortcomings and limitations of their own.[13]

Reframing the problem posed by pure public goods in terms of preference revelation and the incentive incompatibility implicit in the Samuelson condition, instead of nonexcludability and zero marginal cost, not only recharacterizes the fundamental policy problems posed by the economics of copyright. It also suggests new solutions. From the very beginning, critics have pointed out that private goods and pure public goods represent polar cases and that many, if not

[10] Samuelson, *Pure Theory*, *supra* note 6, at 388-89; *accord* Samuelson, *Diagrammatic Exposition*, *supra* note 6, at 355 (noting the difficulty of getting consumers to reveal their preferences for pure public goods so that optimal production can be determined); Samuelson, *Aspects*, *supra* note 7, at 334 (same).

[11] Samuelson, *Aspects*, *supra* note 7, at 335-36.

[12] For examples of such systems, see Edward H. Clarke, *Multipart Pricing of Public Goods*, 11 PUB. CHOICE 17 (1971); Theodore Groves & John Ledyard, *Optimal Allocation of Public Goods: A Solution to the "Free Rider" Problem*, 45 ECONOMETRICA 783 (1977); Theodore Groves & Martin Loeb, *Incentives and Public Inputs*, 4 J. PUB. ECON. 211 (1975); William Vickrey, *Counterspeculation, Auctions, and Competitive Sealed Tenders*, 16 J. FIN. 8 (1961). For earlier solutions that predate the formalization of public good economics, see Erik Lindahl, *Just Taxation—A Positive Solution* (1919), *reprinted in* CLASSICS IN THE THEORY OF PUBLIC FINANCE 168 (Richard A. Musgrave & Alan T. Peacock eds., Elizabeth Henderson trans., 1958); Knut Wicksell, *A New Principle of Just Taxation* (1896), *reprinted in* CLASSICS IN THE THEORY OF PUBLIC FINANCE, *supra*, at 72 (J.M. Buchanan trans.).

[13] For surveys of this literature, see CORNES & SANDLER, *supra* note 1, at 198-239; Jean-Jacques Laffont, *Incentives and the Allocation of Public Goods*, *in* HANDBOOK OF PUBLIC ECONOMICS, *supra* note 1, at 537, 554-66; Oakland, *supra* note 1, at 522-30.

most, goods fall somewhere in between these two extremes.[14] Samuelson himself recognized the existence of such intermediate cases, but questioned the tractability of the problems they posed.[15]

Notwithstanding Samuelson's pessimism about the likely fruitfulness of the enterprise, a major literature has emerged exploring "impure public goods."[16] The best-developed literature on impure public goods focuses on the economics of congestion, derived largely from Charles Tiebout's work on "local public goods" and James Buchanan's pioneering work on "club goods."[17] Although congestion costs are sometimes described as reintroducing a degree of rivalry, they do not in fact prevent the same quantity of production from appearing as an argument in more than one person's consumption function. Put another way, optimal production of impure public goods must still satisfy the Samuelson condition.

Instead, congestion costs are better understood as introducing a new dimension along which a particular public good's contribution to economic welfare can vary. In addition to varying according to *price* and *quantity*, characteristics such as congestion acknowledge that public goods can vary in terms of their *quality*. In this sense, the theory of impure public goods can be regarded as relaxing the assumption underlying pure public goods theory that the relevant

---

[14] *See, e.g.*, Stephen Enke, *More on the Misuse of Mathematics in Economics: A Rejoinder*, 37 REV. ECON. & STAT. 131, 132 (1955) (noting the existence of a large number of intermediate goods that do not fit into Samuelson's theory); Julius Margolis, *A Comment on the Pure Theory of Public Expenditure*, 37 REV. ECON & STAT. 347, 347-48 (1955) (observing that governments provide many goods that do not conform to Samuelson's strict definition).

[15] Samuelson, *Aspects*, *supra* note 7, at 335-36.

[16] For surveys of the literature on impure public goods, see CORNES & SANDLER, *supra* note 1, at 255-72, 347-479; Oakland, *supra* note 1, at 499-509.

[17] James M. Buchanan, *An Economic Theory of Clubs*, 32 ECONOMICA (n.s.) 1, 2 (1965); Charles M. Tiebout, *A Pure Theory of Local Expenditures*, 64 J. POL. ECON. 416 (1956). For an overview of this literature and an application of the economics of congestion to the Internet, see Christopher S. Yoo, *Network Neutrality and the Economics of Congestion*, 94 GEO. L.J. 1847, 1863-1900 (2006).

goods are homogenous.[18] Subsequent work has moved beyond congestion to explore other dimensions along which quality can vary.[19]

As Tiebout first pointed out, variations in quality create the possibility that individual consumers will reveal their preferences by reallocating their purchases to different providers in order to maximize quality. This mobility can give rise to de facto markets for public goods in which consumers reveal the intensity of their preferences spatially, even when they lack the means to do so through the quantities they consume and lack the incentive to do so through the prices they pay.[20] Depending on the shape of the congestion function, it is theoretically possible that markets will provide and allocate impure public goods in an efficient manner.[21] The systematic bias toward underproduction disappears.

A shift to an impure public goods perspective on copyright thus would have the potential to transform the basic policy inferences generally regarded as inherent in the economics of copyright. This is not to say that a shift to an impure public goods approach would be a panacea. To say that markets *can* support the optimal production and allocation of impure public goods is

---

[18] This assumption is usually made only implicitly. For examples in which this assumption is made explicitly, see Robert B. Ekelund, Jr. & Joe R. Hulett, *Joint Supply, the Taussig-Pigou Controversy, and the Competitive Provision of Public Goods*, 16 J.L. & ECON. 369, 381 (1973); Earl A. Thompson, *The Perfectly Competitive Production of Collective Goods*, 50 REV. ECON. & STAT. 1, 2 (1968).

[19] *See infra* Part III.B.

[20] Tiebout, *supra* note 17, at 419-21, 424.

[21] *See, e.g.*, CORNES & SANDLER, *supra* note 1, at 351 ("Under a wide variety of circumstances, these clubs can achieve Pareto-optimal results without resorting to government provision."); Samuelson, *Aspects*, *supra* note 7, at 335 (noting that in impure public goods, "we might find just the right conditions of scarcity of space and of independence of consumptions" so that ordinary pricing "happens . . . to pick up each indirect external marginal utility"); Suzanne Scotchmer, *Public Goods and the Invisible Hand*, *in* MODERN PUBLIC FINANCE 93, 94 (John M. Quigley & Eugene Smolensky eds., 1994) ("The thrust of the modern literature on clubs is that admissions to clubs are private goods like any others, and that we should therefore expect the market to perform well in the sense of the first welfare theorem . . . ."). The same holds true for the strand of impure public goods known as spatial competition. *See* B. Curtis Eaton & Myrna Holtz Wooders, *Sophisticated Entry in a Model of Spatial Competition*, 16 RAND J. ECON. 282, 289-92 (1985) (analyzing circumstances under which spatial competition models achieve efficiency); Oakland, *supra* note 1, at 529 ("Under certain idealized conditions . . . mobility can lead to efficient levels of spatial public goods."); Joseph E. Stiglitz, *The Theory of Local Public Goods*, *in* THE ECONOMICS OF

not to say that they always *will*. Again, depending on the shape of the relevant congestion function, it is quite possible for markets to reach equilibrium with either too many or too few impure public goods. Unlike in the case of private goods, there is no "invisible hand" inexorably guiding the equilibria for impure public goods toward efficiency.[22] The equilibria for impure public goods thus fall somewhere between the polar cases of efficient production (as is the case with perfect competition for private goods) and systematic market failure (as is the case with pure public goods). Instead, the policy inferences are more ambiguous and fact specific, in that both efficient production and market failure are possible. Thus, to the extent that the resulting equilibrium tends toward too few impure public goods, policy responses exist that simultaneously promote optimal production and utilization. It is only when the market reaches equilibrium with too many impure public goods that a tension exists between optimal production and utilization. The impure public goods approach thus contradicts the conventional wisdom that access and incentives are always and inherently in tension. It also suggests, again in sharp contrast to the conventional approach, that the more difficult policy problem is the potential for overproduction, rather than underproduction.

Despite the potential insights of returning to the fundamentals of public good economics by analyzing copyright through the lens of the Samuelson condition, an extended exploration of the connection has yet to appear in the literature.[23] This Article seeks to rectify that state of

---

PUBLIC SERVICES 274, 311, 312 (Martin S. Feldstein & Robert P. Inman eds., 1977) (noting that spatial models can reach equilibria that maximize social welfare).

[22] *See* B. Curtis Eaton & Richard G. Lipsey, *Product Differentiation*, *in* 1 HANDBOOK OF INDUSTRIAL ORGANIZATION 723, 742 (Richard Schmalensee & Robert D. Willig eds., 1989) (noting the absence of an "invisible hand" with respect to spatial competition); Scotchmer, *supra* note 21, at 99 (finding the same with respect to club goods).

[23] A search of the Westlaw JLR database identifies only five articles that even mention the Samuelson condition. A similar search of the ALLREV database in the LAWREV library of Lexis turns up only four references. None of those articles addresses copyright law. Furthermore, only two copyright articles mention the

affairs. Part I describes the conventional approach to the economics of copyright, demonstrating how it has been based on nonexcludability and zero marginal cost and then analyzing how it has been applied in the context of copyright to such issues as fair use, duration, compulsory licenses, databases, digital rights management (DRM), and derivative works. Part II examines the true source of market failure identified by Samuelson's foundational work on public good economics, which is the inability of markets to induce consumers to reveal their true preferences. Part III explores the major strands of the literature on impure public goods theory, focusing first on the economics of congestion and second on spatial competition. It shows how market-based outcomes can approach first-best solutions that the conventional approach suggests are unattainable, while also discussing the ways that private ordering can fall short of optimality. Part IV applies the insights from impure public goods theory to the copyright doctrines introduced in Part I. In offering this analysis, I do not purport to offer a definitive resolution of any particular area of copyright law. My discussion is simply intended to demonstrate how embracing a different set of intuitions could reorient the way questions about copyright law are framed.

---

difficulty in determining the intensity of consumers' preferences for public goods, and neither of those articles analyzes the problem at any depth or even refer to the Samuelson condition: David J. Brennan, *Fair Price and Public Goods: A Theory of Value Applied to Retransmission*, 22 INT'L REV. L. & ECON. 347, 367 (2002); Alfred C. Yen, *The Legacy of* Feist*: Consequences of the Weak Connection Between Copyright and the Economics of Public Goods*, 52 OHIO ST. L.J. 1343, 1367 (1991). The only previous paper of which I am aware that explicitly links copyright and the economics of impure public goods is Stanley M. Besen & Sheila Nataraj Kirby, *Private Copying, Appropriability, and Optimal Copying Royalties*, 32 J.L. & ECON. 255, 257, 264-70, 280 (1989). That article models congestion simply by positing the presence of constantly increasing marginal cost, which fails to capture the problems of incentive incompatibility associated with the Samuelson condition. *See infra* note 164.

I. THE CONVENTIONAL APPROACH TO APPLYING PUBLIC GOOD ECONOMICS TO COPYRIGHT

The theory of pure public goods has undergone a fairly radical transformation since it was first expounded by Paul Samuelson in 1955. What began as a framework for determining the proper scope of public expenditure has evolved into a technical term of art that is no longer coterminous with goods that must be provided by the government.[24]

This Part lays out the way that public good economics has traditionally been applied in the copyright literature. Section A focuses on the role of nonexcludability. Section B explores the implications of modeling nonrivalry as zero marginal cost. Section C reviews the commentary applying the conventional approach to current copyright-related issues, including fair use, the copyright term, compulsory licenses, protection of databases, DRM, and protection of derivative uses. Section D discusses the analytical shortcomings of the conventional approach.

A. Nonexcludability

As noted above, nonexcludability is often held up as one of the defining characteristics of a pure public good. Consider lighthouses, which have long been regarded as a classic example of a nonexcludable good.[25] Nonexcludability means that lighthouse services cannot be provided

[24] That Samuelson initially envisioned his work as a comprehensive theory is underscored by the fact that he titled his initial exposition "The Pure Theory of Public Expenditure." Samuelson, *Pure Theory*, *supra* note 6, at 387. Samuelson later expressed regret over formulating the title in this manner, recognizing that his theory was underinclusive in that governments often provide goods and services for reasons aside from those addressed by his theory. Samuelson, *Diagrammatic Exposition*, *supra* note 6, at 355-56. The scholarship on impure public goods, discussed in Part III, *infra*, reveals that Samuelson's theory was also overinclusive in that markets can efficiently provide certain types of public goods without government intervention.

[25] *See* JOHN STUART MILL, *Principles of Political Economy*, *in* 3 COLLECTED WORKS OF JOHN STUART MILL 968 (J.M. Robson ed., 1965) (using the lighthouse example); A.C. PIGOU, THE ECONOMICS OF WELFARE 183-84 (4th ed. 1938) (same); HENRY SIDGWICK, THE PRINCIPLES OF POLITICAL ECONOMY 406 (3d ed. 1901) (same). Other oft-cited examples of nonexcludable goods include fire and police protection, fireworks displays, and national defense.

to ships that have paid for those services without simultaneously providing them to other ships in the area that have not paid for them. As a result, nonexcludability gives rise to a positive externality that can cause systematic market failure.[26] Standard economic theory dictates that lighthouses should be created whenever the social benefits they would generate exceed the costs needed to create and operate them. If the revenues captured by lighthouses accurately reflect the social benefits they create, private ordering would effectively ensure that this condition is met. A profit-maximizing lighthouse owner would compare the revenue it would receive to the costs it would incur and would operate the lighthouse so long as doing so would generate net profits. Nonexcludability causes the revenue generated by lighthouses to fall short of their social benefits. For example, if two ships find a mechanism for coordinating their activities, they could agree to pay a single fee and then both benefit from a lighthouse's services. Alternatively, a ship could try to free ride on lighthouse services for which another ship has paid. In either event, the total revenue captured by the lighthouse would understate the social value of the lighthouse. This inevitably causes some lighthouses to cease operating or fail to be constructed even though the benefits they would have created would have exceeded their costs.

Some commentators accept the notion that copyrightable works are nonexcludable, in that the ready availability of copying technologies keeps authors who have once sold their works from preventing nonpaying customers from obtaining access to those works.[27] The market failures associated with nonexcludability have traditionally provided one of the central

[26] *See, e.g.*, PIGOU, *supra* note 25, at 331 (offering the classic discussion of how externalities can cause some industries to produce suboptimal levels of output); Francis M. Bator, *The Anatomy of Market Failure*, 72 Q.J. ECON. 351, 370 (1958) (describing how nonappropriability can cause market failure).

[27] Indeed, a number of leading law and economics textbooks analyze the economics of copyright in terms of nonexcludability. *E.g.*, ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 42-43, 108-09 (3d ed. 2000); HOWELL E. JACKSON ET AL., ANALYTICAL METHODS FOR LAWYERS 362-63 (2003).

justifications for copyright. By providing legal remedies against those who copy works without paying for them, copyright makes works at least somewhat excludable, although the costliness of enforcement dictates that the exclusion that copyright provides will inevitably remain somewhat imperfect.

B. Nonrivalry as Zero Marginal Cost

As noted earlier, the other characteristic generally thought to define a pure public good is nonrivalry, which occurs when consumption by one person does not reduce the supply available for consumption by others. Again, the lighthouse is often used to illustrate the concept.[28] The fact that one ship benefits from a lighthouse's services does not reduce the supply of lighthouse services available to other ships.

Copyrightable works are generally considered to be nonrival in this manner. Once the fixed costs needed to create the first copy of a particular work have been incurred, any number of copies of the original can be made without reducing the supply available for additional copies. As noted earlier, the copyright literature has typically modeled nonrivalry by assuming that the marginal cost is zero across all volumes of production.[29]

Zero marginal cost gives rise to a classic pricing problem. One of the basic principles of welfare maximization is that individuals should be permitted to consume a good whenever the benefits they would derive from consuming the good exceed the costs of permitting them to do so. Assuming that the prices individuals pay provide an accurate reflection of the benefits they derive, economic welfare is maximized if price is set to equal marginal cost. Thus, if a creative

work that could be costlessly copied were priced efficiently in terms of access (i.e., priced at marginal cost), it would be priced at zero.[30] Pricing at zero, however, would cause the work to generate no revenue whatsoever, in which case the author would have no incentive to produce the work in the first place.[31] This implies that providing authors with sufficient incentive to produce creative works requires giving them the means to set prices that exceed marginal cost. Any such means, however, would necessarily reduce access below efficient levels by excluding some consumers even though the benefits they would have derived from consuming the work would have exceeded the costs of allowing them to do so. In other words, any attempt to provide additional incentives for the creation of copyrightable works necessarily exacerbates the welfare losses associated with insufficient access.

---

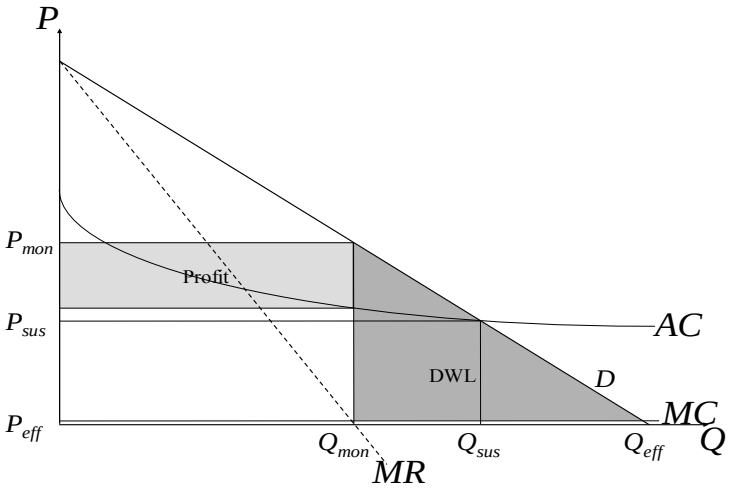
[28] For illustrations, see DENNIS W. CARLTON & JEFFREY M. PERLOFF, MODERN INDUSTRIAL ORGANIZATION 82 (4th ed. 2005); PAUL A. SAMUELSON & WILLIAM D. NORDHAUS, ECONOMICS 37-38 (18th ed. 2005); JOSEPH E. STIGLITZ, ECONOMICS OF THE PUBLIC SECTOR 128 (3d ed. 2000).

[29] *See supra* note 3 and accompanying text.

[30] *See, e.g.*, RICHARD A. MUSGRAVE & PEGGY B. MUSGRAVE, PUBLIC FINANCE IN THEORY AND PRACTICE 51 (2d ed. 1976) (concluding, in the case of nonrival social goods, that "[e]fficient resource use requires that price equal marginal cost, but marginal cost . . . is zero, and so should be price"); Samuelson, *Aspects*, *supra* note 7, at 335 (noting that the marginal cost of public broadcasting is zero, and implying that the cost of listening should also be zero).

[31] This conclusion does not depend on the extreme assumption that marginal cost of reproduction is zero. Indeed, the same problem arises under positive marginal cost so long as the fixed costs are sufficiently large that production falls on the declining portion of the average cost curve. When that is the case, the average cost necessarily lies above the marginal cost curve, and any price that equals marginal cost will necessarily fall below average cost, and fail to allow the work to break even.

Figure 1: The Conventional Approach to Modeling the Economics of Copyright



These effects can be illustrated using Figure 1, which has become standard in the copyright literature.[32] The exclusivity provided by copyright permits authors to charge prices that exceed marginal cost. Left to their own devices, authors will produce at the point where the revenue they would generate from selling an additional copy (*MR*) no longer exceeds the cost of making an additional copy (*MC*), leading them to set price and quantity at P_{mon} and Q_{mon}. Because this is the point that maximizes authors' profits, it is also necessarily the point that maximizes their incentives to create copyrightable works. At the same time, economic welfare would be maximized if price were set equal to marginal cost, which would lead to the price and

[32] This Figure is adapted from Yoo, *supra* note 5, at 227 fig.1. For examples of similar figures appearing in other work, see Julie E. Cohen, *Copyright and the Perfect Curve*, 53 VAND. L. REV. 1799, 1802 fig.A (2000); William W. Fisher III, *Property and Contract on the Internet*, 73 CHI.-KENT L. REV. 1203, 1236 fig.2 (1998); Fisher, *supra* note 1, at 1701 n.201 fig.1, 1708 n.232 fig.2; Edmund W. Kitch, *Elementary and Persistent Errors in the Economic Analysis of Intellectual Property*, 53 VAND. L. REV. 1727, 1732 fig.1 (2000); S.J. Liebowitz, *Copyright Law, Photocopying, and Price Discrimination*, 8 RES. L. & ECON. 181, 185 fig.1 (1986); Robert P. Merges, *Intellectual Property Rights and the New Institutional Economics*, 53 VAND. L. REV. 1857, 1858 (2000); Stewart E. Sterk, *Rhetoric and Reality in Copyright Law*, 94 MICH. L. REV. 1197, 1206 fig.1 (1996).

Figure 1 differs from the graphs appearing in previous commentary in one important respect: the sources cited above portray the entire difference between price and marginal cost (i.e., the 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

quantity represented by P_{eff} and Q_{eff}. The exclusion of consumers who would derive net benefits from consuming the work creates deadweight loss (represented by the dark grey triangle). Thus, from the standpoint of allocative efficiency, copyright allows authors to charge prices that are too high (represented by the difference between P_{mon} and P_{eff}) and to sell quantities that are too low (represented by the difference between Q_{mon} and Q_{eff}). To the extent that this price also exceeds average cost, exclusivity also allows the monopolist to earn supracompetitive returns (represented by the light grey rectangle).

This analysis suggests that markets can be expected to exhibit a systematic bias toward underutilization of goods with zero marginal cost. One classic policy response promotes access and reduces supracompetitive returns by calibrating copyright doctrine to force authors to bring their production closer to efficient levels. Absent price discrimination, the lowest sustainable price and quantity is where the demand curve crosses the average cost curve, represented by P_{sus} and Q_{sus}. This price (indeed any price that would allow authors to break even) necessarily exceeds marginal cost and thus is inevitably second-best in terms of access. Because any such intervention would necessarily force authors to charge less than their profit-maximizing price, this solution also inevitably reduces incentives.

