2000

Shaping Competition on the Internet: Who Owns Product and Pricing Information

Maureen A. O'Rourke
Boston University School of Law

Shaping Competition on the Internet: Who Owns Product and Pricing Information?

Maureen A. O'Rourke[*]

---

* Professor of Law, Boston University School of Law. Many thanks to Jerry Reichman for inviting me to the conference, "Taking Stock: The Law and Economics of Intellectual Property Rights," and to the partic ants at that Conference for their comments and suggestions. Thanks also to Bob Bone, Joe E lley, Dan Burk, Jane Ginsburg, Keith Hylton, Vic Khanna, Mark Lemley, Mike Meurer, James Molloy, Raquel Ortiz, Rebecca Troendle Hewitt, Fred Yen, and to my research assistants Scott Katz, Aaron Feinberg, and Tamarah Belczyk for their help. The author provided comments on and signed an amicus brief in support of Bidder's Edge, Inc. in its litigation against eBay. She received no remuneration for this activity.

INTRODUCTION

Historically, markets have almost always fallen short of satisfying the conditions for and providing consumers with the benefits of perfect competition. Certain characteristics of electronic markets, however, enhance the possibility that e-commerce[1] will be conducted in an environment that comes closer to attaining the perfectly competitive ideal than that of most conventional markets.

Essentially, technology analogous to that which users already employ to search the Internet can also enable this retailing revolution by allowing consumers easily to obtain comparative product and pricing information. However, for a number of reasons, on-line merchants (e-tailers), are asserting a variety of legal claims that, if successful, will hamper consumers' abilities to use the most efficient tools to obtain this information. In particular, if courts adjudicating these claims apply existing property law governing

1. The stakes are high. While business-to-consumer commerce remains a small part of overall retail sales (.64% in the fourth quarter of 1999 ("4Q99")), it approximated an absolute value of $5.3 billion in 4Q99 and continues to grow. *See* U.S. DEP'T. OF COMMERCE, DIGITAL ECONOMY 2000 9 (2000).

tangible items to Internet activities without considering that medium's unique nature, they may inadvertently, but nevertheless effectively, confer upon web site owners exclusive rights to their product and pricing information tantamount to ownership.

One of the requisites for perfect competition is the costless exchange of information. If e-tailers can control the most efficient means of access to and the accompanying dissemination of their product and pricing information, they may impede movement toward the ideal. Why would they seek this control and how should the law respond? The answers to these questions, particularly the latter, will largely determine the nature of competition on the Internet, including the balance of power between producers and consumers.

Part I of this Article explains the theory of perfect competition, and both why the Internet could facilitate it and why it apparently has not yet done so. Part II considers why and how some sites are seeking to protect their otherwise publicly available product and pricing information from particularly those users obtaining it through automated means. Some of these sites' motivations are anti-competitive while others are not. The law must consider how best to structure competition on the Internet in light of both these conflicting purposes and the interests of consumers in having easy access to such information.

Part III examines the legal landscape, analyzing some of the claims sites are raising. This evaluation reveals the inadequacies of simply applying existing law to new technology. Part IV proposes a test to address complaints arguing that unwanted visits to a web site should be legally prohibited. It then briefly considers the broader normative question: "What should e-commerce look like?" Part IV concludes by contending that, at this early state of Internet technology, policymakers would do better to err on the side of weaker rather than stronger property rights. This decisional framework may change over time as the market develops, but provides a reasonable starting point.

I. PERFECT COMPETITION—THEORY AND THE INTERNET REALITY

The textbook definition of perfect competition is "[a] market structure in which all firms in an industry are price takers and in which there is freedom of entry into and exit from the industry."[2] A

2. RICHARD G. LIPSEY ET AL., ECONOMICS 915 (8th ed. 1987).

perfectly competitive market is one characterized not only by no barriers to entry or exit, but also homogeneous products and perfectly informed buyers and sellers able to meet with each other without incurring search or other transaction costs.[3] In such a market, price equals marginal cost,[4] keeping supply and demand in equilibrium.[5]

Most markets are not perfectly competitive. Real barriers to entry and exit exist, products are differentiated, information asymmetries persist, and transaction costs are non-trivial.[6] The relative strength of these factors' variations from the ideal determines where a market fits in the range between perfect competition and the other end of the spectrum—monopoly, where industry output is controlled by a single source that may price above marginal cost, imposing deadweight losses on consumers.[7] A change to a market that brings it closer to perfect competition is generally viewed as desirable.[8]

---

3. *See Frictions in Cyberspace*, ECONOMIST, Nov. 20, 1999, at 94 (describing an "ideal competitive market" as one in which "products are identical, customers are perfectly informed, there is free market entry, a large number of buyers and sellers and no search costs"); *How to Be Perfect*, ECONOMIST, Feb. 12, 2000, at 82

> [A perfect market] would allow all buyers and sellers to meet together, with full information about supply and demand. There would be no barriers to entering or leaving the market. And every buyer would be matched with the supplier that could best meet his needs . . . . And there would be no 'transaction costs,' such as time wasted seeking the right product.

Id.

4. Marginal cost is "[t]he increase in total cost resulting from raising the rate of production by one unit. Mathematically, [it is] the rate of change of cost with respect to output." LIPSEY ET AL., *supra* note 2, at 913.

5. *See Frictions in Cyberspace*, *supra* note 3; *How to Be Perfect*, *supra* note 3.

6. *See How to Be Perfect*, *supra* note 3 (explaining that while some financial markets approach perfect competition, defects exist even in those markets because all parties are not equally informed, and also noting that other markets are often characterized by slow-to-change prices that are offered by sellers on a take-it-or-leave-it basis).

7. *See* ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 39-40 (1988) (illustrating supply and demand curves under monopoly conditions); LIPSEY ET AL., *supra* note 2, at 914. Under antitrust law, an economic monopoly is defined as the ability to raise price above the competitive level or to exclude competitors. *See* United States v. E. I. Du Pont De Nemours & Co., 351 U.S. 377, 391 (1956) ("Monopoly power is the power to control prices or exclude competition.").

8. *See generally* JACK HIRSHLEIFER, PRICE THEORY AND APPLICATIONS 234 (4th ed. 1988) (illustrating that monopoly can be regarded as a hindrance to trade, and that the resulting reduction in the volume of production and exchange results in an efficiency loss compared to the results in a competitive market). Economists have, bowever, formulated a theory of the "Second Best." Briefly, "[t]he General Theory of Second Best [ ] holds that correction for one market imperfection will not necessarily be efficiency-enbancing unless we also simultaneous[ly] correct for all other market imperfections. This is because one correction may have unintended and unanticipated consequences that adversely impact market efficiency or transaction costs elsewhere." Thomas S. Ulen, *Courts, Legislatures, and the General Theory of Second Best in Law*

A. Toward Perfect Competition? Internet Technology and Markets

The Internet employs technology that makes realization of the conditions for perfect competition more likely. For example, search engine software allows users to sort through the vast amounts of information on the Internet quickly and easily, enabling them to make more informed purchasing decisions. By decreasing transaction costs and making information more accessible, search engines help Internet markets to more closely approximate the zero transaction cost and perfect information hallmarks of textbook perfect competition than "real" world markets.

Search engines work by periodically sending software called a "spider" or a "robot" out to crawl the web, making copies of web sites from which the software culls relevant information to use in building a searchable database.[9] When a user enters a query, the search engine scans its database, returning a list of sites likely to match the user's interests. Search engines vary in their quality, but the technology continues to evolve toward increasing accuracy.[10]

Specialized search engines called metasites can aggregate the search process in more or less depth.[11] Metasites offer relatively little original content, instead providing value by organizing the content of others in one or more ways. Some metasites provide indices into specific types of information. For example, TotalNEWS.com's web site consists primarily of hyperlinks to other news organizations.[12] It functions as a one-stop portal from which a user can travel to a number of different sites that actually perform the service of providing the news.

Metasites, including TotalNEWS.com, often combine indices with a search capability. For example, AuctionWatch.com, a metasite that provides an index into sites that conduct auctions (while

---

and Economics, 73 CHI.-KENT L. REV. 189, 220 (1998). Ulen also notes the sheer difficulty of quantifying the second best problem and integrating it into legal analysis, contending that institutional characteristics both militate against judges considering second best effects in common law cases and suggesting that the legislature is best suited to address them. *See id.*

9. *See* Maureen A. O'Rourke, *Fencing Cyberspace: Drawing Borders in a Virtual World*, 82 MINN. L. REV. 609, 623-24 (1998) (explaining how search engines operate and citing authorities).

10. While techniques for more accurately matching a user's query with sites that will actually be of interest are improving, the sheer volume of information on the web prevents search engines from being comprehensive. *See* Ryan Davis, *Study: Search Engines Can't Keep up with Expanding Net*, BOSTON GLOBE, July 8, 1999, at C1 (noting that "catch-all" search engines are having difficulty keeping up with the proliferation of web pages); *see also Search Engine Watch*, (last visited June 8, 2000), http://www.searchenginewatch.com (rating search engines and providing guidance on their use).

11. *See* Mike France, *Copyright on the Net: Who 'Owns' a Price*, BUS. WK., Dec. 13, 1999, at EB14.

12. *See TotalNEWS*, (last visited June 8, 2000), http://www.tetalnews.com.

not holding auctions itself), lets a user search for items of interest.[13] A user can enter the desired item, say, "Mickey Mantle autographed baseball." AuctionWatch will search its database built by copying information from actual auction web sites, and return a list of those sites that offer the autographed baseball and at what price.[14]

Because it offers comparison pricing data, AuctionWatch.com is also called a "shopbot." Generally, shopbots are software agents that use spiders to amass product and pricing information to allow comparison shopping.[15]

Other metasites combine a search capability with additional data aggregation, functioning as "aggregators of aggregators." For example, a user who visits metacrawler.com[16] may enter a search request that metacrawler will then pass on to a number of popular search engines, returning results from all of them.[17] Users would find this service valuable because any *one* search engine indexes only a fraction of web sites.[18] By aggregating the search process, metacrawler.com obviates the need for users to travel to a number of individual search engine web sites to enter the same query and sort through the results manually. Similarly, a shopbot aggregator is one that amasses the information of many shopbots, allowing a user to go to one site to obtain comparison pricing information from a number of different shopbots.[19]

More advanced intelligent agent software also decreases search costs by allowing a user to create a personalized

---

13. *See AuctionWatch.com*, (last visited June 5, 2000)http://www.auctionwatch.com.

