

2003

Why (And How) Fairness Matters at the IP/ Antitrust Interface

Daniel A. Farber

Brett H. McDonnell

Follow this and additional works at: <https://scholarship.law.umn.edu/mlr>

Part of the [Law Commons](#)

Recommended Citation

Farber, Daniel A. and McDonnell, Brett H., "Why (And How) Fairness Matters at the IP/Antitrust Interface" (2003). *Minnesota Law Review*. 780.

<https://scholarship.law.umn.edu/mlr/780>

This Article is brought to you for free and open access by the University of Minnesota Law School. It has been accepted for inclusion in Minnesota Law Review collection by an authorized administrator of the Scholarship Repository. For more information, please contact lenzx009@umn.edu.

Why (and How) Fairness Matters at the IP/Antitrust Interface

Daniel A. Farber[†] and Brett H. McDonnell^{††}

INTRODUCTION

This Article calls into question a widespread view about the goals of intellectual property and antitrust law.¹ Economists usually take for granted that economic efficiency should be the goal of the intellectual property and antitrust laws.² Legal scholars in these fields, many of whom have been heavily influenced by economics, often agree that the law should be crafted to maximize total surplus.³ We question this

[†] McKnight Presidential Professor of Public Law, University of Minnesota; Sho Sato Professor of Law, Boalt Hall School of Law, University of California, Berkeley.

^{††} Associate Professor of Law, University of Minnesota. We would like to thank Dan Gifford, Mark Lemley, David McGowan, Paul Rubin, and participants at the *Minnesota Law Review* Volume 87 Symposium for their helpful criticisms.

1. The antitrust/intellectual property interface has given rise to extensive, economically oriented literature. See, e.g., William F. Baxter, *Legal Restrictions on Exploitation of the Patent Monopoly: An Economic Analysis*, 76 YALE L.J. 267, 343 (1966); Louis Kaplow, *The Patent-Antitrust Intersection: A Reappraisal*, 97 HARV. L. REV. 1813, 1816-17 (1984).

2. In principle, efficiency should include allocative efficiency (immediate changes in total surplus in the relevant markets), productive efficiency (producing a given output with the lowest cost of input feasible), and dynamic efficiency (effects on growth, innovation, and firm organization). See Joseph F. Brodley, *The Economic Goals of Antitrust: Efficiency, Consumer Welfare, and Technological Progress*, 62 N.Y.U. L. REV. 1020, 1025 (1987); F.M. Scherer, *Antitrust, Efficiency, and Progress*, 62 N.Y.U. L. REV. 998, 998-1002 (1987); see also Richard J. Gilbert & Steven C. Sunshine, *Incorporating Dynamic Efficiency Concerns in Merger Analysis: The Use of Innovation Markets*, 63 ANTITRUST L.J. 569, 574-81 (1995).

3. See, e.g., ROBERT H. BORK, THE ANTITRUST PARADOX: A POLICY AT WAR WITH ITSELF 90-91, 405 (1978); HERBERT HOVENKAMP, FEDERAL ANTITRUST POLICY: THE LAW OF COMPETITION AND ITS PRACTICE 61-63 (2d ed. 1999) [hereinafter HOVENKAMP, FEDERAL ANTITRUST POLICY]; 1 HERBERT HOVENKAMP ET AL., IP AND ANTITRUST: AN ANALYSIS OF ANTITRUST

view and argue that the law should also encourage a fair division of the economic surplus, at least as a secondary goal.

In making this claim, we must confront two kinds of skepticism. The first is the skepticism of the economically uninitiated, who may wonder why anyone would write an article to prove such an obvious proposition. The second is the skepticism of the economically sophisticated, who may wonder why anyone would write an article to defend such a ridiculous proposition. Our discussion is pitched more to the latter group; we wish to make fairness a more respectable concept in law and economics circles and to show how it can usefully add to economic analysis.⁴

Part I presents a chain of increasingly specific arguments against exclusive reliance on economic efficiency. It begins with general criticisms of welfarism (the ultimate basis for the efficiency standard). Then it argues that welfarists themselves should also consider the distributive impacts of intellectual property rules. Finally, it contends that even if efficiency should be the global goal of antitrust and intellectual property law, this goal may not scale down to the operational level: Designing rules that maximize the total surplus from specific transactions may be infeasible or may even be counter-productive by stimulating wasteful rent-seeking.

One of the main arguments for economic efficiency is that it mimics the results that rational people would voluntarily agree to, *ex ante*. A very similar defense can be made of fairness, which we define in terms of game theory. Furthermore, given that a fair division of surplus is a strongly held social value, there is no principled reason why courts should ignore this value in interpreting antitrust and intellectual property laws.

Part II makes the case for fairness as an independent legal goal, but also argues that using fairness as a tiebreaker may, in the long run, promote economic efficiency. Even if society adopts efficiency as the overall goal of antitrust and intellectual

PRINCIPLES APPLIED TO INTELLECTUAL PROPERTY LAW 1-11 to 1-13 (2002) (quoting WARD BOWMAN, JR., *PATENT AND ANTITRUST LAW: A LEGAL AND ECONOMIC APPRAISAL* (1973)) [hereinafter HOVENKAMP ET AL., 1 IP AND ANTITRUST]; RICHARD A. POSNER, *ANTITRUST LAW: AN ECONOMIC PERSPECTIVE* 8 (1976).

4. This is a project we share with another participant in this Symposium. See Michael J. Meurer, *Fair Division*, 47 *BUFF. L. REV.* 937, 937-44 (1999).

property law, a presumption against one-sided divisions of surplus can still be defended. Where there are increasing returns to scale, potential producers and customers would agree *ex ante* to a fair division of surplus. Such an agreement would make it easier for the producer to gain a critical toehold in the market, foster expansion, and allow consumers to receive more benefits from economies of scale. Part II also argues that tilting the division of surplus toward producers is likely to cause them to engage in rent-seeking. Thus, a meta-rule in favor of fairness when direct evidence of efficiency is ambiguous may itself be efficient.

Part III is a preliminary exploration of how this meta-rule would affect the analysis of intellectual property issues. The first conclusion is that there should be a legal presumption in favor of open standards except where efficiency concerns clearly dictate otherwise. The second conclusion is that the law should disfavor price discrimination and similar conduct by rights holders, again with the qualification that efficiency concerns may override this presumption.

This Article is very much a work in progress. Much remains to be done in clarifying the meaning of fairness; the solution we borrow from game theory is appealing, but certainly open to dispute. Our applications of the theory are particularly tentative, given the complexity of the issues at the intellectual property/antitrust interface. But we hope that our arguments will at least put fairness on the intellectual agenda of economically oriented analysts of law.

I. REJECTING "EFFICIENCY" AS AN EXCLUSIVE CRITERION

In the study of both antitrust and intellectual property public policy, scholars with a law and economics bent most often assume that selecting the most efficient legal rule is the proper goal.⁵ Though "efficiency" has a number of possible meanings, it is most often characterized as maximizing the total wealth or surplus predicted to be created by the alternative rules being considered.⁶ In this Part, we argue that

5. See *supra* text accompanying notes 2-3.

6. See RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 14-15 (5th ed. 1998). Again, if one thinks of long-run behavior, this can be broken down into allocative efficiency, productive efficiency, and dynamic efficiency. See *supra* note 2.

the standard arguments made in favor of this assumed goal are far from indisputable. In section A, we confront the leading theoretical—economic and philosophical—arguments for the standard position and find them wanting. In section B, we confront the standard legal arguments and find them wanting as well. In section C, we argue that even to the extent that wealth or surplus maximization is accepted as a goal, that criterion often does not give clear directions as to how to choose among possible alternatives, so that some other criterion is needed at least as a tiebreaker. In all three sections, we give some arguments for fairness as an alternative criterion that deserves some independent consideration.⁷

A. THEORETICAL ARGUMENTS FOR EFFICIENCY ARE NOT INDISPUTABLE

One arrives at maximizing wealth or total surplus as the exclusive goal of the law only at the end of a long chain of reasoning. Each link in the chain raises difficult questions. In this section, we examine some of those questions. We start at the highest level of generality and work our way down. We start with welfarism generally. Maximizing total surplus makes sense as an exclusive objective only within a welfarist approach. Such an approach can be, and has been, debated at great length.⁸ We just pause on this point long enough to note that welfarism can be denied, suggesting a role for a variety of non-economic values. For the rest of this section we work within a welfarist perspective. Even within welfarism, though, one need not focus only on efficiency—one may well also care about how increases or decreases in welfare are distributed. We next look at a variety of justifications for focusing on efficiency and ignoring distributive effects in making most policy choices. We note important counterarguments that have recently been made on this point. Figuring out the distributive

7. In the next Part we will go into more detail as to the meaning of fairness and give further arguments in its favor. The final Part on applications will discuss how to operationalize fairness in the area of antitrust and intellectual property. For now we just state that by advocating fairness as a goal we are suggesting that policy makers and judges should give some degree of independent consideration to maximizing consumer surplus as opposed to total surplus.

8. See generally, e.g., R.M. HARE, *MORAL THINKING: ITS LEVELS, METHOD AND POINT* (1981); JOHN RAWLS, *A THEORY OF JUSTICE* (rev. ed. 1999); J.J.C. SMART & BERNARD WILLIAMS, *UTILITARIANISM: FOR AND AGAINST* (1973).

effects of various antitrust and intellectual property policies can be quite complicated. For the most part we focus on the distinction between consumer surplus and producer surplus, although we do quickly note tough questions as to how producer surplus may be distributed among producers and consumer surplus distributed among consumers.

Even if one takes efficiency in the sense of maximizing overall social welfare as one's policy goal, that does not necessarily lead to an exclusive focus on maximizing total surplus created within the relevant product market affected by a legal rule. First, there may be major general equilibrium effects generated outside the specific market under consideration. Because of such effects, a policy that maximizes surplus generated within one market may not maximize overall surplus. Second, even focusing on the specific market in question, there may be problems with attempting to maximize total surplus. Total surplus equals producer surplus plus consumer surplus. Consider a policy that tends to reduce consumer surplus but increase producer surplus more. On its face, such a policy would seem to increase total surplus. That may not be right, however. Producer surplus may tend to generate more unproductive rent-seeking behavior among those who would like to capture that profit, causing the surplus to be dissipated. Consumer surplus typically generates less such rent-seeking. If this rent-seeking effect is strong enough, then the increase in producer surplus at the expense of consumer surplus may not be worth it.⁹

1. Arguments for Welfarism

Maximizing wealth as an exclusive goal would seem to make sense only within a welfarist perspective. Welfarism maintains that social policies should be evaluated only according to their effects on human welfare.¹⁰ Maximizing wealth is a welfarist goal, although it is far from the only welfarist goal. Thus, if one does not accept welfarism, one should not accept maximizing wealth as an exclusive policy goal. The converse is not true: One can accept welfarism

9. One could argue that one could take maximizing efficiency as an exclusive goal but take rent-seeking into account. If rent-seeking is pervasive, though, it may make sense as an economizing device to deemphasize producer surplus relative to consumer surplus. See *infra* Part I.A.5.

10. See Tim Chappell & Roger Crisp, *Utilitarianism*, in 9 ROUTLEDGE ENCYCLOPEDIA OF PHILOSOPHY 551, 552 (Edward Craig ed., 1998).

without accepting maximizing wealth as an exclusive goal. Thus, a threshold question is whether welfarism is the best approach to evaluating public policy.

Philosophers have argued for centuries over the merits of welfarism generally and its most prominent variant, utilitarianism, in particular.¹¹ Modern welfare economists do deal with related questions. Many of them have actually been skeptical about utilitarianism because of doubts about the possibility of making interpersonal comparisons of utility.¹² Instead, economists have generally relied mainly on Pareto efficiency,¹³ a weak welfarist criterion, in making policy evaluations.¹⁴ It is difficult to rank alternatives according to Pareto efficiency in most real world applications, however. In its place economists tend to use Kaldor-Hicks efficiency, a stronger but more controversial measure.¹⁵

Several decades ago, Richard Posner made a sustained argument that legal analysts should focus exclusively on wealth maximization.¹⁶ He got his head handed to him on a plate by a variety of critics, most notably Ronald Dworkin.¹⁷ Recently, Louis Kaplow and Steven Shavell have argued for welfarism as the exclusive proper normative approach to legal analysis.¹⁸ It is too soon to know whether their heads will join Posner's on the plate, but we believe that their arguments are

11. Welfarism maintains that policies should be evaluated considering only their effects on human welfare or well being. Utilitarianism maintains that the goal of policy should be to maximize the total or average level of utility generated. Utilitarianism is a variant of welfarism, but other variants of utilitarianism disagree with utilitarianism over distributive issues. *See id.* at 552-53.

12. A key early work is LIONEL ROBBINS, *AN ESSAY ON THE NATURE AND SIGNIFICANCE OF ECONOMIC SCIENCE* (2d ed. 1935).

13. An economic distribution is Pareto efficient if there is no way to make at least one person better off without making anyone worse off. HAL R. VARIAN, *MICROECONOMIC ANALYSIS* 198 (2d ed. 1984).

14. *See, e.g., id.* at 198-206.

15. *See, e.g.,* ROBERT COOTER & THOMAS ULEN, *LAW AND ECONOMICS* 43-44 (3d ed. 2000); POSNER, *supra* note 6, at 14-15.

16. *See* Richard A. Posner, *The Ethical and Political Basis of the Efficiency Norm in Common Law Adjudication*, 8 *HOFSTRA L. REV.* 487, 487-88 (1980).

17. For an overview of the debate, see NEIL DUXBURY, *PATTERNS OF AMERICAN JURISPRUDENCE* 397-406 (1995). For Dworkin's leading contribution, see Ronald M. Dworkin, *Is Wealth a Value?*, 9 *J. LEGAL. STUD.* 191, 191-94 (1980).