Another classic policy response to market bias toward underutilization of zero marginal cost goods is facilitating authors' ability to engage in price discrimination in the hope that allowing them to charge low-value users a lower price than high-value users will induce authors to serve the inefficiently excluded consumers between Q_{sus} and Q_{eff}. Indeed, commentators have long acknowledged that perfect price discrimination can help allocate public goods in an

---

of the producer surplus lying above the average cost curve properly can be regarded as profit. Yoo, *supra* note 5, at 226 n.46.

efficient manner.[33] Although forcing high-value users to pay more than low-value users may seem unfair, wealth transfers from consumers to producers have no impact on efficiency.

Other commentators have taken a less sanguine view of price discrimination. Perfect price discrimination is a practical impossibility, and the welfare implications of imperfect price discrimination are ambiguous.[34] Any system of price discrimination also requires the incurrence of implementation costs.[35] In addition, to the extent that price discrimination enhances authors' ability to extract consumer surplus from inframarginal consumers (consumers represented by the quantities purchased from the origin to Q_{sus}), it will increase authors' ability to earn supracompetitive returns.[36] Rather than facilitating price discrimination, these commentators would prefer the more traditional approach of calibrating copyright to trade off second-best outcomes in terms of both access and incentives. Some also point out that the shortfall between Q_{sus} and Q_{eff} can be redressed through a wide range of alternative institutional forms, such as secondary markets and libraries, which can enhance low-value users' ability to obtain access to

[33] *See* JACK HIRSHLEIFER & JOHN G. RILEY, THE ANALYTICS OF UNCERTAINTY AND INFORMATION 259 (1992) (identifying perfect price discrimination as one of the classic solutions to the problems posed under the "traditional analysis" of information as a public good). For an overview of these arguments, see Yoo, *supra* note 5, at 230.

[34] For the seminal analysis of the welfare implications of imperfect price discrimination, see JOAN ROBINSON, THE ECONOMICS OF IMPERFECT COMPETITION 188-202 (2d ed. 1969). For a more contemporary discussion, see JEAN TIROLE, THE THEORY OF INDUSTRIAL ORGANIZATION 137-39, 142-49 (1988). For an application to copyright, see WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 339-40, 378, 389 (2003); Michael J. Meurer, *Copyright Law and Price Discrimination*, 23 CARDOZO L. REV. 55, 78-79, 100 (2001).

[35] *See* Benkler, *supra* note 3, at 2072, 2079 (arguing that implementing price discrimination is costly); Meurer, *supra* note 34, at 101-02 (observing that price discrimination "induces . . . wasteful rent-seeking costs"); Neil Weinstock Netanel, *Market Hierarchy and Copyright in Our System of Free Expression*, 53 VAND. L. REV. 1879, 1914-15 (2000) (noting that price discrimination requires investments to identify and sort consumers into different segments).

[36] *See* Boyle, *supra* note 3, at 2025-26 (arguing that perfect price discrimination simply transfers surplus from consumers to producers); Meurer, *supra* note 34, at 92-93, 98-102 (suggesting that price discrimination may cause undesirable redistribution of consumer surplus, decrease output, and induce rent-seeking); Michael J. Meurer, *Price Discrimination, Personal Use and Piracy: Copyright Protection of Digital Works*, 45 BUFF. L. REV. 845, 877-80 (1997) (observing that price discrimination leads to more profits for producers); Netanel, *supra* note 3, at 293 n.31 (noting that price discrimination allows copyright owners to capture a larger share of the consumer surplus).

copyrighted works.[37] Indeed, with respect to customers who would otherwise be inefficiently excluded from purchasing, permitting them to free ride completely and granting them access to the work is arguably a better choice.

Finally, a number of commentators have entertained the possibility of using government subsidies to solve the marginal cost pricing problem.[38] This would obviate the need for authors to recover their first-copy costs through the prices they charge and would allow copyrighted works to be sold at marginal cost. Indeed, as Harold Hotelling noted in his classic 1938 article, declining average costs caused by large, up-front fixed costs represent one of the classic justifications for government subsidies.[39]

C. Applications of the Conventional Approach to Specific Copyright Doctrines

Commentators have relied on the putative tradeoff between access and incentives implicit in the conventional approach when using public good economics to analyze a wide range of copyright-related issues. These include the fair use doctrine, copyright duration, compulsory licenses, database protection, DRM, and derivative works.

---

[37] *See* Cohen, *supra* note 32, at 1806 (arguing that price discrimination theories do not account for alternate means of access, such as second-hand markets and libraries); Jonathan Weinberg, *Hardware-Based ID, Rights Management, and Trusted Systems*, *in* THE COMMODIFICATION OF INFORMATION 343, 357-59 (Niva Elkin-Koren & Neil Weinstock Netanel eds., 2002) (observing that price discrimination would cut off sharing and secondary markets).

[38] For classic copyright articles discussing the use of subsidies to permit copyrighted works to be distributed at marginal cost, see Arrow, *supra* note 1, at 623; Breyer, *supra* note 1, at 306-07; Robert M. Hurt & Robert M. Schuchman, *The Economic Rationale of Copyright*, 56 AM. ECON. REV. 421, 426 (1966); William R. Johnson, *The Economics of Copying*, 93 J. POL. ECON. 158, 171-72 (1985); Arnold Plant, *The Economic Aspects of Copyright in Books*, 1 ECONOMICA (n.s.) 167, 193 (1934).

[39] *See* Harold Hotelling, *The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates*, 6 ECONOMETRICA 242, 242 (1938) (arguing that "taxes might well be applied to cover the fixed costs of electric power plants, waterworks, railroads, and other industries in which fixed costs are large, so as to reduce to the level of marginal cost the prices charged for the services and products of these industries"). For a modern analysis applying Hotelling's insights to intellectual property, see John F. Duffy, *The Marginal Cost Controversy in Intellectual Property*, 71 U. CHI. L. REV. 37 (2004).

1. Fair Use

The dominant economic justification for fair use regards it as a means of compensating for market failures induced by transaction costs. Under this rationale, fair use is justified by the fact that transaction costs can prevent low-value users from obtaining access to copyrighted works even though economic welfare would increase if they were permitted to do so.[40] In the tradition of the analysis of liability rules pioneered by Guido Calabresi and Douglas Melamed,[41] this argument would create what amounts to a compulsory license priced at zero whenever friction in the bargaining process prevents low-value, welfare-enhancing transactions from occurring. Consistent with this interpretation, courts have limited fair use to copying that does not adversely affect the market for the copyrighted work.[42] The Supreme Court has called the effect on the potential market for the copyrighted work "undoubtedly the single most important element of fair use."[43]

Over time, the emergence of new institutional arrangements (including performing rights organizations, such as Broadcast Music, Inc. (BMI) and the American Society for Composers,

---

[40] For the seminal statement of this argument, see Gordon, *supra* note 1, at 1614-22, 1627-30. For later, similar arguments, see Timothy J. Brennan, Harper & Row v. The Nation, Inc.*: Copyrightability and Fair Use*, 33 J. COPYRIGHT SOC'Y U.S.A. 368, 382 (1986); Landes & Posner, *supra* note 1, at 357-58.

[41] Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1106-07, 1119-21 (1972).

[42] *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 593 (1994) ("Evidence of substantial harm to [the relevant market] would weigh against a finding of fair use . . . ."); Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 566-67 (1985) ("'Fair use, when properly applied, is limited to copying by others which does not materially impair the marketability of the work which is copied.'" (citation omitted)); Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 450-51 (1984) (rejecting fair use when "the particular use is harmful, or that if it should become widespread, it would adversely affect the potential market for the copyrighted work"); Princeton Univ. Press v. Mich. Document Servs., Inc., 99 F.3d 1381, 1386-88 (6th Cir. 1996) (en banc) (refusing to find fair use for copying that impaired the market for licensing photocopies for coursepacks); *cf.* 17 U.S.C. § 107(4) (2000) (including "the effect of the use upon the potential market for or value of the copyrighted work" as one of the factors to be considered in determining the scope of fair use).

[43] *Harper & Row*, 471 U.S. at 566. *But see* Am. Geophysical Union v. Texaco Inc., 60 F.3d 913, 926 (2d Cir. 1994) (suggesting that the Supreme Court may no longer regard the effect on the potential market as being of paramount importance).

Authors, and Publishers (ASCAP),[44] and copyright collectives, such as the Copyright Clearance Center (CCC);[45] new distribution and communication technologies, such as the Internet;[46] and the advent of self-help technologies, such as DRM[47]) have reduced the transaction costs of licensing low-value uses of copyrighted works. Were transaction costs the only economic justification for fair use, these developments would support a significant contraction of its scope.[48]

These developments have placed renewed importance on alternative economic justifications for fair use, including those based on public good economics. Some have argued that fair use is needed to mitigate the welfare losses associated with allowing authors to charge prices that exceed marginal cost by making it possible for low-value users who are inefficiently excluded by the price mechanism to obtain access to copyrighted works.[49] Others have justified fair use as a way to prevent copyright holders from earning supracompetitive returns.[50] The fact

---

[44] For descriptions of BMI and ASCAP, see Broadcast Music, Inc. v. CBS, Inc., 441 U.S. 1, 4-5 (1979); Stanley M. Besen et al., *An Economic Analysis of Copyright Collectives*, 78 VA. L. REV. 383, 385-86, 401-02 (1992). For a general description of performing rights organizations, see Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CAL. L. REV. 1293, 1328-40 (1996).

[45] For descriptions of the CCC, see *Am. Geophysical Union*, 60 F.3d at 930; Besen et al., *supra* note 44, at 386-87.

[46] For observations that these technologies have decreased transaction costs, see Jane C. Ginsburg, *Authors and Users in Copyright*, 45 J. COPYRIGHT SOC'Y U.S.A. 1, 15 (1997); 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, 881 (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).

[47] *See infra* Part I.C.4 (describing DRM).

[48] Note that the emergence of new markets for low-value uses would not redress market failures that arise with respect to uses such as parody, in which bargaining fails because the would-be parodist is locked into a bilateral monopoly with the original author. For arguments to this effect, see *Cardtoons, L.C. v. Major League Baseball Players Ass'n*, 95 F.3d 959, 975 (10th Cir. 1996); LANDES & POSNER, *supra* note 34, at 158-59; Robert P. Merges, *Are You Making Fun of Me?: Notes on Market Failure and the Parody Defense in Copyright*, 21 AIPLA Q.J. 305, 308-12 (1993).

[49] John Cirace, *When Does Complete Copying of Copyrighted Works for Purposes Other Than for Profit or Sale Constitute Fair Use? An Economic Analysis of the* Sony Betamax *and* Williams & Wilkins *Cases*, 28 ST. LOUIS U. L.J. 647, 657-58, 660-62 (1984); Fisher, *supra* note 1, at 1700-19.

[50] Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1135-36 (1990); Sterk, *supra* note 32, at 1211-12.

that market failure is endemic under the theory of pure public goods has led some scholars to question the usefulness of market failure as a benchmark for determining the scope of fair use.[51]

2. Duration

Commentators have also invoked public good economics as support for limitations on the duration of the copyright term. These commentators accept the access/incentives tradeoff implicit in the conventional approach, acknowledging that although authors must be given the exclusivity necessary to charge the supramarginal cost prices required to support the creation of the work in the first instance, those rights inevitably impose deadweight losses. The need to balance these two considerations implies a copyright term of limited duration that provides sufficient incentive to induce the creation of the work, but thereafter allows the work to become freely available to all at marginal cost.[52]

3. Compulsory Licenses

As noted earlier, high transaction costs have provided the traditional justification for compulsory licenses,[53] as evidenced by the consistency with which the government has invoked this rationale when enacting compulsory licenses.[54] The reduction in transaction costs associated

---

[51] *See* Glynn S. Lunney, Jr., *Fair Use and Market Failure:* Sony *Revisited*, 82 B.U. L. REV. 975, 996 (2002) ("Because market failure is inevitable, the concept of market failure cannot serve as a useful guide in determining which uses of a copyrighted work should be fair . . . .").

[52] *See* William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 475-76 (2003). For other evaluations of copyright duration in terms of public good economics, see Wendy J. Gordon, *Authors, Publishers, and Public Goods: Trading Gold for Dross*, 36 LOY. L.A. L. REV. 159, 164-66, 170-77 (2002); Avishalom Tor & Dotan Oliar, *Incentives To Create Under a "Lifetime-Plus-Years" Copyright Duration: Lessons from a Behavioral Economic Analysis for* Eldred v. Ashcroft, 36 LOY. L.A. L. REV. 437, 446-49 (2002).

[53] *See supra* notes 40-41 and accompanying text.

[54] H.R. REP. NO. 94-1476, at 89 (1976), *reprinted in* 1976 U.S.C.C.A.N. 5659, 5704 (supporting compulsory licenses for cable retransmission of broadcast signals because "it would be impractical and unduly burdensome to require every cable system to negotiate with every copyright owner whose work was retransmitted by a cable

with digital transmission, networking, and the emergence of copyright collectives has undercut this justification, which has placed renewed emphasis on alternative justifications for compulsory license.[55]

For example, some commentators justify compulsory licenses as a means to force copyright owners to allow greater access to their works.[56] Envisioning compulsory licenses as a way to promote access suggests that compulsory licenses can also be viewed as a way to resolve the tradeoff between access and incentives implicit in the traditional approach to pure public goods.[57] Other scholars have implicitly drawn on arguments favoring the use of liability rules when valuation is difficult[58] to theorize that the difficulties in getting customers to reveal their preferences for pure public goods in a truthful manner justifies making the work available for a standard royalty set by the government. The government would base this royalty on new survey methodologies that are better able to ascertain the intensity of consumers' preferences for particular public goods.[59]

---

system"); INFO. INFRASTRUCTURE TASK FORCE, INTELLECTUAL PROPERTY AND THE NATIONAL INFORMATION INFRASTRUCTURE: THE REPORT OF THE WORKING GROUP ON INTELLECTUAL PROPERTY RIGHTS 52 (1995) ("In certain circumstances, particularly where transaction costs are believed to dwarf per-transaction royalties, Congress has found it necessary to provide for compulsory licenses.").

[55] *See supra* notes 44-48 and accompanying text; INFO. INFRASTRUCTURE TASK FORCE, *supra* note 54, at 52 (concluding that "[t]echnology will facilitate individual licensing schemes" and that "under current conditions, additional compulsory licensing of intellectual property rights is neither necessary nor desirable").

[56] *E.g.*, Jane C. Ginsburg, *Creation and Commercial Value: Copyright Protection of Works of Information*, 90 COLUM. L. REV. 1865, 1926 (1990) ("[T]he real purpose of a compulsory license is to reduce the extent to which the copyright ownership of the covered work conveys monopoly power, so that the copyright owner must make the work available to all who wish to access and exploit it.").

[57] *See* Charles Lubinsky, *Reconsidering Retransmission Consent: An Examination of the Retransmission Consent Provision (47 U.S.C. § 325(b)) of the 1992 Cable Act*, 49 FED. COMM. L.J. 99, 128-30, 140 (1996) (describing the public goods problem inherent in television programming and noting the argument that compulsory licenses are designed "to resolve the standard public goods problem dealing with the makers of creative works").

[58] *See* Calabresi & Melamed, *supra* note 41, at 1106-07 (arguing that liability rules are preferable to property rules when consumers have incentives to conceal their true valuations to appropriate a higher proportion of the available surplus).

[59] *See* Brennan, *supra* note 23, at 367-75.

4. Database Protection

Commentators have also drawn on the economics of pure public goods when analyzing the proper scope of database protection. For example, Alfred Yen employs public good economics to critique the Supreme Court's *Feist* decision limiting copyright protection to databases in which creators have exercised creativity in the selection and arrangement of data.[60] Yen takes the traditional approach to public good economics as his starting point, defining pure public goods in terms of nonexcludability and nonrivalry. Because databases satisfy these conditions, Yen notes that they will be subject to systematic underproduction, but is concerned that any attempt to increase incentives for their production will run afoul of the tradeoff between access and incentives. He thus proposes limiting any protection for databases to those that are unlikely to recoup their costs of production.[61] The problem is that creativity in selection and arrangement bears no relation to the likelihood of recoupment, which depends on factors such as the magnitude of the first-mover advantage, the ratio of sales to development costs, the extent to which the database could be financed through the sale of complementary goods and advertising, the availability of copy protection, and the database producer's ability to engage in price discrimination.[62] As a result, Yen recommends abandoning creative selection as the touchstone of copyright protection for databases in favor of an approach that bears a stronger relation to the economics of public goods.[63]

---

[60] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 348-49, 362-63 (1991).

[61] Yen, *supra* note 23, at 1365-69.

[62] *Id.* at 1369-73. This aspect of Yen's argument is reminiscent of a classic line of articles analyzing the economics of copyright, which suggests that first-mover advantages, threats of retribution, and other methods might be sufficient to permit authors to recover their fixed costs even in the absence of copyright protection. *See* Breyer, *supra* note 1, at 299-306; Hurt & Schuchman, *supra* note 38, at 427-29; Plant, *supra* note 38, at 173-75.

[63] Yen, *supra* note 23, at 1374, 1377.

James Gibson similarly associates markets for databases with the problems posed by public good economics. Like Yen, Gibson begins his analysis by positing that databases confront the classic problems of nonexcludability and supramarginal cost pricing.[64] Although the fact that databases once had to be reified in a concrete form has historically struck a balance between access and incentives by necessarily creating a degree of excludability and rivalry, the digitization of data and the advent of copy and access protection have upset this balance by giving database creators greater control over their works.[65] Thus, at least with respect to sole source data that cannot be independently compiled by others,[66] Gibson proposes requiring that database creators deposit copies of their databases in a central repository, available for privileged uses and ready for release into the public domain once the term of protection has expired.[67]

5. Digital Rights Management

Public good economics has also influenced the debate about DRM, in which sellers of creative works use license terms and technological copy protection to impose restrictions greater than those established by copyright law. Some have lauded this development, arguing that by facilitating price discrimination, DRM will increase access to copyrighted works.[68] Others have taken a less sanguine view, arguing that DRM allows parties to alter the balance between access and incentives struck by the copyright statute.[69] Opponents contend that not only is there no

---

[64] James Gibson, *Re-reifying Data*, 80 NOTRE DAME L. REV. 163, 172-74 (2004).
[65] *Id.* at 179-81, 189-98.
[66] *Id.* at 216-20.
[67] *Id.* at 233-39.
[68] *E.g.*, 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, 587-90 (1998); Fisher, *supra* note 32, at 1234-40; 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, 1169 (1998).
[69] *E.g.*, Niva Elkin-Koren, *Copyright Policy and the Limits of Freedom of Contract*, 12 BERKELEY TECH. L.J. 93, 101 (1997); Mark A. Lemley, *Intellectual Property and Shrinkwrap Licenses*, 68 S. CAL. L. REV. 1239, 1277-78

guarantee that DRM will necessarily lead to greater access,[70] but also that it introduces bias toward certain types of content and content providers[71] and enhances copyright holders' ability to capture supracompetitive returns.[72]

6. Derivative Uses

Commentators have also invoked public good economics when analyzing the proper scope of derivative use rights. As with the other aspects of copyright protection, derivative uses pose the familiar tradeoff implicit in the conventional approach: giving broad protection to derivative uses increases the incentives for creating copyrightable works, but introduces a degree of allocative inefficiency by denying some customers access to those works even when it would be welfare enhancing to permit access. In the context of derivative uses, however, these arguments receive an additional twist because the derivative use right also prevents subsequent authors from creating new works based on prior material. Economic welfare is thus reduced not only by the static efficiency losses resulting from the inability of some consumers to obtain

---

(1995); Netanel, *supra* note 3, at 385; David A. Rice, *Public Goods, Private Contract and Public Policy: Federal Preemption of Software License Prohibitions Against Reverse Engineering*, 53 U. PITT. L. REV. 543, 544-46, 591, 619-21 (1992); *cf.* Maureen A. O'Rourke, *Copyright Preemption After the* ProCD *Case: A Market-Based Approach*, 12 BERKELEY TECH. L.J. 53, 80 (1997) (describing this argument without endorsing it).

[70] *See* LANDES & POSNER, *supra* note 34, at 40 (noting the absence of a firm theoretical or empirical basis for believing that imperfect price discrimination is likely to increase output); Benkler, *supra* note 3, at 2079 (arguing that imperfect price discrimination's impact on aggregate social welfare is an empirical question that cannot be determined a priori); Meurer, *supra* note 36, at 894-98 (concluding that price discrimination made possible by contract in addition to broad copyright protection may decrease output). *See generally* Yoo, *supra* note 5, at 230 & n.59 (collecting sources on the ambiguous impact of imperfect price discrimination on output).

[71] *See* Cohen, *supra* note 32, at 1811 (arguing that price discrimination will not encourage access to goods for which there are few substitutes); Netanel, *supra* note 35, at 1915 (stating that "[f]irms with extensive content inventories and an established customer base" are better able to exploit the advantages of price discrimination).

[72] *See* Boyle, *supra* note 3, at 2021-23 (providing an example of price discrimination leading to increased profits); Meurer, *supra* note 36, at 877-80 (noting that "[m]ore price discrimination means more profit to the sellers of digital works").

access to the works that exist today, but also by the dynamic efficiency losses resulting from the works that would be created in the future.[73]

For example, Glynn Lunney argues that public good economics justifies drawing a distinction between derivative users and ordinary users of a copyrighted work. Each ordinary user typically must purchase a copy of the work in order to obtain its benefits. Derivative users, in contrast, seek to incorporate elements of the original work into a new work of authorship, thereby exploiting a work's "public good aspect": they typically need to purchase only a single copy of the original work in order to serve multiple customers of the derivative work.[74] Absent a separate derivative use right, authors would have no choice but to charge the same price to both ordinary and derivative users, which would inefficiently exclude some ordinary users. The derivative use right allows authors to charge derivative users prices that reflect the fact that the single sale to the derivative user actually serves as a proxy for multiple purchases of the original work.[75] Lunney would limit the scope of copyright's derivative work protection to uses that implicate the original work's public good aspects, which are those with a transformative component.[76]

---

[73] *See* Lemley, *supra* note 1, at 994-99 (describing the costs associated with limiting follow-on innovators' ability to access existing works); Glynn S. Lunney, Jr., *Reexamining Copyright's Incentives-Access Paradigm*, 49 VAND. L. REV. 483, 629 (1996) (arguing that when derivative works are involved, welfare maximization must take into account the production of new works as well as the allocation of existing works).

[74] Lunney, *supra* note 73, at 635-38.

[75] *Id.* at 639-40.