14. *See* Victoria Slind-Flor, *E-Shoppers Gain, but EBay Loses*, NAT'L L.J., Jan. 31, 2000, at B6 (explaining that auctionwatch.com "uses an automated program, a 'bot' or 'spider,' to crawl through the millions of listings on eBay and other online auction houses, and assemble them on its own site"); *see also* Greg Stohr, *Ebay Faces Antitrust Probe over Curbs on Auction Search Sites: Wants Pay for Access*, NAT'L POST, Feb. 5, 2000, at D09 (explaining how software "crawlers" search other websites and giving an example of how a consumer might use AuctionWatch.com to find a low priced Pez dispenser).

15. *See* Daniel Fisher, *The Web: Spiders and Lies*, FORBES, Jan. 24, 2000, at 148, 148.

16. *See Metacrawler*, (last visitod June 8, 2000), http://www.metacrawler.com.

17. *See Customize Your MetaCrawler Search Options*, (last visited June 8, 2000), http://www.metacrawler.com/customize (allowing users to search one, some, or all of the following search engines—About.com, AltaVista, DirectHit, Excito, FindWhat, Google, GoTo.com, Infoseek, Kanoodle, LookSmart, Lycos, RealNames, Sprinks, Thunderstone, and WebCrawler).

18. *See* Davis, *supra* note 10 (describing a February 1999 study showing that "the overlap between engines remains relatively low").

19. *See* France, *supra* note 11 (explaining the operation of priceman.com, a shopbot metasite).

experience.[20] The user will receive only that information of interest to him or her, decreasing the cost to sort through the extraneous data that a less discriminating search engine may return.[21]

The advent of all of this technology is causing a paradigmatic shift in methods of retailing. Historically, the slow movement of information in the mass market led producers to make decisions before they had accurate data on consumer preferences: The consumer was a price and product taker.[22] By speeding the flow of information, the Internet can eliminate many inefficiencies in the conventional supply chain.[23] It also offers the possibility of "dynamic pricing" in which retail prices move freely to reflect the then current state of supply and demand.[24] Additionally, "choiceboard" software can allow consumers to become "product makers," designing their desired products from a menu of options presented by the supplier.[25]

Software tools can also enable buyers and sellers to aggregate into groups, enhancing the bargaining power of buyers with like needs.[26] At the same time, individual buyers can take advantage of shopbot technology to select the provider offering the lowest price.

By decreasing the costs for users to find and become informed about products in which they are interested, all of the aforementioned technology makes Internet markets more likely to

---

20. *See Intelligent Agents Increasingly Used in E-Commerce*, (last visited June 8, 2000), http://www.bot-spot.com/news/000327ecommerce.html (collecting a number of announcements explaining how companies are planning to deploy intelligent agents).

21. *See id.* (explaining how technology will allow consumers to select their favorito foods from neighborhood restaurants); *see also* Wayne Kawamoto, *The Good News About News Bots, at* http://www.botspot.com/dailybot/newsbot.html (last visitod June 8, 2000) (explaining how newsbot technology can "search the websites of newspapers, magazines, and e-zines to find articles that are of specific interest to [the searcher]").

22. *See* Adrian J. Slywotzky, *The Age of the Choiceboard*, HARV. BUS. REV., Jan.-Feb. 2000, at 40, 40 (noting that traditionally, manufacturing had to begin before suppliers had reliable information about demand); *see also* Chip Bayers, *Capitalist Econstruction*, WIRED, Mar. 2000, at 210, 212 (noting that an unpredictable supply and demand chain "can easily break down," leading producers to "hedge against fluctuations . . . [by using] a fixed list price").

23. *See* Bayers, *supra* note 22, at 212 ("A seamless network of consumers, vendors, and manufacturers means increased efficiency and communications, real-time sales analysis, a diminished need to stockpile excess inventory, and the eradication of many hurdles in the supply chain.").

24. *See id.; see also* Michael D. Smith et al., *Understanding Digital Markets: Review and Assessment, in* UNDERSTANDING THE DIGITAL ECONOMY 5 (Erik Brynjolfsson & Brian Kahin, eds., forthcoming), *available at* http://ecommerce.mit.edu/papers/ude (citing empirical studies finding that e-tailers change prices significantly more often than their bricks and mortar counterparts).

25. *See* Slywotzky, *supra* note 22, at 41 (discussing choiceboards and arguing that they will dominato in many industries, reshaping the nature of competition).

26. *See How to Be Perfect*, *supra* note 3.

approach perfect competition than conventional ones. The result should be both lower and more uniform prices. The reality, though, is something different.

B. *Price Competition on the Internet—Pricing Dispersion and the Power of the Brand*

Conflicting studies on price make it difficult to draw a definitive conclusion about price levels on the Internet other than that they are not systematically lower than in the bricks and mortar world.[27] Additionally, price dispersion on the Internet is not lower than in conventional markets: Prices vary by as much as 50% between e-tailers, with an average difference of 33% for books and 25% for CDs.[28] Several factors account for this disparity.

1. Barriers to Entry

Barriers to entry exist even on the Internet. There are "huge marketing and technical costs [in] setting up a business."[29] Further, once a consumer begins to buy from one e-tailer, it may face significant costs in switching to another.[30] Some e-tailers customize the shopping experience, leaving consumers reluctant to switch to another site that may not offer the same features.[31] Even if another online merchant were to offer a comparable experience, consumers may not wish to invest time in replicating the personalization already in effect on the first site. Some sites offer loyalty rewards

---

27. *See* Karen Clay et al., *Retail Strategies on the Web: Price and Non-price Competition in the Online Book Industry*, 4, 9-15, (last visited June 8, 2000) http://dnet.heinz.cmu.edu/dcsrg/books/papers/paper1.pdf (finding, in a week-long study of 107 hook titles, that online and real world book stores charged essentially the same prices, but that real world stores were cheaper overall because sales tax is generally lower than shipping costs). *But see* Smith et al., *supra* noto 24, at 3-4 (reviewing empirical studies showing that prices tended to be higher online in 1996-97 but that by 1998-99 they had dropped to a level 9-16% below that of bricks and mortar stores, and suggesting that this change could be a result of the Internet's becoming a more efficient market).

28. *See* Clay et al., *supra* note 27, at 4, 26 (finding large price disparities across booksellers ranging up to 73%); Smith et al., *supra* noto 24, at 6 (noting also that airline ticket prices differ by as much as 20%).

29. *Internet Economics: A Thinkers' Guide*, ECONOMIST, Apr. 1, 2000, at 64, 64 [hereinafter *Internet Economics*]; *see also* Clay et al., *supra* note 27, at 7 (quoting Amazon's CEO as saying, "[w]hile it is true that if all you want to do is put up something for sale, the barriers of entry are extremely low on the Internet. If you actually want to sell a lot of that stuff, they are quite high and getting higher all the time").

30. *See Frictions in Cyberspace*, *supra* note 3.

31. *See id.*; Clay et al., *supra* note 27, at 8-9 (describing various ways merchants try to avoid price competition by differentiating themselves).

(like the conventional airline frequent flyer programs), to repeat purchasers.[32] Also, the network effects that characterize some products like software can keep consumers "locked into" a single supplier that, as a result, can exercise considerable market power, including power over price.[33] Secondcomers then, may face material barriers to entry that partly account for the observed price dispersion.

2. Imperfect Information

For consumers to take advantage of comparison pricing tools like shopbots, they have to know that such tools exist, and be willing to spend time using them.[34] If either of these conditions does not hold, the consumer will not be perfectly informed about alternatives. Moreover, even if the consumer avails his- or herself of a shopbot, he or she is still unlikely to be perfectly informed because a particular shopbot indexes only a portion of the web.[35]

3. Search costs

Non-trivial search costs also prevent perfect competition from materializing on the Internet. Search costs are relatively lower on the Internet than in conventional media but are not, in an absolute sense, anywhere near zero. Search engines are less than totally accurate, tending toward indiscriminacy. One search for online bookstores returned 5,173,884 web pages that might be of interest.[36] However, search costs should continue to decline as tech-

---

32. *See* Clay et al., *supra* note 27, at 8-9.

33. *See Internet Economics*, *supra* note 29, at 64 ("In some industries, low marginal costs . . . and network effects . . . will result in increasing returns to scale, and thus the emergence of monopolies."); *see generally* Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CAL. L. REV. 479 (1998) (explaining network effects). Digital distribution, however, may require that software providers rethink their pricing strategies. *See* Smith et al., *supra* note 24, at 23-24 (noting the strategy of providing an "open source" and also stating that "value-oriented pricing strategies are likely to be more effective"); *see also* J. Bradford DeLong & A. Michael Froomkin, *Speculative Microeconomics for Tomorrow's Economy*, (Nov. 22, 1999) http://personal.law.miami.edu/~froomkin/articles/spec.htm (discussing the evolution of software marketing).

34. *See* Clay et al., *supra* note 27, at 6.

35. *See supra* note 10; *see also* Clay et al., *supra* note 27, at 6 n.7 ("It is important to note that search engines do differ in the amount of information provided and [sites] covered. Thus, even consumers who use a search engine may not have perfect information.").

36. *See* Smith et al., *supra* noto 24, at 11 n.5 (conveying the results of a search using the AltaVista search engine).

nology evolves and consumers become both increasingly aware of its existence and adept at its use.

4. Homogeneous products

Perfect competition also requires homogeneous products. Product differentiation on the Internet might account for the perceived variety in price. However, studies show that while heterogeneity may account for some dispersion, a substantial amount is "due to other unobserved retailer-specific factors such as brand, trust, and awareness."[37] Indeed, shopbot executives have indicated that their customers often buy from the brand they recognize even when it is not the low price leader: Amazon.com and CDNow, well-known e-tailers, "charge prices that are 7-12% higher than" those of other, lesser known e-tailers.[38]

As search engine technology improves, transaction costs will decline, potentially decreasing the power of a recognizable brand. Some commentators predict that this power will inevitably wane as users' trust moves to an intelligent agent personalized for their needs.[39] If trust shifts in this way, the only brands with value will be those of the software agent and the manufacturer of the product: The site from whom the agent makes the ultimate purchase and any others employed by it during the purchasing process will be transparent to the user.[40]

At the same time, however, e-tailers with established brands are unlikely to be receptive to erosion of their brands' power. In fact, a desire to maintain whatever market power the brand confers explains, at least in part, why some sites are seeking to control the flow of their product and pricing information.