18. LOUIS KAPLOW & STEVEN SHAVELL, *FAIRNESS VERSUS WELFARE* 3-4 (2002).

also unsuccessful for reasons that we have identified elsewhere.¹⁹ We simply assert two conclusions here. First, it is very hard to make convincing, non-tautological arguments for fundamental philosophical positions such as welfarism, and Kaplow and Shavell have not succeeded in doing so. Second, Kaplow and Shavell do not show us how to choose among competing welfarist positions, and it would appear that non-welfarist arguments are needed to make such choices.

What is the significance of these rather abstract debates for discussion of the interface between antitrust and intellectual property? Lawyers, judges, and scholars can and do rely on a variety of non-welfarist values in this area. Liberty and a variety of notions of fairness are frequently invoked.²⁰ Although these values can be given welfarist justifications, they often have non-welfarist roots, and many will want to follow such values even when their welfarist justifications appear to give out. Our take on the philosophical debate is that there are no good knockdown arguments against relying on such non-welfarist values.

2. Efficiency and Distributive Fairness

For the remainder of this section we take welfarism as a given. Acceptance of welfarism does not by any means lead to acceptance of wealth maximization as a goal. Most importantly, one may very well care not only about the total amount of surplus or wealth created, but also about how it is distributed. An exclusive goal of wealth or surplus maximization ignores distributive concerns. One must justify doing so.

The leading traditional justification is the second fundamental theorem of welfare economics. This theorem states that, given certain assumptions, any Pareto efficient distribution can be achieved by first distributing initial endowments of wealth by a system of lump-sum taxes and

19. See Daniel A. Farber, *What (If Anything) Can Economics Say About Equity?*, 102 MICH. L. REV. (forthcoming 2003) (manuscript at 2-5, on file with authors); Brett H. McDonnell, *The Economists' New Arguments* (Feb. 19, 2003) (unpublished manuscript, on file with authors).

20. See, e.g., *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 560 (1985) (explaining that fairness determinations with respect to "verbatim copying from a public figure's manuscript" requires the application of "traditional equities of fair use"); *Motion Picture Patents Co. v. Universal Film Mfg. Co.*, 243 U.S. 502, 519 (1917) (explaining that public interest "is more a favorite of the law than is the promotion of private fortune").

transfers, and then allowing production and trade to occur so that a competitive equilibrium is reached.²¹ The general strategy that this suggests is to use taxes and transfers to achieve the distribution one wants, and then let the market work. All policies other than tax and transfers should be focused on achieving an efficient market.

The well-known problem with the second theorem is that lump-sum taxes and transfers are not achievable in the real world. Any actual system of taxing and transferring will affect the behavioral incentives of some persons, and hence lead to some degree of inefficiency.²² This reopens the question as to whether other policy instruments should be used to promote distributive equity. The work of Kaplow and Shavell, however, suggests that even in the imperfect real world we should only use tax and transfer policy to achieve our distributive goals.²³ They argue that a broad wealth or income tax policy is a less inefficient way of redistributing than any other policy, because it is hard for persons to evade a broad tax policy, and hence such a policy is likely to have a weaker effect on behavior than other policies that achieve a similar distributive effect.²⁴ Moreover, using non-tax rules for distributive purposes has a double-distortionary effect. First, setting an antitrust rule that differs from the efficient rule has a direct impact on efficiency through its effect on behavior covered by that rule. Second, there is an indirect effect. In considering how much to work and earn, there will be less incentive to earn more because doing so will make one a target for the disadvantageous antitrust rule.²⁵ Hence, it is (second) best to use only tax policy to distribute wealth—tax policy will have some efficiency cost, but a lower cost than any other policy.

That argument is the current state of the art in law and economics.²⁶ Chris Sanchirico, however, has recently launched a powerful challenge.²⁷ Sanchirico uses theory of the second

21. See HAL R. VARIAN, *INTERMEDIATE MICROECONOMICS: A MODERN APPROACH* 497-99 (2d ed. 1990).

22. See Chris William Sanchirico, *Deconstructing the New Efficiency Rationale*, 86 CORNELL L. REV. 1003, 1018-31 (2001).

23. See Louis Kaplow & Steven Shavell, *Why the Legal System Is Less Efficient than the Income Tax in Redistributing Income*, 23 J. LEGAL STUD. 667, 667-69, 677 (1994).

24. *Id.*

25. See *id.*

26. See COOTER & ULEN, *supra* note 15, at 111-12.

27. See Sanchirico, *supra* note 22, 1069-70 (concluding that equity should

best arguments against Kaplow and Shavell. Suppose one starts from a state where one has used only tax policy to distribute wealth, and has used all other available policy instruments simply to promote efficiency, ignoring their distributive effects. Now, consider the following change. Relax the tax policy a bit to achieve a bit less redistribution, but also to reduce the inefficiency of the tax policy. The marginal decrease in inefficiency will be notable, as that was the only instrument used for redistribution, and hence its marginal efficiency cost is probably fairly large. Now, use some other policy instruments to gain back the lost redistribution. The marginal effect on efficiency of those changes will be zero, as all of the other policy instruments have been set to maximize efficiency alone, and hence the marginal effect of a small policy change on efficiency must be set at zero. Thus, one should be able to use the other policy instruments to achieve the same amount of redistribution with a lower efficiency cost than if one used only tax policy to redistribute.²⁸

Like all arguments, Sanchirico's needs some assumptions to work. There is an ongoing debate as to the likely practical effects of his argument. Kaplow and Shavell claim the practical effects are limited;²⁹ Sanchirico forcefully denies this.³⁰ Ultimately it is an empirical question. Empirical questions within the theory of the second best are notoriously difficult to answer.³¹ Our sense is that Sanchirico gets the best of this debate so far, though only time will tell. Of course, the answers may be different for different jurisdictions. For instance, tax and transfer policies are less developed in the European Union; hence other areas of policy in the European Union may be legitimately more focused on distributive effects than is the case with U.S. policy.³²

There are other reasons that redistribution through tax

be a "criterion for evaluating legal rules").

28. See *id.* at 1022-23.

29. See Louis Kaplow & Steven Shavell, *Should Legal Rules Favor the Poor? Clarifying the Role of Legal Rules and the Income Tax in Redistributing Income*, 29 J. LEGAL STUD. 821, 832-34 (2000).

30. See Sanchirico, *supra* note 22, at 1057-64.

31. For an overview of the theory of the second best, see Richard S. Markovits, *Second-Best Theory and Law & Economics: An Introduction*, 73 CHI.-KENT L. REV. 3, 3-10 (1998).

32. Brett H. McDonnell & Daniel A. Farber, *Are Efficient Antitrust Rules Always Optimal?*, 48 ANTITRUST BULL. (forthcoming 2003) (manuscript at 1-2, 15-18, on file with authors).

and transfer policies may not work well. They may not be politically feasible.³³ Monetary transfers may get diverted to the wrong persons due to corruption.³⁴ One may care about characteristics other than income and wealth.³⁵

The implications for antitrust and intellectual property policy are that one cannot dismiss distributive effects as a policy concern simply by saying that taxation will answer any and all distributive concerns.³⁶ That does not mean, however, that there may not be arguments more particular to antitrust or intellectual property that dictate ignoring distributive effects. We turn now to such possible concerns.

3. Distribution in Antitrust and Intellectual Property

Economic analysts of antitrust often assume that the appropriate goal of policy is to maximize total surplus created within the product market under consideration.³⁷ Total surplus equals producer surplus plus consumer surplus. Producer surplus is the difference between how much revenue producers receive and how much it cost them to produce the good or service (i.e., it basically is profit).³⁸ Consumer surplus is the difference between how much consumers pay for the good or service and how much they would have been willing to pay for it.

The core distributive argument for favoring consumer surplus over producer surplus is that on average the recipients of consumer surplus are poorer than the recipients of producer surplus, and hence moving surplus from producers to

33. See McDonnell, *supra* note 19, at 23-24; Meurer, *supra* note 4, at 970 n.117.

34. See Gilles Saint-Paul, *Are Intellectual Property Rights Unfair?*, IZA Discussion Paper No. 639, 20 at http://ssrn.com/abstract_id=358285 (last visited Feb. 24, 2003).

35. See Meurer, *supra* note 4, at 970 n.117.

36. Saint-Paul, *supra* note 34, at 4-9, has formally modeled the trade-off between weakened intellectual property rights and tax/transfer policies as ways to achieve redistribution. He finds that tax and transfer policies have less distortionary effects on the economy than weakened intellectual property rights. *Id.* at 20. It is unclear, however, how robust those results are to particular modeling forms. Furthermore, the model does not account for concerns expressed *supra* notes 27-35 and *infra* Part I.A.4-5.

37. See *supra* text accompanying notes 2-3.

38. With the caveat that the cost of production includes a normal return for capital and entrepreneurial effort, which return is generally included in accounting, but not economic, definitions of "profit."

consumers improves distributive equity.³⁹ Of course, this is at most true only on average. Some recipients of consumer surplus are richer than some recipients of producer surplus. Thus, an important empirical question becomes how strong is the correlation between consumer surplus and lower wealth. The stronger the correlation, the stronger the distributive argument for favoring consumer surplus. Also, the argument applies much more to markets for final consumer products. For intermediate products, where the consumers are other companies using the product as an input for their own production, this redistributive concern is unlikely to apply.

We thus need to look just a bit more closely at consumer surplus and producer surplus. Consider the former first. Less wealthy people will tend to consume a higher percentage of their income than more wealthy people.⁴⁰ Thus, a tax on consumption will tend to be more regressive than a tax on income, which in turn is more regressive than a tax on profits. Of course, in many particular industries consumers will be wealthier than the national average—the luxury yacht industry, for instance. But on average the correlation does hold, with moderate strength.⁴¹

Producer surplus is divided among a variety of parties, the most important of which are employees and the providers of capital. In the U.S., employees receive about two-thirds of industrial revenue, and providers of capital receive about one-third.⁴² Income and wealth distribution among employees probably resembles distribution among consumers, though this naturally varies among industries. Providers of capital, though, tend to be skewed to the wealthy. Even in the U.S.,

39. This is true because less wealthy people tend to consume a greater fraction of their income. See Robin Cooper Feldman, *Consumption Taxes and the Theory of General and Individual Taxation*, 21 VA. TAX REV. 293, 329 n.85 (2002).

40. See *id.*

41. See *id.* In a more detailed analysis one will sometimes have to be more careful about how different practices may affect the distribution of surplus among consumers—we shall see an example of this in our discussion of price discrimination, *infra* Part III.B. For now, though, we shall pass over that point.

42. In 2001, net national income was \$8.122 trillion. Bureau of Economic Analysis, U.S. Department of Commerce, *The National Income and Product Accounts Tables tbl. 1.14* (2001), <http://www.bea.gov/bea/dn/nipaweb/TableFixed.asp>. Of that, 5.875 trillion, or 72%, was compensation of employees, while the rest (proprietors' income, rental income, corporate profits, and net interest) went to providers of capital. See *id.*

where about half of the population owns stock⁴³ (an extremely high proportion by international and historical standards), this skew is significant. It is basically the wealthy half of the population that owns shares, and among that half, share ownership is still heavily concentrated among the wealthier.⁴⁴ Bond ownership is even more heavily skewed towards the wealthy.⁴⁵

Thus, the correlation between wealth and consumer surplus versus producer surplus definitely exists, although it is imperfect. How attractive an instrument antitrust policy is for achieving redistribution depends in part on how inefficient tax policy is. This is an open question, but there is room for at least some concern over distributive fairness as a proper part of making antitrust policy.⁴⁶

The same sorts of arguments apply to intellectual property as well. Intellectual property policies have predictable effects on the distribution between producer and consumer surplus. Indeed, such effects have been an object of study and concern among intellectual property scholars.⁴⁷ Thus, distributive fairness arguments are relevant for intellectual property as well. Most holders of valuable intellectual property rights today are large institutions, not starving artists. A wrinkle in the area of intellectual property is that long-run effects become of more importance; the long-run encouraging of innovation being the core purpose of intellectual property law.⁴⁸

43. See NYSE, *Shareownership 2000*, at 16, <http://www.nyse.com/market/info/shareownersurvey.html> (last visited Feb. 24, 2003) (interpreting research to imply that "51.8 percent of all individuals who are either household heads, or spouses of household heads, own stock").

44. See *id.* at 25, tbl. 8 (showing that 35.3% of all shares are owned by families with income over \$250,000, and 63.0% of all shares are owned by families with income over \$100,000).

45. See E-mail from Shawgi Tell, Niagara County Community College (May 27, 1998) ("The top 0.5% of wealth holders still own 32% of stocks—double the 16% share held by the bottom 90%. Bond ownership is even more top-heavy. The top 0.5% holds 46% of the total, while the bottom 90% holds just 10%." (quoting an unpublished Federal Reserve paper)), at <http://www.fullerton.edu/sociology/orleans/incomedist.htm> (last visited Feb. 24, 2003) (on file with the authors).

46. See *infra* Part I.B.1.

47. IP law functions by allowing monopolistic practices, thereby reducing consumer surplus, in order to increase profit to induce innovation. See, e.g., Kaplow, *supra* note 1, at 1821.

48. See *infra* Part I.A.5.

4. Efficiency and Partial Versus General Equilibrium Analysis

Suppose one is not convinced by our arguments so far, and is committed to a consequentialist analysis that addresses distributive concerns only through tax and transfer policy. Surely then, antitrust and intellectual property policy should focus on maximizing total surplus within the product markets being studied. That is, however, not necessarily the case. Even if one accepts maximizing wealth or total surplus as a policy goal, reasons may remain for giving greater weight to consumer surplus than producer surplus. One such set of reasons relates to partial versus general equilibrium analysis. A second set of reasons relates to rent-seeking. We explore the former in this section, and the latter in the next section. In some sense, the arguments within these two sections are thus already included within an efficiency goal, broadly understood. Most scholarly and judicial analyses of efficiency in antitrust, however, focus more narrowly on an immediate analysis of total surplus created in the product market in question. For their part, intellectual property analyses consider long-run dynamic efficiency via incentives to innovate.