[76] *Id.* at 641-45. Lunney views the quantum of additional creative expression required for a follow-on work to fall outside the scope of the derivative use right as quite small. In Lunney's words, "any significant transformation of or variation from the underlying work should preclude a finding of infringement even if the underlying work remains recognizable." *Id.* at 650. Lunney makes his point about derivative uses as part of a larger claim that strengthening copyright protection can impose opportunity costs by diverting resources from more economically beneficial activities. *Id.* at 488-89. This argument presumes that the overall economy is already in general equilibrium, which would only be true if the level of copyright protection were already calibrated correctly. Furthermore, if the market is not in general equilibrium, it is theoretically possible that strengthening copyright could cause economic welfare to increase as well as decrease. Yoo, *supra* note 5, at 241 n.95.

Mark Lemley similarly concludes that public good economics can help delineate the proper scope of the derivative use right. Although copyright typically balances the tradeoff between access and incentives implicit in the conventional approach to public good economics by carefully calibrating the scope and duration of copyright protection,[77] additional complexities arise when a copyrightable work is simultaneously a good consumed by ordinary users and an input used by derivative users to create new works of authorship. As a theoretical matter, the holder of the copyright in the original work has every incentive to license the work in a way that maximizes its value.[78] As a practical matter, however, markets for licensing copyrighted works are often impeded by a number of imperfections—such as transaction costs, uncertainty, externalities, strategic behavior, and noneconomic incentives—that can inefficiently limit access and tip the balance away from the access side of the tradeoff by preventing welfare-enhancing licensing from occurring.[79]

Accordingly, Lemley argues that giving the author of the original work complete control over all derivative works would not strike the proper balance. Instead, he advocates a system of divided entitlements similar to the system used in patent law, in which the initial author would retain a copyright over the original work, while authors of derivative works would control the additional copyrightable expression that they have added. Under this approach, follow-on authors cannot publish or otherwise commercialize their derivative works without obtaining a license from the original author. At the same time, the original author cannot use the additional

[77] Lemley, *supra* note 1, at 994-99.
[78] *Id.* at 1047.
[79] *Id.* at 1048-67.

creative contribution embodied in the derivative work without first receiving the follow-on author's permission.[80]

According to Lemley, this system of divided entitlements would recalibrate the balance between access and incentives by giving copyright holders greater incentive to reach licensing agreements with authors of derivative works. Dividing entitlements in this manner encourages licensing agreements by dictating that absent such an agreement, neither the initial author nor the follow-on author will be able to take advantage of the improvements.[81]

In addition, giving follow-on authors a degree of copyright protection offers a solution to Arrow's information paradox.[82] The absence of such protection places authors of derivative works in a Catch-22. Negotiating licenses requires follow-on authors to disclose the nature of their derivative works in order to allow the initial authors to assess their value. The absence of any independent copyright protection in derivative works leaves initial authors free to appropriate them without the follow-on authors' consent once their content has been disclosed. The risk of losing the entirety of the derivative work makes follow-on authors understandably reluctant to engage in licensing negotiations, which reduces access below optimal levels and forces initial authors to take a greater role in identifying potential innovators. Giving derivative works a degree of independent copyright protection would allow follow-on authors to initiate contact with initial authors with greater impunity, which would in turn promote access by making welfare-enhancing licensing agreements easier to reach.[83]

---

[80] *Id.* at 1062, 1074-77.
[81] *Id.* at 1062-63.
[82] *See* Arrow, *supra* note 1, at 615 (describing the paradox).
[83] *See* Lemley, *supra* note 1, at 1062, 1068-69.

D. The Analytical Shortcomings of the Conventional Approach

The basic policy inferences that follow from the conventional approach to applying public good economics to copyright have exerted significant influence over the economic analysis of copyright. It is thus interesting that Samuelson did not regard either nonexcludability or zero marginal cost to be essential attributes of pure public goods.[84] If well taken, Samuelson's challenge to the conventional approach would have sweeping implications for the economic analysis of copyright.

Consider first the role of nonexcludability. The claim that nonexcludability inevitably causes market failure for copyrightable works has increasingly come under empirical and conceptual attack. As an empirical matter, the emergence of copy protection and DRM has greatly increased authors' ability to employ self-help in preventing nonpaying customers from obtaining access to their works.[85] Indeed, it has long been recognized that exclusion is typically possible, with the costs of exclusion depending on the state of technology.[86] Restated in terms of the lighthouse example, the problem is not that the exclusion of nonpaying ships is impossible, but rather that excluding them would be prohibitively costly.[87]

As a conceptual matter, the work of Ronald Coase has shown that private ordering may be better able than previously thought to correct for the market failures caused by externalities. The renowned Coase theorem holds that so long as transaction costs are low, the parties may be able to bargain around externalities to reach the efficient result without government

---

[84] *See* Samuelson, *Aspects*, *supra* note 7, at 335-36.
[85] *See supra* Part I.C.4 (discussing DRM).
[86] *E.g.*, Bator, *supra* note 26, at 374-75.
[87] *Id.* at 376 n.5.

intervention.[88] This insight complemented Coase's earlier work showing how the choice of institutional form can minimize transaction costs.[89] He later applied this approach to the problem of public goods in his classic critique of the claim that the nonexcludability of lighthouse services required that they be provided by the government. He pointed out that private entities were able to provide English lighthouses for decades by identifying another, more easily excludable good—specifically port usage—that could serve as a proxy for the consumption of lighthouse services.[90] Subsequent questions about the applicability of this insight to early English lighthouses[91] have not blunted Coase's core insight about the ability of alternative institutional forms to reduce transaction costs to the point where markets can emerge. This has been demonstrated eloquently by the emergence of performing rights organizations, such as BMI, ASCAP, and the CCC.[92]

On a more fundamental level, a close analysis of the literature on public goods reveals that whether or not a good is excludable does not eliminate the need to satisfy the Samuelson condition, which as noted earlier is generally recognized as the distinguishing characteristic of public goods. I will postpone detailed analysis of the different roles that nonexcludability and

---

[88] R.H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 15 (1960). Coase illustrated his point through the classic scenario in which a factory's smokestack imposes negative externalities on residents living nearby. One solution is for the government to impose a tax on the factory equal to the amount of harm it imposes on the residents. Another solution is for the private residents to pay the factory not to pollute. *Id.* at 41-42.

[89] R.H. Coase, *The Nature of the Firm*, 4 ECONOMICA (n.s.) 386, 392 (1937).

[90] R.H. Coase, *The Lighthouse in Economics*, 17 J.L. & ECON. 357, 360-61 (1974), *reprinted in* FAMOUS FABLES OF ECONOMICS 32, 32 (Daniel F. Spulber ed., 2002); *cf.* Yoo, *supra* note 17, at 1873-85 (drawing on Coase's critique of the lighthouse market failure to show how alternative institutional arrangements can provide market-based solutions to externalities in the context of the Internet).

[91] *See* Richard A. Epstein, *The Libertarian Quartet*, REASON, Jan. 1999, at 61, 64-65 ("The only way the fee can be charged is through the exercise of state monopoly power at the port."); Andrew Odlyzko, *The Evolution of Price Discrimination in Transportation and Its Implications for the Internet*, 3 REV. NETWORK ECON. 323, 325-26, 341-42 (2004) (noting that English lighthouse fees were set by government charters, not commercial negotiations); David E. Van Zandt, *The Lessons of the Lighthouse: "Government" or "Private" Provision of Goods*, 22 J. LEGAL STUD. 47, 48 (1993) (arguing that the provision of lighthouse services could not be characterized as a "private enterprise").

[92] *See supra* notes 44-45 and accompanying text.

the Samuelson condition play in the analysis of pure public goods until after the discussion of the foundations of public good economics appearing in the next Section. For now, it suffices to point to Samuelson's observation that the fundamental problems surrounding public goods would remain even if those goods were rendered completely excludable.[93] As a result, a number of leading public good theorists have questioned whether nonexcludability should be regarded as part of the definition of a pure public good.[94]

Conceptual problems also surround the fact that the conventional approach models nonrivalry with the assumption that marginal cost is zero. Indeed, were zero marginal cost the only problem, public good economics would simply be an application of the general problem of declining average cost that dominates the study of public utilities and natural monopoly, and the solution would simply require allocating fixed costs across different outputs.[95] Samuelson clearly rejected this claim, arguing that public goods raised concerns that are distinct from and independent of the problems of joint supply.[96] Although Samuelson acknowledged that his theory did have implications for declining average costs,[97] he did not regard deviations from marginal cost pricing as the central problem posed by pure public goods. Indeed, as Samuelson

---

[93] Samuelson, *Aspects*, *supra* note 7, at 335 ("Being able to limit a public good's consumption does not make it a true-blue private good."); *see also* J.G. Head, *Public Goods and Public Policy*, 17 PUB. FIN. 197, 215 (1962) (tracing the decreasing role played by nonexcludability in Samuelson's work).

[94] In the words of one leading commentator, "the significance of exclusion rests with the characteristics of private market provision of public goods and the financing options open to the government should it decide to provide the public good, but not with the fundamental properties of public goods themselves." Oakland, *supra* note 1, at 491. For other examples of this view, see STEPHEN SHMANSKE, PUBLIC GOODS, MIXED GOODS, AND MONOPOLISTIC COMPETITION 7, 17-20 (1991); Bator, *supra* note 26, at 374-75; Oakland, *supra* note 1, at 486. For similar conclusions appearing in the commentary on copyright, see Brennan, *supra* note 23, at 350; Brennan, *supra* note 3, at 686 n.43.

[95] James Buchanan, *Joint Supply, Externality and Optimality*, 33 ECONOMICA (n.s.) 404, 408 (1966); Harold Demsetz, *The Private Production of Public Goods*, 13 J.L. & ECON. 293, 293, 304-06 (1970).

[96] Paul A. Samuelson, *Contrast Between Welfare Conditions for Joint Supply and for Public Goods*, 51 REV. ECON. & STAT. 226 (1969); Samuelson, *Aspects*, *supra* note 7, at 355; *see also* Ekelund & Hulett, *supra* note 18 at 387; Oakland, *supra* note 1, at 490-91.

pointed out, the problems he identified would remain even if one used government subsidies to allow producers to price at marginal cost.[98] For reasons that I will subsequently explore in greater detail, pricing copyrighted works at marginal cost would not solve the essential difficulty in getting consumers to reveal the intensity of their preferences. In other words, even if the government used subsidies to allow producers to price at marginal cost, it would still face insuperable problems when determining how big those subsidies should be.

II. THE SAMUELSON CONDITION AS THE TRUE FOUNDATION OF PUBLIC GOOD ECONOMICS

There is thus reason to question whether nonexcludability and zero marginal cost capture the essence of public good economics. Why this is the case can be best understood by returning to the original conception of public good economics articulated by Samuelson. Framing the issues in terms of nonexcludability and zero marginal cost overlooks what Samuelson regarded as the defining characteristic of pure public goods. Specifically, the fundamental problem is that consumers of pure public goods have both the motivation and the ability to understate the intensity of their preferences. This incentive incompatibility is what Samuelson saw as the true root of the market's tendency to underproduce public goods.

A. The Baseline Case of Private Goods

The economic problems posed by pure public goods are most easily understood by comparing a two-person economy involving two private goods with a two-person economy

---

[97] *See* Samuelson, *Diagrammatic Exposition*, *supra* note 6, at 356 (acknowledging that his initial analysis "did not go far enough in claiming for it relevance to the vast area of decreasing costs").

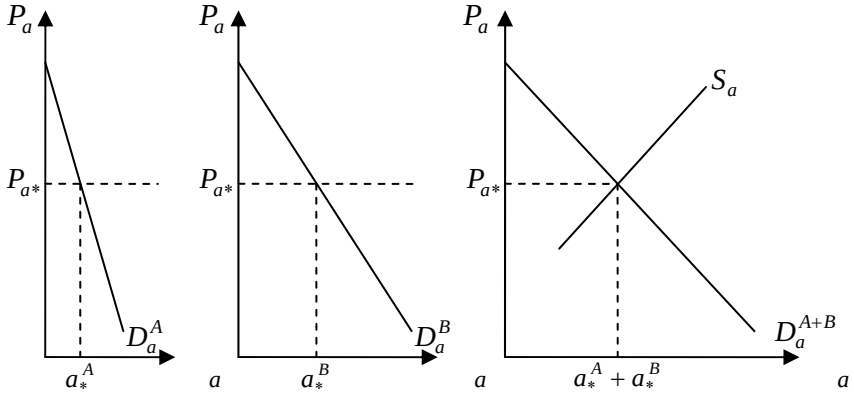
[98] Samuelson, *Aspects*, *supra* note 7, at 335-36.

involving a private good and a public good.[99] For the first economy, assume that the society is populated by two people, Adam and Beth, who each have a demand for apples and oranges. Both goods are clearly rival, in that Adam's consumption of apples and oranges reduces the supply of each available for consumption by Beth and vice versa. Both goods are also clearly divisible, in that Adam's decision to consume a particular quantity of apples or oranges does not require that Beth consume the same quantity. Figure 2 represents both Adam's and Beth's demand curves for apples, with the quantity of apples (a) depicted on the horizontal axis and the price of apples (P_a) depicted on the vertical axis. Adam's demand curve is denoted by D_a^A, while Beth's demand curve is denoted by D_a^B.

The market demand curve can be derived simply by adding together the quantity of apples that Adam and Beth would demand at any particular price. In other words, the market demand curve is the *horizontal* summation of each consumer's individual demand curves. The equilibrium can be determined by superimposing a market supply curve on the market demand curve, which leads to an equilibrium price of P_a* (as depicted in Figure 2). At this point, Adam consumes a_*^A and Beth consumes a_*^B. Note that there is no reason to assume that a_*^A and a_*^B will be equal. In other words, both Adam and Beth pay the same price and reveal the intensity of their respective preferences by consuming different quantities.

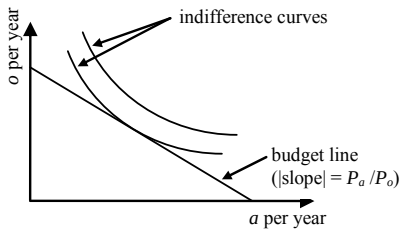
[99] This specific example is adapted from HARVEY S. ROSEN, PUBLIC FINANCE 58-63 (7th ed. 2005). For a more mathematical treatment of the distinction between private and pure public goods, see HAL R. VARIAN, MICROECONOMIC ANALYSIS 144-57 (3d ed. 1992).

Figure 2: Aggregation of Demand for Private Goods (Horizontal Summation)



The resulting equilibrium has the significant property of allocating apples in a Pareto-efficient manner. According to standard consumer theory, Adam and Beth adjust their purchases until both of their marginal rates of substitution of apples for oranges (MRS_{ao}) equal the price of apples divided by the price of oranges (P_a/P_o).[100] Because only relative prices matter, the price of oranges can be arbitrarily set equal to \$1 without loss of generality. If so, the condition for maximizing surplus is $MRS_{ao} = P_a$, and the price of apples represents the rate at which an individual is willing to substitute apples for oranges. Because Adam's demand curve (D_a^A)

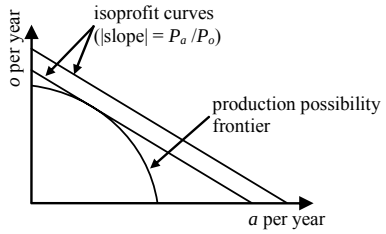
[100] Proofs of this relation appear in every standard microeconomic textbook and can be easily illustrated in the context of the two-good economy involving apples and oranges discussed above. A particular consumer's willingness to trade apples for oranges can be used to generate a set of indifference curves. The farther the indifference curve is from the origin, the higher the level of utility achieved. The slope at any point along the indifference curve is the consumer's marginal rate of substitution of apples for oranges (MRS_{ao}). Consumers are also subject to budget constraints, represented by a straight line with a slope of $-P_a/P_o$. The consumer has the incentive to try to reach the indifference curve farthest from the origin given the relevant budget constraint. Optimal consumption will occur at the point where the indifference curve is tangent to the budget line. At this point, $MRS_{ao} = P_a/P_o$. If $MRS_{ao} > P_a/P_o$, then Adam could increase his utility by increasing his consumption of apples and decreasing his consumption of oranges. The marginal utility he derives from apples will fall and the marginal utility he derives from oranges will rise until $MRS_{ao} = P_a/P_o$.



shows the maximum price he would pay to consume a particular quantity of apples, it also represents his MRS_{ao} for any particular level of apple consumption. Similarly, D_a^B represents Beth's MRS_{ao} schedule. In equilibrium, both Adam and Beth set $MRS_{ao} = P_{a*}$. At the same time, the supply curve for apples (S_a) represents the marginal rate of transformation of apples for oranges (MRT_{ao}) at any particular level of production. In equilibrium, the producer sets $MRT_{ao} = P_{a*}$.[101]

Thus, in equilibrium $MRS_{ao}^A = MRS_{ao}^B = MRT_{ao}$. All consumers receive the same marginal utility from each good, but they consume different quantities. Because neither the consumers nor the producers can make themselves better off by moving to any other point, the resulting equilibrium is Pareto optimal. Equally importantly, neither Adam nor Beth has any incentive to misrepresent the value each places on apples. Because the uniform price is determined by the market, the only way they can increase the utility that they derive is by varying the quantities of the good that they purchase. They could purchase less than or more than their preferred quantity of the good, but doing so would simply have the effect of lowering

[101] On the supply side, the scarcity of inputs necessitates a tradeoff between the number of apples and the number of oranges that can be produced. This is depicted by the production-possibility frontier, which represents the maximum amount of the two products that can be jointly produced given the existing resource constraints. The slope of the production-possibility frontier at any point is the marginal rate of transformation of apples or oranges (MRT_{ao}). At the same time, producers will be willing to forgo selling apples so long as they can make up for the lost revenue by selling additional oranges. This tradeoff can be used to generate isoprofit curves, with the curves located farther from the origin representing higher levels of profit. The slope of the isoprofit curves is necessarily $-P_a/P_o$. The producer would like to reach the highest level of profit permitted by its resource constraints. This occurs where the production-possibility frontier is tangent to the isoprofit curve, which necessarily implies that $MRT_{ao} = P_a/P_o$.



the utility they derive. Thus, absent other sources of market failure, when goods are divisible the first welfare theorem of neoclassical economics indicates that markets are likely to support efficient levels of production and consumption of private goods.

B. Pure Public Goods

A starkly different situation arises with respect to pure public goods. In the original Samuelsonian conception of nonrivalry, the central feature is not jointness in production, typically modeled by zero marginal cost, but rather jointness in consumption, which, as noted earlier, means that consumption by one person does not reduce the supply available for consumption by others. Stated somewhat more formally, nonrivalry allows the same quantity to serve as an argument in both Adam's and Beth's consumption functions.[102] Consumption of a good is fully joint when everyone who purchases the good necessarily consumes the entire industry output, although they may pay different prices. When that is the case, the good is described as being *indivisible*,[103] which means that if both Adam and Beth purchase the good, each necessarily consumes a good of the same magnitude.

In addition, even though a producer of a pure public good cannot provide one consumer with a different level of services than any other consumer, the producer typically can increase or decrease the total amount of services provided to all consumers by varying the amount of resources put into producing any particular pure public good. For example, even though

---

[102] *See supra* note 7 and accompanying text.

[103] HOWARD R. BOWEN, TOWARD SOCIAL ECONOMY 172-73 (1948) [hereinafter BOWEN, SOCIAL ECONOMY]; JAMES M. BUCHANAN, THE DEMAND AND SUPPLY OF PUBLIC GOODS 174-76 (1968); SAMUELSON & NORDHAUS, *supra* note 28, at 372; Bator, *supra* note 26, at 374; Howard R. Bowen, *The Interpretation of Voting in the Allocation of Economic Resources*, 43 Q.J. ECON. 27, 27 (1943) [hereinafter Bowen, *Voting*]. Indeed, the leading book-length analysis of public good economics regards the terms "nonrivalry of consumption" and "indivisibility of benefits" as synonymous. CORNES & SANDLER, *supra* note 1, at 8.

lighthouse owners cannot vary the amount of lighthouse services provided on a customer-by-customer basis, they can alter the total services provided to all customers by increasing the lighthouse's height or its brightness. Similarly, although the government cannot provide strategic defense to one household without simultaneously providing it to all neighboring households, it can increase the amount of strategic defense provided by increasing the size of the defense forces or by investing more resources on equipment and training.

Copyrighted works are often described as being indivisible in precisely this manner.[104] For example, although film studios or record companies may vary the amount of resources devoted to producing any particular movie or song, once the work has been completed, all viewers and listeners of that movie or song must necessarily consume a product of the same magnitude. Similarly, software developers can vary the amount of resources to increase or decrease the level of sophistication of any particular software package. Once that level has been set, however, all users necessarily consume a software package of the same size.[105]

At first glance, the assertion that copyrighted works are indivisible may appear to be inconsistent with the fact that different people purchase different numbers of copies of particular works. For example, some customers opt to see a particular movie multiple times, while others choose to view it only once. Similarly, some users may purchase multiple copies of a particular software package, while others may purchase only a single copy.

---

[104] 1 PAUL GOLDSTEIN, COPYRIGHT § 1.14.1, at 1:45 (2d ed. 1996); Arrow, *supra* note 1, at 615; J.H. Reichman, *Legal Hybrids Between the Patent and Copyright Paradigms*, 94 COLUM. L. REV. 2432, 2434 n.1, 2442 n.43 (1994).

[105] Of course, developers may choose to market different versions of the same software package. As a formal matter, new versions based on the original program are more properly regarded as derivative works that are conceptually distinct from the original. On a more general level, versioning is probably best understood as a way to separate the intensity of different consumers' preferences than as introducing a degree of divisibility.

The key to unraveling this conundrum is to keep in mind the distinction between the creative expression itself and the medium in which it is stored. Copyright protects only the former.[106] The intangible aspects of the creative expression that is the copyrightable work are nondepletable, in that one can make an infinite number of copies of it without reducing the supply available for consumption by others. Recognizing that copyright protects only the intangible aspects of a creative work also makes it easier to characterize copyrighted works as indivisible, since nothing prevents the same intangible property from appearing as an argument in more than one consumer's consumption function. In addition, the resources that went into producing that intangible component are necessarily the same for all consumers of the creative work.