---

37. Smith et al., *supra* note 24, at 10.

38. *Id.* at 12.

39. *See* Bayers, *supra* note 22, at 211.

40. *See id.* at 215.

> "In fact, the entire notion of a retail brand is out of place [on the Internet] . . . On the Web, you can collect user feedback and have an objective third party [intelligent agent] steer you toward the products you want to buy as an all-in-one service . . ." The only brands that survive . . . will belong to manufacturers and intermediaries. . . ."The value is at the edges of the network—the product on one end and the customer interface on the other."

Id.

II. CONTROLLING PRODUCT AND PRICING INFORMATION

Recently, the number of sites objecting to indexing by search engines, particularly metasites, has increased. Formerly, sites welcomed the search engines as tools making it easier for users to find them. As some sites have had success in establishing their brand names in the consumer consciousness, however, their need for the search engines has correspondingly decreased. In fact, the best-known sites may prefer that users travel directly to them rather than first going to a search engine from which they may choose to go to a competitive site.

In 1997, one-third of the sites from which the early generation shopbot BargainFinder sought product and pricing information took technological steps to block the shopbot's access.[41] Indeed, some shopbots today search only sites with whom they have agreements granting permission to index.[42] Of course, the value to consumers of a shopbot that searches only selected sites is lower than that of a more comprehensive one.[43]

EBay, the online auction site, has been one of the most vociferous opponents of metasites. It has taken both technological and legal steps to block auction indexing sites like AuctionWatch.com[44] and Bidder's Edge from using spiders to obtain information on the products offered at the eBay site.[45] Similarly, Ticketmaster has sued Tickets.com, trying to prevent Tickets.com from indexing in-

---

41. *See* DeLong & Froomkin, *supra* note 33 ("As of March 1997, one third of merchants targeted by BargainFinder locked out its queries."). One merchant objected on competitive grounds while another two indicated that they could not process the volume of "hobbyist" requests coming from Bargain Finder. *See id.*

42. *See* Jim McLollan, *Web Life: Shopping Around*, GUARDIAN, Oct. 14, 1999, *available at* 1999 WL 25738124 (noting that two reasons account for the failure of a shopbot to search all web sitos: (i) some sites block shopbots from searching their content; and (ii) some shopbots "will search only the online shops they have deals with"); Tricia Serju-Harris, *Net Shopping Firm Sues Priceman.com*, HOUS. CHRON., Sept. 16, 1999, at 2 (describing a lawsuit filed by a shopbot, mySimon.com against a shopbot aggregator, priceman.com, and quoting the president of the company that owns Priceman.com as saying, "[s]ome sites enter into merchant agreements and only search the merchants they sign up with").

43. *See* Serju-Harris, *supra* note 42 (noting that Priceman.com contends that if shopbots search only those sites with which they have agreements, "[t]hat doesn't give consumers the equal opportunity of other values on the Net").

44. *See* Louise Kehoe, *Networking—Online Information: Free, or Not So Free*, BUS. DAY, Feb. 10, 2000, at 20, *available at* 2000 WL 7451027 (describing how eBay and AuctionWatch engaged in a technological war as eBay blocked AuctionWatch's access and AuctionWatch circumvented the block).

45. *See* First Amended Complaint at ¶ 28, eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N.D. Cal 2000) (No. C-99-21200) (filed Jan. 14, 2000) (commencing legal action against Bidder's Edge for indexing the eBay site and describing how Bidder's Edge circumvented "technical measures" eBay had put into place to keep Bidder's Edge out).

formation about what tickets are available for purchase at the Ticketmaster web site, and from providing consumers with hyperlinks allowing them to travel to the interior pages of the Ticketmaster web site.[46]

"[A] virtual war over access to and control of price information is raging in the online marketplace."[47] In evaluating what the legal response should be, it is first helpful to consider why these disputes are occurring, and the non-legal steps that sites are taking to order competition on the Internet.

A. Why Object to Indexing?

One might intuitively think that online merchants would welcome any publication of their product and pricing information (whatever the source) because it increases the chance that consumers will visit their sites and make purchases. For a number of reasons, however, e-tailers are trying to control whether and how consumers obtain product and pricing information.

1. Maintaining Market Power

From a social perspective, perfect competition reflects the most efficient marketplace ordering and provides consumers with the greatest benefits. Firms, however, are concerned primarily with their own profits rather than aggregate social good. Given a choice, they would prefer to exercise some degree of market power and avoid competing solely on price. Market power allows a producer to capture a greater percentage of the consumer surplus than under perfect competition.[48]

The empirical data on price dispersion shows that firms with established brands are exerting some market power not attributable to quality differences. Such firms have an incentive to use that power to withhold information from consumers that, if provided to them, would bring the relevant markets closer to perfect competi-

---

46. *See* Second Amended Complaint at 16-24, Ticketmaster Corp. v. Tickets.Com, Inc., No. 99-7654, 2000 U.S. LEXIS 12987 (C.D. Cal. Aug. 10, 2000) (filed Apr. 21, 2000) (alleging a number of causes of action, including copyright infringement, breach of contract, and unfair competition for the manner in which Tickets.com indexes Ticketmaster's event information and hyperlinks users to the Ticketmaster site).

47. DeLong & Froomkin, *supra* note 33.

48. *See* HIRSHLEIFER, *supra* note 8, at 234 (demonstrating that part of the consumer surplus in a competitive market is transferred to the monopolist in the form of producer surplus in a monopoly).

tion.[49] Because technology is evolving in a direction that may erode brand power, this strategy may not be successful in the long-run. In the short-term, however, it may allow established sites to extract supra-competitive profits by erecting a barrier to entry and keeping consumer search costs higher than they would otherwise be. That this strategy may not endure is hardly a reason to allow use of the law as a sword to stop automated indexing in the interim, unless there is an offsetting, pro-competitive rationale justifying limits on the availability of product and pricing information.

2. Product Differentiation and Misleading Information

Price alone may not always be an accurate signal. An e-tailer might argue that consumers are misled when faced with only its high price compared to competitors' lower ones. The consumer may purchase from the provider with the lowest price because he or she is unaware that a higher price reflects better product attributes like warranties or services for which the consumer, if informed, would be willing to pay.[50] If unfettered price-only competition proceeds, the market will degenerate, offering low-quality products with shoddy service.[51] Thus, pricing information should be kept from the consumer because he or she lacks the capability to disaggregate it into its component parts.

This argument ignores both theory and fact. Theory would hold that if a shopbot does not provide all of the data relevant to a purchasing decision, consumers will not use it. Market forces should therefore lead shopbots to improve their technology to more accurately inform consumers. Indeed, over time, BargainFinder developed from a bare-bones system to one that provides output "that can be sorted by price, by speed of delivery, or by merchant. It can be further customized to take account of shipping costs by zip code, and results can be sorted by price or by speed of delivery."[52]

Available data also indicate that the product differentiation story simply may not be true, at least for highly homogeneous products like books and CDs. Also, surprisingly, one study found

---

49. *But see* HIRSHLEIFER, *supra* note 8, at 234-35 (noting the Theory of Second Best which holds that one change to a market, such as solving problems of information asymmetries, may not enhance efficiency if all other market imperfections are not simultaneously corrected).

50. *See* DeLong & Froomkin, *supra* note 33 (calling this the " 'kindly service' stery"); *see also How to Be Perfect*, *supra* note 3 (discussing dimensions other than price on which suppliers compete, including speed of delivery, reputation, and warranty, and describing technology that allows competition across many factors).

51. *See id.*

52. *Id.*

that "pricier retailers offer service terms, such as flexible returns rules, that are no better—and sometimes worse—than cheaper ones."[53]

In short, the consumer is better off knowing something about the product rather than nothing at all.[54] The Supreme Court shares this view. In *Bates v. State Bar of Arizona*, the Court addressed the claim that advertising by lawyers should be restricted:

> [It is true that a]dvertising does not provide a complete foundation on which to select an attorney. But it seems peculiar to deny the consumer, on the ground that the information is incomplete, at least some of the relevant information needed to reach an informed decision . . . . [T]he prohibition of advertising [ ] serves only to restrict the information that flows to consumers. Moreover, the argument assumes that the public is not sophisticated enough to realize the limitations of advertising, and that the public is better kept in ignorance than trusted with correct but incomplete information. We suspect the argument rests on an underestimation of the public. In any event, we view as dubious any justification that is based on the benefits of public ignorance.[55]

Likewise, in *National Society of Professional Engineers v. United States*, a Sherman Act case involving a ban on competitive bidding by an association of engineers, the Court said:

> It may be, as petitioner argues, that competition tends to force prices down and that an inexpensive item may be inferior to one that is more costly . . . . [C]ompetitive bidding for engineering projects may be inherently imprecise and incapable of taking into account all the variables which will be involved in the actual performance of the project. Based on these considerations, a purchaser might conclude that his interest in quality . . . outweighs the advantages of achieving cost savings by pitting one competitor against another.[56]

As the Court implies, quality conscious consumers may simply avoid the competitive process. But that is not a reason to withdraw from them, or others who are less concerned about quality, the opportunity to participate in that process if they so choose. Product differentiation then, is not a strong basis for granting control over product and pricing information to a web site.

Sites may claim, however, that shopbots mislead consumers in other ways. When a user queries a shopbot, the shopbot usually does not search against the contents of the entire web, but rather against its own database that is comprised of contents drawn from

---

53. *Frictions in Cyberspace, supra* note 3.

54. Certainly, if the consumer makes repeat purchases, he or she will discover relevant product differences besides price and adjust his or her behavior accordingly. Moreover, consumers will invest the time and money appropriate to the magnitude of their purchase in understanding how to allocate the price to product features.

55. Bates v. State Bar of Ariz., 433 U.S. 350, 374-75 (1977).

56. National Soc'y of Prof'l Eng'rs v. United States, 435 U.S. 679, 694 (1978) (citations omitted).

the web at a specific time.[57] Shopbots update their databases, but not in real time.[58] If sites change their prices, a consumer may receive an old quotation from a shopbot, and be disappointed when he or she tries to make a purchase only to discover that the quoted price is not the actual one. The consumer may blame the destination site for the discrepancy rather than the shopbot, eroding the destination site's goodwill.