Traditional economic analysis of both antitrust and intellectual property concentrates on the effects that the law has on the particular product markets being regulated. The law may have effects that go beyond the immediate increase or decrease of consumers and producer surplus in those product markets, however. Among these broader effects of antitrust law,⁴⁹ at least two are relevant to the relative weight we should give producers and consumer surplus.

First, producer surplus generated within a product market may be used to lobby politicians and gain favor.⁵⁰ Greater concentration among producers may make such lobbying easier, and hence increase its bad effects on society. This has a variety of implications for antitrust policy.⁵¹ In particular, it may be a reason for favoring consumer surplus, as producer surplus is much more likely to be used to lobby, since producer surplus recipients are typically much more concentrated and organized than recipients of consumer surplus.⁵²

49. See generally McDonnell & Farber, *supra* note 32 (manuscript at 2-23) (discussing the broad effects of antitrust law on political institutions, corporate governance, distributions of income, and risk-bearing).

50. See *id.* (manuscript at 3-4).

51. See *id.* (manuscript at 3-6).

52. See *id.* This is very closely related to the rent-seeking arguments that

Second, antitrust policy has important effects on disciplining agency misbehavior by persons within businesses.⁵³ This effect is potentially quite large,⁵⁴ but its implications for antitrust policy are still unclear and underexplored. This may add another argument for favoring consumer surplus. Large producer surplus may give rise to more opportunity for business insiders to misbehave, and hence give rise to greater agency costs. Agency costs among consumers, on the other hand, are typically non-existent or trivial.

Where there are multiple imperfect markets, there may be another general equilibrium effect of note. Allowing producers in one market to capture more surpluses will induce greater investment in that market. If producers in another market can capture less of the surplus generated in their market, investment may shift from that market, even if the net social benefit generated there is greater than in the former market.⁵⁵

5. Efficiency and Rent-Seeking

Finally (among the theoretical arguments for wealth or total surplus maximization as a goal), suppose one takes a consequentialist approach, handles distributive equity through tax policy, and ignores general equilibrium effects outside of the product market, either because the general equilibrium effects are believed to be small or because their effect is too radically uncertain. Surely then one is committed to maximizing total surplus created in the product market as the appropriate goal of antitrust and intellectual property law. That, too, is still not necessarily the case. The greater susceptibility of producer surplus to dissipation may still be a good reason to favor consumer surplus.

As noted above, producer surplus in the form of profit tends to be more concentrated among its recipients than does consumer surplus.⁵⁶ High potential profits will often induce

we make in the next section. *See infra* Part I.A.5.

53. *See* McDonnell & Farber, *supra* note 32 (manuscript at 8-11) (noting that the threat of market share loss in a competitive industry helps discourage firm managers from pursuing their own interests as opposed to those of the firm).

54. *See id.* (manuscript at 10).

55. *See* Glynn S. Lunney, Jr., Copyright and the Supposed Efficiency of First-Degree Price Discrimination (unpublished manuscript, on file with authors).

56. *See supra* notes 40-45 and accompanying text.

potential receivers of those profits to make large outlays in order to achieve those profits. For instance, firms may create high barriers to entry by making high outlays in infrastructure, which may allow them to credibly commit to lowering prices in the event of entry by a competitor. Firms may spend money on lobbying and other forms of political influence in order to create laws or subsidies that prevent entry. Firms may commit excessive amounts to research and development in order to beat other firms and receive a patent. Such expenditures will often exceed the social optimum, and indeed may often be large enough to counteract most of the social gain from the producer surplus.⁵⁷

This point has been long recognized in the area of antitrust. Much scholarly and recent judicial writing on antitrust is ambiguous as to whether the proper goal of antitrust is maximizing total or consumer surplus.⁵⁸ In his landmark book on antitrust, Richard Posner advocates maximizing consumer surplus, not total surplus, as the goal of antitrust law precisely because he believes that producer surplus will tend to be dissipated through rent-seeking.⁵⁹ There are some counterarguments to Posner's point. Some expenditures to achieve producer surplus may increase social welfare—for instance, large outlays on infrastructure to commit to lowering prices in the event of entry may also be socially beneficial. Thus, at this point it is still unsettled within antitrust scholarship whether total or consumer surplus maximization is the best policy goal.⁶⁰

The same basic point applies to intellectual property law and policy as well. There, however, the potential positive effects from competition to gain producer surplus loom even more. After all, the whole point of intellectual property law is to lure businesses into innovating through the promise of monopoly profits. Thus, totally ignoring producer surplus does seem more suspect when considering intellectual property law

57. See COOTER & ULEN, *supra* note 15, at 128-35.

58. See Daniel J. Gifford & Robert T. Kudrle, *European Union Competition Law and Policy: How Much Latitude for Convergence with the United States?*, 48 ANTITRUST BULL. (forthcoming 2003) (manuscript at 33, on file with authors); David McGowan, *Innovation, Uncertainty, and Stability in Antitrust Law*, 16 BERKELEY TECH. L.J. 729, 784-86 (2001).

59. See POSNER, *supra* note 3, at 11-12.

60. Our point for the moment is simply to establish that the arguments are unsettled. In Part II we will advance further arguments as to why consumer surplus should be at least somewhat favored over producer surplus.

than when considering antitrust.

Even within intellectual property law, however, the tension between producer and consumer surplus has been noted. First, there is a frequently discussed tension between dynamic efficiency and allocative efficiency. Dynamic efficiency requires some incentive to innovate, hence intellectual property's existence. On the other hand, intellectual property works by granting short-term potential monopoly power. The exercise of that power leads to allocative inefficiency, as the intellectual property right holder reduces output below the competitive level. Intellectual property law strives to balance these two effects, albeit with a greater emphasis on dynamic efficiency than is typical in antitrust law.⁶¹

Moreover, rent-seeking may occur within the context of innovation as well. For instance, two potential innovators may spend too much too soon to achieve an invention and be the first to patent. Thus, overly strong intellectual property rights may lead to overinvestment and innovation that occurs sub-optimally early. This is simply a specific instance of the more general problem of rent-seeking and producer surplus. Moreover, holders and potential holders of intellectual property rights may engage in other forms of rent-seeking, such as lobbying, to strengthen those rights.⁶² To the extent that overinvestment and lobbying by intellectual property rights holders are significant problems, then even within the context of intellectual property it makes sense to give greater weight to consumer surplus than producer surplus.

Of course, to the extent that rent-seeking does dissipate producer surplus, then one could say that a careful analysis based only on maximizing efficiency will take that into account—and that the argument in this section is really not an argument against efficiency as an exclusive goal. That is true, but the analysis may often be difficult to do on a case-by-case basis. If one believes that rent-seeking will often dissipate much producer surplus, then a policy of giving less weight to producer surplus than consumer surplus may make sense as a way of achieving a simpler and yet generally accurate analysis. That is particularly so as we only advocate considering fairness

61. See, e.g., 1 HOVENKAMP ET AL., IP AND ANTITRUST, *supra* note 3, at 1-9 to 1-13.

62. Many people believe that this helps explain recent legislative changes in intellectual property law. See *infra* note accompanying text.

in situations where efficiency gives no clear answer.⁶³

We have so far worked through many of the leading arguments in favor of the maximization of total surplus generated within the product market as the exclusive goal of antitrust and intellectual property law. It takes a chain of arguments to reach that point as the proper policy goal, and each link in the chain has significant weaknesses. As a matter of basic philosophical and economic theory, therefore, it is quite open to argue that other policy goals should be pursued as well, and in particular that courts should give greater weight to consumer surplus than producer surplus. We shall make more arguments in favor of that position in Part II. First, we turn from philosophy and economics to consider some leading legal arguments in favor of maximizing total surplus as a policy goal.

B. LEGAL ARGUMENTS FOR EFFICIENCY ARE NOT IRONCLAD

Various more traditional legal arguments have been promoted supporting efficiency in the sense of maximizing surplus as the proper goal for antitrust and intellectual property law. Some of those arguments arise from examining the legislative history of relevant laws. Others arise from judicial interpretations of the laws. We consider the legislative and judicial history of both antitrust law and intellectual property law in this section. We also consider the relationship of both sets of laws to rent-seeking behavior.

1. Antitrust

A look at the legislative history of United States antitrust law reveals that a hodgepodge of goals was cited by its creators and advocates. Robert Bork has famously argued that efficiency as a goal best explains this legislative history.⁶⁴ Many scholars have replied that this presents a badly oversimplified picture. Many other goals were importantly present. These include protecting small producers from large producers and protecting politics from the influence of big businesses.⁶⁵ Most importantly for our purposes, it has been argued that a leading purpose of the antitrust laws was to stop wealth transfers from consumers to producers—that is, to

63. See *infra* Part II.E.

64. See BORK, *supra* note 3, at 51-71.

65. See HOVENKAMP, *FEDERAL ANTITRUST POLICY*, *supra* note 3, at 47-51; McGowan, *supra* note 58, at 741-65.

protect consumer surplus.⁶⁶ Even among law and economics scholars, a more fashionable explanation of the legislative history than Bork's is that the antitrust laws represent rent-seeking by small producers protecting themselves against more efficient new large producers.⁶⁷

This scholarship on the legislative history of United States antitrust laws, thus, does not point to any one exclusive goal. Maximizing consumer surplus is thus not the exclusive goal. It does appear to have been an important goal, though.⁶⁸ Efficiency may also have been an important goal, but not an exclusive one by any means.⁶⁹ Indeed, efficiency does not appear to have even been the most important goal among those who created the antitrust laws.⁷⁰ Thus, to the extent that legislative history matters in determining how statutes should be interpreted,⁷¹ this suggests that maximizing efficiency should not be the exclusive goal of United States antitrust law.

The history of judicial interpretation of the antitrust laws is also mixed, and has varied over time. For decades, courts seemed to see promoting rivalry as the leading goal of antitrust law.⁷² They may also have seen protecting small producers as an important goal.⁷³ Beginning in about the mid-seventies, courts shifted to more of a focus on efficiency.⁷⁴ This came in response to the Chicago Revolution in antitrust scholarship. Courts using economic concepts have not clearly distinguished between consumer surplus and total surplus as a goal,

66. See Robert H. Lande, *Wealth Transfers as the Original and Primary Concern of Antitrust: The Efficiency Interpretation Challenged*, 50 HASTINGS L.J. 871, 873-76 (1999).

67. See THE CAUSES AND CONSEQUENCES OF ANTITRUST: THE PUBLIC-CHOICE PERSPECTIVE 5 (Fred S. McChesney & William F. Shughart II eds., 1995); Thomas J. DiLorenzo, *The Origins of Antitrust: An Interest-Group Perspective*, 5 INT'L REV. L. & ECON. 73, 81 (1985); George J. Stigler, *The Origin of the Sherman Act*, 14 J. LEGAL STUD. 1, 7 (1985).

68. See Lande, *supra* note 66, at 885-947.

69. See HOVENKAMP, FEDERAL ANTITRUST POLICY, *supra* note 3, at 47-51.

70. See *id.* at 48-49.

71. This itself is certainly a highly disputed point. See WILLIAM N. ESKRIDGE, JR. ET AL., CASES AND MATERIALS ON LEGISLATION: STATUTES AND THE CREATION OF PUBLIC POLICY 669-1098 (3d ed. 2001).

72. See Gifford & Kudrle, *supra* note 58 (manuscript at 12-14). A classic case exemplifying this goal is *Brown Shoe Co. v. United States*, 370 U.S. 294, 343-45 (1962).

73. See Gifford & Kudrle, *supra* note 58 (manuscript at 2).

74. *Id.* (manuscript at 14). Key decisions include *Continental T.V., Inc. v. GTE Sylvania Inc.*, 433 U.S. 36, 53-55 (1977) and *United States v. General Dynamics Corp.*, 415 U.S. 486, 504-10 (1974).

however.⁷⁵ This lack of a clear distinction is understandable. For one, in many cases both goals suggest the same legal answer. For another, the leading Chicago scholarship on point was ambiguous on this question; Bork spoke of consumer welfare, but appears to have meant total surplus;⁷⁶ Posner spoke of efficiency, but argued that most producer surplus will be dissipated, so that efficiency largely dictates maximizing consumer surplus.⁷⁷

Let us view this legal history from a slightly different perspective. Whatever one thinks about the relative weight of the various goals stated as justifications for antitrust laws, small producers organizing to protect themselves against new big businesses at the turn of the century were critical to those laws gaining political support.⁷⁸ Thus, rent-seeking behavior was an important part of the origins of the antitrust laws. Not long after those laws were written, however, the feared new big businesses did indeed come to dominate national politics. Nonetheless, courts continued to use the antitrust laws against those newly powerful businesses.⁷⁹ Sometimes courts interpreted the law as being designed to protect smaller businesses.⁸⁰ Other times courts interpreted the law as being aimed at protecting consumers or society as a whole via an efficiency norm.⁸¹ Either way, a law created by an old dominant economic and political class was used to police the economic power of a new set of institutional actors. Antitrust law had its origins in rent-seeking, but came to be interpreted as a tool for protecting the public against a powerful elite. It forms an interesting counterpoint with intellectual property law, as the next section shows.

2. Intellectual Property

In contrast with antitrust law, intellectual property law reflects one clear major underlying goal: promoting innovation by granting innovators some degree of power over others' use of

75. See McGowan, *supra* note 58, at 784-86.

76. See BORK, *supra* note 3, at 110 (noting that the "consumer welfare model" does not take transfers from consumers to producers into account).

77. See *supra* note 59 and accompanying text.

78. See sources cited *supra* note 65.

79. See, e.g., *United States v. Am. Tobacco Co.*, 221 U.S. 106, 155-77 (1911); *Standard Oil Co. v. United States*, 221 U.S. 1, 70-77 (1911).