Thus, even though some consumers may choose not to read, view, use, or listen to the entirety of a particular work, while other consumers may choose to obtain multiple copies of the same work, the magnitude of the intangible property that each consumes (as measured by the number of resources that went into producing it) is precisely the same. The fact that some consumers choose to purchase multiple copies is better regarded as an indication of the intensity of their preference for the copyrighted work rather than consumption of a different quantity. And as we shall see, the fact that the pure public good is an input rather than a finished good does not materially affect the analysis.[107]

---

[106] *See* 17 U.S.C. § 202 (2000) ("Ownership of a copyright, or of any of the exclusive rights under a copyright, is distinct from ownership of any material object in which the work is embodied."); H.R. REP. NO. 94-1476, *supra* note 54, at 124, *reprinted in* 1976 U.S.C.C.A.N. 5659, 5739 ("The principle restated in section 202 is a fundamental and important one: that copyright ownership and ownership of a material object in which the work is embodied are entirely separate things."). The fact that copyright does not attach until the work is fixed in a tangible medium of expression, 17 U.S.C. § 101, does not alter the fact that copyright protects only the intangible property.

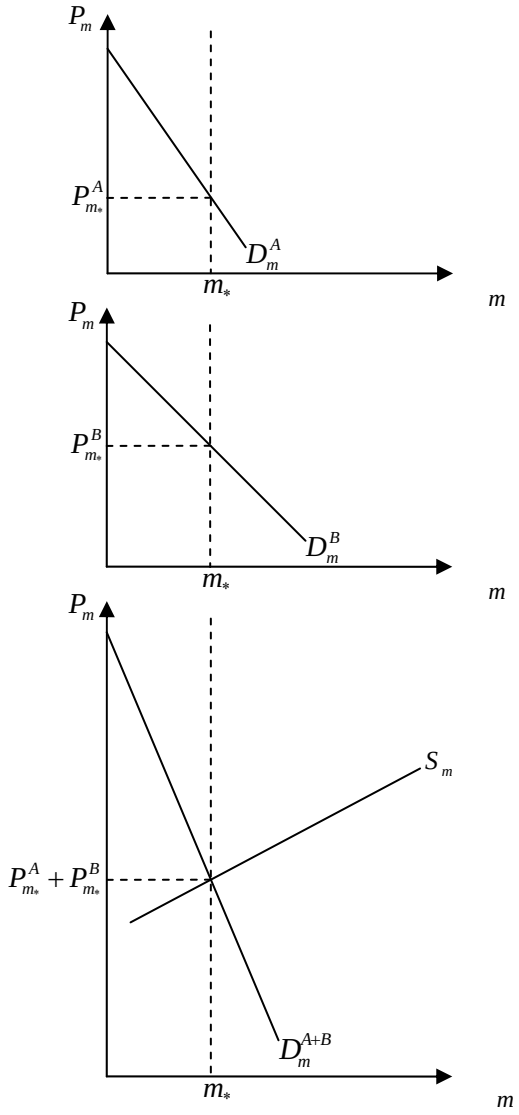
[107] *See infra* notes 112-115 and accompanying text.

The process of deriving the market demand curve for pure public goods differs starkly from the process for deriving the market demand curve for private goods. Figure 3 depicts both Adam's and Beth's demand curves for a particular movie, represented by D_m^A and D_m^B respectively, with the horizontal axis depicting the size of the movie produced (as determined by the number of resources used to produce it) (m)[108] and the vertical axis (P_m) depicting the price of the movie. When everyone consumes exactly the same quantity of goods, deriving the market demand curve requires adding together the prices that each consumer would be willing to pay for a given quantity. For example, Adam is willing to pay P_1^A and Beth is willing to pay P_1^B for a pure public good of a particular size. Their total willingness to pay for that public good is therefore $P_1^A + P_1^B$. Thus, unlike the market demand curve for private goods, which represents the *horizontal* summation of the individual consumers' demand curves, the market demand curve for pure public goods (represented by D_m^{A+E}) is the *vertical* summation of the individual consumers' demand curves.[109]The equilibrium can again be derived by superimposing a supply curve onto the market demand curve, which leads to an equilibrium quantity of m^*. At this quantity, Adam's willingness to pay is $P_{m^*}^A$, while Beth's willingness to pay is $P_{m^*}^B$, creating a total market demand of $P_{m^*}^A + P_{m^*}^B$. Again, there is no reason to assume that Adam and Beth will place the same value on the good. The difference in the intensity of their preferences is reflected by the difference in their reservation prices.

[108] The quantities depicted along the horizontal axis in Figure 3 differ from the quantities depicted on the horizontal axis in Figure 2. In Figure 2, the horizontal axis depicts the allocation of a particular private good. In Figure 3, the horizontal axis depicts the total amount produced of a particular public good. The allocations of the particular quantity of public goods produced are depicted vertically.

[109] *See* Samuelson, *Diagrammatic Exposition*, *supra* note 6, at 353-54 ("[W]e must in the case of public goods add different individuals' curves vertically."). Samuelson acknowledged that this insight was first identified by

Figure 3: Aggregation of Demand for Pure Public Goods (Vertical Summation)



The efficiency of this equilibrium can, again, be analyzed in terms of marginal rates of substitution. Assuming as before that the price of oranges is \$1, P^A_{m*} represents Adam's marginal rate of substitution of movies for oranges (MRS^A_{mo}), and P^B_{m*} represents Beth's marginal rate of substitution of movies for oranges (MRS^B_{mo}). The equilibrium total market price

---

Howard Bowen. Samuelson, *Pure Theory*, *supra* note 6, at 388; *see also* BOWEN, SOCIAL ECONOMY, *supra* note

($P_{m*}^{A} + P_{m*}^{B}$) thus equals $MRS_{mo}^{A} + MRS_{mo}^{B}$. From the standpoint of production, the total market price still represents the marginal rate of transformation of movies for oranges (MRT_{mo}). Instead of having the marginal rate of transformation equal each individual consumer's marginal rate of substitution ($MRS_{ao}^{A} = MRS_{ao}^{B} = MRT_{ao}$), as was the case in the equilibrium for private goods, the equilibrium for pure public goods requires that the marginal rate of transformation equal the *sum* of each individual consumer's marginal rate of substitution ($MRS_{mo}^{A} + MRS_{mo}^{B} = MRT_{mo}$). Stated slightly more generally:

$$\sum_{i} MRS_{mo}^{i} = MRT_{mo} .$$

This has become known as the Samuelson condition and constitutes the key distinction between public and private goods.[110]

The optimal provision of pure public goods represents an interesting inversion of the situation with respect to private goods. For private goods, individual consumers pay the *same price* and signal the intensity of their preferences by consuming *different quantities*. For pure public goods, conversely, individuals consume the *same quantity* and signal the intensity of their preferences by paying *different prices*.

The fact that individual consumers must signal the value that they place on pure public goods through prices rather than quantities has a dramatic effect on the likelihood that markets will produce and allocate pure public goods in an efficient manner. Because optimality for pure public goods requires that all consumers purchase the same quantity and pay their marginal

---

103, at 176-78 & n.5; Bowen, *Voting*, *supra* note 103, at 30-31 & n.3.

[110] *See supra* note 9 and accompanying text.

valuations, individual consumers have strategic incentives to understate the value that they place on the pure public good in the hope that other consumers will bear a larger share of the costs.[111]

Samuelson regarded this inability to induce consumers to reveal truthfully the intensity of their preferences as the true source of the systematic bias toward underproduction of public goods. Later theorists have employed game theory to evaluate the severity of the underproduction. Under these models, each consumer accepts the spillover benefits created by the conjectured level of spending by other consumers on the pure public good and then adds additional funds of her own until the marginal benefits of further increases in expenditure equal the marginal cost. From this process, one can construct each consumer's best response function for any conjectured level of spending by other consumers. These best response functions can be combined to identify the resulting Nash equilibrium. Because each consumer individually equates her own marginal rate of substitution to the marginal rate of transformation (rather than the aggregation of the marginal rates of substitution of all consumers), the total level of spending necessarily falls short of the levels needed to satisfy the Samuelson condition.[112] Experimental evidence has confirmed the existence of this tendency.[113] Interestingly, it is always rational for individual consumers to contribute some amount toward provision of the pure public good regardless of the spending levels that they assume that other consumers will contribute. Because in equilibrium consumers make suboptimal contributions toward production of the public good rather than zero contribution—and to distinguish it from the systematic bias toward

---

[111] Samuelson, *Aspects*, *supra* note 7, at 336; Samuelson, *Diagrammatic Exposition*, *supra* note 6, at 355; Samuelson, *Pure Theory*, *supra* note 6, at 388.

[112] *See, e.g.*, CORNES & SANDLER, *supra* note 1, at 26-30, 153-61; Richard Cornes & Todd Sandler, *Easy Riders, Joint Production, and Public Goods*, 94 ECON. J. 580, 584-91 (1984).

[113] For reviews of the literature, see DOUGLAS D. DAVIS & CHARLES A. HOLT, EXPERIMENTAL ECONOMICS 317-75 (1993); John O. Ledyard, *Public Goods: A Survey of Experimental Research*, *in* HANDBOOK OF EXPERIMENTAL ECONOMICS 111, 122-69 (John H. Kagel & Alvin E. Roth eds., 1995).

underproduction associated with nonexcludability—the literature refers to this effect as "easy riding" rather than "free riding."[114]

The fact that many copyrightable works are not final goods, but instead must be combined with other inputs before they can be sold to consumers does not fundamentally change the analysis. The outcome still must satisfy the Samuelson condition, although the condition is modified so that the marginal rate of transformation must equal the sum of the marginal valuations of the firms that wish to use the good as an input rather than the sum of the marginal rates of substitution of the consumers who wish to consume the good as an end product.[115] In short, the same incentives to misrepresent the intensity of one's preferences remain.

C. Critique of the Conventional Approach

1. A New Perspective on Nonexcludability

Refocusing the analysis around the fundamental principles distinguishing public goods from private goods sheds new light on the role played by nonexcludability. The Samuelson condition underscores the extent to which the easy riding associated with indivisibility represents a problem that is analytically distinct from the free riding associated with nonexcludability. As noted earlier, the Samuelson condition requires that consumers each pay their full marginal valuation of the public good. However, consumers of pure public goods have no incentive to reveal the true intensity of their preferences. This incentive incompatibility remains even if the good is rendered completely excludable.

[114] *See, e.g.*, CORNES & SANDLER, *supra* note 1, at 30; Cornes & Sandler, *supra* note 112, at 580 n.2.

Restated in terms of the Samuelson condition, complete excludability does not alter the fact that indivisibility requires the vertical summation of demand curves, which in turn requires that the marginal rate of transformation equal the sum of every consumer's marginal rate of substitution. This vertical summation gives consumers the incentive to understate the intensity of their preferences even if the good is completely excludable. Certainly, nonexcludability would worsen the problems of underproduction, and the prospect that consumers might enjoy benefits without having to pay for them would dampen investment incentives. But the incentive and opportunity to easy ride would remain even if the free riding from nonexcludability were completely eliminated.

Viewing the problem posed by pure public goods in this manner reveals why Samuelson did not regard rendering a good excludable as sufficient to eliminate the problems associated with market provision of public goods.[116] This posture also explains why many theorists have questioned whether nonexcludability is properly regarded as an essential feature of pure public goods.[117] The fundamental problem that lies at the heart of public good economics will thus remain no matter how much technological development and innovation in institutional forms increase the excludability of copyrightable works.

2. A New Perspective on Nonrivalry as Zero Marginal Cost

Returning to the fundamentals of public good economics also helps illuminate the analytical deficiencies associated with modeling nonrivalry as zero marginal cost. As noted

---

[115] Keimei Kaizuka, *Public Goods and Decentralization of Production*, 47 REV. ECON. & STAT. 118, 118 (1965); Oakland, *supra* note 1, at 493-94; Agnar Sandmo, *Optimality Rules for the Provision of Collective Factors of Production*, 1 J. PUB. ECON. 149, 153 (1972).

[116] *See supra* note 6 and accompanying text.

[117] *See supra* note 94 and accompanying text.

earlier, the feature generally recognized as distinguishing public goods from private goods is the fact that efficient provision of public goods must satisfy the Samuelson condition: $\sum MRS = MRT$. The left-hand side of the equation, which requires the summation of the marginal rates of substitution of all consumers, is the source of the incentive incompatibility that causes markets for public goods to fail. The market failure inherent in the left-hand side of the Samuelson condition exists regardless of whether marginal cost is zero. Indeed, one could easily incorporate a nonzero marginal cost function into the right-hand side of the Samuelson condition, either by building it in to the marginal rate of transformation or by adding a term to reflect increasing marginal cost. Doing so would not alleviate the difficulties in inducing consumers to reveal their preferences inherent in the left-hand side of the equation. In short, the incentive for consumers to misrepresent the intensity of their preferences will persist regardless of the exact nature of the production function and is thus independent of the problems that arise when marginal cost is zero.

For this reason, Samuelson emphasized that marginal cost pricing is only one of several necessary conditions for the efficient provision of public goods. Even if the shortfall in production associated with the zero marginal cost problem were eliminated (such as by directly subsidizing production), the Samuelson condition would still require the ability to discern each customer's marginal valuation of further increases in the public good.[118]

3. The Proper Scope of Price Discrimination

Reconceiving public good economics in terms of the Samuelson condition provides new insights into the role of price discrimination. First, the quantity range over which price

discrimination is relevant differs depending on whether nonrivalry is modeled as zero marginal cost or as indivisibility. As noted earlier and as depicted in Figure 1, the efficiency loss when nonrivalry is modeled as zero marginal costs results from the exclusion of consumers represented by the difference between Q_{mon} or Q_{sus} and Q_{eff}. Price discrimination can solve this problem by permitting authors to attract those customers by offering them lower prices without also having to offer those lower prices to existing customers. Thus, when nonrivalry is modeled as zero marginal cost, economic efficiency only requires that price discrimination be effective over the range running from Q_{mon} or Q_{sus} to Q_{eff}.[119] It does not matter whether producers are able to exercise perfect price discrimination over inframarginal consumers (those consumers represented by the range of output running from the origin to Q_{mon} or Q_{sus}), since only the behavior of the inefficiently excluded consumers is critical for economic efficiency.

This conclusion contrasts sharply with the policy implications that arise when nonrivalry is viewed through the lens of the Samuelson condition. The Samuelson condition requires that the marginal rate of transformation equal the sum of the marginal valuations of all consumers and not just the marginal consumer. Thus, if a profit-maximizing producer of a pure public good is expected to produce efficient levels of the good, she must be able to charge every consumer a price precisely calibrated to the particular consumer's incremental valuation. In other words, in order to satisfy the Samuelson condition, the producer must be able to price discriminate over the entire range of output and not just with respect to those consumers who would be inefficiently excluded by an author's decision to charge a price that exceeds marginal cost.

---

[118] Samuelson, *Aspects*, *supra* note 7, at 336.

[119] Indeed, this suggests that price discrimination need not be perfect throughout the entire range between Q_{sus} and Q_{eff}. Price discrimination can be imperfect so long as it still permits the consumer with the lowest valuation to purchase the good.

This represents a fairly dramatic expansion of the range over which price discrimination is important. This expansion in turn places greater importance on facilitating price discrimination with respect to all customers. It also contradicts suggestions that alternative institutional arrangements that facilitate low-value users' ability to access copyrighted works can serve as equally effective substitutes for price discrimination.[120]

The second insight is that, contrary to the claims of the conventional approach, price discrimination need not be perfect in order to maximize welfare. The Samuelson condition implies that optimality does not require that producers capture all of the consumer surplus. It is sufficient if they are able to appropriate the marginal rate at which each consumer would substitute further expansion of the public good for other goods. Indeed, this suggests that permitting competitive producers to engage in perfect price discrimination would lead to overproduction of the public good.[121] Interestingly, the tendency toward overproduction that exists under perfect price discrimination by competitive producers of a pure public good disappears when the producer engaging in perfect price discrimination is a monopolist. Because the marginal revenue curve implicit in the industry demand curve is $\sum MRS$, a profit-maximizing monopolist would produce where the marginal revenue equals the marginal rate of transformation, which would, of course, satisfy the Samuelson condition.[122]

That said, both perfect price discrimination and the appropriation of the aggregate marginal valuations of all consumers implicit in the Samuelson condition require complete

---

[120] *See supra* note 37 and accompanying text.

[121] *See* Thompson, *supra* note 18, at 6. For surveys of this literature, see CORNES & SANDLER, *supra* note 1, at 243-55; Oakland, *supra* note 1, at 515-17, 520-22.

[122] Thompson, *supra* note 18, at 7. For an earlier, less technical discussion that makes a similar point, see James M. Buchanan, *Public Goods in Theory and Practice: A Note on the Minasian-Samuelson Debate*, 10 J.L. & ECON. 193, 195 (1967).

information about every consumer's reservation price. They also require a pricing mechanism that is capable of extracting the entirety of that reservation price from each consumer and preventing consumers from using arbitrage to defeat that pricing regime. Under the more conventional assumptions that information and pricing mechanisms are imperfect, the bias toward underproduction reemerges[123] and is more severe under monopoly provision than under competitive provision.[124]

III. THE THEORY OF IMPURE PUBLIC GOODS

One of the first criticisms leveled at Samuelson's work was that public and private goods represent idealized polar cases and that most real-world cases lay somewhere in between.[125] The literature on "impure public goods" emerged from these criticisms, as scholars began to explore intermediate cases between private goods and pure public goods. The predominance of Musgrave's two-part definition of pure public goods has naturally led commentators to identify and categorize these intermediate cases by relaxing the elements of that definition.[126] As a result, many theorists segregate impure public goods into two categories: those that remain nonrival but are excludable and those that remain nonexcludable but are rival.[127] This approach

[123] *See* William H. Oakland, *Public Goods, Perfect Competition, and Underproduction*, 82 J. POL. ECON. 927, 937-38 (1974) (concluding that competitive production of public goods leads to underproduction).

[124] For comparisons of competitive and monopolistic solutions, see Geoffrey Brennan & Cliff Walsh, *A Monopoly Model of Public Goods Provision: The Uniform Pricing Case*, 71 AM. ECON. REV. 196, 201-02 (1981); Dagobert L. Brito & William H. Oakland, *On the Monopolistic Provision of Excludable Public Goods*, 70 AM. ECON. REV. 691, 701-02 (1980).

[125] *See supra* notes 14-17 and accompanying text.

[126] *See supra* note 2 and accompanying text.

[127] For examples of this approach, see MUSGRAVE & MUSGRAVE, *supra* note 30, at 50-51; Eyal Benvenisti, *Collective Action in the Utilization of Shared Freshwater: The Challenges of International Water Resources Law*, 90 AM. J. INT'L L. 384, 388 (1996); Peter Eckersley, *Virtual Markets for Virtual Goods: The Mirror Image of Digital Copyright?*, 18 HARV. J.L. & TECH. 85, 117 (2004); Brett M. Frischmann, *An Economic Theory of Infrastructure and Commons Management*, 89 MINN. L. REV. 917, 942-43 (2005); Inge Kaul et al., *Defining Global*

makes the manner in which nonrivalry is characterized of critical importance because this characterization necessarily limits the ways in which we can regard goods as impure public goods.

Refocusing the analysis on the Samuelson condition fundamentally reframes the way we think about impure public goods. For example, the most highly developed literature on impure public goods is the study of "club goods" and "local public goods." These goods differ from pure public goods in two important ways. First, the shared good is assumed to be excludable. Second, the shared good is subject to congestion, in that the utility enjoyed by each consumer decreases as the total number of people consuming the good rises.[128]

As we shall see, introducing congestion costs does not prevent the same quantity from appearing as an argument in multiple purchasers' consumption functions or allow purchasers to consume different quantities. As a result, impure public goods must still satisfy a form of the Samuelson condition. Thus, problems of incentive compatibility remain, notwithstanding the introduction of congestion costs. It is for this reason that the literature concludes that the introduction of congestion costs does not eliminate the essential problem posed by public good economics.[129]

Strictly speaking, then, congestion is not a relaxation of the assumption that goods are nonrival. Instead, congestion is more properly regarded as a new dimension along which utility can vary that is distinct from both price and quantity. Thus, rather than being regarded as a

---

Public Goods, *in* GLOBAL PUBLIC GOODS: INTERNATIONAL COOPERATION IN THE 21ST CENTURY 2, 5 (Inge Kaul et al. eds., 1999).

[128] For surveys of the literature on the economics of congestion, see CORNES & SANDLER, *supra* note 1, at 272-77, 347-479; Oakland, *supra* note 1, at 499-509; Yoo, *supra* note 17, at 1863-74.

[129] *See* Oakland, *supra* note 1, at 499.

factor that causes variation in the *quantity* of a public good, congestion is more accurately regarded as a factor that causes variation in the *quality* of a public good.[130]

Reconceptualizing the study of impure public goods in this manner yields two distinct insights. First, it expands the range of goods that can properly be regarded as impure public goods to include any that vary in quality. For this reason, the leading overview of public good economics describes impure public goods not as a spectrum along which individual assumptions are relaxed, but rather as a "catch all term for any goods not purely public or private."[131] In particular, recharacterizing impure public goods in this manner illuminates the connection between public goods theory and the literature on product differentiation, which models variations in quality explicitly.

Second, and even more importantly, introducing a new dimension distinct from price and quantity provides a new means through which consumers can signal the intensity of their preferences. Indeed, as Tiebout noted, variations in quality create the possibility of de facto markets in which consumers reveal the intensity of their preferences by reallocating their purchases from one provider to another even when they are unable to reveal their preferences through quantity and are unwilling to reveal their preferences through price.[132] In the process, the introduction of another consideration that is incentive compatible raises the possibility of equilibria in which the Samuelson condition may be satisfied notwithstanding the fact that goods remain indivisible. As a result, efficient market provision of impure public goods becomes quite

[130] CORNES & SANDLER, *supra* note 1, at 348; Eitan Berglas & David Pines, *Clubs, Local Public Goods and Transportation Models: A Synthesis*, 15 J. PUB. ECON. 141, 148 (1981); Oakland, *supra* note 1, at 499; Suzanne Scotchmer, *Two-Tier Pricing of Shared Facilities in a Free-Entry Equilibrium*, 16 RAND J. ECON. 456, 467 (1985).
[131] CORNES & SANDLER, *supra* note 1, at 4-5.
[132] Tiebout, *supra* note 17, at 419-20, 424.

feasible,[133] a conclusion that contrasts sharply with the policy implications of the theory of pure public goods.