While this concern has some merit,[59] it may be addressed in less restrictive ways than by legally barring shopbots from indexing and thereby withholding information from consumers. For example, shopbots could post a notice informing visitors that the shopbot itself is solely responsible for the accuracy of product and pricing information and that such data changes with time. Regardless, as more consumers use shopbots, only the more accurate ones should survive. The law should consider whether it should impose an affirmative disclosure obligation on the shopbot to warn against possible inaccuracies, but the mere fact that some errors occur does not support a legal prohibition against its use of spiders.

3. Protecting Computer Servers from Overburdening

A potentially pro-competitive justification for a site's desire to stop unwanted indexing is based on the burden that spiders place on the indexed system's resources. For example, eBay would argue that it is not seeking unprecedented control over its product and pricing information. It seeks only to limit the *manner* in which a user obtains that information.[60] Ebay is willing to allow unlicensed indexers to conduct real-time queries against its database, and for consumers to obtain product and pricing information from eBay di-

---

57. *See supra* notes 9-10 and accompanying text.

58. *See* First Amended Complaint at ¶ 23, eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N.D. Cal. 2000) (No. C-99-21200) ("Bidder's Edge copies the eBay category listing pages only periodically. The auctions on eBay are ongoing and the items listed, number of bids and the price of any item are subject to change at any time. The auction listings which appear on Bidder's Edge, therefore, frequently do not accurately reflect the current state of the auction on eBay and those listings can in fact be dramatically incorrect due te staleness."); *see also* Charles Piller, *Ebay Moves to Halt Indexing of Its Auctions by Outsiders*, L.A. TIMES, Oct. 9, 1999, at C1 (giving examples of inaccuracies in search results both from AuctionWatch and from eBay itself).

59. One might, of course, query how this is any different from what happens in the "real world." Magazines and other publications often quote prices that may turn out to be inaccurate at a particular time and place.

60. *See eBay*, 100 F. Supp. 2d at 1062 ("It appears that the primary dispute was over the *method* [Bidder's Edge] uses to search the eBay database. eBay wantod [Bidder's Edge] to conduct a search of the eBay system only when the [Bidder's Edge] system was queried by a [Bidder's Edge] user.") (emphasis added).

rectly or from "licensed" indexers.[61] It wants to stop unauthorized indexing because the technology employed by the spiders imposes a burden on its servers, decreasing response time to and availability of the system for "licensed" indexers and "legitimate" users interested in making a purchase at the site. Indeed, any site overburdened in this way may lose customers and eventually be forced out of business, even though it might be the most efficient producer in the marketplace.

If this is true, then it is a cause for concern. This suggests that the law minimally should consider the burden that spiders place on the systems they search and provide a remedy if that burden becomes "unreasonable."[62]

What is an unreasonable burden? EBay argues that Bidder's Edge's spider occupies about 1.53% of its system's capacity.[63] Standing alone, this is trivial. But if many other spiders were to roam the site, they could collectively burden the system enough to cause it to crash, or at least to force eBay into spending money to increase its capacity to forestall just such a crash.[64] Should the law provide a remedy based on an anticipation of future harm if unauthorized spiders proliferate? Internet economics suggests not.

The optimal number of spiders indexing a site may be greater than one, but it is likely also to be less than the number that would cause the system to crash. Many sites still make money by selling advertising.[65] Metasites, because they offer little or no

---

61. *See id.; see also* First Amended Complaint at ¶ 28, *eBay* 100 F. Supp. 2d at 1058 (noting that eBay has licensed others to index its site and offered such a license to Bidder's Edge). It is unclear whether the conduct which eBay would tolerate would, in fact, result in less of a burden for its system than the conduct to which it objects—i.e., unauthorized spidering. Note also that after the court enjoined Bidder's Edge from engaging in spidering, Bidder's Edge provided its users with information on eBay auctions by searching the eBay site in real-time. *See* Steven Bonisteel, *Bidder's Edge Searches eBay Again*, NEWSBYTES, June 8, 2000, http://www.newsbytes.com/pubNews/00/150308.html.

62. Another alternative is simply to provide sites with an entitlement to block all unlicensed spidering. *See infra* Part IV (discussing the normative framework that informs the selection of an entitlement system).

63. *See eBay*, 100 F. Supp. 2d at 1063 (stating that the 1.53% figure refers to the number of requests received by eBay from Bidder's Edge, and noting that eBay also argues that 1.10% of the data transferred by it during certain time frames was transmitted to Bidder's Edge).

64. Indeed, this possibility was a major motivation in the *eBay* court's willingness to grant a preliminary injunction. *See id.* at 1066 ("If [Bidder's Edge's] activity is allowed to continue unchecked, it would encourage other auction aggregaters to engage in similar recursive searching of the eBay system such that eBay would suffer irreparable harm from reduced system performance, system unavailability, or data losses.").

65. *See Internet Advertising Revenues Soar to $4.6 Billion in 1999*, (last visited May 31, 2000), http://www.iab.net/news/content/revenues.html [*hereinafter Internet Advertising Revenues*].

content of their own, rely particularly heavily on this model. Advertising rates are going down, in part because of the large supply of web sites.[66] Chances are high that the number of indexing sites that could attract enough money to remain in business is less than the number that would materially adversely affect system performance.

The law should therefore tread carefully in assessing claims that unauthorized spidering overburdens the indexed site's servers. Such claims may have merit[67] or may be asserted as a pretext to mask an anti-competitive intent. Further, even if unauthorized spidering results in an intolerable burden from a legal perspective, the law should consider the benefit the public derives from easy availability of product and pricing information from even unauthorized sources. The appropriate remedy to best protect consumer welfare may not always be to enjoin the spider. Instead, more creative redress like allowing the spider to continue its conduct but requiring it to pay the indexed site to expand that site's capacity may make sense.

4. Protecting Revenue Models

Another potentially pro-competitive justification for stopping unauthorized indexing is to protect the indexed site's revenue models. Web sites make money in a variety of ways in addition to the sale of products.[68] Besides selling advertising space,[69] many sites now also make money through commissions: They receive a fee from a site when they refer customers to it.[70] Some have also entered into

---

66. *See 10th Interactive Newspapers Conference Isn't Rosy, but Thorns Are Less Sharp*, NEWSINC., Mar. 1, 1999, *available at* 1999 WL 13223580 (attributing declining rates to "huge supply"); *see also* Thomas Hirschmann, *Punters Fancy Yahoo! Prospects*, NAT'L POST, Sept 10, 1999, at D02 ("Ad rates have been in a prolonged slump, with more companies chasing the available advertising dollars."), Donna L. Hoffman & Thomas P. Novak, *How To Acquire Customers on the Web*, HARV. BUS. REV., May-June 2000, at 179, 186 (noting that rates for banner ads declined from $70 CPM in 1994 to around $30 today).

67. *See supra* note 41 (explaining that small sites objectod to Bargain Finder's indexing because they could not process the volume of queries).

68. For an early discussion of web revenue models, see O'Rourke, *supra* note 9, at 625-30. The revenue models discussed there—advertising, subscriptions, cybershopping (or e-commerce), and commission arrangements—continue to be viable. The relative mix among the models as well as the technology implementing a particular model changes over time.

69. The advertising model is based on that pioneered by the television industry. *See* Jeffrey L. Sampler, *Ask the Tough Question Before Entering a Portal*, KNIGHT-RIDDER TRIB. BUS. NEWS, July 8, 1999, *available at* 1999 WL 17355522 ("The current justification for [advertising] revenue is analogous to the television advertising model—the more eyeballs, the higher the advertising rate.").

70. *See* Nicholas G. Carr, *Hypermediation: Commerce as Clickstream*, HARV. BUS. REV., Jan.-Feb. 2000, at 46, 46-47 (giving an example of how a user could pass through nine intermediaries on his or her way to a destination, with money changing hands with each click, and not-

linking agreements of various types. A popular destination site may be able to convince others to pay it for the privilege of hyperlinking to it or indexing its contents.[71] By providing a separate source of information unsanctioned by the destination site, "unlicensed" metasites undercut these revenue models in one way or another.

For example, advertising rates depend in some way on the number of users who see the ad.[72] If the user's entry into the web is through a metasite, the chances that the user will travel to the site on which the ad appears may decrease, thereby lowering the destination site's advertising revenue. Additionally, empirical data shows that advertisers are increasingly using their money to place ads on specialized sites tailored to the interests of their target markets.[73]

That traffic may be diverted before it reaches a particular site is not, however, a new problem. Unless the user types a site's URL directly into his or her web browser, he or she will pass through one or more sites on the way to the destination. When a metasite indexes product and pricing information to present comparison data to consumers, a particular consumer's failure to click-through to the destination site results from its informed decision not to purchase there. That this decision will mean less advertising revenue for the destination site selling the product hardly seems a compelling reason to prohibit indexing.

A commission arrangement is a revenue model for the referring rather than the destination site. In fact, under this model, "unlicensed" metasites could actually effectively increase the destination site's profit. When a user travels to the site through an unaffiliated metasite, the destination owes no commission. However, to protect their revenue streams, affiliate sites would prefer that users

---

ing that the volume of clicks and efficiency in processing transactions make "microtransactions" feasible).

71. EBay has entered into license agreements with some aggregators. *See* First Amended Complaint at ¶ 28, eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N.D. Cal. 2000) (No. C-99 21200).

72. The parties may set the advertising rato in any of a numher of ways, hasing it on the number of users who see the ad, the number wbo click-through, or the number wbo click-through and make a purchase from the sponsoring site. The latter metric most accurately measures the effectiveness of the ad in inducing purchase. According to the Internet Advertising Bureau, "hybrid deals accounted for 52% of revenue transactions for Q4 [1999], 53% for the year [1999] with CPMs or impression-hased deals at 40% for both Q4 and the year and performance-based deals at 8% of Q4 revenues, 7% for the full year." *Internet Advertising Revenues, supra* note 65.