80. See *supra* note 73 and accompanying text.

81. See *supra* notes 68-69 and accompanying text.

their ideas.⁸² What is uncertain, however, is how far the law should go in promoting that goal. The legislative and judicial history show that Congress and the courts have gone back and forth as to how strong and broad intellectual property rights should be.

The legislative history of intellectual property in the United States begins with the Patent and Copyright Clause, which provides that Congress shall have the power to “promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”⁸³ Copyright, patent, and trademark law have all undergone considerable judicial and legislative upheaval over the years in response to new technologies and new economic structures. The fairly explicit goal of legal innovation has been to appropriately protect innovators and give them incentives to create without going so far that consumers are overly hurt by the monopoly power created. Opinions differ as to how well the law has responded to the changes or if it has encouraged a proper balance between incentives to create and limits on monopoly power—some are relatively sanguine,⁸⁴ while others are less so.⁸⁵ Even some of the more sanguine, though, are worried that recently the legislative process has been increasingly dominated by large corporate intellectual property rights holders that have succeeded in promoting legislation that advances their interests but which is more dubious for the general interest. The Sonny Bono Copyright Term Extension Act of 1998⁸⁶ and the Digital Millennium Copyright Act (DMCA)⁸⁷ are frequently cited examples.⁸⁸

The judicial history of intellectual property laws shows courts inventing doctrines to strengthen or weaken the statutes in instances where judges seem to think the law has gone too far or not far enough. The doctrines of patent misuse and

82. See *infra* note 141 and accompanying text.

83. U.S. CONST. art. I, § 8, cl. 8.

84. See Robert P. Merges, *One Hundred Years of Solicitude: Intellectual Property Law, 1900-2000*, 88 CAL. L. REV. 2187, 2240 (2000).

85. See Jessica D. Litman, *Copyright, Compromise, and Legislative History*, 72 CORNELL L. REV. 857, 903-04 (1987).

86. Pub. L. No. 105-298, 112 Stat. 2827 (1998) (codified as amended in scattered sections of 17 U.S.C.).

87. Pub. L. No. 105-304, 112 Stat. 2860 (1998) (codified as amended in scattered sections of 17 U.S.C.).

88. See Merges, *supra* note 84, at 2233-39.

copyright misuse import antitrust concerns into intellectual property doctrine.⁸⁹ Sometimes the statutes allow aggressive holders of intellectual property rights to unduly monopolize markets, leading to lost social welfare because output is constricted. In certain circumstances, courts will invoke the misuse doctrine to limit such behavior.⁹⁰ At other times, courts invoke the doctrine of contributory infringement to stop persons from skirting the statutes and treading on the rights of intellectual property rights holders.⁹¹

It is interesting to examine this legal history from a rent-seeking perspective and compare it to the history of the antitrust laws. The antitrust laws started as rent-seeking by small producers, but morphed in the hands of the courts into public-spirited laws used to protect consumers against powerful producers.⁹² The intellectual property laws move in the opposite direction. They started as public-spirited laws to promote innovation without any powerful, well organized interest groups standing behind the laws. Over time, though, innovation industries became more powerful and better organized, and they pushed Congress for extensions to the laws that became harder to justify as public-regarding laws. Rent-seeking thus increasingly warped the intellectual property laws. The DMCA is a leading example of this process.⁹³ A key current question is how far courts will go in using antitrust and related doctrines such as misuse to limit the potential abuses made possible by these newer laws.⁹⁴

C. EFFICIENCY'S AMBIGUITY AND THE NEED FOR A TIEBREAKER

There is a final problem with efficiency as an exclusive criterion in making and interpreting antitrust and intellectual property law. In many situations, an economic analysis of efficiency does not give clear answers as to the best legal policy. Courts presented with a case cannot just not decide it because their preferred criterion does not tell them what to do under

89. See 1 HOVENKAMP ET AL., *IP AND ANTITRUST*, *supra* note 3, at 3-2.

90. See *id.* at 3-7 to 3-10.

91. See *id.* at 3-16 to 3-20.

92. See *supra* notes 64-81 and accompanying text.

93. See Pamela Samuelson, *Intellectual Property and the Digital Economy: Why the Anti-Circumvention Regulations Need to Be Revised*, 14 *BERKELEY TECH. L.J.* 519, 522-24 (1999).

94. See Dan L. Burk, *Anti-Circumvention Misuse*, 50 *UCLA L. REV.* (forthcoming 2003) (manuscript at 51-63, on file with the authors).

the circumstances. Thus, even if one disagrees with all of the above and believes that efficiency should be the exclusive criterion in making policy, one must still give guidance as to what to do in the many instances where efficiency gives no answer.

Efficiency's unclarity is becoming increasingly clear in antitrust scholarship. In the early years of the Chicago Revolution in antitrust there was optimism that economics would provide relatively definite guidance most of the time.⁹⁵ As people with training in a wider variety of economic approaches got involved in the fray, however, that optimism dissipated. Post-Chicago antitrust scholarship has shown that in many instances economic reasoning is ambiguous, and empirical work in economics is generally not strong enough to weigh opposing effects.⁹⁶ The increased use of game theory in industrial organization and antitrust scholarship has made policy implications particularly ambiguous. In game theory, even more than other areas of economics, conclusions are quite peculiarly sensitive to small differences in assumptions about the structure within which people are acting.⁹⁷

In intellectual property, the difficulty of economics in giving clear answers has been evident even to the most avid proponents of the Chicago school. Frank Easterbrook, for instance, has been quite insistent as to the poverty of our knowledge in this area.⁹⁸ In general, we know the basic thrust of the opposing effects that matter. We have very little knowledge, however, about the strength of those opposing effects or how they vary in different industries. Thus, even as basic a question as the optimal length of copyright or patent terms is incredibly hard to answer using an efficiency-based economic analysis.

So even if one is committed to a pure efficiency criterion, one needs a fallback criterion for those many instances where efficiency analysis yields no answer. In Part II, we argue for a notion of fairness tied to consumer surplus as such a

95. See, e.g., BORK, *supra* note 3, at 81-82.

96. See, e.g., Timothy J. Brennan, *Do Easy Cases Make Bad Law? Antitrust Innovations or Missed Opportunities in United States v. Microsoft*, 69 GEO. WASH. L. REV. 1042, 1090-93 (2001).

97. See *id.* at 1054-56.

98. See Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207; David McGowan, *Copyright Nonconsequentialism* (Mar. 7, 2003) (unpublished manuscript, on file with the *Minnesota Law Review*).

tiebreaker.

II. WHY FAIRNESS?

The question is whether any other norms besides efficiency should be used to assess intellectual property rules. No doubt, a number of norms are possible, some of them discarding welfarism entirely. Our proposal, however, is a relatively modest modification of the total surplus standard.⁹⁹ When the efficient rule can be clearly determined, we are willing to accept it despite our qualms about the efficiency standard. Those doubts are outweighed by the advantages of unifying intellectual property and antitrust law as much as possible. Relying on a single social goal, to the extent that society can do so, simplifies legal analysis and fosters a greater degree of coherence in the legal system. Arguments along these lines have been made for a pure efficiency standard.¹⁰⁰ Although we reject *exclusive* reliance on efficiency for the reasons discussed above, the arguments are sufficiently persuasive to support efficiency as the dominant normative standard. But antitrust and intellectual property law are inevitably an exercise in the theory of the second best, with the often frustrating result that the efficient rule cannot be determined with confidence.¹⁰¹ We propose as a tiebreaker that the presumption should be in favor of rules that fairly divide the transactional surplus between producers and consumers.

One of our goals is to overcome the instinctive response of the economically minded to the term “fairness,” which is likely to strike them as impossibly vague, non-rigorous, and disconnected from social welfare.¹⁰² In section A, we discuss the meaning of fairness and the reasons why economists should take the idea seriously. Sections B, C, and D provide arguments for our fairness presumption. Section B argues that the law should mimic the result of a fair bargaining process

99. In practice, we think that the evidence will often be too ambiguous to provide a clear answer about efficiency, so we expect our default fairness presumption to be operative in many cases.

100. See McGowan, *supra* note 58, at 776-86.

101. Antitrust law seeks to eliminate monopolies to create efficient product markets. Intellectual property law creates monopolies to foster innovation. There is no clear way to calculate the trade-offs between the two.

102. Here, as in other areas of the law, fairness may seem to be “devoid of principled content” to some analysts. See Ian B. Lee, *Fairness and Insider Trading*, 2002 COLUM. BUS. L. REV. 119, 121.

between the parties. Section C suggests that courts legitimately consult social norms in resolving statutory ambiguities. Empirical evidence leaves no doubt that fairness is a widespread norm in our society.¹⁰³ Section D argues that requiring a fair division of surplus serves valuable distributive goals and puts a damper on socially wasteful rent-seeking. Our aim in Part III is not to provide a general legal theory of fairness, let alone an irrefutable one. Instead, we hope to provide overlapping arguments from a variety of perspectives, which cumulatively make a persuasive case for the fairness presumption.

Part III closes by clarifying the nature and scope of our fairness presumption. In Part IV, we will explore how the presumption would apply to more specific problems on the boundary between antitrust and intellectual property law.

A. WHAT IS FAIRNESS?

In his famous book about justice,¹⁰⁴ John Rawls began by hypothesizing the "original position" in which the parties act behind a "veil of ignorance," deprived of knowledge of their own identities.¹⁰⁵ He defined as fair (and therefore just) whatever agreement they would reach under these circumstances.¹⁰⁶ Judging fairness by comparison with a hypothetical bargain is a useful technique. Rawls used his technique to determine the basic institutional structure that should govern a just society.¹⁰⁷ Our concern is more mundane. We are interested in how to divide surplus in the context of intellectual property. To get a sense of how to analyze this problem, we begin with a simple hypothetical.

Suppose that Firm sells software to a variety of consumers and is in a position to assess the demand of each consumer for the good. Mr. Consumer would be willing to pay up to \$100 for the product, and the cost of manufacturing the product is \$20. The Firm is a monopoly. What bargain should we expect between the Firm and Consumer? The surplus is \$80. How would rational parties divide it?

At first blush, one might assume that the Firm (being a

103. See *infra* notes 143-44 and accompanying text.

104. RAWLS, *supra* note 8.

105. *Id.* at 11, 118-23.

106. *Id.* at 10-15.

107. See *id.*

monopolist) can demand \$100, extracting the full surplus. This cannot be quite right, because at \$100 the consumer is completely indifferent and has no reason to accept the deal. But if the offered price is \$99, the consumer can only lose by turning down the deal. So the monopolist seizes almost all of the surplus. This is the Bad Monopolist story.

In fact, this story is too simplistic, as can be shown by turning the tables. Imagine that the consumer makes a "final offer" of \$21. If the Firm refuses the offer, it loses a dollar because it has lost a sale. While it is true that it could sell that particular good to another buyer, it could have made a sale of another unit to that second buyer anyway, so it has lost one sale at the end of the day. So the position of the parties is actually symmetrical: Either one can threaten to deprive the other of any surplus by walking away from the bargaining table.

Presumably, the bargain must lie somewhere between \$20 and \$100, but it might seem impossible to say anything more about where rational bargainers will end up. Yet mathematician John Nash, of *A Beautiful Mind* fame,¹⁰⁸ proved otherwise. He showed that the bargain point would be one that maximized the product of the utility of the two parties.¹⁰⁹ There are two independent ways of getting to his solution of the bargaining game. One is axiomatic.¹¹⁰ The Nash solution follows from three simple assumptions: first, that the solution depends on utility but not on the units in which utility is measured; second, that if the parties have the same utility functions, they split the goods evenly; and third, that the outcome is unaffected by irrelevant alternatives (those that would not be chosen anyway).¹¹¹ The other way of getting to the Nash solution is to model the bargaining process. If the parties place some positive value on time, they will converge on the Nash solution rather than prolong the bargaining.¹¹²

Whether people actually do bargain precisely this way is unclear,¹¹³ but Nash's model has some normative appeal. The

108. *A BEAUTIFUL MIND* (Universal Studios 2001).

109. See BRIAN BARRY, 1 *THEORIES OF JUSTICE* 13-14 (1989).

110. See *id.* at 16-17.

111. *Id.* For example, if the parties would decide on a price of \$70 when the range of outcomes is \$20-\$100, they will also do so when the range is \$40-\$80.

112. *Id.* at 19.

113. See H. PEYTON YOUNG, *EQUITY IN THEORY AND PRACTICE* 124-27

Nash solution tells us how surplus would be divided in a setting of complete rationality and voluntariness.¹¹⁴ That seems as close to operationalizing the idea of a fair division as can be expected.

Contrary to theory, one might well actually see a monopolist making a take-it-or-leave-it offer for almost all of the surplus. Since this tactic in general should not work with rational bargainers, the existence and success of this conduct needs to be explained. One possibility is that the monopolist is taking advantage of the timidity or ignorance of the consumer. The other possibility is that the monopolist has arranged matters to make bargaining impossible with any individual buyer. One way of doing so is to ensure that deviations from the originally demanded price become common knowledge. In that case, the monopolist faces a huge potential loss if it reduces the price to any one consumer, since all other consumers will then demand reductions. Thus, the monopolist has locked itself into the high price with each consumer. Another way of doing this is to raise transaction costs by not allowing the agents who actually deal with consumers to negotiate. Any deviation from the firm's initial demand would require approval from some high corporate official, which is impractical. In short, the monopolist gains from tying its own hands so that bargaining is impossible. Only by exploiting the irrationality of the other party or by artificially tying its own hands does the monopolist succeed in seizing the whole surplus. This observation strengthens the claim that the hypothetical bargain has more ethical weight than the actual one, since the actual bargain may take place under circumstances that have been intentionally distorted to benefit one party.

Thinking of the situation in Nash terms has an implicit normative dimension. The implicit assumption is that the intellectual property holder is entitled to as much surplus as it could obtain from each buyer in separate, individualized

(1994) (suggesting that the expectations about the other party also shape outcomes, in addition to utility functions).