To say that efficient equilibria for impure public goods are feasible is not to say that they are inevitable. The efficiency of the resulting equilibria ultimately turns on the shape of the relevant quality function. In fact, markets for impure public goods can provide incentives for entry that are either too weak or too strong.[134] The maximization of economic welfare would thus require varying the strength of copyright protection on a case-by-case basis in a way that reflects the precise quality function associated with each good. The question then becomes whether conducting such fact-specific inquiries is advisable.

The balance of this Part is organized as follows. Section A discusses the literature on impure public goods that draws on the economics of congestion. After laying out the basic insights of the literature, it considers and ultimately rejects the possibility that the economics of congestion might serve as the basis for modeling copyright. Section B analyzes the literature on impure public goods that draws on the economics of product differentiation, showing how spatial competition can provide new insights into the basic policy issues confronting copyright.

A. The Economics of Congestion

1. A Basic Description of the Economics of Congestion

The first self-conscious attempt to explore when markets might efficiently provide public goods was offered by Charles Tiebout, who proposed a model of local public goods in which

[133] *See supra* note 21 and accompanying text.

[134] For arguments that there is no "invisible hand" guiding markets for impure public goods, see *supra* note 22 and accompanying text.

residents shared a resource that was in fixed supply and for which there was an optimal level of use.[135] Shortly thereafter, James Buchanan offered his theory of club goods as a way to explore the intermediate cases that exist between the polar extremes of pure public goods and private goods.[136] Samuelson acknowledged the connection between these theories and his theory of pure public goods, but questioned whether these alternative approaches could ever be rendered feasible.[137] For purposes relevant to this Article, the analysis of club goods is indistinguishable from the analysis of local public goods.[138]

Club goods differ from the classic definition of pure public goods in two ways. First, though exclusion may be costly, club goods are fully excludable. Second, they are subject to congestion. In other words, while the jointness of supply always permits an additional consumer to enjoy the shared facility, increasing the number of consumers imposes congestion costs on existing users.[139] Although it is sometimes described as rendering a good partially nonrival,[140] congestion does not prevent the same quantity from appearing as an argument in more than one consumer's consumption function. Instead, congestion introduces a new, quality-oriented dimension—distinct from price and quantity—that can serve as a source of variation in the utility derived by individual consumers.

---

[135] Tiebout, *supra* note 17, at 419.

[136] Buchanan, *supra* note 17, at 1-2. Mancur Olson also expounded a theory of club goods at roughly the same time as Buchanan. MANCUR OLSON, THE LOGIC OF COLLECTIVE ACTION 22-43 (1965). For a review of the early history of club goods theory, see CORNES & SANDLER, *supra* note 1, at 351-54.

[137] *See supra* note 15 and accompanying text.

[138] Oakland, *supra* note 1, at 502-03; *see also* CORNES & SADLER, *supra* note 1, at 366-68 (finding club good theory analogous to local public good theory and identifying key differences). In the discussion that follows, references to "club goods" are intended to encompass both theories.

[139] Buchanan, *supra* note 17, at 3-5. Although Tiebout does not use the term "congestion," his assumption of the existence of a resource with a "U"-shaped average cost curve is consistent with the concept. Tiebout, *supra* note 17, at 419.

[140] CORNES & SANDLER, *supra* note 1, at 9; Frischmann, *supra* note 127, at 952-53.

This suggests that providing the efficient number of club goods still requires satisfying a form of the Samuelson condition.[141] In this case, the Samuelson condition requires that

$$\sum_{i=1}^{s} MRS_{mo}^{i} = MRT_{mo} - MC_{congestion}$$

where MRS_{mo}^{i} represents each individual's marginal rate of substitution between the club good, *m*, and an outside private good, *o*, and where MRT_{mo} represents the marginal rate of transformation of the club good and the outside private good. The left-hand side of this equation can be interpreted as the incremental consumer benefits from further increases in the size of the club good. The right-hand side of the equation represents the marginal cost of such increases, which is the incremental cost of expanding the club good less the aggregate decongestion benefits of expanding the capacity. This condition requires that production of club goods be expanded until the aggregate marginal benefits from further increases in size no longer exceed the marginal cost.

Properly speaking, then, congestion does not introduce a degree of rivalry. The same quantity can still appear as an argument in more than one customer's consumption function, and optimality still requires satisfying the Samuelson condition. The addition of the congestion term also does nothing to eliminate the vertical summation of the marginal rates of substitution, which is the source of the incentive incompatibility that represents the core problem associated with public goods. Instead, congestion is more properly regarded as a consideration that is analytically distinct from rivalry.[142]

---

[141] CORNES & SANDLER, *supra* note 1, at 357-58; Eitan Berglas, *On the Theory of Clubs*, 66 AM. ECON. REV. 116, 117, 119 (1976); Robin Boadway, *A Note on the Market Provision of Club Goods*, 13 J. PUB. ECON. 131, 133 (1980); Oakland, *supra* note 1, at 500, 503.

[142] *See supra* note 130 and accompanying text.

At the same time, the presence of congestion costs introduces a second optimality condition. Unlike pure public goods, for which further expansion of the customer base always reduces the costs borne by individual consumers,[143] expansion of the customer base for club goods gives rise to a tradeoff. As with pure public goods, the addition of new members reduces average cost by spreading the fixed costs of creating the shared resource over a larger membership.[144] When a public good is subject to congestion, however, increasing club size also increases the congestion costs borne by each member. Congestion thus creates a diseconomy of scale, which in turn implies the existence of an optimal club size beyond which any further benefits from spreading fixed costs over a larger number of members would be offset by the costs imposed by increases in congestion.[145] It is the presence of such a diseconomy of scale that prevents markets for public goods from collapsing into natural monopolies.[146]

For any representative club, membership should be increased so long as the marginal benefits that the additional club member would derive from joining the club exceed the increase in congestion costs that the additional member would impose on current members.[147] Stated more formally, this requires that for every club member, i,

$$MRS_{sa}^{i} = MRT_{sa}^{i}$$

where s represents club size, MRS_{sa} represents the marginal rate of substitution between increasing club size and additional consumption of an outside private good, and MRT_{sa} equals the

[143] *See* CORNES & SANDLER, *supra* note 1, at 349 ("The entire population is in a single provision association for pure public goods."); Buchanan, *supra* note 17, at 1-2 (noting that for pure public goods, "the optimal sharing group . . . includes an infinitely large number of members").

[144] Buchanan, *supra* note 17, at 8.

[145] *Id.* at 7-8.

[146] Yoo, *supra* note 5, at 232-33, 248-49.

[147] CORNES & SANDLER, *supra* note 1, at 357-58; Berglas, *supra* note 141, at 117; Buchanan, *supra* note 17, at 4-5.

marginal rate of transformation of increasing club size versus producing another unit of the outside good. The left-hand side of the equation can be interpreted as the marginal benefits that the club members would enjoy from admitting an additional member. The right-hand side of the equation can be interpreted as the marginal (congestion) costs of admitting an additional member.

Unlike the Samuelson condition, this second condition is potentially incentive compatible. A club can induce its members to act efficiently simply by charging them a membership fee that equals their marginal contribution to congestion. If customers cannot vary the intensity of their use of club facilities, this can be accomplished simply by charging a lump-sum membership fee calibrated to the average member's contribution to congestion.[148] If the customers can vary the intensity of their use of club facilities, the classic solution is to impose a two-part tariff: one part consisting of a lump-sum membership fee designed to extract consumer surplus, and the other part consisting of a variable fee calibrated to match the congestion costs imposed by the last unit consumed.[149]

The ability of and incentive for consumers to reallocate their purchases from club to club in response to congestion costs provides a basis for revealing consumers' preferences that is distinct from both price and quantity. Clubs that fall below optimal size have the incentive to attract new members, since the benefits of spreading the fixed costs needed to maintain the

[148] *See* Berglas, *supra* note 141, at 117 (deriving the optimal solution, assuming nonvariable intensity, for a swimming pool).

[149] For the classic analysis of two-part pricing, see Walter Y. Oi, *A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly*, 85 Q.J. ECON. 77, 80-81 (1971). For an early application of two-part pricing to club goods, see Berglas, *supra* note 141, at 119 (noting this solution, but concluding that it is inefficient). A more sophisticated model by Suzanne Scotchmer shows that two-part pricing can induce efficient consumption of club resources. Scotchmer demonstrates that the number of clubs is inefficient for finite economies, but converges to the efficient result as the economy becomes increasingly large. Scotchmer, *supra* note 130, at 462-65. This solution

shared resource over a larger number of members would offset the increase in congestion costs associated with adding new members. In addition, to the extent that club size exceeds optimal levels, one would expect members in overly congested clubs to exit and start new clubs.[150] Eventually, club members should redistribute their purchases until the economy is divided into clubs of optimal size and the level of congestion is spread equally across all clubs.[151]

Because both club good conditions must be solved simultaneously, the addition of the second condition can give rise to equilibria with welfare characteristics that are strikingly different from those that prevail under a pure public goods analysis. Club goods do not exhibit the systematic bias toward underproduction associated with pure public goods. Indeed, it is quite feasible that markets will produce the socially optimal number of club goods in equilibrium. In addition, so long as the economy is sufficiently large, equilibrium prices should closely approximate efficient levels and should converge to marginal cost as the economy grows larger.[152] Finally, the possibility of entry by new clubs effectively guarantees that no club earns supracompetitive returns.[153] The efficiency of the club goods equilibrium is subject to a number of technical caveats,[154] but none are central to my argument.

---

assumes the absence of transaction costs. *But see* Yoo, *supra* note 17, at 1865-66 (reviewing the literature relaxing this assumption).

[150] The possibility of entry by new clubs represents one of the characteristics that distinguishes club goods from local public goods. *See* Scotchmer, *supra* note 21, at 94 (briefly discussing costs of entry with respect to clubs as similar to costs of entry with respect to producers of private goods).

[151] Berglas, *supra* note 141, at 117-18; Berglas & Pines, *supra* note 130, at 154; *cf.* Tiebout, *supra* note 17, at 418-20, 424 (drawing a similar conclusion for local public goods).

[152] Berglas & Pines, *supra* note 130, at 146, 156; Scotchmer, *supra* note 130, at 458, 463-65, 468.

[153] The indivisibility of the fixed costs of entry may give rise to an "integer problem," in which n clubs earn supracompetitive returns, while $n + 1$ clubs would not. That said, the number of clubs in equilibrium will fall short of the optimum by no more than one. So long as the economy is sufficiently large, any supracompetitive returns should also be relatively small. Berglas, *supra* note 141, at 118; Scotchmer, *supra* note 130, at 464.

[154] For example, a club good equilibrium will prove stable only if dividing the overall population of club members by the optimal club size results in an integer. When that occurs, every club member lacks the incentive to switch clubs. The resulting equilibrium is said to be in the *core*, which implies that the equilibrium is Pareto optimal. A noninteger result destabilizes the equilibrium, since those excluded from club membership possess the

This is not to say that congestion costs represent an economic panacea. Because the components of the second efficiency condition are not perfectly correlated with the components of the Samuelson condition, markets can reach equilibrium with insufficient, excess, or optimal entry, depending on the shape of the relevant congestion function. Nonetheless, because the club good equilibrium is no longer bounded away from achieving either efficient levels of access to

---

incentive to bid their way into a club by offering to accept a lower payoff than a current club member. The result is a constant shuffling of club composition. Berglas & Pines, *supra* note 130, at 157; Mark V. Pauly, *Clubs, Commonality, and the Core: An Integration of Game Theory and the Theory of Public Goods*, 34 ECONOMICA (n.s.) 314, 323-24 (1967). Fortunately, introduction of a concept known as the *approximate core* renders the nonexistence of a stable equilibrium less problematic than it may seem at first glance. If the number of club members is large relative to the number of nonmembers, club members can make side payments to nonmembers in order to induce them not to destabilize the existing coalitions. So long as the economy is sufficiently large, the resulting utilities should lie very close to core utilities. Scotchmer, *supra* note 21, at 104-05; Myrna Wooders, *The Tiebout Hypothesis: Near Optimality in Local Public Good Economies*, 48 ECONOMETRICA 1467, 1474, 1479-82, 1484 (1980).

Another critical assumption is that consumer preferences within each club are homogeneous. *See* CORNES & SANDLER, *supra* note 1, at 351 (noting that the bulk of the literature on club goods has focused on homogeneous clubs). It has been recognized since Walter Oi's seminal discussion of two-part pricing that consumer heterogeneity can cause two-part prices to become inefficient, since no single, lump-sum fee will be sufficient to allow the club to extract all of the available surplus. Oi, *supra* note 149, at 81-88. To the extent that consumer preferences are heterogeneous, one would expect consumers to partition themselves into different clubs consisting of members with the same preferences. CORNES & SANDLER, *supra* note 1, at 367; Berglas, *supra* note 141, at 116, 120; Berglas & Pines, *supra* note 130, at 152-53; Martin McGuire, *Group Segregation and Optimal Jurisdictions*, 82 J. POL. ECON. 112, 131 (1974); Oakland, *supra* note 1, at 504; Mark V. Pauly, *Cores and Clubs*, 9 PUB. CHOICE 53, 60-64 (1970). If integer problems prevent the total population from segregating itself into homogeneous clubs, individuals with different preferences may have to join together to form a *mixed club*. Early analyses disputed the optimality of mixed clubs equilibria. *Compare* Berglas & Pines, *supra* note 130, at 150-51 (concluding that replicating a mixed club is nonoptimal), *with* Todd Sandler & John T. Tschirhart, *Mixed Clubs: Further Observations*, 23 J. PUB. ECON. 381, 388-89 (1984) (arguing that replicating a mixed club is optimal under certain circumstances). Later work has shown that mixed clubs may be optimal under certain conditions. *See* Suzanne Scotchmer & Myrna Holtz Wooders, *Competitive Equilibrium and the Core in Club Economies with Anonymous Crowding*, 34 J. PUB. ECON. 159, 171-72 (1987) (finding that the mixed club equilibrium approaches efficiency so long as crowding is anonymous and the economy is large).

The efficiency of two-part pricing also depends on the assumption that excluding nonmembers and metering club usage are costless. When transaction costs are taken into account, clubs may find it more economical to charge a flat-rate price based on the contribution to congestion by the average club member. Robert J. Barro & Paul M. Romer, *Ski-Lift Pricing, with Applications to Labor and Other Markets*, 77 AM. ECON. REV. 875, 876-79 (1987); Robert W. Helsley & William C. Strange, *Exclusion and the Theory of Clubs*, 24 CAN. J. ECON. 888, 895-96 (1991). *See generally* CORNES & SANDLER, *supra* note 1, at 387-90 (surveying the literature on the impact of exclusion costs on club goods). Lack of information about the intensity of individual club members' demand for usage of club facilities can give rise to a moral hazard problem, in which members with high demand are able to enjoy benefits that exceed what they pay and are able to impose costs on club members with relatively low demand. Kangoh Lee, *Transaction Costs and Equilibrium Pricing of Congested Public Goods with Imperfect Information*, 45 J. PUB. ECON. 337, 359 (1991). As noted earlier, clubs may avoid transaction costs through a variety of alternative institutional forms. *See supra* notes 89-92 and accompanying text.

the public good or efficient incentives for the creation of the public good, the welfare properties of the result are considerably more attractive and open up the policy space in important ways.

2. The Applicability to Copyright

One can conceive of copyrightable works as club goods, where potential purchasers segregate themselves into groups of consumers with similar preferences in an effort to economize on fixed costs.[155] William Landes and Richard Posner have recently suggested that copyrighted works may be subject to *congestion externalities* that cause the utility derived from consuming a work to decrease as the total number of people consuming the work increases. Analogizing to how overexposure can prematurely exhaust the commercial value of a celebrity's likeness, they argue that additional consumption can depress demand for a copyrighted work.[156] Although Landes and Posner discuss congestion externalities in the context of copyright duration and the dangers surrounding the overuse of resources that are unowned, their analysis has potential implications for a broader range of copyright-related issues.

The ultimate relevance of the congestion externalities for copyright is not completely clear. As an initial matter, there is reason to question whether increases in consumption will degrade the quality of copyrightable works for those who have already purchased them. Concerns about overexposure seem limited to the use of copyrighted characters in commercial advertising.[157] Indeed, it is quite possible that additional exposure would increase the value of

---

[155] *See* Besen & Kirby, *supra* note 23, at 264-70, 280 (discussing the advantages of creating purchasing groups to economize on fixed costs).

[156] *See* Landes & Posner, *supra* note 52, at 484-88.

[157] Michael Abramowicz, *An Industrial Organization Approach to Copyright Law*, 46 WM. & MARY L. REV. 33, 85 (2004).

the work, either by serving as de facto advertising[158] or by tapping into solidarity or associative characteristics.[159] The ambiguousness of the impact of these external effects has led courts and commentators to reject congestion externalities as a source of market failure in the related context of the right of publicity.[160]

Furthermore, the fact that purchasing decisions by one individual may have external effects on other purchasers is not always economically problematic. Consider the classic case in which marginal cost is rising. Since manufacturers in perfect competition set prices along their respective marginal cost curves, any increase in the quantity produced by a given manufacturer to meet additional demand causes the price paid by inframarginal buyers to rise. It is for this reason that Pigou made his famous error in arguing that all markets that did not face constant marginal cost needed to be corrected either through taxes or subsidies.[161]

If Pigou's reasoning were correct, the near universality of nonconstant marginal cost would make government intervention in the economy endemic. The problem is that the external effects that Pigou identified are what we now understand to be pecuniary externalities—external effects that are fully integrated into market mechanisms.[162] Returning to the example of rising

---

[158] *Id.* at 84.

[159] Netanel, *supra* note 35, at 1907-09. For more general analyses of solidarity goods, see H. Leibenstein, *Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand*, 64 Q.J. ECON. 183, 190-99 (1950); Cass R. Sunstein & Edna Ullmann-Margalit, *Solidarity Goods*, 9 J. POL. PHIL. 129 (2001).

[160] Cardtoons, L.C. v. Major League Baseball Players Ass'n, 95 F.3d 959, 975 (10th Cir. 1996); Mathews v. Wozencraft, 15 F.3d 432, 439 (5th Cir. 1994); Michael Madow, *Private Ownership of Public Image: Popular Culture and Publicity Rights*, 81 CAL. L. REV. 125, 222-23 n.445 (1993); Eugene Volokh, *Freedom of Speech and the Right of Publicity*, 40 HOUS. L. REV. 903, 911 n.32 (2003).

[161] Pigou first advanced this argument in A.C. PIGOU, WEALTH AND WELFARE 177-78 (1912).

[162] *See* Jacob Viner, *Cost Curves and Supply Curves*, 3 ZEITSCHRIFT FÜR NATIONALÖKONOMIE 23 (1931), *reprinted in* READINGS IN PRICE THEORY 198 (George J. Stigler & Kenneth E. Boulding eds., 1952) (establishing the distinction between pecuniary and technological externalities). For critiques in the same vein, see, for example, Howard S. Ellis & William Fellner, *External Economies and Diseconomies*, 33 AM. ECON. REV. 493, 494-503 (1943); F.H. Knight, *Some Fallacies in the Interpretation of Social Cost*, 38 Q.J. ECON. 582, 584-92 (1924); Tibor Scitovsky, *Two Concepts of External Economies*, 62 J. POL. ECON. 143, 146 (1954). For recent discussions, see John F. Duffy, *Intellectual Property Isolationism and the Average Cost Thesis*, 83 TEX. L. REV. 1077, 1081-83

marginal cost introduced above, the increase in price resulting from an increase in the quantity produced simply transfers surplus from one market actor to another. Thus, when markets are functioning well, pecuniary externalities are fully internalized by market transactions. Such externalities may transfer wealth, but they have no impact on efficiency.[163] As a result, they are not properly regarded as the type of externality that leads to market failure. On the contrary, this type of market-mediated external effect is a necessary feature of a properly functioning market.[164] Technological externalities, in contrast, are effects that are external to the parties to the transactions and thus are not mediated through a price mechanism, such as occurs when pollution imposes costs on neighbors that are not reflected in the polluter's costs or revenue. As such, technological externalities can lead to market failure, although, as Coase pointed out, the parties can internalize any externality so long as a market exists in which the parties can bargain

---

(2005); S.J. Liebowitz & Stephen E. Margolis, *Are Network Externalities a New Source of Market Failure?*, 17 RES. LAW & ECON. 1, 4-10 (1995).

[163] *See, e.g.*, William J. Baumol, *On Taxation and the Control of Externalities*, 62 AM. ECON. REV. 307, 312 n.8 (1972) (noting the "well known" proposition that "pecuniary externalities do *not* lead to resource misallocation"); Bruce C. Greenwald & Joseph E. Stiglitz, *Externalities in Economies with Imperfect Information and Incomplete Markets*, 101 Q.J. ECON. 229, 229 (1986) (noting that "pecuniary externalities by themselves are not a source of inefficiency"); Louis Makowski & Joseph M. Ostroy, *Appropriation and Efficiency: A Revision of the First Theorem of Welfare Economics*, 85 AM. ECON. REV. 808, 824 (1995) ("The moral drawn from Pigou's error was that pecuniary externalities should be distinguished from welfare-relevant ownership externalities."). When markets are not functioning well, pecuniary externalities do not necessarily lead to efficiency. Even though market transactions tend to internalize pecuniary externalities, the existence of other market imperfections may cause the resulting equilibrium to be inefficient. This theoretical result is not thought to provide any general policy implications. The magnitude and direction of the effect of both the market imperfections and the pecuniary externalities are ambiguous. Absent some reason to think that either would bias the market in a particular direction, there is no reason to believe that internalizing pecuniary externalities would yield systematic benefits. *See* Lee Hsien Loong & Richard Zeckhauser, *Pecuniary Externalities Do Matter When Contingent Claims Markets Are Incomplete*, 97 Q.J. ECON. 171, 171-79 (1982) (arguing that while pecuniary externalities lead to inefficiencies where markets are incomplete, the direction of that inefficiency cannot be predicted).