73. *See Secondary Portals Getting More Focused*, IR. TIMES, Feb. 11, 2000, at 59 (citing forecasts tbat 57% of advertising revenue will move to more targeted sitos but noting that large portal sites like AOL, Yaboo!, and MSN can withstand some erosion in their profits).

travel through them rather than other, unaffiliated metasites. The destinations might therefore oppose unaffiliated sites simply to keep their current network happy.[74]

Some sites currently make money by charging other sites for the privilege of linking and/or indexing. Metasites that engage in such conduct for free clearly undercut the revenue stream the destination site enjoys.

Existing legal authority suggests that the mere act of including a hyperlink on a site does not implicate any legally cognizable right of the linked site.[75] That some sites pay for the "privilege" of linking to another may reflect conservative legal advice in an environment of some uncertainty and/or that the linking agreement provides it with benefits other than simply the right to link.[76] If, however, there is no legal right to stop a site from including a hyperlink, then the "unlicensed" link of the metasite, while it *does* undercut revenue derived from linking agreements, does not in itself produce a harm meriting legal redress.

The instant issue is whether there is a legal right to stop unauthorized indexing. If there were, sites could protect the revenue they derive from licensing the right to index. But to argue that *because* sites make money in this way, they *must* have an entitlement to stop "unlicensed" indexers bootstraps the result without any consideration of the normative implications of granting such a right. The point here is that the law should critically examine claims that unauthorized indexing constitutes impermissible free-riding on the indexed site's information simply because it undercuts a site's source of revenue. Generally, the law is primarily concerned with free-riding when it threatens the incentive to create and market the original product (as in the case of piracy or close similarity) than when it does not (as when the free-rider adds substantial value).

---

74. Interestingly, the commission model provides an incentive for metasites to seek an agreement with another site before indexing it. If the metasite attracts a sufficient volume of traffic, it might find the destination site willing to pay it for indexing it. As the commission model gains acceptance and other technological trends erode the brand power of established sites, "unlicensed" metasites may decline. On the other hand, metasites may lose the value of their independence if they affiliato with sites they index.

75. "[H]yperlinking does not itself involve a violation of the Copyright Act . . . since no copying is involved." Ticketmaster Corp. v. Tickets.Com, Inc. ("Ticketmaster I"), 54 U.S.P.Q. 2d (BNA) 1344, 1346 (C.D. Cal. 2000) (order on defendant's Motion to Dismiss the First Amended Complaint). However, this order was superseded by a minute order issued August 10, 2000 that did not discuss the validity of hyperlinking. *See* Ticketmaster Corp. v. Tickets.Com, Inc. ("Ticketmaster II"), No. CV 99-7654, 2000 U.S. Dist. LEXIS 12987 (C.D. Cal. Aug. 10, 2000).

76. *See* O'Rourke, *supra* note 9, at 643-44 (discussing how businesses use linking agreements "to establish complementary business and marketing relationships, provide for reciprocal links, establish performance requirements, and clarify liability issues").

Ultimately, of course, courts cannot divorce the resolution of the relevance of such diversion from the normative question of how entitlements should be defined on the web—a question to which this Article later returns.[77]

The preceding evaluation of sites' motivations for blocking spiders does suggest that courts should take a skeptical view of claims that such control is necessary to protect consumers. However, courts should consider the real nature of the burden "unlicensed" spidering places on the system as well as the impact the spider has on the site's revenue stream. They should balance these private costs against the public benefit accruing from easy availability of comparison product and pricing information. The more threatening the spidering is to the very existence of the web site, the more likely it should be policed in some way.

B. How to Prevent Indexing Technologically

Sites seeking to block indexers do not have to rely on the law to enjoin a metasite. They can take technological measures to block the access of those whom they wish to exclude. These technological measures can be more or less complex. For example, a site may include a robot exclusion header in its robots.txt file.[78] This header signals a site's unwillingness to be searched by spiders. A spider encountering it should, therefore, desist from indexing the site. For this technique to be successful, the spider must voluntarily choose to respect the header.[79]

Other technological measures are less dependent on the spider's goodwill. A site is usually able to detect a spider's presence if an unusually large number of requests for information are coming from one server. Every server has a unique "Internet Protocol" ("IP") address. The site can investigate the particular IP address associated with a large number of queries, then implement technological steps to block requests originating there from being proc-

77. *See infra* Part IV (discussing the normative framework that informs the selection of an entitlement system).

78. *See* eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1061 (N.D. Cal. 2000) (defining a robot exclusion header as "a message, sent to computers programmed te detect and respond to such headers, that [the site] does not permit unauthorized robotic activity").

79. *See id.; see also id.* at 1062 (quoting an e-mail stating that "well-behaved robots are still expected to check the robots.txt file" and respect any exclusion).

essed.[80] However, by routing requests through another's sever (called a "proxy server"), the indexing site can evade an IP block.[81]

A site could use password or encryption technology more effectively to protect its content from unwanted visitors. However, in their current form, these devices can't discriminate between spiders and other, "legitimate" users. Moreover, while they may slow the spider down, they may also make access difficult enough for "legitimate" users that these visitors will instead travel to an easier to access open site.

The risk with any technological measure is that the blocked site will simply break through the technological barrier, causing the original site to build a higher wall, and so on, until the parties reach some equilibrium point in this technological arms race. A site blocking access would likely argue that this arms race is not the best use of resources. Rather, it may assert that even if a spider may enter a publicly available web site, once that site takes some technological step to prevent access, the spider cannot legitimately circumvent that protective measure. By bifurcating the questions of access and use, the blocking site has a better chance of effectively stopping use even of uncopyrightable data: Impermissible access contaminates any use, even those the copyright law would allow.

If the law upholds such access constraints, it may inadvertently hamper competition and create exclusive rights akin to intellectual property rights regardless of the status of the "locked" information under the intellectual property laws. A brief analysis of the causes of action that sites are asserting reinforces the contention that the crucial policy issue for the immediate future is less

---

80. *See id.* at 1061 (describing how eBay investigates robotic activity and blocks it by ignoring requests from the originating IP address).

81. *See id.*

> Organizations often install "proxy server" software on their computers. Proxy server software acts as a focal point for outgoing Internet requests . . . Typically, organizations limit the use of their proxy servers to local users. However, some organizations, either as a public service or because of a failure to properly protect their proxy server through the use of a "firewall," allow their proxy servers to be accessed by remote users. Outgoing requests from remote users can be routed through such unprotected proxy servers and appear to originate from the proxy server. Incoming responses are then received by the proxy server and routed to the remote user. Information requests sent through such proxy servers cannot easily be traced back to the origination IP address and can be used to circumvent attempts to block queries from the originating IP address. Blocking queries from innocent third party proxy servers is both inefficient, because it creates an endless game of hide-and-seek, and potentially counterproductive, as it runs a substantial risk of blocking requests from legitimate, desirable users who use that proxy server.

Id. (citations omitted).

what constitutes permissible use of information than what constitutes allowable access.

III. LEGAL CHALLENGES TO ACCESS AND USE OF PRODUCT AND PRICING INFORMATION

There are a variety of both federal and state claims that sites could tenably assert to protect their product and pricing information. This Article focuses on those that go to the claim that indexing such information obtained from a publicly available web site without permission is an unauthorized access and/or use of that information.

A. Federal Claims

1. Copyright Infringement

At first glance, the claim that the copying of product and pricing information is copyright infringement is untenable.[82] Such information is an uncopyrightable fact.[83] Reproduction of it therefore cannot constitute copyright infringement.

However, the manner in which search engines work likely implicates copyright rights. Most web sites contain at least some copyrightable data. Search engines make "wholesale" copies of web site content, including the copyrightable, then extract the data they want in creating their own databases. This intermediate copying may constitute infringement of the exclusive right to reproduce.

But a court may excuse an infringement as fair.[84] In evaluating a claim of fair use, a court balances a number of factors,[85] in-

82. *See* Delong & Froomkin, *supra* note 33 ("The claim that E-bay's auction prices or the details of the sellers' offers to sell are protected in the U.S. by copyright, trade secret, or other intellectual property law is bogus."). Noto also that eBay doesn't even own all of the information it presents: Much of it is provided by the parties who offer items for bid. *See* David Lazarus, *AuctionWatch Rams Through EBay's Information Blockade*, S. F. CHRON., Jan. 19, 2000, at D1 (quoting an AuctionWatch spokesman as stating that eBay's content is "not even generated by eBay . . . [but] generated by their user base[,]" and noting eBay's response "conceding that descriptions and photes submitted by users are indeed the property of the submitters," but arguing that "additional information, such as number of bids received and length of bidding, is produced hy eBay and remains eBay's property").

83. *But see* CDN, Inc. v. Kapes, 197 F.3d 1256, 1260 (9th Cir. 1999) (holding prices in a guide to collectible coins copyrightable because the publisher used a creative, original process to arrive at the valuations).

84. *See* 17 U.S.C. § 107 (1994).

85. Under the Copyright Act, courts may weigh "(1) the purpose and character of the use . . .; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion

cluding market harm to the copyright owner and the public benefit. Under the video game cases of *Sega Enterprises. Ltd. v. Accolade, Inc.*,[86] *Atari Games Corp. v. Nintendo of America, Inc.*,[87] and *Sony Computer Entertainment, Inc. v. Connectix Corp.*,[88] when an infringer copies copyrightable material as a step in extracting the uncopyrightable, its copying may be excused as fair. A fair use finding is particularly likely if the infringer's end use is "transformative" in some way, and courts remain free to consider the public benefit.[89]

For example, in both *Sega* and *Atari*, the copying was a step in the defendants' marketing of independently created video games that ran on the plaintiffs' game consoles.[90] In *Sony*, the copying resulted in the creation of software allowing video game owners to play their games on a PC in addition to Sony's Playstation.[91] Although this copying could undercut Sony's Playstation sales, the court found it to be transformative because it allowed users to play games on a new platform.[92]

Courts have just begun to consider fair use claims relating to the use of copyrighted information on the Internet. In *Kelly v. Arriba Soft Corp.*, the district court held that a search engine that returned images copied from the indexed web sites in response to a user's search query made a fair use of the copyrighted images.[93] The court emphasized that the defendant's use was quite different from the plaintiff's because its purpose was functional rather than aesthetic, and was designed to make access to Internet information easier.[94]

In contrast, the same district court refused to make a finding of fair use in allowing a different case to proceed. The Los Angeles Times and The Washington Post sued Free Republic for operating a

---

used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work." *Id.*

86. Sega Enters. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1527 (9th Cir. 1992).

87. Atari Games Corp. v. Nintendo of Am., Inc, 975 F.2d 832, 842 (Fed. Cir. 1992).

88. Sony Computer Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 602 (9th Cir. 2000).

89. *See, e.g., Sega*, 977 F.2d at 1523 ("[W]e are free to consider the puhlic benefit resulting from a particular use notwithstanding the fact that the alleged infringer may gain commercially.").