114. An alternative solution was proposed by Braithwaite. See BARRY, *supra* note 109, at 37-40. We will not consider his solution here, because it is more complicated and because the data needed to calculate either solution exactly is unlikely to be available. As Barry says, Braithwaite's solution "has a certain perverse ingenuity that makes it worth investigating for those who like that sort of thing." *Id.* at 40. Such readers are invited to pursue the subject.

bargaining. Thus, the seller is entitled to offer its product to each individual buyer on whatever conditions it likes; the seller's property rights are fully respected. Although the seller is entitled to the benefit of its property rights, it is not entitled to the extra advantages that it can obtain because of the happenstance that it can coordinate its bargaining strategy against all buyers simultaneously while the buyers cannot coordinate with each other.¹¹⁵ The seller's superior ability to coordinate its strategy is extraneous to its property interest rather than an intrinsic aspect of ownership. The Nash solution thus recognizes the seller's property entitlement but not all of the strategic advantages that come along with monopoly status.

The details of the Nash solution are not very relevant for present purposes. It is unclear how often the utility data needed to calculate the solution generally will be available. The point is that it is possible to talk about fairness with the kind of rigor that economists use to discuss efficiency. In particular, this example shows that a rational division of surplus will never be completely one-sided and will be equal if the parties have similar utility functions. More fundamentally, Nash's work confirms that a hypothetical bargain between rational actors can be used as a guideline for assessing bargaining outcomes, not only as to efficiency but also as to the distribution of surplus.

B. BARGAINING OVER SURPLUS

Philosophers like Rawls ask us to judge rules by asking whether individuals would have agreed to them in advance—in Rawls's case, from a position that is imagined to be prior to the formation of society. Economists are also used to evaluating rules by asking whether the parties would have agreed to them *ex ante*. We argue in this section that application of our proposed fairness presumption to intellectual property and antitrust issues can be defended as the product of such a hypothetical bargain—a bargain that would be not only fair but also more efficient *ex ante* than unconstrained seller conduct.¹¹⁶

115. Were it not for the antitrust laws, the buyers could form a cartel and bargain collectively, such as in labor unions.

116. This argument is loosely inspired by one of the classics in the field, Farrell and Katz's article on compatibility and innovation. See Joseph Farrell & Michael L. Katz, *The Effects of Antitrust and Intellectual Property on Compatibility and Innovation*, 43 *ANTITRUST BULL.* 609, 613-21 (1998).

1. The Consumer's Bargain and Increasing Returns to Scale

In using a hypothetical bargain as the basis for setting rules, the critical question is how to specify the *ex ante* position. In this Article, much of the focus is on situations involving increasing returns to scale. That is, it is assumed that the good or service in question is either cheaper to produce in large quantities or has more value to each consumer when other consumers are also using the product. The *ex ante* position is that facing the manufacturer and the first potential adopters, before the product has taken hold on the market. In such a situation, the initial adopters care not only about the terms on which they can buy the product originally, but also about what terms will later be available to them and to other consumers. If the terms proposed to later consumers are unfair, sales will be diminished, which will either increase the cost of the good (because production costs otherwise would have gone down more with scale) or reduce its value to consumers (because there will be fewer later consumers to share with). Thus, in the pre-takeoff position, consumers would favor a regime in which a fair division of surplus is guaranteed to all consumers, not just themselves. (They would actually prefer that buyers get all of the surplus but would not be able to negotiate such a bargain.) In the pre-takeoff position, the seller would also favor such rules in order to draw in the first set of consumers, thereby setting the stage for the seller to reap economies of scale as the business grows.

Because of the existence of economies of scale, potential consumers act as if they were behind a veil of ignorance, not knowing whether they will be early or late adopters. Those who may be late adopters would favor fair treatment of the early adopters, without whom the business will never get off the ground and make scale economies available to them later on.¹¹⁷ Those who may be early adopters must be concerned because, if potential late adopters anticipate unfair terms, they

Farrell and Katz use a formal model to show how producer behavior in network industries is shaped by the prospect of establishing a large network, which in turn is partly driven by consumers' expectations about each other's conduct. *See id.*

117. Often, early consumers are especially valuable (compared to late adopters) for two reasons. First, the early consumers are often more knowledgeable, so their adoption sends a signal to possible later consumers that the product is valuable. Second, early consumers help the seller to create a sufficiently large network to make entry by future competitors unappealing.

may not adopt the product, limiting the availability of scale economies to early consumers. So expectations about the fairness of the terms for later consumers can affect the expected value of the product to earlier consumers. Thus, each consumer has an interest in how other consumers are treated, having the same effect as a veil of ignorance.

This scenario raises two immediate questions. First, does fairness to the later consumer really matter to the early adopters? One might argue that late buyers will still sign on if they are allowed to retain *any* of their consumer surplus, a smidgeon of surplus being better than no surplus at all. Once they come on board, earlier adopters will benefit from the resulting economies of scale. This argument ignores the fact, though, that people care about fairness: The empirical evidence shows clearly that people are willing to pass up some benefits if they feel that the terms of the offer are unfair.¹¹⁸ Thus, the later consumers may not come on board at all. In addition, the smaller surplus provided by an unfair deal creates less enticement for later consumers, thereby slowing the expansion of the network even if ultimately these consumers do all come on board.

The second question is whether the seller will unilaterally offer the optimum terms (from the point of view of early adopters) to potential late adopters. Like the early adopters, the seller also has an interest in network expansion in order to take advantage of economies of scale. But the seller cannot fully internalize the benefits of network expansion, and therefore will expand less quickly and fully than is optimal for consumers. Expansion provides benefits to existing customers. The seller cannot fully extract these benefits for several reasons: the difficulty of employing perfect price discrimination;¹¹⁹ the impracticality of adjusting prices constantly as new customers join;¹²⁰ and the possibility that existing consumers may drop out if they feel abused by the seller, even if allowed to retain some shred of surplus. Thus,

118. See *infra* notes 143-44 and accompanying text.

119. Depending on the identity of new consumers, their value to current consumers will vary. In a phone network, the new consumers might be people one would not want to call anyway. In the operating system context, the new consumers may want only software that current consumers do not value, so current consumers experience no benefit from the increased incentive to design software compatible with the operating system.

120. This is most obvious with software, where existing consumers purchase upgrades only episodically.

the seller has an incentive to abuse later consumers even at the expense of slowing expansion, particularly after the network has become dominant enough to deter new entry.¹²¹

The bottom line is that the seller has an incentive to offer unfavorable terms to late adopters, slowing expansion. If earlier buyers cannot lock in their own contract terms permanently, they face the threat of such opportunistic conduct themselves. Anticipating this possibility, potential early adopters may falter, particularly if the product's ultimate value is uncertain in other respects. It would be to the seller's advantage to commit in advance to limit future opportunistic conduct to entice the initial nucleus of consumers and obtain the critical mass sufficient to deter new entry. Thus, if bargaining *ex ante*, the seller would find it advantageous to agree to a long-term relational contract guaranteeing buyers a fair share of surplus. Legal rules can provide the equivalent of this hypothetical *ex ante* contract.

In situations with increasing returns to scale in production, this problem is often solved by requiring potential suppliers to receive a franchise from a public authority.¹²² The authority represents the interests of all consumers, those initially desiring the service as well as those entering later (for example, in an expanding municipality). In this situation, the *ex ante* bargain can actually be made explicit, backed by public regulatory authority. As a representative of the early adopters, the municipality can guarantee that the utility will offer suitable rates to later adopters, thus allowing expansion of service and eventual lower rates for everyone. But in other markets with increasing returns, such as many informational goods, consumers deal with the seller individually, making such an explicit bargain impractical.

2. The Case of Network Goods

Consider the not uncommon situation in which an inventor

121. The discussion in the text concerns situations where there are economies of scale in consumption. Where there are economies of scale in production, the converse problem exists: early adopters may not receive price reductions given to later ones. If early adopters seek to solve this problem with Most Favored Nation (MFN) clauses, the producer will offer above-optimal prices to later adopters.

122. MARK SEIDENFELD, MICROECONOMIC PREDICATES TO LAW AND ECONOMICS 61-2 (1996) (discussing government licenses and rate regulation).

offers a new product in a network market.¹²³ The initial group of consumers would like to know three things: first, how good the product is (how much surplus it can produce); second, how much of the surplus they will receive, as opposed to the seller; and third, whether other consumers will adopt it (so they can benefit from network externalities). Ideally, the seller would like to commit in advance to a fair division of surplus. Such a commitment would have two advantages from the point of view of the initial consumer. First, it would obviously increase that specific consumer's share of surplus. Second, the availability of fair terms will encourage other later consumers to adopt, feeding the network effect that produces much of the value of the good. It is risky for the early buyers to simply count on the seller's good faith. Once the network effects have taken hold, it will be difficult for consumers to exit, creating a major incentive for opportunistic conduct by the seller in the end game.

In particular, later consumers may be subject to price discrimination in an effort to extract more of the surplus from them. This does not *directly* concern the initial group of consumers, except to the extent that they anticipate having to reenter the market themselves (either to purchase upgrades or because the product requires repairs or replacement). It does, however, concern them indirectly. The less surplus that is provided to later buyers, the slower and less certain the growth of the network will be. This, in turn, diminishes the expected value of the good to the first consumers due to network effects. Thus, a commitment to treat later consumers fairly would be valuable to those consumers who are first considering the product. Consequently, since the seller needs this nucleus of first consumers in order to get the network effects rolling and make the product a success in the market place, the seller would be willing to commit to a fair sharing of surplus in order to make this happen. For these reasons, the seller should be bound on estoppel principles to the promises that it makes to early adopters in order to get positive feedback going.¹²⁴

123. By this we mean a market such as telephones or operating systems, in which the value of the good to any one consumer is partly a function of how many other consumers also use the good. See Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CAL. L. REV. 479, 488-500 (1998).

124. See Mark A. Lemley & David McGowan, *Could Java Change Everything? The Competitive Propriety of a Proprietary Standard*, 43 ANTITRUST BULL. 715, 740-43 (1998).

As in many relational contracts, however, the contract is likely to be incomplete in important ways. This is because it is very difficult for the seller to make a binding contractual commitment that covers all of the potential opportunities for opportunistic behavior. Just as private law recognizes a general duty of good faith as a backstop to explicit contractual terms, the fairness presumption enforces the implicit undertaking of the network seller to its buyers.

There are substantial barriers to the negotiation of explicit contractual protections. Even the initial adopters have some direct reason to be concerned that their current contract may not protect them against opportunism. In many situations, it will not be feasible to lock in contractual terms for the original group of consumers, because they will have to recontract when obtaining upgrades. At that point, they will be at the mercy of the seller, who may well have a dominant market position by then. It would be even more difficult for the initial adopters to enter into a contract that would prevent opportunism toward later adopters. It is hard to imagine a contract that would govern not only the initial terms with the original group but also the terms of the seller's later contracts with others. Such a contract would require the parties to foresee at the very early stages of a developing market what business terms will be desirable at later stages. It would also present serious problems of enforcement.¹²⁵

In some other contexts, reputation can serve as a substitute for legal enforcement. But the seller of a new network product may not have had a prior opportunity to establish such a reputation. Furthermore, the benefits of opportunistic conduct may simply outweigh the possible loss of reputation. Compare the situation of the network seller to that of an employer that needs its workers to make substantial firm-specific investments. Once the workers have made such

125. There are two problems with such an enforcement action, even assuming a suitable contract could be drafted to begin with. The first concerns the availability of an adequate remedy. Damages would be difficult to calculate, while specific performance is an extraordinary remedy in a contract case. Second, the plaintiff would be producing benefits for other consumers collectively. Such a suit is likely to require either collective action by a large group of plaintiffs or a procedural structure allowing a class action and attorney's fees. The class action may be frustrated by questions about the typicality of the plaintiff and the suitability of the issue for class action treatment given the individualized nature of damages, while attorney's fees are not available under the American rule in contract cases.

investments, they open themselves up to opportunistic conduct by the employer. In many situations, even without formal contractual protection, workers may be protected by the employer's need to hire new workers—if it betrays its commitments to older workers, newer workers will be reluctant to commit to the firm. The seller of a network good does not always face this problem, however. Once it establishes critical mass, the good may dominate the market simply because it is *there*.¹²⁶ Thus, the seller of a network good can be in somewhat the same position as an employer that plans to hire and train all of its workers when it first starts business, with no future hiring. Reputational effects are less useful here, and, to the extent they exist, they distort the market by favoring established firms presenting new products rather than new firms.

Given the inability of the parties to enter into such a contract themselves, the law can provide the equivalent through a presumption against an unfair division of surplus. This flexible standard is hard to specify *ex ante* by contract but easier to enforce *ex post* through litigation. Such a standard encourages initial consumers to take a chance on new network goods rather than holding back, thus making it more likely that society will realize the benefits of these goods and increasing the incentive for inventing them in the first place.

Part III discusses in more detail how fairness plays out in the intellectual property arena. For now, two lines of argument are worth noting. First, using an open standard provides some guarantee to consumers against future mistreatment. Any individual seller's possible opportunistic behavior is checked by the threat of competition from other sellers. Second, anti-discrimination rules also limit the ability of the seller to extract surplus from particular buyers. Some buyers will inevitably be in a better position to protect themselves against opportunistic behavior than others; a rule limiting discrimination in effect extends the benefits of this protection to other, more vulnerable consumers. Thus, when an open standard is not available, consumers will obtain some protection from a suitable nondiscrimination commitment by the owner of a closed standard.