[164] Therefore, the parallels that Besen and Kirby drew between their model and a club goods model are not completely apt. The diseconomies of scale in the Besen and Kirby model arose from the assumption that marginal cost was increasing. Besen & Kirby, *supra* note 23, at 257. Since increasing marginal cost is fully internalized in the price mechanism, the price increase for other consumers associated with movement along the marginal cost curve is more properly considered a pecuniary externality fully internalized by the market, rather than a technological externality that can create market failure.

around the problem and the transaction costs are not so high as to prevent the parties from reaching agreement.[165]

As Landes and Posner note, determining whether a change in value associated with congestion externalities is the result of technological or pecuniary externalities can be difficult, if not impossible.[166] On the one hand, a drop in value associated with an increase in consumption might be the result of a technological externality that operates outside of the market. The correction for this problem would be straightforward: create a property right to cover all uses in order to facilitate the creation of the missing market needed to internalize the technological externality. On the other hand, a drop in value associated with an increase in consumption might instead be the result of a pecuniary externality. For example, an author whose copyright gave it a true monopoly would set price and output to maximize its profits. By necessary implication, any deviation in output would cause a reduction in value. In that case, the reduction in value from further increases in output would not be the result of a technological externality in need of potential redress, but rather would be an inframarginal effect completely mediated by the monopolist's ability to set price so as to maximize its profits.

The difficulty is that decreases in value as consumption increases are consistent with either scenario. Absent additional information, one cannot determine whether any reduction in value stems from a technological externality or a pecuniary externality, such as a deviation from the profit-maximizing quantity or a reduction in price in the face of declining average costs.[167] There is thus reason to doubt whether increases in consumption of copyrighted works will in fact

[165] *See supra* note 88 and accompanying text.
[166] Landes & Posner, *supra* note 52, at 486-88.
[167] *See* Liebowitz & Margolis, *supra* note 162, at 7-9 (advancing a similar argument in the related context of network economic effects).

decrease the quality enjoyed by those who have already purchased such works and whether courts will be able to determine with any certainty whether any such decreases would be the result of technological rather than pecuniary externalities. The absence of any reliable way for making these determinations limits the utility of applying the club goods branch of the theory of impure public goods as a tool for analyzing copyright.

B. Spatial Competition

What is less well recognized is that the economics of impure public goods can also be modeled through an approach pioneered by Hotelling known as spatial competition.[168] Under spatial competition, goods are completely excludable, and producers vie for business not on the basis of price, but rather by choosing a location along a linear geographic space.[169] At the same time, transportation costs cause the utility derived by each customer to vary. If the revenue captured by a producer exceeds its costs, the supracompetitive returns attract entry by new producers. The consumers who are located the closest to the new entrant reallocate their purchases to the new producer in an effort to minimize their transportation costs. Absent sunk costs in location, incumbent producers accommodate the new entrant by shifting their positions until they are spread evenly across the geographic space. This process reaches equilibrium when

[168] Harold Hotelling, *Stability in Competition*, 39 ECON. J. 41 (1929).

[169] The discussion that follows extends my previous analysis of the implications of spatial competition for copyright. Yoo, *supra* note 5, at 241-46, 260-72. For overviews of the literature on spatial competition, see JOHN BEATH & YANNIS KATSOULACOS, THE ECONOMIC THEORY OF PRODUCT DIFFERENTIATION 11-34 (1991); Eaton & Lipsey, *supra* note 22, at 734-61; Jean J. Gabszewicz & Jacques-François Thisse, *Location*, *in* 1 HANDBOOK OF GAME THEORY WITH ECONOMIC APPLICATIONS 281 (Robert J. Aumann & Sergiu Hart eds., 1992). For a somewhat less technical survey, see JEFFREY CHURCH & ROGER WARE, INDUSTRIAL ORGANIZATION: A STRATEGIC APPROACH 379-411 (2000). For other analyses drawing the connection between spatial competition and copyright, see Abramowicz, *supra* note 157, at 45-68; Gerald R. Faulhaber, *File Sharing, Copyright, and the Optimal Production of Music*, 13 MICH. TELECOMM. & TECH. L. REV. 77 (2006), *available at* http://www.mttlr.org/volthirteen/faulhaber.pdf.

consumers can no longer reduce their transportation costs by reallocating their purchases to another producer and entry has dissipated all of the available supracompetitive returns.[170]

Hotelling himself recognized that the same framework could be extended to model competition among products distributed along a characteristics space rather than a geographic space. For example, as Hotelling explained, one can envision manufacturers of apple cider as deciding where to produce along a spectrum of product characteristics running from sweet to sour.[171] In a characteristics space, the decline in utility results not from transportation costs, but rather from a particular product's divergence from each consumer's ideal product characteristics.[172]

As with other forms of public good economics, spatial competition permits the same quantity to appear as an argument in multiple customers' consumption functions. In addition, the fact that one customer consumes a particular quantity necessarily means that others must consume the same quantity. It is for this reason that Samuelson saw the connection between public good economics and spatial competition, concluding that they posed nearly identical analytical problems.[173]

---

[170] The localized nature of competition and the indivisibility of benefits do create a limited possibility of sustainable supracompetitive returns. Such supracompetitive returns should be trivially small so long as the economy is sufficiently "large." Yoo, *supra* note 5, at 240, 244 n.102, 250-51, 279.

[171] Hotelling, *supra* note 168, at 53-54. For a listing of other articles drawing the same connection, see Richard Schmalensee, *Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry*, 9 BELL J. ECON. 305, 309 n.7 (1978).

[172] Hotelling, *supra* note 168, at 54.

[173] Samuelson, *Aspects*, *supra* note 7, at 336 (invoking Hotelling's example of spatial competition between sweet and sour cider producers and concluding that such competition "*is analytically almost exactly like my model of public expenditure*"). For other commentators describing spatial competition as a way to model impure public goods, see BUCHANAN, *supra* note 103, at 53-54; Stiglitz, *supra* note 21, at 309-12. Spatial competition is related to another form of competition among differentiated products known as "monopolistic competition." *See also* EDWARD HASTINGS CHAMBERLIN, THE THEORY OF MONOPOLISTIC COMPETITION 196-99, 260-65 (8th ed. 1962) (drawing the connection between monopolistic competition and spatial competition); BEATH & KATSOULACOS, *supra* note 169, at 5-6 (describing spatial competition and monopolistic competition as alternative ways to model product differentiation); Eaton & Lipsey, *supra* note 22, 727-28 (same). The connection between product differentiation and public good economics is further underscored by the fact that economics theorists also regard

1. A Basic Description of Spatial Competition

The connection between spatial competition and the theory of impure public goods is drawn most explicitly in the work of Nobel laureate Joseph Stiglitz. Stiglitz follows Hotelling's approach by assuming that individual consumers are evenly distributed across a linear geographic space.[174] Individuals must incur linear transportation costs to travel to the closest producer.[175] Entry by another producer requires the incurrence of an up-front fixed cost.

It is tempting to think about transportation costs as analogous to congestion costs. On closer inspection, however, such a characterization would be inapt, since in the transportation cost context the addition of another customer to the purchasing group does not degrade the utility derived by existing customers. The framework advanced in this Article suggests that it may be more helpful to think of transportation as introducing an independent source of variation in utility that is distinct from both price and quantity. Indeed, like Hotelling, Stiglitz recognizes that his model can easily be generalized to encompass competition among differentiated products that compete by choosing a location along a spectrum of product characteristics.[176] Spatial competition thus captures the essence of impure public goods theory in that access to particular

---

monopolistic competition as a form of competition among impure public goods. *See* Oakland, *supra* note 1, at 505; P.A. Samuelson, *Pure Theory of Public Expenditure and Taxation*, *in* PUBLIC ECONOMICS, *supra* note 2, at 98, 119.

[174] Stiglitz, *supra* note 21, at 309-10. The model of spatial competition described in this Article is called a *discrete choice model*, in which each consumer buys only one of the available product varieties and purchases only from the closest provider. More general models can accommodate the possibility that consumers will purchase multiple products from multiple providers. For examples of such models, see V. Bhaskar & Ted To, *Is Perfect Price Discrimination Really Efficient? An Analysis of Free Entry*, 35 RAND J. ECON. 762, 767-69 (2004); Eaton & Lipsey, *supra* note 22, at 751-52. For further discussion, see *infra* notes 212, 214 and accompanying text.

[175] Stiglitz assumes that transportation costs are linear. Stiglitz, *supra* note 21, at 310. Later work has shown that they can be nonlinear as well. *See* Bhaskar & To, *supra* note 174, at 764-65 (offering a general model of spatial competition in which transportation costs can take a variety of linear and quadratic shapes); C. d'Aspremont et al., *On Hotelling's "Stability in Competition"*, 47 ECONOMETRICA 1145, 1148-49 (1979) (modeling transportation costs as quadratic); Faulhaber, *supra* note 169, at 94-95 (offering illustrations of differently shaped transportation cost functions in the copyright context).

goods is fully excludable, all individuals consume a good of the same magnitude, and variations in quality provide a dimension aside from price and quantity along which the utility derived by individual consumers can vary.

Beginning with the case in which each producer charges a uniform price,[177] economic efficiency requires the provision of the optimal number of spatially competitive goods. Entry reduces aggregate transportation costs, but requires the incurrence of additional fixed costs. Thus, the optimal number of spatially competitive goods occurs where the aggregate improvement in utility exactly offsets the cost of creating another spatially competitive good. In other words, as Stiglitz points out, the resulting equilibrium must still satisfy the Samuelson condition for determining the optimal number of public goods,

$$\sum MRS^{i}_{mo} = MRT_{mo}$$

where MRS_{mo} is the marginal rate of substitution between the impure public good, m, and an outside private good, o, and MRT_{mo} is the marginal rate of transformation of the impure pubic good and the outside good.[178] The left-hand side represents the aggregate increase in marginal utility caused by the reduction in transportation costs associated with entry by an additional producer. The right-hand side represents the incremental cost of adding an additional producer.[179] As with previous forms of public good competition, the *sum* of every consumer's

---

[176] Stiglitz, *supra* note 21, at 309-10 (recognizing "an obvious slight modification to the analysis in which the differences are not with respect to location but with respect to preferences for different public goods").

[177] Stiglitz actually discusses the uniform pricing scenario second and begins his analysis by focusing on perfectly discriminating producers. Stiglitz, *supra* note 21, at 310-11. For purposes of this Article, it makes more sense to discuss these scenarios in the reverse order.

[178] *See id.* at 311-12. In order to remain consistent with the earlier portions of this Article, I use notation that differs from Stiglitz's. The basic intuitions nonetheless remain the same.

[179] As noted earlier, spatial competition differs from the economics of congestion in that allowing additional consumers to purchase a good does not degrade the utility enjoyed by other consumers of the same good. As a result, the right-hand side of this equation differs from the right-hand side of the equation for club goods in that it does not include a term to represent the degradation in quality resulting from congestion.

marginal rate of substitution of the impure public good must equal the marginal rate of transformation. Thus, the fundamental incentive incompatibility problem inherent in public good economics remains.

Optimality requires not only that the efficient number of goods be produced, but also that those goods be efficiently allocated. Thus, spatial competition must also satisfy a second efficiency condition, which is analogous to the second condition that applied to club goods.[180] This condition requires that each producer of a spatially competitive good serve additional consumers until the marginal utility of providing the good to another person (taking transportation costs into account) equals the marginal cost of doing so.[181] As was the case with club goods and local public goods, decisions about whether to purchase and from whom to purchase can reflect variations in the utility that different customers derive, which in turn depends on the number of producers as well as the shape of the transportation cost function.[182] Furthermore, consumers have no incentive to misrepresent their preferences, either by purchasing when the transportation costs exceed the utility they would derive or by refusing to purchase when their marginal utility exceeds the transportation costs they would have to bear. By giving customers the incentive to reveal their preferences by reallocating their purchases to different producers, spatial competition can thus give rise to de facto markets in much the same manner as club goods and local public goods.

---

[180] *See supra* note 147 and accompanying text.

[181] *See* Stiglitz, *supra* note 21, at 312. Note that Stiglitz's treatment differs from the standard Hotelling set-up in that Stiglitz allows the median purchaser to determine the price and the level of provision. *Id.* at 311.

[182] *Compare* Hotelling, *supra* note 168, at 53-54 (showing how products exhibit minimal differentiation when transportation costs are linear), *with* d'Aspremont et al., *supra* note 175, at 1148-49 (showing how products exhibit maximal differentiation when transportation costs are quadratic). For an analysis of the level of entry under different transportation cost functions, see Bhaskar & To, *supra* note 174, at 764-66.

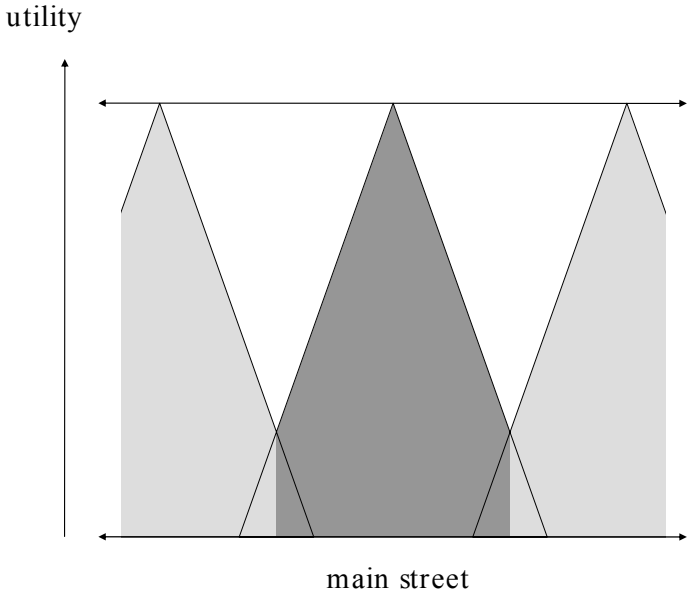
Spatial competition can be depicted graphically, as in Figure 4 below. In this Figure, consumers are distributed uniformly across a geographic space, transportation costs are linear, and each consumer purchases from the closest producer so long as the transportation costs needed to get to the producer's location do not exceed the utility the consumer would derive from her purchase.[183] The dark grey, pentagonal region represents the utility of the consumers served by each producer. Entry continues until the revenue captured by a producer equals the fixed costs of entry, at which point the market reaches equilibrium.

The result is an interesting inversion of Tiebout's model of local public goods. Under Tiebout's approach, the local public good occupies a fixed location, and consumers relocate so as to maximize their utility.[184] Under the spatial competition approach, the potential consumers occupy fixed locations and the providers of the public good adjust their locations in response to the distribution of potential customers and the location of their competitors. The resulting model is also quite similar to the classic conception of club goods. The key difference is that the decrease in utility, which determines purchasing patterns, results from transportation costs rather than from congestion costs.

[183] By depicting the competition as taking place along an infinite linear product space, Figure 4 represents an oversimplification. If spatial competition were to take place along an infinite linear space, no equilibrium would exist since a new entrant would always find it possible to enter to the outside of the existing players. This problem is usually solved by assuming a finite linear product space, by assuming a circular product space, or by assuming sunk costs in location and analyzing the impact of interior entry. Figure 4 should thus be taken as a representation of a portion of a larger model in which equilibria exist. In addition, by depicting that the producer captures all of the available surplus, Figure 4 in effect presumes that the producer is engaging in perfect price discrimination. Spatial competition models can be adjusted fairly easily to take into account the fact that price discrimination is inevitably somewhat imperfect. *See* Yoo, *supra* note 5, at 261-62 & fig.6.

[184] Tiebout, *supra* note 17, at 419.

Figure 4: A Graphical Representation of Spatial Competition in a Geographic Space



Spatial competition that depicts differentiated products vying for consumers by varying their attributes along a spectrum of product characteristics would seem a natural way to model competition among copyrighted works. Spatial competition also captures the effect of entry by imperfect substitutes that characterizes the market for copyrighted works. Furthermore, it allows for equilibria to be determined by variations in product characteristics and by different consumers' taste for those characteristics rather than by price or quantity.

2. The Policy Implications of Spatial Competition for Copyright

a. The Feasibility of Promoting Optimal Access

Most importantly for our purposes, spatial competition offers new solutions to problems that appear to be intractable under the pure public goods theory.[185] Consider first the problem of inefficient access associated with supramarginal cost pricing. As a preliminary matter, spatial competition calls into question whether the spread between price and marginal cost represents an appropriate measure of access efficiency. When products have different characteristics, consumer surplus—the difference between a consumer's reservation price and the actual price paid—is only one source of economic welfare. Consumers can also derive welfare from consuming goods that fit better with their preferences. Thus, the fact that markets for copyrighted works reach equilibrium at a point where price exceeds marginal cost is not necessarily an indication of market failure; rather, it may be nothing more than a side effect of the fact that products are differentiated. Indeed, when both sources of economic welfare are taken into account, an equilibrium in which price exceeds marginal cost may in fact be optimal.[186]

---

[185] The discussion that follows extends my previous analysis of these issues in Yoo, *supra* note 5, at 252-56, 264-76.

[186] For the classic statement, see CHAMBERLIN, *supra* note 173, at 94 (conceding that the equilibrium under monopolistic competition could be regarded as "a sort of ideal"). For more contemporary statements of the same principle, see BEATH & KATSOULACOS, *supra* note 169, at 61-63; KELVIN LANCASTER, VARIETY, EQUITY, AND EFFICIENCY 14 (1979). For similar conclusions offered in the related context of monopolistic competition, see Robert L. Bishop, *Monopolistic Competition and Welfare Economics*, *in* MONOPOLISTIC COMPETITION THEORY: STUDIES IN IMPACT 251, 261 (Robert E. Kuenne ed., 1967); E.H. Chamberlin, *Product Heterogeneity and Public Policy*, 40 AM. ECON. REV. 85, 89-92 (1950); Harold Demsetz, *The Nature of Equilibrium in Monopolistic Competition*, 67 J. POL. ECON. 21, 22 (1959); Avinash K. Dixit & Joseph E. Stiglitz, *Monopolistic Competition and Optimum Product Diversity*, 67 AM. ECON. REV. 297, 300-02 (1977); N. Gregory Mankiw & Michael D. Whinston, *Free Entry and Social Inefficiency*, 17 RAND J. ECON. 48, 49, 54-55 (1986); Michael Spence, *Product Differentiation and Welfare*, 66 AM. ECON. REV. 407, 407-08, 411 (1976).

At the same time, spatial competition reveals the important role that entry can play in promoting access. As noted earlier, the presence of supracompetitive returns attracts entry by other producers of close substitutes until those supracompetitive returns have been dissipated.[187] Whether entry causes price to rise or fall depends on its effect on the elasticity of demand. Because entry by close substitutes should cause demand to become more elastic, entry should have the effect of pushing price toward marginal cost in much the same manner as entry does under perfect competition.[188] Under any pricing regime, the spread between price and marginal cost should be relatively small if the economy is sufficiently large. Indeed, if the economy is made infinitely large, either by letting the utility that customers derive approach infinity or by letting the fixed costs of entry approach zero, prices will asymptotically converge to marginal cost.[189]

This observation implies that increasing the size of the market by increasing the number of welfare-generating activities encompassed by copyright represents an alternative approach to promoting access. Producers will divide the available surplus until the surplus captured by each

---

[187] *See supra* note 170 and accompanying text.

[188] Yoo, *supra* note 5, at 253. Such price competition would be particularly intense if spatial competition were to occur along more than one dimension. In one-dimensional spatial competition, every producer competes with no more than two competitors. If the competitive space is expanded to three dimensions, each producer may compete with as many as six adjacent neighbors. If competition expands to four dimensions, each producer may theoretically compete with as many as half the firms operating in the product group. G.C. Archibald & G. Rosenbluth, *The "New" Theory of Consumer Demand and Monopolistic Competition*, 89 Q.J. ECON. 569, 576-84 (1975). Empirical studies have largely confirmed this effect. *See, e.g.*, Robert C. Feenstra & James A. Levinsohn, *Estimating Markups and Market Conduct with Multidimensional Product Attributes*, 62 REV. ECON. STUD. 19, 36-41 (1995) (concluding that automobiles compete with each other spatially along at least four different product characteristics and that the average car competed with 5.90 other models).

[189] BEATH & KATSOULACOS, *supra* note 169, at 145-47; Eaton & Lipsey, *supra* note 22, at 761; Eaton & Wooders, *supra* note 21, at 289-91, 292, 294. For similar results derived in the related context of monopolistic competition, see Jean-Pascal Benassy, *Market Size and Substitutability in Imperfect Competition: A Bertrand-Edgeworth-Chamberlin Model*, 56 REV. ECON. STUD. 217, 231-32 (1989); Oliver D. Hart, *Monopolistic Competition in a Large Economy with Differentiated Commodities*, 46 REV. ECON. STUD. 1, 11-13, 20 (1979); Larry E. Jones, *The Efficiency of Monopolistically Competitive Equilibria in Large Economies: Commodity Differentiation with Gross Substitutes*, 41 J. ECON. THEORY 356, 358, 372, 375 (1987); Mankiw & Whinston, *supra* note 186, at 56-57.

individual producer no longer exceeds the fixed costs of entry.[190] Increasing the surplus available should thus stimulate a higher level of competitive entry. The increased level of entry should drive price closer to marginal cost.

Price discrimination can help bring the equilibrium level of access under spatial competition even closer to the optimum in two distinct ways. First, as Stiglitz notes, price discrimination can increase total output by allowing producers to sell to low-value users without having to sacrifice any revenue from sales to high-value users.[191] Second, price discrimination has the added benefit of promoting entry by increasing producers' ability to appropriate surplus. The resulting increase in price competition should provide an independent force driving prices closer to marginal cost.

Modeling copyright as an impure public good thus suggests an alternative approach to promoting access that differs starkly from the conventional approach. Rather than promoting access directly by limiting the level of copyright protection, the impure public goods approach promotes access indirectly by facilitating entry and allowing the ensuing increase in competition to drive price closer to marginal cost. In the process, it contradicts the conventional approach's presumption that any solution that allows authors to recover their first-copy costs is bounded away from providing optimal levels of access to copyrighted works. Indeed, because price converges asymptotically to marginal cost as entry increases, the systematic bias toward underutilization simply disappears. In addition, the fact that entry will continue until all

[190] Again, the "integer problem" created by fixed cost indivisibilities and the localized nature of spatial competition can allow producers to earn supracompetitive returns. So long as the economy is sufficiently large, any such profits should be relatively small. *See supra* notes 153, 170.

[191] Stiglitz, *supra* note 21, at 310-11; *see also* Severin Borenstein, *Price Discrimination in Free-Entry Markets*, 16 RAND J. ECON. 380, 392, 394-95 (1985) (concluding that price discrimination under spatial competition can increase total quantity sold).

supracompetitive returns are dissipated undercuts any suggestion that increasing the total surplus encompassed by copyright will enhance authors' ability to earn supracompetitive returns.