90. *See Atari*, 975 F.2d at 842; *Sega*, 977 F.2d at 1527.

91. *See Sony*, 203 F.3d at 607.

92. *See id.*

93. Kelly v. Arriba Soft Corp., 77 F. Supp. 2d 1116, 1121 (C.D. Cal. 1999).

94. *See id.* at 1119. *But see* Infinity Broad. Corp. v. Kirkwood, 150 F.3d 104, 108 (2d Cir. 1998) (holding that while a difference in purpose may seem to support fair use, the legal inquiry focuses on the question of transformation). A "difference in purpose is not quite the same thing as transformation, and [the Supreme Court] instructs that transformativeness is the critical inquiry under this [first fair use] factor." *Id.*

bulletin board site at which members posted news stories from the papers along with their comments and criticisms for the purpose of inducing discussion.[95] Free Republic argued that it had a fair use right to post the news to encourage debate and criticism, and also disputed that it undercut the newspapers' revenues.[96] The judge indicated that the combined commercial nature of Free Republic's enterprise, its effect on the newspapers' revenues, and its copying of the stories in their entirety made its fair use claim unlikely to succeed.[97] In particular, the court found "little transformative about copying the entirety or portions of a work verbatim[,]" despite the addition of comments, criticism, and discussion.[98]

A spider's copying is also non-transformative as its value inheres not in adding content but in accurately reproducing product and pricing information. However, as in *Arriba Soft*, the spider adds value by making the Internet easier to navigate. Moreover, the spider's copying of web site contents as a step in extracting product and pricing information is more like that involved in the video game cases than in either *Arriba Soft* or *Free Republic*: The information used in the spider's eventual offering to consumers is only the uncopyrightable. The search engine never displays the copyrighted information to the public. Instead, it links to the copyright owner's site. This situation is distinct from that of *Free Republic* where the infringing site offered verbatim copies of copyrighted information. It more closely resembles the video game cases involving intermediate copying. That the uncopyrightable information is not "transformed" before its end use should not bar a finding of fair use particularly since the infringer's end use significantly benefits the public by decreasing consumers' search costs. Courts are therefore likely to excuse this copying as fair unless it so burdens the system as significantly to impair the value of the overall copyrighted work. Indeed, in a minute order denying a preliminary injunction, the Ticketmaster court indicated that, by analogy to *Sony*, fair use would likely shelter Tickets.com's copying of the Ticketmaster site to extract uncopyrightable, factual information.[99]

---

95. *See* Los Angeles Times v. Free Republic, 54 U.S.P.Q. 2d (BNA) 1453, 1454 (C.D. Cal. 2000) (granting plaintiffs' Motion for Partial Summary Judgment and denying defendants' Motion for Summary Judgment).

96. *See id.* at 1461.

97. *See id.* at 1466-72.

98. *Id.* at 1460.

99. *See* Ticketmaster, Corp. v. Tickets.Com, Inc. ("Ticketmaster II"), No. CV 99-7654, 2000 U.S. Dist. LEXIS 12987, at *4, *11-13 (C.D. Cal. Aug. 10, 2000) (noting that the court did not intend its order to be a published opinion, but nevertheless reporting the court's statement that

2. The Digital Millennium Copyright Act ("DMCA")

The DMCA makes it illegal to "circumvent a technological measure that effectively controls access to a work protected" by copyright.[100] As noted above, most sites will contain some copyrighted material. If the technical solution a site adopts to keep an indexer out is considered "a technological measure that effectively controls access," then the indexer's bypassing of that measure could violate the DMCA.[101]

A technological measure is an access control if it "requires the application of information, or a process or a treatment, with the authority of the copyright owner, to gain access to the work."[102] The meaning of this definition is far from clear but it seems primarily geared toward hacking[103] around technological protection devices that require some sort of "key" to open them that is within the control of the copyright owner.[104] Neither a robot exclusion nor blocking an IP address seems to be an access control: Gaining access to the work by "going around" either does not require information from the copyright owner or replicating a key provided by the copy-

---

because the information Tickets.com copied from Ticketmaster is not used in competition with Ticketmaster and is destroyed after the factual, unprotected information is extracted, a fair use defense is likely to succeed).

100. 17 U.S.C. § 1201(a)(1)(A) (Supp. IV 1998).

101. *Id.* While the prohibition against the distribution of devices enabling circumvention took effect with the Act's passage, the ban on the act of circumvention will not be effective until October 2000, two years after the Act's enactment. *See id.*

102. *Id.* §1201(a)(3)(B). Note that the definition of "circumvent[ing] a technological measure" includes conduct like bypassing or avoiding it. *See id.* § 1201(a)(3)(A). Ignoring a robot exclusion or avoiding an IP block does not violate the Act, however, because neither implicates an access control measure.

103. Unfortunately, the term hacking is not susceptible of a precise definition. To programmers, a hacker is a "clever programmer." *See whatis?com, at* http://www.whatis.com (last visited June 9, 2000) (definition found by searching the IT-specific encyclopedia under the term "hacker"). A "cracker" is one who attempts to break into a computer system. *See whatis?com, at* http://www.whatis.com (last visited June 9, 2000) (definition found by searching the IT-specific encyclopedia under the term "cracker" or using link under the torm "hacker"). However, in common parlance, a hacker is one who tries to break into a computer system. *See whatis?com, at* http://www.whatis.com. This Article uses the term "hacker" in this latter sense.

104. *See, e.g.*, CSC Holdings, Inc. v. Greenleaf Elecs., Inc., No. 99 C 7249, 2000 WL 715601, at *6 (N.D. Ill. June 2, 2000) (holding the plaintiff authorized to maintain an action under the DMCA where defendants provided unauthorized converter-decoders to unscramble cable signals); Universal City Studios, Inc. v. Reimerdes ("Universal City Studios I"), 82 F. Supp. 2d 211, 214, 216-17 (S.D.N.Y. 2000) (holding that hacking of the Content Scramble System, "an encryption-based security and authentication system that requires the use of appropriately configured hardware such as a DVD player or a computer DVD drive to decrypt, unscramble and play back, but not copy, motion pictures on DVDs," violated the DMCA); Real Networks, Inc. v. Streambox, Inc., No. 2:99CV02070, 2000 WL 127311, at *7 (W.D. Wash. Jan. 8, 2000) (holding that circumventing a system that supplies information for an authentication sequence likely violates the DMCA).

right owner. Thus, ignoring the exclusion or using a proxy server is unlikely to violate the DMCA.

If, however, "going around" these devices *was* a violation of the access provisions, or if the spider were more obviously to fall within the statute by breaking into a password-protected or encrypted system, the indexing web site would likely violate the Act. The DMCA ostensibly does not excuse an unauthorized access even when the eventual use of the information is fair.[105] The legislative debates characterized unauthorized access as "the electronic equivalent to breaking into a locked room to obtain a copy of a book."[106] That one might make a fair use of the book's content does not excuse the unauthorized access.

Essentially, through its definitions, the DMCA has protected a web site from unauthorized access when the site owner has taken relatively costly steps (both in implementing the technology and in risking the loss of customers), to exclude the unwanted.[107] By adopting a "fence" of a certain height then, a web site signals that unwanted access imposes costs on it that justify the DMCA's respecting that fence. The public benefit is not obviously considered, but one might argue that the DMCA implicitly assumes that it is not so great as to warrant permitting circumvention of the access controls. An information provider might not post a web site at all if it cannot erect a legally enforceable barrier to entry.

As technology becomes increasingly sophisticated and it becomes easier for a site to block access selectively using a technological measure as defined by the Act, the DMCA may warrant a second look. As access control technologies become easier to implement, the DMCA's original balance may be upset in a way that unintentionally and unduly impedes competition. However, as written, it

---

105. *See* Universal City Studios, Inc. v. Reimerdes ("Universal City Studios II"), 55 U.S.P.Q. 2d 1873 (S.D.N.Y. 2000), *available at* No. 00 Civ. 0277, 2000 WL 1160678, at *17-18 (stating that the DMCA's legislative history indicates that fair use is not a defense to a claim under the anti-circumvention provisions); Pamela Samuelson, *Intellectual Property and the Digital Economy: Why the Anti-Circumvention Regulations Need to Be Revised*, 14 BERKELEY TECH. L.J. 519, 539-40 (1999) (explaining how the DMCA's statement that it does not affect fair use rights, as stated in 17 U.S.C. § 1201(c)(1) (1994), squares with its provisions forbidding circumvention of access protection devices).

106. *See Samuelson, supra* note 105, at 539 n.111 (quoting the House Manager's Report).

107. Of course, while these steps may be costly today, they may not be so tomorrow. Note also that the approach of requiring that an entity take steps to block access as a condition of its right to exclude is used in trade secret law. Under the Uniform Trade Secrets Act, information cannot be a trade secret and protected against misappropriation unless it "is the subject of efforts that are reasonable under the circumstances to maintain its secrecy." UNIF. TRADE SECRETS ACT § 1(4)(ii) (amended 1985), 14 U.L.A. 438 (1990).

should not prevent unauthorized spidering achieved through ignoring a robot exclusion or using a proxy server.

3. The Computer Fraud and Abuse Act ("CFAA")—A DMCA for Uncopyrightable Data?

The CFAA's legislative history reveals that it was intended to address hacking to access private, confidential information.[108] The statute's language, however, is much broader:[109] "Whoever . . . intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains . . . information from any protected computer if the conduct involved an interstate or foreign communication" can be held criminally and civilly liable under the CFAA.[110] " 'Exceed[ing] authorized access' means to access a computer with authorization and to use such access to obtain or alter information in the computer that the accessor is not entitled so to obtain or alter."[111] To prevail in a civil action, the aggrieved party must suffer "damage," defined as "any impairment to the integrity or availability of data, a program, a system, or information, that [ ] (A) causes loss aggregating at least $5,000 in value during any 1-year period to one or more individuals."[112]

A spider's access is "intentional." Even if the indexed site has not taken steps to block a spider's entry, it may argue that the spider "exceeds authorized access" because the site is intended to attract traffic interested in purchasing goods and services, not in indexing the site's contents. Thus, the spider uses its authorized access to the publicly available site to gather information it is not entitled to obtain. This argument is not particularly persuasive. The

---

108. *See* S. REP. NO. 104-357, at 7 (1996) (stating, in the context of explaining what "information" is, that "the premise of this suhsection is privacy protection"); *see also id.* (referring to "hacker" intrusions).