The point is not just that open standards or certain

126. The Microsoft Windows operating system is often considered to be an example.

nondiscrimination rules are desirable for consumers, but that ex ante they are also desirable for sellers. Given increasing returns to scale, sellers need to attract early adopters with fair terms (including fair terms for upgrades), and early adopters will be more willing to join if the guarantee of fair treatment increases the likelihood of adoption by later consumers.¹²⁷ Thus, where the parties cannot credibly commit to such future fair treatment, they would ex ante favor legal mandates that protect the fair division of surplus.¹²⁸

If predictable trade-offs exist between fair division of surplus and the size of surplus, the parties' ex ante bargain might change. Each side might be willing to agree to a provision that increases surplus, even if it receives a smaller fraction of surplus. This complication is irrelevant in the current context. This Article only proposes that fairness be used as a tiebreaker when efficiency effects are unpredictable. This means that the parties have no reason to anticipate that a diminished share of surplus will be outweighed by an increase in the total amount. Thus, with the size of surplus being unaffected (at least in any predictable way) by the terms of the agreement, preferences about the division of surplus are the dominant factor.

Enforcing these ex ante preferences seems morally appealing. An enforcement rule benefits both sides of the deal, sellers as well as consumers, by facilitating the introduction of new goods into the market. Moreover, once consumers have "signed on" to the good, they may be, to some extent, stuck because of the network seller's ability to attain market dominance. It seems wrong for the seller to take advantage of their vulnerable position. Furthermore, consumers as a whole are probably less affluent than owners of intellectual property,¹²⁹ so ensuring consumers a fair share of surplus also prevents a perverse redistribution of wealth.

Such rules in industries with increasing returns to scale also benefit society as a whole. Goods with increasing returns to scale, such as the Internet, can contribute greatly to social wealth. A fairness-based legal regime makes it easier for such products to get their initial foothold, and therefore, decreases the risks facing the inventors of such products.

127. In particular, early adopters will hold back if they think that later adopters will get a greater share of surplus.

128. Ex post, of course, the seller may have other ideas.

129. See *supra* notes 38-48 and accompanying text.

C. STATUTORY INTERPRETATION AND THE BARGAINING SURPLUS

Both intellectual property law and antitrust law are based on statutes.¹³⁰ The field of statutory interpretation is highly contentious, with intentionalists, dynamic interpreters, legal process theorists, and textualists all doing battle.¹³¹ Although we have our own views on these matters, this is not the place to address them. Under the first three of these views, however, judges legitimately consult social norms in interpreting statutes.

Intentionalists believe that statutes should be construed in accord with the intent of the legislature.¹³² To the extent that we lack specific information about the intentions of the legislators, it is reasonable to assume that they shared the widespread norms of our society. It is clear that fair division of surplus is one of those norms.¹³³

Dynamic interpreters and legal process theorists both give independent weight to social norms. Dynamic interpreters maintain that statutes should be construed in light of current social values, even at the expense of original intent.¹³⁴ Legal process theorists contend that the legislature should be presumed to be composed of reasonable people with reasonable purposes, and they argue that legislation should be construed in light of those purposes.¹³⁵ Broad social norms have a *prima facie* claim to be considered as part of the definition of reasonableness here.

There are several reasons why law should presumptively

130. See Copyright Act of 1976, Pub. L. No. 94-553, 90 Stat. 2541 (codified as amended in scattered sections of 17 U.S.C. (2000)); Patent Act of 1952, ch. 950, 66 Stat. 792 (codified as amended at 35 U.S.C. §§ 1-293 (2000)); Lanham Act, ch. 540, 60 Stat. 427 (1946) (codified as amended at 15 U.S.C. §§ 1051-1127 (2000)); Clayton Act, ch. 323, 38 Stat. 730 (1914) (codified as amended at 15 U.S.C. §§ 12-27 and 29 U.S.C. §§ 52-53 (2000)); Sherman Act, ch. 647, 26 Stat. 209 (1890) (codified as amended at 15 U.S.C. §§ 1-7 (2000)).

131. The best collection of materials on the subject is ESKRIDGE ET AL., *supra* note 71, at 669-1098.

132. See John F. Manning, *Textualism as a Nondelegation Doctrine*, 97 COLUM. L. REV. 673, 677-89 (1997).

133. See *infra* notes 143-44 and accompanying text. This view of legislative intent may be too optimistic when the process is dominated by unopposed special interests.

134. See WILLIAM N. ESKRIDGE, JR., *DYNAMIC STATUTORY INTERPRETATION* 48-80 (1994).

135. HENRY M. HART, JR. & ALBERT M. SACKS, *THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW* 1374-80 (William N. Eskridge, Jr. & Philip P. Frickey eds., 1994).

track social norms.¹³⁶ One is that norms may have a tendency to reflect efficiency, giving them presumptive status even for the devout believer in efficiency as the “One True Guide to Social Policy.”¹³⁷ Another is that people resent deviations from social norms; this is a social cost. They may also resist government efforts to dislodge existing norms, making enforcement more difficult and expensive—another social cost that stems from conflicts between laws and norms. Indeed, though an economist might wish that all judges were concerned only with rigorous economic analysis, it will be easier to obtain full judicial enforcement of rules that correspond to the judges’ own intuitions about justice. It also bears remembering that this is, after all, a democracy. Presumably, what the people as a whole believe to be morally right is entitled to some consideration in interpreting the laws made by their representatives.¹³⁸ As Joseph Brodley puts it, “to hold that producers perform a civic duty when they systematically take from buyers the entire economic surplus is an Orwellian result that no democratic government could long sustain.”¹³⁹

Textualism has not played a major role in antitrust law, where the words of the Sherman Act and other statutes have been more or less smoothed away in the name of social

136. See Avery Katz, *Taking Private Ordering Seriously*, 144 U. PA. L. REV. 1745, 1749-53 (1997); Jody S. Kraus, *Legal Design and the Evolution of Commercial Norms*, 26 J. LEGAL STUD. 377, 381-83 (1997); Richard H. McAdams, *The Origin, Development, and Regulation of Norms*, 96 MICH. L. REV. 338, 345-47 (1997). See generally Mark A. Lemley, *The Law and Economics of Internet Norms*, 73 CHI.-KENT L. REV. 1257, 1257-65 (1998) (providing a useful overview of the extensive scholarship on this subject matter).

137. As one scholar put it,

Lacking time to educate themselves in every aspect of market and culture, the public tends to be skeptical when an industry claims that expert opinion shows that what is good for the company will also be good for the nation, and that state aid in enforcing its desires will produce an economically efficient result. And you know what? Given the arguments reviewed in this Paper, I would say that the public has a point.

James Boyle, *Cruel, Mean, or Lavish? Economic Analysis, Price Discrimination and Digital Intellectual Property*, 53 VAND. L. REV. 2007, 2039 (2000).

138. Cf. Stephen F. Ross, *The Political Economy of the Efficiency Defense* (Mar. 8, 2003) (unpublished manuscript, on file with the *Minnesota Law Review*) (applying a similar argument against exclusive reliance on efficiency in the context of mergers).

139. See Brodley, *supra* note 2, at 1036.

welfare.¹⁴⁰ Intellectual property law has also shied away from rigid textualism—a deviation that is invited by the intellectual property clause of the Constitution itself, which firmly cements this body of law to the dominant purpose of advancing socially valuable innovation. According to the Supreme Court, benefit to the general public rather than to inventors and writers—or to put it another way, consumer surplus rather than producer surplus—is the ultimate goal of the intellectual property clauses.¹⁴¹ In any event, it would be surprising if statutes passed for a variety of reasons happened to contain language that was precisely directed to maximizing total surplus at the complete expense of other important social norms.¹⁴²

We have so far presumed that a fair division of surplus is indeed a widely shared norm. Fortunately, empirical support for this proposition comes readily to hand. A striking example is provided by the ultimatum game. In the ultimatum game, the subjects must agree on how to divide a fixed sum of money. The first person proposes a division. Either the second person accepts the division or the money goes to neither of them. The rational strategy is clear. For example, suppose the amount in question is a dollar. If the second person is offered even one cent out of the dollar, he should accept, since the alternative is to reject and get nothing. Knowing this, the first person should keep as large a share as possible for herself. So she should propose ninety-nine cents for herself and one cent for the other person; rationally, the other person should accept the proposal. Hundreds of experiments have been conducted with variants of this game. The results are clear.¹⁴³ Unless the first person

140. See *Spectrum Sports Inc. v. McQuillan*, 506 U.S. 447, 454 (1993); Frank H. Easterbrook, *Statutes' Domains*, 50 U. CHI. L. REV. 533, 544 (1983); Frank H. Easterbrook, *Workable Antitrust Policy*, 84 MICH. L. REV. 1696, 1702-05 (1986); McGowan, *supra* note 58, at 762-63

141. See *Eldred v. Ashcroft*, 123 S. Ct. 769, 802 (2003) (Breyer, J., dissenting) (the reward to the creators is “a means, not an end”); *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975). In *Eldred*, Justice Breyer cites legislative history to the same effect. *Eldred*, 123 S. Ct. at 803. For example, in 1909, Congress said that copyright was designed for “the benefit of the great body of people.” *Id.*

142. There is a reason that the copyright doctrine is called “fair use” and not “uses involving prohibitive transaction costs.”

143. See RICHARD H. THALER, *THE WINNER'S CURSE: PARADOXES AND ANOMALIES OF ECONOMIC LIFE* 21-35 (1992); Ernst Fehr & Klaus M. Schmidt, *Theories of Fairness and Reciprocity—Evidence and Economic Applications*, CESIFO WORKING PAPER SERIES 4-9 (Dec. 2000), at <http://www.cesifo.ed/> (last visited Apr. 24, 2003).

makes a “fair” offer—at least twenty percent of the total—it is very likely to be rejected.¹⁴⁴ Individuals are willing to make a sacrifice, rejecting a possible gain, in order to punish someone who makes an unfair offer. People do indeed care about the fair division of surplus. This strong social norm has some claim to recognition in interpreting the statutes passed by their representatives.

D. FAIRNESS AND RENT-SEEKING

Our final supporting argument for presumptively giving consumers a fair share of surplus is that the allure of producer surplus results in wasteful rent-seeking of various kinds. This wasteful rent-seeking can take a number of forms.

First, the availability of high rents may encourage a wasteful race by firms. If only one firm can capture the market, it is worthwhile for several firms to invest in the race, but the efforts of all but one contestant will be wasted. The classic example involves patent races, in which multiple firms seek to perfect a discovery earlier than their competitors, when delaying the invention a little would not have been costly to society as a whole.¹⁴⁵ Another example is the race to become the industry standard. Producer surplus can easily be dissipated in such socially wasteful activities.¹⁴⁶

Second, the opportunities for extracting surplus vary. Some products may not lend themselves as well to this purpose as others, depending on technology, market structure, and other factors. For example, one market may have low barriers to entry, making it harder to exploit consumers. Investment in creating new products will be skewed toward industries where producers get a higher share of surplus, even if total surplus is higher somewhere else.

Third, the prospect of collecting rents will spur producers to influence the political process. They may seek legislation authorizing what would otherwise have been prohibited actions or legislation excluding possible competitors. Unwarranted expansions of intellectual property statutes may fall into either category, either by encouraging practices such as price discrimination or by allowing the use of intellectual property

144. See Fehr & Schmidt, *supra* note 143, at 5.

145. See, e.g., Richard A. Posner, *The Law and Economics of Intellectual Property*, DAEDALUS 5, 9 (2002).

146. See *supra* Part I.A.5.

rights as a basis for monopolistic exclusion. It is particularly likely that producers (rather than consumers) will engage in this political rent-seeking. Producer groups are often smaller and have more concentrated interests, making effective organization much simpler and thus more likely.¹⁴⁷ Moreover, current rents can provide a war chest for further rent-seeking expeditions. To put it as bluntly as possible, producers tend to be piggy than consumers, in political terms.

What may well become the paradigm of such piggy behavior is the Sonny Bono Copyright Term Extension Act, a statute named in honor of a former performer and member of Congress who wanted copyright terms to last forever.¹⁴⁸ The Act was just recently held to be constitutional by the Supreme Court in *Eldred v. Ashcroft*.¹⁴⁹ As Justice Breyer's dissent points out, and as the majority does not seriously contest,¹⁵⁰ the main result of the statute was to enrich existing copyright holders:

This statute will cause serious expression-related harm. It will likely restrict traditional dissemination of copyrighted works. It will likely inhibit new forms of dissemination through the use of new technology. It threatens to interfere with efforts to preserve our Nation's historical and cultural heritage and efforts to use that heritage, say, to educate our Nation's children. It is easy to understand how the statute might benefit the private financial interests of corporations or heirs who own existing copyrights. . . . [W]ith respect to existing works, the serious public harm, and the virtually nonexistent public benefit could not be more clear.¹⁵¹

The majority opinion in *Eldred* makes it clear that Congress has the constitutional power to adopt such a rent-seeking statute.¹⁵² This does not mean that courts should actively encourage this conduct. Instead, they should apply the intellectual property statutes to favor the public interest rather

147. See DANIEL A. FARBER & PHILIP P. FRICKEY, *LAW AND PUBLIC CHOICE: A CRITICAL INTRODUCTION* 23-24 (1991).

148. Pub. L. No. 105-298, 112 Stat. 2827 (1998) (codified as amended in scattered sections of 17 U.S.C. (2000)).

149. See *Eldred v. Ashcroft*, 123 S. Ct. 769, 790 (2003).

150. See *id.* at 786-87. The Court upheld the law on the theory that a "rational person" *might* have thought that the law served some public purpose, but the majority opinion gives little ground for confidence that Congress in fact thought so or that such a belief would have been correct. See *id.* at 782. The Court also relied heavily on the fact that Congress had previously passed retroactive extensions of intellectual property terms, which only shows that similar rent-seeking efforts had also succeeded in the past. See *id.* at 783.

151. *Id.* at 813.