A spatial competition approach also avoids the tendency in the existing literature to represent all of the different aspects of copyright protection with a single variable and to speak in general terms about the overall strength of copyright protection.[192] Instead, it suggests that access would best be promoted if copyright protection were relatively strong along certain dimensions and relatively weak along others. Specifically, the fact that maximizing entry can simultaneously increase incentives to create copyrightable works and promote efficient access to those works favors making the copyright relatively "large," in that it contains a large number of surplus-generating activities within its scope. Entry would further be promoted if copyright were relatively "intense," in that it permits authors to appropriate a significant proportion of the available surplus. At the same time, the desire to promote such entry counsels both against imposing any artificial restrictions on the level of entry by close substitutes and in favor of a copyright that is relatively "narrow," in that a competing product may come relatively close in the characteristics space to existing works without constituting infringement.[193]

---

[192] *See* Christian Koboldt, *Intellectual Property and Optimal Copyright Protection*, 19 J. CULTURAL ECON. 131, 136 (1995) (representing all aspects of copyright protection with a single variable, *P*); Landes & Posner, *supra* note 1, at 333-36 (representing all aspects of copyright protection with a single variable, *z*); Ian E. Novos & Michael Waldman, *The Effects of Increased Copyright Protection: An Analytic Approach*, 92 J. POL. ECON. 236, 238-39 (1984) (representing all aspects of copyright protection with a single variable, *H*).

[193] Yoo, *supra* note 5, at 265-72. For related arguments in the context of patents, see Richard Gilbert & Carl Shapiro, *Optimal Patent Length and Breadth*, 21 RAND J. ECON. 106, 106-07 (1990); Paul Klemperer, *How Broad Should the Scope of Patent Protection Be?*, 21 RAND J. ECON. 113, 120-24 (1990).

b. The Feasibility of Optimal Incentives

As noted earlier, spatial competition also opens up the possibility that markets might provide optimal levels of access.[194] Early work suggested that in the absence of perfect price discrimination, markets for differentiated products would exhibit a systematic bias toward producing too few goods.[195] Samuelson and Stiglitz both recognized that, even in the absence of price discrimination, spatial competition can produce the optimal number of goods.[196]

Again, the key to understanding why this is the case is the Samuelson condition. As noted earlier, the optimal provision of an impure public good requires the producer to

---

[194] *See supra* note 21 and accompanying text. The discussion that follows extends the analysis first discussed in Yoo, *supra* note 5, in 256-64.

[195] Michael Spence, *Product Selection, Fixed Costs, and Monopolistic Competition*, 43 REV. ECON. STUD. 217, 217-20 (1976) (concluding that the complete appropriation of consumer surplus is a necessary condition for optimal provision). 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 the appropriability of information goods likely leads to an overall increase in the production of distinct works).

The reasons are well illustrated by the following example. Consider the extreme case when marginal cost is zero. Economic welfare would increase if a work were created whenever the prospective benefits of that work exceed the fixed costs needed to produce it. For example, society would benefit from a work that created $10 million in surplus and required $7.5 million in fixed costs to make. If the author were only able to appropriate 70% of the available surplus, however, she would receive only $7 million in revenue. The inability to capture all of the surplus created would lead the author not to create the work even though economic welfare would have increased had she done so. Enabling the author to appropriate 80% of the available surplus would allow works that cost $7.5 million to be created, but would still leave out works that cost $9 million despite the fact that creating that work would also enhance economic welfare. The only way to ensure the creation of the marginal welfare-enhancing work (the work whose cost is just below the total benefits of $10 million) is to enable perfect price discrimination. Yoo, *supra* note 5, at 257.

In addition, price discrimination may well be a necessary condition for the existence of equilibria. Phillip J. Lederer & Arthur P. Hurter, Jr., *Competition of Firms: Discriminatory Pricing and Location*, 54 ECONOMETRICA 623, 623-24 (1986); W.B. MacLeod et al., *Price Discrimination and Equilibrium in Monopolistic Competition*, 6 INT'L J. INDUS. ORG. 429, 429 (1988).

[196] *See* Samuelson, *Aspects*, supra note 7, at 335 (reasoning that in mixed cases between the polar extremes of private and pure public goods, "we might find just the right conditions of scarcity of space and of independence of consumptions so that ordinary market pricing could lead to the optimum," provided that such ordinary pricing "happens to pick up each indirect external marginal utility"); Stiglitz, *supra* note 21, at 312 (noting that spatial competition in the absence of price discrimination can reach equilibria that are social-welfare maximizing). For analogous findings in the context of monopolistic competition, see BEATH & KATSOULACOS, *supra* note 169, at 61-66; G.C. Archibald, *Chamberlin Versus Chicago*, 29 REV. ECON. STUD. 2, 7-14 (1961); Oliver D. Hart, *Monopolistic Competition in the Spirit of Chamberlin: Special Results*, 95 ECON. J. 889, 901, 903 (1985); Roger W. Koenker & Martin K. Perry, *Product Differentiation, Monopolistic Competition, and Public Policy*, 12 BELL J. ECON. 217, 226-

appropriate the sum of *marginal* benefits derived by all of the consumers of that good. The problem is that markets do not provide an incentive-compatible mechanism for determining consumers' marginal benefits. On the contrary, consumers have the incentive to understate the intensity of their preferences in an attempt to easy ride on contributions made by others. It is for this reason that markets tend to produce too few public goods.

This tendency toward underproduction is mitigated in the case of spatial competition by the fact that the surplus captured by those who enter comes from two different sources. Part of the surplus captured by the new entrant results from *demand creation*—that is, new surplus generated either by inducing consumers who were otherwise not purchasing to enter the market or by providing greater utility to those who were already purchasing by allowing them to obtain goods that lie closer to their ideal preferences. Because demand creation represents an incremental increase in welfare, it tends to push the market equilibrium toward the welfare-maximizing result.

At the same time, some of the surplus captured by the new entrant is the result of *demand diversion*—that is, surplus cannibalized from other producers already in the market. Because this surplus was already being satisfied by a prior entrant, its appropriation by the new entrant represents a wealth transfer from one producer to another that makes no incremental contribution to economic welfare.

The problem is that profit-maximizing entrants base their entry decisions on a comparison of total revenue with total costs without regard to whether the revenue captured results from demand creation or demand diversion. Because the presence of demand diversion

---

27 (1981); MacLeod, *supra* note 195, at 430; Mankiw & Whinston, *supra* note 186, at 55; Spence, *supra* note 186, at 413.

causes total revenue to exceed total social benefits, entrants may find it profitable to enter even when it would be socially wasteful for them to do so. As such, demand diversion can promote the production of additional goods even when further entry would be economically inefficient.[197] Concern about excess entry may strike some people as somewhat unorthodox, given that greater product choice and greater price competition are generally regarded as economically beneficial.[198] It remains an important issue when products are differentiated and entry requires the incurrence of fixed costs, since entry would be inefficient if the economic benefits associated with an additional product do not exceed the fixed costs of entry.[199]

The systematic bias toward producing too many goods inherent in demand diversion can compensate for the systematic bias toward producing too few works inherent in the Samuelson condition as well as imperfections in the ability to appropriate surplus. Whether markets will reach equilibrium with too many or too few goods depends on which of these two effects

---

[197] Borenstein, *supra* note 191, at 388-89, 393; *see also* Mankiw & Whinston, *supra* note 186, at 54-55 (offering a similar discussion in the context of monopolistic competition); Spence, *supra* note 195, at 410 (same). The following example may help illustrate the point. Suppose that a new entrant exactly duplicates the position of an existing product. Because all of the revenue captured by the entrant would consist exclusively of demand diversion, entry would simply waste resources without providing any compensating welfare benefits. Steven T. Berry & Joel Waldfogel, *Free Entry and Social Inefficiency in Radio Broadcasting*, 30 RAND J. ECON. 397, 397-98 (1999).

A recent paper by Bhaskar and To identifies a different source of demand diversion. Rather than following the standard assumption in the literature that all firms enter simultaneously, they presume that entry occurs after existing firms have already evenly distributed themselves across the product space. Under this approach, demand diversion arises not as a result of direct business stealing, but rather from the fact that existing firms must relocate in order to accommodate the new entrant. This in turn allows the new entrant to appropriate surplus that was previously captured by one of the incumbents before it was forced to move to a different location. Bhaskar & To, *supra* note 174, at 775. Despite the differences in formulation between these two approaches, the policy implications are largely the same for both. To the extent that revenue consists of demand diversion, it drives markets toward excess entry.

[198] *See, e.g.*, J. MacKie-Mason et al., *Service Architecture and Content Provision*, 20 TELECOMM. POL'Y 203, 207 (1996) (describing excess entry as "fairly unconventional for an economic problem," given that "more choice over available goods is routinely assumed to be unambiguously desirable").

[199] *See* Eaton & Lipsey, *supra* note 22, at 731 (calling "whether there are too few or too many products in equilibrium" the "major issue" in the case of the monopolistic competition branch of product differentiation); Mankiw & Whinston, *supra* note 186, at 48 (noting that "[e]conomists typically presume that free entry is desirable for social efficiency," but that entry can be inefficient when it requires the incurrence of fixed costs); Spence, *supra*

dominates. Indeed, if the tendency toward overproduction caused by demand diversion happens to offset exactly the tendency toward underproduction associated with easy riding and nonappropriability, markets would produce the optimal number of goods. Spatial models thus do not necessarily exhibit the systematic tendency toward underproduction characteristic of pure public goods models.

That said, there is no reason to suppose that these forces will counterbalance each other so precisely. Whether conditions are such that markets will reach equilibrium with too many or too few impure public goods has important policy implications. Consider first the case in which demand diversion comprises a relatively small amount of the surplus appropriated by the entrant, either because producers are able to appropriate only a relatively small proportion of the available surplus created by their goods or because the good at issue has relatively few close substitutes. When this is the case, one would expect the market to reach equilibrium at a point where there are too few works, thus leaving no reason for policymakers not to make copyright protection as large, intense, and narrow as possible. Doing so would maximize incentives for creation and thereby bring the level of product variety as close to optimal as possible. At the same time, the increase in entry would promote efficient levels of access to the works by maximizing the level of competition among close substitutes.

For these types of works, then, the tension between access and incentives generally thought to underlie much of copyright policy disappears. Instead, the same policy instruments

---

note 186, at 408 (noting that too much and too little entry represent important, but oft-ignored sources of welfare loss when products are differentiated).

can promote both interests simultaneously. As a result, the justification for regarding copyright as a "necessary evil" collapses.[200]

The policy implications are somewhat more complicated when works serve as reasonably good substitutes for one another and when producers are able to appropriate a large proportion of the available surplus. Under these circumstances, it is likely that demand diversion will overcompensate for the tendency toward underproduction caused by easy riding and that markets will produce too many works in equilibrium. When that is the case, providing proper incentives for creating the optimal number of works necessarily requires limiting either the size or the intensity of copyright protection. The concomitant reduction in the degree of entry necessarily causes some reduction in the level of access. Although the access/incentive tradeoff reappears for these goods, it bears emphasizing that it is no longer an endemic problem confronting all of copyright. Instead, it is a more limited problem that is significantly more restricted in scope and contingent upon the existence of certain factual predicates.

Because the magnitude of the tradeoff depends on the level of demand diversion associated with the level of substitutability and appropriability inherent in a particular work, a first-best solution would require calibrating the level of copyright protection on a case-by-case basis. Making such evaluations should prove no easy matter. If spatial competition occurs in a geographic space, it may be possible to observe and parameterize the relevant transportation cost

[200] For the classic statement of this position, see Thomas B. Macaulay, Speech Before the House of Commons (Feb. 5, 1841), *in* 8 THE WORKS OF LORD MACAULAY 195, 199 (Lady Trevelyan ed., 1900) ("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."). For examples of modern restatements of this position, see, for example, Brennan, *supra* note 3, at 687-88; Mark A. Lemley, *Ex Ante Versus Ex Post Justifications for Intellectual Property*, 71 U. CHI. L. REV. 129, 131 (2004); Mark S. Nadel, *How Current Copyright Law Discourages Creative Output: The Overlooked Impact of Marketing*, 19 BERKELEY TECH. L.J. 785, 787 (2004); Neil Weinstock Netanel,

functions. If spatial competition occurs in a characteristics space, the problem is considerably more difficult. Unlike price and cost, individual preferences for particular product characteristics cannot be observed directly. Moreover, preference functions can take a much broader range of shapes and magnitudes than transportation cost functions, which are constrained by the cost characteristics of the inputs needed to provide the necessary transportation. For this reason, two leading spatial competition theorists candidly acknowledge that with respect to spatial competition in a characteristics space, "we believe that we would be quite unable to recognize an optimum if we saw one."[201]

Despite these difficulties, a small literature has emerged attempting to assess the potential welfare losses from excess entry in the context of spatial competition among differentiated products.[202] Some studies suggest that so long as the relevant economy is relatively large, any welfare losses resulting from excessive entry are likely to be quite small.[203] Other studies have found the welfare losses from excess entry to be more significant.[204]

Asserting Copyright's Democratic Principles in the Global Arena, 51 VAND. L. REV. 217, 248-49 (1998); Yen, *supra* note 23, at 1368. *See generally* Yoo, *supra* note 5, at 216 n.9 (collecting other similar sources).

[201] Eaton & Lipsey, *supra* note 22, at 760; *see also* Berry & Waldfogel, *supra* note 197, at 417 (suggesting that empirically modeling entry in a characteristics space would require techniques that exceed the current state of the art).

[202] *See* Yoo, *supra* note 5, at 274-76 (offering a preliminary review of the literature).

[203] For example, Ronald Goettler and Ron Shachar study spatial competition among major broadcast television networks, concluding that the equilibrium nearly achieved the optimal level of product differentiation, with the shortfall explained by bounded rationality and the networks' adherence to certain rules of thumb about scheduling. Ronald L. Goettler & Ron Shachar, *Spatial Competition in the Network Television Industry*, 32 RAND J. ECON. 624, 647-52 (2001). For other estimates suggesting that the welfare losses from excess entry are relatively small, see Eaton & Wooders, *supra* note 21, at 291 (calling the resource misallocation from excess entry "vanishingly small" in large economies); 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); *cf.* Spence, *supra* note 195, at 411-13 (conducting a series of illustrative calculations in the related context of monopolistic competition and concluding that "the equilibrium is often a reasonably good approximation to the constrained optimum"). A very different result obtains in the case of small economies. *See* Yarrow, *supra*, at 521-23 (concluding that welfare losses are much greater in such situations); BEATH & KATSOULACOS, *supra* note 169, at 64-66 (reaching a similar conclusion).

[204] Steven Berry and Joel Waldfogel study spatial competition for advertising among radio stations, finding excess entry of 74% with an annual deadweight loss of $2.3 billion. They acknowledge that these welfare losses

It thus appears that the empirical record is not yet sufficiently developed to infer how often entry will be optimal, excessive, or insufficient. As a result, decision makers seeking to adjust the scope of copyright protection to achieve optimal entry will struggle to determine the appropriate direction and magnitude of such adjustments. In addition, any welfare losses resulting from excess entry would necessarily be counterbalanced by welfare gains from increased product variety and increased price competition.

Decisions about the scope of copyright protection would thus depend on a careful assessment of a number of case-specific considerations, including the availability of substitutes for the work in question, consumers' preferences for the work, and the author's ability to appropriate surplus. The transaction costs of making these determinations would inevitably be significant. In addition, the workability of a case-by-case approach varies greatly based on whether the scope of copyright protection is determined on an ex ante or an ex post basis. Consider first the problems of making such an assessment ex ante. Determining the level of appropriability and substitutability is likely to be particularly difficult before the work is actually created. In addition, copyright law would have to devise some method to address the moral hazard problems caused by the possibility that authors might shirk in the quality of their works once the level of copyright protection has been set. The legal system would also have to find a way to credibly commit to the level of protection established ex ante in order to protect authors against the dangers of ex post opportunism once the first-copy costs have been sunk. Problems

---

might be offset by welfare benefits to radio listeners (rather than advertisers) and by welfare gains from increases in the diversity in radio programming, neither of which they were able to measure directly. Berry & Waldfogel, *supra* note 197, at 411-17.

Gerald Faulhaber employs a spatial competition model to study whether file sharing and DRM have caused excess entry in the music industry, concluding that excess entry likely has occurred. At the same time, Faulhaber recognizes that entry might create additional benefits not taken into account by his model if such entry inspires the creation of follow-on works. Faulhaber, *supra* note 169, at 92-102.

would also surround any attempt to assess the level of appropriability and substitutability associated with a particular work ex post. The ex ante uncertainty would force authors to discount their expectations about the likely value of their works against the possibility of a change in valuation ex post.

The analysis is further complicated by the fact that appropriability and substitutability tend to change over time. Thus, even if policymakers managed to identify those situations in which access and incentives are in tension and managed to calibrate copyright so as to strike the proper balance between those considerations, the resulting balance between access and incentives is unlikely to prove stable.

These concerns suggest that the better alternative may be to forego case-by-case analysis in favor of a simpler approach that, despite being a bit Procrustean, establishes general rules that apply to all copyrightable works. Two possible approaches immediately come to mind. On the one hand, Congress and the courts could ignore the potential welfare losses from excess entry and instead maximize access by making copyright as large, intense, and narrow as possible. On the other hand, Congress and the courts could ignore the case-by-case variations and attempt to calibrate a uniform copyright to strike a rough balance between the welfare losses from excess entry and the welfare losses from insufficient access. Although a uniform approach would overprotect some works and underprotect others, the overall result may be preferable both to evaluating copyright on a case-by-case basis and to simply maximizing entry.

Between these two alternatives, I would favor fostering a copyright that is large, intense, and narrow over trying to strike a balance. Any attempt to calibrate the level of copyright protection would confront the same dearth of empirical evidence regarding the pervasiveness and the magnitude of the welfare losses associated with excess entry discussed above. This in turn

would force Congress and the courts to base their resolution of the underlying tradeoff largely on conjecture. In contrast, making copyright as large, intense, and narrow as possible would have the institutional advantage of allowing decentralized decisions by market actors all over the economy to determine the proper level of entry. Not only should this improve the mechanism for incorporating information about costs and consumer preferences at any particular point, it should also accommodate technological change without incurring the delay biases inherent in governmental processes.[205] This also would have the advantage of giving legislators and courts a mandate that is relatively clear and easy to implement when compared with the type of empirically speculative and indeterminate balancing implicit in the other approach. The maximization of entry and access, even when entry may be economically excessive, should also have some appeal to those who favor maximizing access to and diversity of creative works for noneconomic reasons.[206]

---

[205] For a sampling of the literature claiming that political biases are distorting the copyright system, see, for example, LANDES & POSNER, *supra* note 34, at 403-19; Tom W. Bell, *Escape from Copyright: Market Success vs. Statutory Failure in the Protection of Expressive Works*, 69 U. CIN. L. REV. 741, 786-87 (2001); Mark A. Lemley, *The Constitutionalization of Technology Law*, 15 BERKELEY TECH. L.J. 529, 531-33 (2000); Jessica Litman, *Copyright Legislation and Technological Change*, 68 OR. L. REV. 275, 359 (1989); Joseph P. Liu, *Copyright and Time: A Proposal*, 101 MICH. L. REV. 409, 448-51 (2002); Merges, *supra* note 32, at 1868-74; Sterk, *supra* note 32, at 1244-46; Timothy Wu, *Copyright's Communications Policy*, 103 MICH. L. REV. 278, 344 (2004). For examples of commentators arguing that the biases in the political process justify more intrusive judicial review, see LAWRENCE LESSIG, FREE CULTURE 215-18 (2004); Neil Weinstock Netanel, *Locating Copyright Within the First Amendment Skein*, 54 STAN. L. REV. 1, 69 (2001). Others have cogently observed that copyright is not the only area of the law supposedly affected by public choice failures and that if accepted, this argument would justify intrusive judicial review of all economic legislation in a manner similar to the now-discredited approach associated with the *Lochner* era. Thomas B. Nachbar, *Judicial Review and the Quest To Keep Copyright Pure*, 2 J. ON TELECOMM. & HIGH TECH. L. 33, 53-54 (2003); Paul M. Schwartz & William Michael Treanor, Eldred *and* Lochner*: Copyright Term Extension and Intellectual Property as Constitutional Property*, 112 YALE L.J. 2331, 2400-09 (2003). It thus comes as no surprise that the Supreme Court has firmly rejected calls for more exacting judicial scrutiny of copyright laws. Eldred v. Ashcroft, 537 U.S. 186, 204-05 & n.10, 208, 217-21 (2003).

[206] Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354, 377-81 (1999); Neil Weinstock Netanel, *Copyright and "Market Power" in the Marketplace of Ideas*, *in* ANTITRUST, PATENTS AND COPYRIGHT 149, 161 (François Lévêque & Howard Shelanski eds., 2005).

3. Potential Limitations to Spatial Competition

The policy implications of analyzing copyright through the lens of spatial competition are quite striking. The systematic biases toward underproduction and underutilization and the danger of supracompetitive returns largely disappear. Instead, spatial competition shows how entry can promote the ability of low-value consumers to access works as well as prevent authors from earning supracompetitive returns. At the same time, spatial competition provides a basis for determining when entry is excessive. Thus, rather than providing consistent support for the expansion of copyright protection, spatial competition introduces notions of optimality that can serve as a basis for distinguishing the dimensions along which copyright protection is too strong and too weak. Other portions of the literature on spatial competition add additional nuances, such as how sunk costs of entry[207] and multilocation entry by a single firm,[208] can foreclose entry by later players.

At the same time, the fit between copyright and spatial competition is not necessarily perfect. For example, spatial models work only if consumers can organize the available products into a set of ordinal rankings. Preferences for certain creative works (such as music, which spans formats including classical, jazz, Top 40, oldies, rock, country, contemporary Christian, and easy listening) may prove insusceptible to being arranged into a coherent linear spectrum. Indeed, as

---

[207] 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-Deterring Strategies in Spatial Competition*, 28 OXFORD ECON. PAPERS 240 (1976); W.J. Lane, *Product Differentiation in a Market with Endogenous Sequential Entry*, 11 BELL J. ECON. 237, 239 (1980); 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).

[208] James A. Brander & Jonathan Eaton, *Product Line Rivalry*, 74 AM. ECON. REV. 323, 330-32 (1984); B. Curtis Eaton & Richard G. Lipsey, *The Theory of Market Pre-emption: The Persistence of Excess Capacity and*

Arrow's theorem points out,[209] consumer preferences can actually be structured in such a way that makes it impossible to talk meaningfully about an overarching hierarchy of preferences.