109. *See* Shaw v. Toshiba Am. Info. Sys., Inc., 91 F. Supp. 2d 926, 936-37 (E.D. Tex. 1999). The Shaw Court upheld a claim under the CFAA based on the sale of defective disk controllers by defining "transmission" as not requiring computer to computer communication but rather as extending to selling controllers with faulty microcode. *See id.* The court rejected Toshiba's argument that Congress intended the CFAA to apply to hackers, not to manufacturers of devices. *See id.* ("[T]his court does not see a blanket exemption for manufacturers in [the statute]; nor does it see the term 'hacking' anywhere in this statute.").

110. 18 U.S.C. § 1030(a)(2)(C) (Supp. II 1996). "[T]he term 'protected computer' means a computer . . . which is used in interstate or foreign commerce or communication." *Id.* § (e)(2)(B).

111. *Id.* § 1030(e)(6).

112. *Id.* § 1030(e)(8) (also defining damage to include actual or potential modification or impairment to medical care, physical injury, and threats to public health or safety).

spider's entrance is authorized and its use of the information would not violate the Copyright Act.[113] Therefore it would be difficult for the site to contend that the spider is not entitled to obtain the information it does. The CFAA claim would likely only succeed if the law were to accord some users (i.e. those unwanted by the site), of a publicly available web site less than all of the rights they would have under copyright law. This seems unlikely because there is simply no wording in or legislative intent behind the CFAA that allows it to discriminate between users in this way.

An indexed site might alternatively argue that a spider (particularly after the site objects to its presence or erects a barrier to entry), does not have authorized access to the site. It then violates the CFAA merely by accessing the site to obtain information. In the absence of some other law giving the spider an entitlement to enter, this claim is quite tenable, particularly because there is no provision in the CFAA that permits considering the public benefit that may accrue from allowing even certain "unauthorized" accesses. Whether a spider violates the CFAA then, largely depends on how the law defines property rights in a web site and therefore what constitutes "unauthorized access."[114] Arguably, regardless of how the law defines those rights, liability should not accrue under the CFAA, because the statute was not intended to deal with developing issues in e-commerce. Indeed, the risk of CFAA liability may chill the development of new tools to conduct e-commerce.

Even if the indexed site could establish a violation of the CFAA, it would still have to prove damages. The site would likely argue that by burdening the system, the spider's use impairs the system's availability to "legitimate" users. This argument, along with the question of whether the site suffers the $5,000 in damage required to sustain a civil action, presents factual questions on which the site may or may not be able to sustain its burden.

Some sites derive revenue from indexing agreements under which they charge spiders for indexing their sites. A site might prove damage by alleging that "unlicensed" spiders deprive it of the revenue it has a right to expect. But this is not the type of damage that the CFAA envisions. Moreover, arguing that this damage justifies a remedy simply begs the question: The site has no entitlement to charge for indexing unless the law gives it one.

---

113. *See supra* Part III.A.1.
114. *See infra* Part IV.

B. State Claims

That the critical issue determining liability for indexing is how to define entitlements in web sites becomes even clearer when one considers the state claims that an indexed web site is likely to raise. While federal intellectual property law has historically provided the primary set of rules geared toward optimizing innovation, state common law also affects innovation incentives, particularly in times of rapid technological development. Because the speed of technological change far outpaces the congressional ability to address the propriety of amendments to the federal scheme, flexible state common law may be quite influential in shaping the direction of new technology. Such law may set the rules regarding both access to web sites and use (particularly of uncopyrightable data), unless and until Congress decides to act.

1. Trespass to Chattels

The ancient cause of action of trespass to land holds a person liable for "[a]ny intentional use of another's real property, without authorization and without a privilege by law to do so . . . without regard to harm . . . . Any physical entry upon the surface of the land is a trespass . . ."[115] This broad cause of action is justified, at least in part, as required to maintain the exclusive right of possession and use of one's real property.[116]

Indeed, in the physical world, a shopkeeper can bar a customer from entry for any reason,[117] including because the customer would like to inventory the store's prices for comparison shopping

---

115. W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS 70 (5th ed. 1984).

116. *See id.*

> The right to the exclusive use of property necessarily implies the correlative right to demand that others who wish to use the property should pay for that use . . . Various policy reasons have been given to justify the position that unprivileged and unauthorized intentional entries should be actionable without damage, including the prevention of the acquisition of a prescription right, to settle a dispute regarding title, to vindicate a property right, and to avoid breaches of the peace.

Id.

117. Public accommodation laws may limit a store's ability to exclude customers based on, for example, race or gender. *See* Joseph William Singer, *No Right to Exclude: Public Accommodations and Private Property*, 90 NW. U. L. REV. 1283, 1289-90 (1996); *see also* James L. Fennessy, Comment, *New Jersey Law and Police Response to the Exclusion of Minority Patrons from Retail Stores Based on the Mere Suspicion of Shoplifting*, 9 SETON HALL CONST. L.J. 549, 552-53 (1999) (noting that New Jersey may be the only state te hold that "businesses which are open to the general public cannot arbitrarily or unreasonably abridge any individual's common law right te access the property").

purposes.[118] By analogy, an indexed site might argue that it has an absolute right to exclude whomever it wants from its "store" and to revoke its invitation to enter should a patron engage in activity to which the storeowner objects.

At least some logic supports the notion that if the law respects physical boundaries, it should also respect virtual ones. However, the policy that supports protecting against physical intrusions is based, at least in part, on preventing breaches of the peace. One often has some financial and psychic attachment to one's land, particularly if the land is used for personal rather than business purposes. The landowner may view a physical intrusion as threatening, and be tempted to respond with force. That the intruder will be liable for trespass both deters his or her conduct, and makes it less necessary for the landowner to resort to violence. The virtual intrusion of the spider is not directly analogous to the physical invasion against which trespass to real property protects. It lacks the immediacy and opportunity for physical confrontation that provides a policy basis for the trespass cause of action.

Litigants may recognize the strained nature of this analogy.[119] They have, of late, been relying on a trespass to chattels theory of liability to regulate unwanted use of system resources, particularly that arising from unsolicited bulk e-mail ("spam").[120] A trespass to chattels occurs whenever one intentionally "dispossess[es] another of [a] chattel, or [ ] us[es] or intermeddle[s] with a chattel in the possession of another."[121] Dispossession occurs when another intentionally takes possession of the chattel or assumes

118. *See* Mosher v. Cook United, Inc., 405 N.E.2d 720, 720-22 (Ohio 1980) (holding that a store had the right to revoke a comparison shopper's license to enter the store's land even though the shopper did not engage in disorderly conduct).

119. Ticketmaster, however, does appear to be suing Tickets.com on a theory akin to virtual trespass. *See* Second Amended Complaint at 21-22, Ticketmaster Corp. v. Tickets.Com, Inc. ("Ticketmaster II"), No. 99-07654, 2000 U.S. Dist. LEXIS 12987 (C.D. Cal. Aug. 10, 2000) (alleging that Tickets.com committed trespass against not only Ticketmaster's computer systems hut also against the "Ticketmaster Web Site"). Note also that one amicus brief filed on behalf of eBay in its suit against Bidder's Edge argues that the analogy to real property is persuasive rather than strained, justifying the importation of real property principles to order relationships on the Internet. *See* Brief of Amici Curiae, Reed Elsevier, Inc., The National Association of Realtors and The e-commerce Coalition, Supporting Affirmance of the Order and Opinion of the District Court at 6-12, eBay Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N.D. Cal 2000) (No. C-99 21200) (filed July 12, 2000).

120. For a summary of the spam cases, see Dan L. Burk, *The Trouble With Trespass*, 4 J. SMALL & EMERGING BUS. L. 27, 28-32 (2000).

121. RESTATEMENT (SECOND) OF TORTS § 217 (1965).

"complete control and dominion" over it.[122] "Intermeddling" means intentionally bringing about a physical contact with the chattel.[123] For liability for intermeddling to accrue, the interference must be "harmful to the possessor's materially valuable interest in the physical condition, quality, or value of the chattel, or [ ] the possessor [must be] deprived of the use of the chattel for a substantial time . . ."[124] Trespass to chattels provides the aggrieved party with a cause of action for an interference not substantial enough to constitute conversion.[125]

Unlike trespass to land, trespass to chattels requires that there be some actual damage suffered by the plaintiff as a result of the trespass.[126] Generally, that the "dignitary interest in the inviolability of chattels" is somewhat less than that associated with exclusive rights in land, accounts for the more stringent requirement of harm that characterizes the trespass to chattels action.[127] However, "that one person is committing a trespass to another's chattel, while it may not be actionable because it does no harm to the chattel or to any other legally protected interest of the possessor, affords the possessor a privilege to use force to defend his interest in exclusive possession."[128]

In the spam cases, courts have held the electronic signals of unwanted e-mail to be sufficiently tangible to interfere with the Internet Service Provider's ("ISP's") system that receives those signals.[129] Although these signals fail to dispossess the ISP of its system or to harm that system physically, the courts have found a legally cognizable harm in the impairment of the system's value: It is not available to process "legitimate" messages of system subscribers

---

122. *Id.* § 221 cmt. c. Dispossession also occurs when one intentionally obtains possession of the chattel by fraud or duress, bars the rightful possessor's access to it, destroys it while in another's possession, or takes it into the law's custody. *See id.* § 221(d)-(g).

123. *See id.* § 217 cmt. e.

124. *Id.* § 218 cmt. e.

125. *See* KEETON ET AL., *supra* note 115, at 85.

126. *See id.* at 87. Under the Restatement, liability accrues if (i) the trespasser takes possession of or deprives the rightful possessor of use of the chattel for a substantial time; or (ii) the chattel itself is impaired in value; or (iii) the rightful possessor suffers bodily harm; or (iv) another "person or thing in which the possessor has a legally protected interest" suffers harm. RESTATEMENT (SECOND) OF TORTS § 218.