152. *Id.* at 777-78.

than colluding with legislators and lobbyists to maximize producer surplus.¹⁵³ Here, the antitrust laws provide an excellent precedent. The Sherman Act was originally a poorly focused effort to protect certain interest groups such as farmers and small businesses; over time, the courts have interpreted this statute instead to further the overall welfare of society.¹⁵⁴ Indeed, the Court has even been willing to narrowly construe later statutes that might conflict with the public-oriented goals of the Sherman Act.¹⁵⁵ It makes sense to construe the antitrust laws and the intellectual property laws together so as to moderate the rent-seeking incentives created by legal protection for intellectual property.¹⁵⁶

The only way to eliminate completely the potential for such rent-seeking would be to deprive producers of *any* share of surplus, restricting their return to their costs (including the cost of capital). But intellectual property law is committed to providing rents to inventors and authors as an incentive for innovation. For this strategy to work, the producers must be allowed some share of surplus. If the government had perfect information, it could set the return on intellectual property at just the right level to maximize social welfare, taking wasteful rent-seeking efforts into account. This information is not available, and it is not likely to be available at any time in the near future. What is known is that, as between two alternatives, the one with a fair share of surplus going to consumers should be favored, assuming that the two produce the same total surplus (or more realistically that it is not known which one produces the greater total surplus).

153. See Jonathan R. Macey, *Promoting Public-Regarding Legislation Through Statutory Interpretation: An Interest Group Model*, 86 COLUM. L. REV. 223, 268 (1986).

154. See *supra* text accompanying notes 72-75.

155. See *Group Life & Health Ins. Co. v. Royal Drug Co.*, 440 U.S. 205, 231-32 (1979) (adopting a narrow construction of the McCarran-Ferguson Act of 1954); *Fed. Mar. Comm'n v. Seatrain Lines, Inc.*, 411 U.S. 726, 733 (1973) (narrowly construing section 15 of the Shipping Act); *United States v. McKesson & Robbins, Inc.*, 351 U.S. 305, 316 (1956) (strictly construing the Miller-Tydings Act and the McGuire Act).

156. Thus, we are skeptical of the view that the intellectual property laws should automatically trump the antitrust laws, a view which is defended in Daniel J. Gifford, *The Antitrust/Intellectual Property Interface: An Emerging Solution to an Intractable Problem* (2002) (unpublished manuscript, on file with authors).

E. A MODEST NORMATIVE SUGGESTION

Our argument has really proceeded at two levels. On the most basic level, we question whether even in principle economic efficiency should be the sole purpose of antitrust and intellectual property law. Social welfare may be an appealing concept, but the case for making it the exclusive moral standard has yet to be made,¹⁵⁷ nor has the case for relegating all distributive issues to tax law been made.¹⁵⁸ Particularly in markets with increasing returns to scale, a fair division of surplus mimics the agreement that the parties would have made *ex ante*. The hypothetical willingness of both sides to consent to a reasonable division of surplus is one ground for considering such a division fair to both sides. It also provides a long-term efficiency argument for protecting fairness: A legal regime favoring fairness will make it easier for innovative products to find markets and for society to realize the resulting benefits from increasing returns to scale. Another argument derives from the greater tendency of producer surplus to spark socially wasteful rent-seeking, political and otherwise.¹⁵⁹

These arguments help explain the specific parameters of our proposal. The proposal is tailored to the area of overlap between these various arguments for fairness. Some of the arguments would support a broader role for fairness, but our proposal is limited to cases in which direct evidence of efficiency is inconclusive regarding the specific market in question. In those cases, our proposed rule would disfavor practices that give producers an unfair share of surplus. Of course, the meaning of the term “unfair share” of surplus is less than crystalline. In principle, the Nash solution is the best benchmark for fairness, but in practice this information may not be available. Social norms of equity are an acceptable rough substitute. Given that notions of fair division are widely shared in our culture,¹⁶⁰ judges should be capable of doing a reasonable job in applying this standard. At least some outcomes will be so one-sided as to leave little doubt of their unfairness, while others (such as a roughly even division of surplus) will clearly pass muster.

A substantial argument could be made for going farther,

157. See *supra* notes 8-20 and accompanying text.

158. See *supra* notes 22-36 and accompanying text.

159. See *supra* notes 145-56 and accompanying text.

160. See *supra* note 143 and accompanying text.

with a default presumption that consumers should get the entire surplus unless the firm can present clear evidence of net efficiency gains.¹⁶¹ There is a good deal in antitrust lore, for example, that could support a consumer surplus standard. Moreover, because the courts have done so much to move antitrust law toward the public interest end of the spectrum,¹⁶² there is also a reasonable argument that the antitrust laws should trump the intellectual property laws in case of conflict. There is more reason to be concerned that the intellectual property laws are tainted by special interest influence, which by now has been filtered out of antitrust law. Although these arguments have some appeal, for present purposes it is enough to promote a compromise between the consumer-oriented antitrust laws and the more producer-oriented intellectual property laws. A presumption in favor of a fair division of surplus provides such a compromise.

This Article focuses on where to draw the line regarding permissible behavior by the seller. It does not address in any detail the further question of legal implementation. Limits could be imposed on seller behavior through an interpretation of the Sherman Act, through restrictive interpretations of the seller's intellectual property rights, through common law restrictions on the seller's contract rights, or through consumer protection statutes. Which route to take is obviously important. Implementing the fairness principle through the antitrust laws would be consistent with the thrust of those laws, but a full analysis would have to consider the utility of private, treble-damages suits and government antitrust litigation versus the other potential methods of implementation such as expansion of the intellectual property misuse doctrines.

The remainder of this Article applies the fairness approach to some significant problems involving standards. These examples will not only illustrate the application of our approach, but also demonstrate the strong arguments for taking account of fairness.

161. See Brodley, *supra* note 2, at 1037-39 (explaining how such an approach would work).

162. See *supra* notes 72-75 and accompanying text.

III. APPLYING FAIRNESS TO STANDARDS ISSUES

A. OPEN VERSUS CLOSED STANDARDS

Our first application of the fairness presumption is the setting of standards. Standards come in a variety of forms with a variety of uses. We focus on standards that allow two or more distinct components of a system to work together. Standards can be either closed or open; that is, either a single party owns the intellectual property rights that control access to the standard (closed) or no party owns those rights (open). Our fairness presumption suggests that, except where closed standards can be shown to have net efficiency benefits, the law should favor and encourage open standards over closed standards. Since deciding the efficiency benefits of open versus closed standards is frequently a complicated and difficult empirical question, this presumption will often come into play.¹⁶³

1. The Meaning and Significance of Closed Versus Open Standards

Standards are of particular importance in the modern computer and telecommunications industries and can give rise to the sort of network effects discussed in Part II. The value of a product whose usefulness depends, at least in part, on interaction with other products will depend on whether or not it can be used with the most commonly owned versions of its complementary products. Once a standard becomes more widely accepted than its alternatives, network effects will frequently result in that standard becoming dominant and the weakening or elimination of its rivals.

To be more concrete, consider two examples of standards. The Windows operating system is a standard that allows a variety of application programs to be run on personal computers that use Windows. Similarly, Internet protocols allow programmers writing Internet applications to write programs that can be used on computers around the world.

Again, standards come in two basic forms, closed and open. The intellectual property rights to a closed standard are owned by a person or entity that can exclude or set conditions on the use of the standard and can charge for the right to use the

163. See *infra* notes 173-78 and accompanying text.

standard. The Windows operating system is a closed standard. An open standard is not owned in such a fashion. Internet protocols are typically open standards, although the Internet Engineering Task Force, which controls some such protocols, now allows for proprietary standards.¹⁶⁴ Intermediate situations between closed and open standards are also possible.¹⁶⁵

One other distinction among standards is worth noting. It relates to how standards are set. "De facto standards" are set by the marketplace, as individual actors choose which standard to use.¹⁶⁶ Windows is a de facto standard.¹⁶⁷ Standards may instead be set by a standard-setting organization (SSO) composed of a variety of private actors. Internet protocols are typically set by SSOs.¹⁶⁸ Finally, standards may be set by governmental agencies. Early on, for instance, Internet protocols were set by such agencies.¹⁶⁹

Standard-setting raises a variety of questions for both antitrust and intellectual property law. In the remainder of this section we shall consider the policy guidelines that should guide judges and legislators in answering such questions.

2. Open Versus Closed Standards and the Efficiency Issue

Open and closed standards have offsetting advantages and disadvantages, assuming a goal of achieving efficiency in the sense of maximizing total surplus.¹⁷⁰ A key advantage of closed standards is that they may be better at encouraging innovation. This is because the owner of a dominant closed standard will often be able to realize monopoly profits, and the prospects of such profits encourage innovation.¹⁷¹ The advantage of closed standards thus rather closely tracks the

164. See Mark A. Lemley, *Intellectual Property Rights and Standard-Setting Organizations*, 90 CAL. L. REV. 1889, 1893 n.7 (2002).

165. See *infra* notes 193-94 and accompanying text.

166. See Lemley, *supra* note 164, at 1899.

167. As with Windows, de facto standards tend to be closed. See *id.* at 1900.

168. See generally *id.* (providing a wide-ranging analysis of SSOs).

169. See *id.* at 1899. Government standard-setting raises issues mostly beyond the scope of this discussion, except to note the different legal and policy questions that arise for de facto standards and for SSOs. See *infra* notes 185-92 and accompanying text.

170. The following analysis draws heavily on Farrell & Katz, *supra* note 116, at 637-42.

171. See *id.* at 638.

advantage of strong intellectual property rights.

Even the innovation advantage of closed standards is ambiguous, however. Under some circumstances the private advantages from innovating and achieving a dominant closed standard may exceed the social advantages, leading to an excessive incentive to innovate. In other circumstances, a closed standard may encourage initial innovation, but then discourage subsequent innovation by competitors.¹⁷²

Open standards, in contrast, tend to be better at encouraging price competition among competitors.¹⁷³ They thereby tend to increase consumer surplus while reducing producer surplus. This effect is also ambiguous, however. Under some circumstances the degree of price competition may vary little depending on the type of standard. On some occasions, price competition among competing owners of incompatible standards may lead to greater price competition than if there were an open standard.¹⁷⁴

How these and other relevant economic factors balance out overall is a complicated question that varies widely depending on specific circumstances. This has led to much debate over the relative merits of open and closed standards.¹⁷⁵ As Lemley characterizes this debate, “[t]here is a voluminous literature on the relative value of open and closed standards, especially in network industries, and a vociferous debate over the merits of both approaches.”¹⁷⁶ Even within one person’s argument the net balance of the competing effects may depend on a variety of controversial fact determinations. For instance, in the recent model of Farrell and Katz, the net benefits of closed versus open standards depend on a variety of factors, importantly including how consumers form expectations about what future product share will be.¹⁷⁷ The Farrell and Katz approach does suggest that under many circumstances it will be good policy to encourage open standards, and to discourage efforts by owners of dominant standards to manipulate expectations so as to make it harder for competitors to replace their product.¹⁷⁸ In

172. *See id.* at 642.

173. *See id.* at 638.

174. *See id.*

175. *See* Lemley, *supra* note 164, at 1902 n.40 (listing some of the articles in this literature).

176. *See id.* at 1902.

177. *See* Farrell & Katz, *supra* note 116, at 621-36.

178. *See id.* at 646.

other circumstances, however, closed standards or product incompatibility may be better—for instance, because it better encourages innovation.¹⁷⁹ Thus, in many cases an efficiency analysis will often lead to unclear policy prescriptions. This leaves significant room for using the fairness presumption.

3. Open Versus Closed Standards and the Fairness Issue

Looking again at the relative advantages and disadvantages of the two sorts of standards shows that the fairness presumption favors open standards. Open standards typically lead to greater price competition,¹⁸⁰ which benefits consumers. The main problem with closed standards is that they encourage too much monopoly, allowing the dominant producers to squeeze consumer surplus and convert it into profit. That is not to deny that closed standards will not sometimes be more efficient, and maybe even better for consumers. Where, however, efficiency does not clearly favor closed standards, then the fairness presumption suggests that antitrust and intellectual property law should be used to encourage open standards.

All three lines of argument for fairness in Part II support this application of the presumption. First, given the importance of network effects for standards, the analysis of the temptation of dominant producers to extract surplus from later adopters is relevant here. Continued competition among many producers using an open standard would help commit producers to providing a fair surplus to consumers even after the standard has become dominant.¹⁸¹ Such commitment may be harder to achieve with a closed standard. Second, the argument as to statutory interpretation and social norms applies as well. Successful dominant standard-setters are often seen as having an unfair amount of power over their markets.¹⁸² Third, rent-seeking can be an issue in standard-setting. Companies may have an incentive to overinvest in developing a product with promise of becoming a proprietary standard.¹⁸³ Companies may engage in costly behavior to discourage rivals that threaten their ability to attain and

179. See *supra* note 171 and accompanying text.

180. See *supra* note 173 and accompanying text.

181. See *infra* notes 193-94 and accompanying text for a way to commit producers even within closed standards.

182. Consider the resentment towards Microsoft and Windows.

183. See *supra* note 57 and accompanying text.

maintain a dominant standard—the *Microsoft* case, for example, is filled with such behavior.¹⁸⁴ Companies may also lobby the government to help maintain their standard. The *Microsoft* settlement may be an example of lobbying money well spent by the owner of a dominant standard, for instance.

The legal issues raised by standard-setting are varied and complex, but we will touch on a couple of them to give a brief sense of how our framework could be applied. Where a company seeks to impose a *de facto* standard in the marketplace, our analysis suggests that courts should be relatively vigilant in applying antitrust scrutiny while being careful about granting sweeping use of intellectual property rights.