Moreover, much of the attractiveness of the spatial competition equilibrium follows from the assumption that consumers are uniformly distributed across the product space. Relaxing this assumption can weaken competition in some portions of the product spectrum and can allow firms in those areas to earn sustainable supracompetitive returns.[210] The model also assumes that entry is open to anyone willing to incur the fixed costs needed to enter. In so doing, it must recognize that the current definition of copyright infringement places some legal limits on how closely one work can resemble another.[211] In assuming that entry is free, the spatial competition model also downplays the possibility that some authors may have unique abilities to generate high levels of utility at particular locations on the product spectrum, which would limit the degree of competition faced by those authors' works.

Moreover, spatial competition is a discrete choice model, in that it assumes consumers purchase a single product from the one producer positioned closest to their respective locations. Thus, the model ignores the possibility that consumers may wish to purchase goods from multiple providers in multiple locations. In addition, it ignores the possibility that consumers may not want to buy their entire amount of a particular product from the closest provider—they may instead want to "crossover" and purchase small quantities from relatively distant

---

Monopoly in Growing Spatial Markets, 46 ECONOMICA (n.s.) 149 (1979); Kenneth L. Judd, *Credible Spatial Preemption*, 16 RAND J. ECON. 153 (1985); Schmalensee, *supra* note 171.

[209] KENNETH J. ARROW, SOCIAL CHOICE AND INDIVIDUAL VALUES 2-3 (2d ed. 1966).

[210] The seminal analysis of the impact of preference asymmetries was offered by Kaldor. Nicholas Kaldor, *Market Imperfection and Excess Capacity*, 2 ECONOMICA (n.s.) 33, 37-40 (1935). 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); Michael Waterson, *The Economics of Product Patents*, 80 AM. ECON. REV. 860 (1990).

[211] Lemley, *supra* note 3, at 1057.

providers.[212] Because they look at each purchaser's decision in isolation, discrete choice models can also have difficulty capturing demand interdependencies, such as those associated with solidarity goods and network economic effects, that exist when one individual's purchasing decisions depend on the purchasing decisions of others.

None of these obstacles are necessarily insuperable. For example, even if the particular product characteristics defy categorization into a coherent spectrum, it might well be possible to organize products based on their appeal to different demographic groups (e.g., organizing types of music into a coherent spectrum based on the average age of their respective audiences). In addition, sophisticated econometric techniques exist that can abstract unobservable product characteristics from the underlying data.[213] Furthermore, practical limitations on entry can be modeled by hypothesizing that follow-on entry into certain locations is impossible or by assuming that certain works will achieve a higher level of utility than others. Lastly, the possibility that consumers might want to consume multiple products might be best accommodated by shifting to a model of differentiated products known as monopolistic competition. This model is less wedded to the discrete choice framework and more easily accommodates concerns such as the inability to organize products into a spectrum of ordinal rankings, entry asymmetries, multiple purchases and crossover, and demand interdependencies.[214]

---

[212] *See* Eaton & Lipsey, *supra* note 22, at 751 (citing LANCASTER, *supra* note 186) (noting Kelvin Lancaster's use of the term "crossover" to describe when a consumer purchases goods outside of the local market).

[213] *See* Goettler & Shachar, *supra* note 203, at 641-43 (inferring that television programming competes in a four-dimensional characteristics space).

[214] *See* Yoo, *supra* note 5, at 236-41 (describing the monopolistic competition approach to modeling product differentiation). Monopolistic competition allows for the possibility that consumers may want to consume multiple products by assuming that all producers are in equal competition with one another. Thus, entry by a new producer will divert sales from all incumbent producers symmetrically. The tradeoff, however, is that the symmetry

More importantly, the insights of an impure public goods approach should remain clear even if the precise arguments and parameters for any particular approach to modeling impure public goods cannot be resolved. Markets for copyrighted works are subject to numerous variations in quality that each can serve as an equilibrating force despite the fact that indivisibility forces each customer to consume the same quantity. Thus, the overall promise of shifting to an impure public goods approach should remain apparent even if spatial competition ultimately proves to be an unsatisfactory way to model markets for copyrighted works.

IV. THE POLICY IMPLICATIONS OF APPLYING IMPURE PUBLIC GOODS THEORY TO COPYRIGHT

Recognizing that copyright should be regarded as an impure (as opposed to a pure) public good carries with it a number of important policy implications. In this Part, I apply the insights gained from the foregoing analysis to evaluate the series of copyright-related doctrines introduced in Part I.C. The absence of any systematic tendency toward underproduction or underutilization suggests that many of the current justifications for these doctrines based in pure public goods theory need to be reconsidered.

A. Fair Use

As noted earlier, the economics of fair use has traditionally focused on the impact of transaction costs. As transaction costs have receded in importance, commentators have suggested that fair use may serve the alternative purpose of striking a balance between access and incentives.[215]

---

assumption fails to capture the possibility that competition among differentiated products might be localized. CHAMBERLIN, *supra* note 173, at 196-98.

[215] *See supra* Part I.C.1.

Reformulating the application of public good economics in the manner I propose would raise doubts about the viability of these alternative economic justifications for fair use. Impure public goods theory reveals that supramarginal cost pricing might not be as problematic as these commentators suggest. The theory that supramarginal cost pricing represents endemic market failure might give way to the broader notion that the welfare losses associated with supramarginal cost pricing might be offset by welfare gains from the increased product diversity made possible by such pricing. In other words, what appears to be a welfare loss from the standpoint of the price-quantity space that dominates conventional microeconomic analysis might in fact be a constrained optimum.

In addition, current justifications for fair use overlook the fact that narrowing the scope of fair use and allowing entry to bring prices closer to marginal cost might actually promote access by stimulating entry and allowing the ensuing increase in competition to reduce prices indirectly. These justifications also fail to consider the possibility that restricting fair use would promote the most efficient spatially competitive equilibrium by enhancing authors' ability to appropriate surplus through price discrimination.

This underscores the key difference between the role that price discrimination plays in the conventional approach to public good economics and in the more fundamental approach that I propose in this Article. In the former, the purpose of price discrimination is to provide low-value users with access to creative works. This favors giving fair use a broad scope and justifies measures that would solve the underutilization problem by facilitating low-value users' ability to obtain access to copyrighted works. In the latter, the purpose of price discrimination is to help authors appropriate more of the available surplus across the entire range of production. Under

this perspective, measures that simply facilitate access by low-value users would not represent a complete solution to the problems posed by public goods.

A shift to an impure public goods approach would also narrow and recast the scope of the access/incentives tradeoff. Specifically, it would no longer be true that any solution would necessarily be second best in terms of both access and incentive. It would also no longer be true that promoting one consideration would necessarily come at the expense of the other. Instead, the central policy problem would be identifying cases in which substitutability and appropriability render excess entry likely. Implementation of the case-by-case approach implicit in this analysis would pose considerable practical difficulties, as demonstrated by the controversy and uncertainty surrounding the implementation of the current fair use doctrine. The lack of clear empirical evidence to help determine precisely where the relevant lines should be drawn and the relative ease with which policymakers and judges could implement a mandate to promote entry suggest that copyright policy might be better served if the scope of fair use were allowed to contract as transaction costs continue to fall.

B. Duration

The theory of impure public goods also calls into question the commentary that regards the access/incentives tradeoff implicit in the conventional approach as a justification for limiting copyright duration.[216] Specifically, shifting to an impure public goods perspective suggests that incentives for creating copyrightable works can be promoted without necessarily sacrificing access, since entry by imperfect substitutes should help drive prices toward marginal cost. Indeed, this reasoning suggests that access would be best promoted if copyright duration were

made as long as constitutionally permitted, since doing so would maximize entry and in the process maximize the price competition that minimizes deadweight loss. The welfare gains from product diversity should further offset the welfare losses from supramarginal cost pricing. Indeed, this suggests that the supposedly irreconcilable conflict inherent in the access/incentives tradeoff may be overstated.

Such a solution is subject to an important caveat. The theory of impure public goods suggests that entry may well be excessive if works are highly substitutable and authors are able to appropriate a high proportion of the available surplus. The fact that appropriability and substitutability are likely to vary widely suggests that the first-best solution would require a copyright term that varies from work to work. Although suggestions that the copyright term vary on a case-by-case basis have appeared from time to time in the literature,[217] implementation difficulties render such a regime impractical.

This leaves Congress with a choice between second-best alternatives. It can promote a copyright that is as large, intense, and narrow as possible, or it can attempt to calibrate copyright to balance the welfare losses associated with reduced access against those associated with excess entry. As noted earlier, the empirical record is not sufficiently well developed to permit a clear assessment of this tradeoff. The important insight is that the ultimate balance need not be as inherently suboptimal as the conventional approach would lead one to believe.

---

[216] *See supra* Part I.C.2.

[217] RALPH S. BROWN & ROBERT C. DENICOLA, CASES ON COPYRIGHT, UNFAIR COMPETITION, AND RELATED TOPICS BEARING ON THE PROTECTION OF WORKS OF AUTHORSHIP 507 (8th ed. 2002); Robert L. Bard & Lewis Kurlantzick, *Copyright Duration at the Millennium*, 47 J. COPYRIGHT SOC'Y U.S.A. 13, 68 n.126 (2000); *cf.* WILLIAM D. NORDHAUS, INVENTION, GROWTH, AND WELFARE: A THEORETICAL TREATMENT OF TECHNOLOGICAL CHANGE 79 (1969) (suggesting a model in which the length of the patent term varies with the elasticity of demand); F.M. SCHERER, INNOVATION AND GROWTH: SCHUMPETERIAN PERSPECTIVES 133 (1984) (calling case-by-case determination of a patent term "not inconceivable").

C. Compulsory Licenses

As noted earlier, scholars have invoked the theory of pure public goods as support for employing compulsory licenses. Some view compulsory licenses as a way to calibrate the balance between access and incentives, while others emphasize the difficulties in inducing consumers to reveal the value that they place on public goods.[218]

Shifting to an impure public goods perspective raises doubts about both of these rationales. With respect to the former justification, the foregoing analysis calls into question the extent to which access and incentives are truly in tension. Indeed, when low substitutability and appropriability cause demand diversion to represent a relatively small amount of the surplus appropriated by an author, both access and incentives can be promoted simultaneously by making copyright as large, intense, and narrow as possible. In these cases, imposing a compulsory license would be counterproductive. Compulsory licenses may be more justifiable when high substitutability and appropriability render excess entry more likely. If compulsory licenses were used to reduce excess entry, they would need to be redesigned to limit their scope to these cases.

With respect to the second justification, allowing consumers to allocate their purchases spatially among different goods can create de facto markets through which consumers can reveal the intensity of their preferences despite the incentive incompatibility problems inherent in the Samuelson condition. The impure public goods approach thus offers an attractive alternative mechanism for determining consumers' valuations for public goods. Determining the optimality of the resulting equilibrium would depend on an assessment of the availability of close

[218] *See supra* Part I.C.3.

substitutes and the ability of producers to appropriate surplus. Although such assessments would doubtlessly pose significant difficulties, such challenges seem more tractable than attempting to measure consumer preferences directly.

D. Database Protection

The economics of impure public goods also offer new insights into the proper scope of database protection.[219] Reconceiving copyright as competition among differentiated products suggests that the tradeoff between access and incentives may not represent as central a problem as previous analyses suggest. Furthermore, requiring database owners to instantiate their intangible property into a tangible form would do little to solve the incentive incompatibility with respect to the revelation of preferences inherent in the Samuelson condition. As noted earlier, the fact that the public good may be an input that must be combined with other rival inputs does not change its character as a public good.[220]

Equally importantly, the possibility for consumers to allocate their purchases spatially opens up new avenues for determining the intensity of their preferences. In addition, protecting databases against copying would foster entry that, in many cases, would simultaneously promote both the access and the incentives sides of the tradeoff envisioned by the conventional approach.[221]

---

[219] *See supra* Part I.C.4.

[220] *See supra* note 115 and accompanying text.

[221] Sole-source data may present an exception to the free entry assumption implicit in the impure public goods approach. The existence of such an exception does not justify denying copy protection to data that are freely available.

E. Digital Rights Management

Shifting to an impure public goods perspective would also shed new light on the debate about DRM.[222] First and foremost, an impure public goods approach recasts the range of quantities over which price discrimination is relevant. Under the conventional approach, the primary purpose of price discrimination is to prevent the exclusion of low-value users (represented in Figure 1 by the difference between Q_{mon} or Q_{sus} and Q_{eff}). The more fundamental, impure public goods approach that I propose suggests that price discrimination is relevant over the entire range of output.[223] Indeed, it suggests that price discrimination can promote efficient levels of provision even if total output decreases.

The impure public goods approach should also effectively eliminate concerns that DRM will enhance authors' ability to earn supracompetitive returns. As noted earlier, competition from close substitutes should effectively dissipate any abnormal profits.[224] In the process, the impure public goods approach reveals an alternative method for promoting access to copyrightable works. So long as the economy is sufficiently large, competition from close substitutes should drive prices close to marginal cost.

The only basis for caution is the possibility that markets for impure public goods might reach equilibrium with excess entry. Indeed, some degree of imperfection in the ability to price discriminate may be necessary to compensate for the impetus toward excess entry provided by demand diversion. Although the empirical record does not permit a definitive resolution of this issue, there seems little reason to adopt a default hostility toward DRM. The institutional

[222] *See supra* Part I.C.5.
[223] *See supra* Part II.C.3.
[224] *See supra* notes 153, 170, 187 and accompanying text.

considerations discussed above favoring a copyright that is large, intense, and narrow would militate in favor of facilitating price discrimination through the use of DRM.

F. Derivative Uses

The issues surrounding derivative uses are somewhat more complex.[225] The Supreme Court has drawn a distinction between derivative uses that are "superseding" and those that are "transformative." Superseding uses are those that simply displace the original work. As such, they are thought to be more likely to compete directly with the original work on which they are based. Transformative uses combine the existing work with other creative elements to create a new work. Derivative uses that are transformative are often thought to be less likely to compete with the original.[226] In addition, protecting transformative uses is often regarded as being more

---

[225] *See supra* Part I.C.6.

[226] In the words of the Court, when discussing the first statutory fair use factor, which focuses on "the purpose and character of the use," 17 U.S.C. § 107(1),

> [t]he central purpose of this investigation is to see, in Justice Story's words, whether the new work merely "supersede[s] the objects" of the original creation, or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message; it asks, in other words, whether and to what extent the new work is "transformative."

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994) (citations omitted) (quoting Folsom v. Marsh, 9 F. Cas. 342, 348 (C.C.D. Mass. 1841) (No. 4,901) and Leval, *supra* note 50, at 1111); *see also* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 562 (1985) (inquiring whether a derivative use had the intended purpose of "supplanting" the original).

Similarly, the third statutory fair use factor, which asks about "the amount and substantiality of the portion used in relation to the copyrighted work as a whole," 17 U.S.C. § 107(3), is taken as a proxy for whether the derivative work is likely to serve as a replacement for the original. *See Campbell*, 510 U.S. at 587-88 ("[A] work composed primarily of an original, particularly its heart, with little added or changed, is more likely to be a merely superseding use, fulfilling demand for the original.").

Finally, in evaluating the fourth statutory fair use factor, which focuses on "the effect of the use upon the potential market for or value of the copyrighted work," 17 U.S.C. § 107(4), the Court concluded that

> when a commercial use amounts to a mere duplication of the entirety of an original, it clearly "supersede[s] the objects" of the original and serves as a market replacement for it. But when, on the contrary, the second use is transformative, market substitution is at least less certain, and market harm may not be so readily inferred.

consistent with the goals of copyright, since such uses necessarily involve additional creativity.[227] Narrowing the derivative use right with respect to transformative works would arguably foster new creativity while having less of an adverse impact on the incentives to create the original work.

In recent years, it has become increasingly clear that the world is not as simple as the distinction between superseding and transformative uses might lead one to believe. As an initial matter, the fact that a transformative use is a work of new authorship does not necessarily mean that it does not compete with the original work.[228] In addition, the courts have increasingly recognized that regardless of whether a derivative use directly interferes with the market for the original, markets for derivative uses can be important in and of themselves.[229] For example, in *Campbell v. Acuff-Rose Music, Inc.*, the Court presciently recognized the importance of a distinct market in rap derivatives.[230] Two courts of appeals have similarly recognized the emergence of a market for licensing photocopies of scholarly journals.[231] Although courts have struggled with

---

Campbell, 510 U.S. at 591 (citations omitted) (quoting *Folsom*, 9 F. Cas. at 348); *see also* Princeton Univ. Press v. Mich. Document Servs., Inc., 99 F.3d 1381, 1386 (6th Cir. 1996) (en banc) (noting that the presumption of harm to the market for the original "disappears entirely where the challenged use is one that transforms the original work into a new artistic creation").

[227] *See Campbell*, 510 U.S. at 579 ("Although such transformative use is not absolutely necessary for a finding of fair use, the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works." (citation and footnote omitted)); Am. Geophysical Union v. Texaco Inc., 60 F.3d 913, 923 (2d Cir. 1995) (noting that to the extent a secondary use is not transformative, it adds nothing to the advancement of the arts and sciences).

[228] *See* Paul Goldstein, *Derivate Rights and Derivative Works in Copyright*, 30 J. COPYRIGHT SOC'Y U.S.A. 209, 217 (1983) (noting the possibility of market competition between an original novel and a transformative use with overlapping expressive content).

[229] *See, e.g.*, *Harper & Row*, 471 U.S. at 568 (noting that copyright "must take account not only of harm to the original but also of harm to the market for derivative works").

[230] 510 U.S. at 593.

[231] *Am. Geophysical Union*, 60 F.3d at 930; *Princeton Univ. Press*, 99 F.3d at 1387-88. Interestingly, the *American Geophysical Union* court noted the absence of a market for individual journal articles. 60 F.3d at 927. The emergence of JSTOR, Science Direct, HeinOnline, and individual article sales through Amazon.com suggests that such markets are beginning to appear as well.

how far they should go in protecting markets for potential derivative uses,[232] they have increasingly recognized that including derivative uses within the copyright protection afforded to an initial work can have a significant effect on the incentives for that work's creation.

From the standpoint of the conventional approach, these developments have once again sharpened the tradeoff between access and incentives. On the one hand, giving authors a broader derivative use right increases the surplus captured by the author of the initial work, which in turn provides greater incentives to create copyrightable works. On the other hand, a broader derivative use right limits the ability of follow-on authors to create new works. The conventional approach would attempt to balance these considerations by permitting follow-on works that include a significant contribution of additional creativity to fall outside the derivative use right.[233]

Shifting to an impure public goods perspective would counsel against such an outcome. Strengthening the derivative use right would increase the incentive to create copyrightable works. The greater incentive for entry would increase the number of close substitutes with which each work competes. The increase in competition would in turn foster the ability of follow-on authors to obtain access to the original work. Under these circumstances, the original work would serve both as an input into another product and as an end product in its own right. The literature on transfer pricing indicates that so long as all of the relevant markets are

---

[232] As the *American Geophysical Union* court noted, "were a court automatically to conclude in every case that potential licensing revenues were impermissibly impaired simply because the secondary user did not pay a fee for the right to engage in the use, the fourth fair use factor would *always* favor the copyright holder." 60 F.3d at 929 n.17. As a result, courts generally have placed some limits on the potential derivative markets that fall within the scope of the fourth statutory fair use factor. *See, e.g.*, *Campbell*, 510 U.S. at 592 (limiting consideration to markets for potential derivative uses "that creators of original works would in general develop or license others to develop"); *Am. Geophysical Union*, 60 F.3d at 930 (limiting consideration to markets for potential derivative uses that are "traditional, reasonable, or likely to be developed").

[233] *See supra* Part I.C.5.

sufficiently competitive, when a good constitutes both an end product and an input into another product, revenue and economic welfare are maximized if the producer charges the same price regardless of whether the good is sold as one or as the other.[234]

This suggests that, rather than promoting access by follow-on authors directly by decreasing the scope of the derivative use right, it is possible to accomplish the same goals indirectly by promoting entry and allowing the ensuing increase in price competition to increase follow-on authors' ability to obtain access to the original work. Again, at some point, the derivative use right may become so strong that it eventually induces excess entry. Unlike under the conventional approach, such market failure is not endemic. Moreover, any welfare losses from excess entry would be offset by the welfare gains from the increase in price competition, including those created by the additional follow-on expression made possible by the drop in price that follow-on authors must pay to obtain access to the original work.

CONCLUSION

The conventional approach to the economics of copyright has created a key misunderstanding about the relevance of public good economics. Framing the issues in terms of nonexcludability and zero marginal cost overshadows the true challenge posed by nonrivalry in consumption, which is the difficulty in getting consumers to reveal their true preferences implicit in the Samuelson condition.

The conventional approach also obscures the relevance of the theory of impure public goods for copyright policy. Framing nonrivalry in terms of zero marginal cost causes impure public goods theory to appear relevant only when the purchase of an additional unit of a public

[234] PAUL MILGROM & JOHN ROBERTS, ECONOMICS, ORGANIZATION AND MANAGEMENT 79-83 (1992).

good increases the costs borne by those who have already purchased that good, an assumption that does not seem to hold for most copyrighted works. The more fundamental approach to public good economics that I have proposed reveals how introducing variations in quality can turn a pure public good into an impure public good just as effectively as can variations in congestion cost. Although this Article focuses primarily on one source of variation in quality—product differentiation along a spectrum of product characteristics—one need not embrace any particular source of quality variation in order to appreciate the significance of this insight.

Thus, to the extent that public good economics has implications for copyright, it is through the theory of impure public goods rather than the theory of pure public goods. The key difference is that, in sharp contrast to pure public goods, markets for impure public goods do not exhibit a systematic tendency toward underproduction and underutilization. On the contrary, impure public goods are susceptible to efficient market production under a wide range of circumstances.

Saying that markets *can* provide impure public goods efficiently does not necessarily mean that they *will* do so in every circumstance. Unlike private goods, impure public goods lack an "invisible hand" that steers market outcomes toward optimality. Determining the best policy response thus depends upon a careful analysis of the underlying empirics and the possible institutional solutions. Even so, impure public goods equilibria have the advantage of not being bounded away from efficient outcomes. Under an impure public goods approach, copyright policy is no longer an exercise in second-best outcomes, but rather a more promising space in which near optimality may be a real possibility.