An example of the Department of Justice and the courts applying relatively strict scrutiny to a *de facto* standard-setter is the *Microsoft* case. The court was properly skeptical of *Microsoft's* invocation of its copyright power to justify licensing restrictions designed to inhibit the adoption of the Netscape browser.¹⁸⁵ The court was also properly suspicious about *Microsoft's* efforts to deceive programmers into believing its version of Java would be easily used on non-Windows platforms.¹⁸⁶

An example of how intellectual property law might be used to control *de facto* standard-setting is *Wang Laboratories, Inc. v. Mitsubishi Electronics America, Inc.*¹⁸⁷ The court used the equitable estoppel doctrine, whereby a court will refuse to enforce an intellectual property right when the right holder has led the alleged infringer to believe that the holder does not intend to enforce its right.¹⁸⁸ In *Wang Laboratories* the company had represented to its customers that its standard would remain open.¹⁸⁹ A similar doctrine is that of implied license, whereby the court infers from the expectations of the parties that a right to use the patent should be applied.¹⁹⁰ Both equitable estoppel and implied license can be used to help protect consumers who have built up expectations early in the

184. See *United States v. Microsoft Corp.*, 253 F.3d 34, 58-59 (D.C. Cir. 2001) (*per curiam*), *cert. denied*, 534 U.S. 952 (2001).

185. See *id.* at 59-64.

186. See *id.* at 76-77.

187. 103 F.3d 1571 (Fed. Cir. 1997).

188. *Id.* at 1581.

189. *Id.* at 1575.

190. See Lemley, *supra* note 164, at 1918-27.

adoption process from being exploited after a producer has become dominant in a market with network effects.¹⁹¹

Different issues arise for standard-setting by SSOs. Since SSOs are a common way to achieve open standards, our framework suggests that courts should generally try to encourage SSO standard-setting. Courts will still want to police against some potential abuses, though. For instance, they will want to ensure that a company with a proprietary standard is not dominating an SSO and using it as a way to impose its product on the market.¹⁹² Courts will also want to ensure that a group of producers are not using an SSO to cartelize a market.

Lemley points to an interesting common practice among SSOs that suggests something of a middle way between open and closed standards. SSO governing documents frequently provide that holders of an intellectual property right in a standard created by the organization must license that copyright to members (and occasionally non-members) on “reasonable and nondiscriminatory” terms.¹⁹³ This may provide a way of ensuring that rights holders receive enough of a return to induce innovation, while protecting against exploitation of other producers and consumers made possible once dominance in a market with network effects has been achieved. As such, the practice seems to be a generally positive one that should be encouraged. A role for courts remains, though. For one, if disputes arise courts may need to interpret what terms are “reasonable” and “nondiscriminatory.”¹⁹⁴ For another, courts should examine each case to ensure that the interests of consumers are properly represented by the SSO agreements. Where many producers with varying interests are represented in an SSO, that will probably be true most of the time, but it

191. See *supra* Part II.B.2.

192. See 2 HOVENKAMP ET AL., IP AND ANTITRUST, *supra* note 3, at 35-33 to 35-46.

193. See Lemley, *supra* note 164, at 1906.

194. See *id.* at 1966-67. As both Lemley and Carl Shapiro recognize, “reasonableness” should be analyzed in terms of ex ante circumstances, not the circumstances that prevail after a standard has become dominant. See *id.* at 1967 n.332; Carl Shapiro, *Setting Compatibility Standards: Cooperation or Collusion?*, in EXPANDING THE BOUNDARIES OF INTELLECTUAL PROPERTY at 31, 96-7 (Rochelle Cooper Dreyfus et al. eds., 2001) (discussing modern standards litigation on what is “reasonable”); see also Mark R. Patterson, *Inventions, Industry Standards, and Intellectual Property*, 17 BERKELEY TECH. L.J. 1043, 1054-55 (2002); *supra* Part II.B.2.

will not always be so.

B. PRICE DISCRIMINATION

Our second application of the fairness presumption is price discrimination. Our presumption indicates that, except when price discrimination and equivalent practices can be shown to have efficiency benefits, they should be disfavored because of the transfer of surplus to the producer. As it turns out, given the unclear efficiency effects of price discrimination in many situations, there is plenty of room for the presumption to operate.

1. The Meaning and Significance of Price Discrimination

A quick review of the basic economics may be helpful. In a competitive market, every producer must price at marginal cost, and the prices must all be the same in equilibrium (otherwise sellers with higher prices would lose business to others).¹⁹⁵ A monopolist charges above marginal cost.¹⁹⁶ At the monopoly price, the monopolist trades off the loss of sales from the higher price against the greater profit from each remaining sale. In the basic model, the monopolist charges the same price to all customers, but it would prefer not to do so. It could make a higher profit by charging different prices to different customers,¹⁹⁷ attempting to extract as much consumer surplus as possible from each.¹⁹⁸

Economists define three separate varieties of price discrimination. First-degree price discrimination is the ideal from the seller's point of view. The seller knows each buyer's willingness to pay for the goods, and can prevent buyers with low valuations from reselling to those with higher valuations. This enables the seller to capture virtually the entire surplus.¹⁹⁹ Second-degree price discrimination occurs when the seller cannot directly observe the buyer's valuations but can induce them to signal their preferences. For example, the seller might sell an inferior version of the good at a lower price to buyers who place less value on the good. Inducing buyers to

195. See, e.g., SEIDENFELD, *supra* note 122, at 35-38.

196. See *id.* at 41-42.

197. See *id.* at 43.

198. See *id.* at 42-44.

199. See Michael J. Meurer, *Copyright Law and Price Discrimination*, 23 CARDOZO L. REV. 55, 68-69 (2001).

signal their preferences is usually costly, so second-degree discrimination can reduce total surplus.²⁰⁰ In third-degree price discrimination, the seller cannot directly observe the buyer's preferences but can observe some characteristic that correlates with a consumer's willingness to pay for the product. For example, the seller may offer a good at a lower price in less affluent geographical areas.²⁰¹

Price discrimination and its equivalents are actually quite widespread. For example, the higher price for first-run movies is designed to extract surplus from those consumers who are the most eager to see a movie and unwilling to wait for its appearance in second-run theaters, on television, or in VHS or DVD format. Moreover, a number of other practices are substantially equivalent to price discrimination in economic terms. Tying arrangements, in which the buyer must purchase another product from the same vendor, often serve as forms of price discrimination. Sometimes the tie is technological rather than contractual. For instance, the manufacturer of a game console may use intellectual property laws to protect the interface between the console and the games, so that players will be forced to purchase games from it. The manufacturer can then sell the console cheap, capturing a wide segment of the market, while in effect charging a higher price to the most avid players, who will play the most different games and who would presumably have been willing to pay the most for a console in the first place.²⁰² A related practice is bundling, in which the buyer cannot purchase individual items except as parts of sets.²⁰³ Exclusive dealing arrangements can also be analogized to ties; in effect they require the purchaser who wants one product to buy all other products within the scope of the agreement from the same seller or forego them entirely.²⁰⁴

Both buyers and competitors often complain about tying arrangements and other forms of price discrimination, leading to antitrust disputes. On the other hand, a number of features of intellectual property law are conducive to price discrimination.²⁰⁵

200. *Id.* at 71-78.

201. *Id.* at 69-71.

202. *See id.* at 117-18.

203. *Id.* at 123-25.

204. *See* Richard A. Posner, *Antitrust in the New Economy*, 68 ANTITRUST L.J. 925, 936 (2001).

205. *See* Meurer, *supra* note 199, at 82-90.

2. Price Discrimination and the Efficiency Issue

The efficiency argument is superficially clear-cut.²⁰⁶ A monopolist normally imposes a uniform price that is higher than the competitive price.²⁰⁷ This maximizes the monopolist's profits but also reduces output, since he cannot sell as many units at the monopoly price as at the competitive price. The efficiency loss takes the form of lost consumption by buyers who would have been willing to pay the competitive price but are not willing to pay the monopoly price.²⁰⁸ If the seller can engage in first-degree price discrimination, however, he can sell goods to every buyer who is willing to pay above marginal cost (the competitive price), charging each one precisely what that specific buyer is willing to pay. So output rises back to the competitive level (though not at the competitive price), and the efficiency loss of monopoly pricing is averted.²⁰⁹ In the intellectual property area, the increased profit to the seller also adds to the incentive to innovate and thus benefits society as a whole.²¹⁰

In reality, however, the efficiency benefits of price discrimination are far less clear. There are at least five reasons why, to quote Judge Posner, "price discrimination has no general tendency to increase efficiency."²¹¹ The first reason, and the one he primarily relies on, is that perfect price discrimination is impossible, and imperfect price discrimination may either increase output above the monopoly level (good) or decrease it even further (bad).²¹²

Second, price discrimination is clearly to the disadvantage of high valuation consumers, who can benefit greatly if they can find a way to engage in arbitrage with low valuation consumers. This can result in socially wasteful expenditures—for example, in smuggling to overcome the seller's geographical pricing scheme or creating software patches to convert the

206. See Brodley, *supra* note 2, at 1033 (pointing to perfect price discrimination as a paradigm of the possible divergence between efficiency and consumer welfare).

207. See *supra* notes 195-98 and accompanying text.

208. See Boyle, *supra* note 137, at 2022.

209. See *id.* at 2022, 2035.

210. *Id.* at 2032 (recounting but not endorsing "the assumption that increasing the pricing power of the producer increases the amount of innovation and information produced").

211. Posner, *supra* note 204, at 932.

212. *Id.* at 932-33 n.10.

inferior version of a product into the more desirable form. The closer the scheme approximates perfect price discrimination, the greater the incentive for buyers to seek to evade it.

Third, in the context of intellectual property, price discrimination does increase the incentive to innovate, but there can be too much of a good thing. Just how much of its resources should society invest in innovation as opposed to providing existing goods and services is not clear, but it is at least arguable that we have pushed beyond the break-even point in some areas.

Fourth, price discrimination and related techniques like tying can be used to exclude rivals, who might otherwise offer superior products or lower prices to consumers. The antitrust orthodoxy of a few years ago was to the contrary. It now seems to be agreed that such anticompetitive effects are possible when these forms of behavior make entry costlier or riskier.²¹³ Price discrimination can also be used to finance predatory conduct.²¹⁴

Fifth, price discrimination is not equally feasible in all markets. Producers respond to their own expected surplus, not the social surplus. Thus, the effect can be to move resources from uses with higher social value to those with lower social value but greater amenability to price discrimination. These distortionary effects can be substantial and can outweigh the direct efficiency gain from price discrimination.²¹⁵

The upshot is that the efficiency of price discrimination is likely to vary in different situations and is also quite often unclear. In those situations, the proposed fairness presumption becomes relevant.

3. Price Discrimination and the Fairness Issue

Attempts to extract surplus from later adopters in network industries may hamper the growth of the network, limiting the benefit of the network to early adopters.²¹⁶ Ex ante, the seller would agree to forego such conduct, but opportunism becomes tempting ex post. The individual attempting to establish a standard has every reason to appear too "nice" to engage in exploitative conduct when the standard is still being established. Different incentives exist afterwards, however.

213. See Meurer, *supra* note 199, at 131.

214. See Posner, *supra* note 204, at 936.

215. See Lunney, *supra* note 55.

216. See *supra* notes 127-28 and accompanying text.

Holding the seller to the bargain he would have rationally agreed to *ex ante* is morally appealing.²¹⁷

It is also noteworthy that first-degree price discrimination can only take place because the seller disables himself from negotiating with high valuation buyers, who could otherwise threaten to walk away unless price concessions were made.²¹⁸ This effort to forestall a fair negotiation process with buyers seems itself to be ethically questionable.

The statutory interpretation argument also seems to apply to price discrimination.²¹⁹ Price discrimination is generally viewed with disdain. Consider the following comment about price discrimination:

Lay people often react to differential pricing for the same good with a sense of unfairness. No matter how many times they are lectured by the economists that it is actually to the benefit of all that producers be able to charge different prices to groups with different ability and willingness to pay, the popular reaction is normally "that's not fair."²²⁰

Indeed, this attitude is reflected in the Robinson-Patman Act, which singles out some forms of price discrimination for condemnation.²²¹ As a matter of statutory interpretation, then, a presumption against price discrimination except where justified on efficiency grounds seems warranted.

Finally, the possibility of price discrimination clearly feeds rent-seeking efforts. Recall the recent efforts of the content industry to obtain legislation enhancing its ability to engage in price discrimination.²²² Monopolies become more valuable where price discrimination is allowed,²²³ and the fight to obtain monopoly status becomes all the fiercer. This battle for legislative advantage is socially wasteful, not to mention disruptive of the democratic process.

Price discrimination by intellectual property holders should not necessarily be considered actionable. In construing intellectual property laws, however, and in dealing with borderline antitrust cases, there is no reason to give price discrimination the benefit of the doubt. On the contrary, the

217. *See supra* text accompanying note 129.

218. *See supra* text accompanying notes 114-18.

219. *See supra* Part II.C.

220. *See Boyle, supra* note 137, at 2038.

221. 15 U.S.C. § 13(a) (2000).

222. *See Burk, supra* note 94 (manuscript at 56).

223. *See supra* notes 197-201 and accompanying text.

presumption should be against expanding the ability of sellers to engage in price discrimination.

CONCLUSION

Although the language of fairness is a common part of the legislative and judicial history of American antitrust and intellectual property law, it has gone out of fashion in recent years among legal scholars influenced by economics. We hope we have shown that this shift in fashion has no firm grounding, not even within economics itself. There are good reasons for caring about distributive effects in fashioning legal policy to guide both areas of law. Considerations of philosophy, economic theory, statutory interpretation, and politics all suggest that producers should not receive an unfair share of surplus. That baseline presumption should be ignored only where it can be shown that efficiency requires giving a greater share to producers. We have briefly suggested how this rule can be applied to several questions on the interface between antitrust and intellectual property, particularly on issues related to industry standards.

Our analysis does not suggest any great sea change in the law. Indeed, considerations of equity and fairness guide judges working in these areas as well as others all the time. Contemporary scholars often tend to dismiss such language. We think the judges, not those scholars, have got it right. Our approach provides a better way of making sense of such legal language and more refined guidance to help apply such equitable concerns. We hope that our analysis will reinforce the judicial bulwark against both a distorted scholarly analysis of efficiency and the political pressures that come from rent-seeking by the producers of intellectual property.