127. KEETON ET AL., *supra* note 115, at 87.

128. RESTATEMENT (SECOND) OF TORTS § 217 cmt. a; *see also id.* § 218 cmt. e (restating the privilege to use force and also noting the disparity in treatment between possessors of land and chattels). "The interest of a possessor of a chattel in its inviolability, unlike the similar interest of a possessor of land, is not given legal protection by an action for nominal damages for harmless intermeddlings with the chattel." *Id.*

129. *See* Burk, *supra* note 120, at 30.

if it is processing spam.[130] Additionally, courts may find harm in the increased costs of consumers. Consumers incur additional costs in dealing with the unwanted mail, including the payment of higher fees to those ISPs that charge incrementally for e-mail access.[131] ISPs themselves may suffer lost revenue attributable to subscribers who cancel their accounts because of the unwanted e-mail.[132] Some ISPs also expend funds to install filtering software that the spammers then defeat through technological means of their own.[133]

The analogy to unwanted indexing is easy to see.[134] The communication between the spider and the indexed site is unwanted (at least once the site informs the spider of its objections), and involves tangible electronic signals that interfere with the system's availability to process requests originating with visitors that the site wants to encourage. Damage could take the form of the costs to build technological blocks to keep the spider out, to add more capacity, and/or to repair any lost goodwill because the system is slower than it otherwise would be, or because the information provided by the spider to the consumer is inaccurate. The spider would thus be liable for trespass to chattels. There is simply no room under the conventional trespass cause of action for a court to consider any public benefit to allowing the spider's conduct. Moreover, even if there were no damage, because the trespass has occurred, the rightful possessor has a privilege to use force to defend his or her possession. Although "force" seems intended to mean physical contact, in this context, it could take the form of a robot exclusion header or other technological measure designed to control access.

Professor Dan Burk has written a devastating criticism of the propriety of applying trespass to chattels to regulate access to and use of the Internet.[135] As he points out, the initial premise holding electronic signals to be tangible enough to support a tres-

---

130. *See id.*

131. *See, e.g.*, Compuserve Inc. v. Cyber Promotions, Inc., 962 F. Supp. 1015, 1023 (S.D. Ohio 1997).

132. *See id.*

133. *See id.* at 1019 (explaining how CompuServe installed programs to screen out spam but the spammers circumvented these blocks by falsifying the header information in their e-mails to disguise their true origin).

134. *See* Burk, *supra* note 120, at 46-47 (discussing the dispute between eBay and Auction-Watch, the auction aggregator, and how the elements of a trespass to chattels claim would be met).

135. *See id.* at 32-37 (criticizing, *inter alia*, courts' willingness to extrapolate from trespass to land cases in holding electronic signals to be tangible and to hunt for damage to sustain the claim, regardless of how remote that damage is from any physical harm to the chattel).

pass to chattels claim is faulty.[136] Cases holding such items as smoke or dust to be tangible did so in the context of trespass to land, a cause of action based on a stronger interest in inviolability than that of trespass to chattels.[137]

At the same time, an actionable trespass has traditionally been defined as, "[a]n unprivileged use or other intermeddling with a chattel which results in actual impairment of its physical condition, quality or value to the possessor."[138] It is difficult to see how the physical condition, quality, or value of the system's servers are impaired by their receiving and processing exactly those signals for which they were designed.[139]

Moreover, under the Restatement test, the trespasser is liable for "the loss thus caused" by the impairment of the chattel.[140] Thus, there is some causal connection between impairment and damages. In the spam cases, the courts have hunted for damage that does not clearly evidence this causal relationship. For example, employee time spent to block unwanted messages has constituted damage in at least one spam case.[141] But this time does not relate to the already tenuous impairment of the chattel.

As Burk points out, by loosening the strict requirements for trespass to chattels, the law, perhaps inadvertently, may confer an extremely broad entitlement on web site owners to exclude any communication they deem unwanted. This right applies regardless of the actual harm (or lack thereof) to the system, the public benefit accruing from the prohibited activity, and the lack of any public debate over the propriety of this new entitlement.[142]

2. Misappropriation

In contrast, a state claim grounded in a misappropriation theory modeled on that of *NBA v. Motorola, Inc.* at least allows a court to consider whether the "ability of other parties to free-ride on

---

136. *See id.* at 33-34.

137. *See id.* at 33 (noting that the effect of "collapsing the separate doctrines of trespass to land and trespass to chattels back into their single common law progenitor . . . effectively creates a brand new cause of action, unknown te modern jurisprudence").

138. RESTATEMENT (SECOND) OF TORTS § 218 cmt. h (1965); *see also* Burk, *supra* note 120, at 34-35.

139. *See* Burk, *supra* note 120, at 34-35, 43-46.

140. RESTATEMENT (SECOND) OF TORTS § 218 cmt. h.

141. *See* Intel Corp. v. Hamidi, No. 98-AS05067, 1999 WL 450944, at *1 (Cal. App. Dep't Super. Ct. April 28, 1999) ("The intrusion by Hamidi into the Intel e-mail system has resulted in the expenditure of company resources . . . te block his 'mailings,' and to address employee concerns about the mailings.").

142. *See generally* Burk, *supra* note 120, at 32-37, 43-47.

the efforts of the plaintiff or others would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened."[143] A metasite free-rides because it uses the information of others to generate its own income. If the spiders cause a burden on the system so great as to threaten the indexed site's existence, the metasite sending the spiders may be liable for misappropriation.

Under *NBA*, a defendant is liable for misappropriation when, in addition to free-riding in a way that threatens the plaintiff's existence, "(i) a plaintiff generates or gathers information at a cost; (ii) the information is time-sensitive; . . . [and] (iv) the defendant is in direct competition with a product or service offered by the plaintiffs."[144] Many metasites do not contain time-sensitive information nor do they compete with the sites they index. Recall, for example, that AuctionWatch.com indexes auction sites but does not conduct auctions itself.[145]

The indexed site might argue that it competes directly with the metasite for advertising revenue. Because a metasite offers more effective target marketing than advertising at one particular indexed site, the metasite, while not directly competing with a *product* offered by the indexed site, does directly compete in offering the *service* of providing advertising space.

Elsewhere, I have suggested that a court should not find direct competition unless the plaintiff and defendant compete with each other both in providing the particular good and for the same pool of advertising money.[146] Under this formulation, many metasites would not be liable for misappropriation.[147]

Of course, *NBA v. Motorola, Inc.* is the governing law only in the Second Circuit that rendered the opinion and other jurisdictions

---

143. NBA v. Motorola, Inc., 105 F.3d 841, 845 (2d Cir. 1997).

144. *See id.*

145. *See supra* note 13 and accompanying text.

146. *See* O'Rourke, *supra* note 9, at 700.

147. However, this test may not allow a site that offers a product also to offer comparative pricing information. For example, amazon.com and barnesandnohle.com compete in both the market for the sale of books and for many of the same advertisers. Under the proposed test, the two sites would be direct competitors. Would this prevent amazon.com from telling a customer what the price of a desired hook is on barnesandnoble.com? It might, if barnesandnohle.com is driven out of business as a result. But shouldn't it be driven out of business if it can't meet the price of the competition? Such indexing still might not be misappropriation if the information is not "time-sensitive" as required by *NBA*. *See NBA*, 105 F.3d at 845. Furtber, the real question is what should be labeled "free-riding" and what should not. When a site sells its own products and offers data on other sites' prices as a point of comparison, it is simply reporting a fact. It is not like the metasite that makes its living by indexing the content of others.

choosing to adopt it. Some jurisdictions take a much broader view of what constitutes misappropriation and use the term "free-riding" in a pejorative sense to refer to any value one derives from the efforts of others.[148]

These broader views, often based on some sense of the defendant's having done something morally wrong, run the risk of copyright preemption. The data that the spiders use to build their databases is not copyrighted. A misappropriation claim based simply on use of uncopyrightable data without the additional elements of *NBA* replicates rights the spider would have under copyright and therefore may face preemption. However, it might be useful in assessing the propriety of the means of access used to obtain the information.

3. Breach of Contract

One of the cheaper technological means a site may take to block unwanted visitors is to require *all* visitors to agree to a click-wrap contract that enumerates forbidden uses. For example, eBay's User Agreement to which one must consent before being allowed to register, states: "Our web site contains robot exclusion headers and you agree that you will not use any robot, spider, other automatic device, or manual process to monitor or copy our web pages or the content contained herein without our prior expressed written permission."[149] There are, however, at least two problems for a site seeking to stop unauthorized indexing by suing for breach of a contractual term like this.

First, the spider may never see the contractual terms. Its manner of entry may allow it to avoid clicking on the "I agree" button that indicates its consent to the terms of the click-wrap. More imaginative modes of seeking agreement—like placing a legend on the home page providing that any use of the site constitutes agreement to the terms and conditions regardless of the user's actual knowledge of them—may not be enforceable under contract law.[150]

---

148. *See, e.g.*, Board of Trade of Chicago v. Dow Jones & Co., 456 N.E.2d 84, 90 (Ill. 1983) (holding, in a pre-*NBA* case, that a board of trade could not base a stock index future contract on the Dow Jones index without permission from Dow Jones, despite the lack of direct competition).

149. *User Agreement*, ¶ 7, (Mar. 10, 2000), http://pages.ebay.com/help/community/png-user.html.

150. It might, however, be enforceable in jurisdictions that have adopted the Uniform Computer Information Transactions Act ("UCITA"). *See* O'Rourke, *supra* note 9, at 691-92 (discussing the enforceability of boilerplate legends forbidding linking under draft UCC Article 2B, UCITA's predecessor).

Further, even if a contract could be formed in this way, a court may hold a particular term unconscionable. Professor Reichman and Jonathan Franklin suggest a doctrine of public interest unconscionability under which "[a]ll mass-market contracts, non-negotiable access contracts, and contracts imposing non-negotiable restrictions on uses of computerized information goods must be made on fair and reasonable terms and conditions, with due regard for the public interest in . . . preservation of competition."[151] The blanket exclusion of spiders employed by shopbots may hamper competition by restricting the flow of pricing information.

Second, even if the contractual term passes muster under contract law, it may be preempted by copyright law, although at least one recent decision suggests otherwise.[152] Since there is no copyright right to employ spiders, surrendering one's ability to use a spider under contract law does not implicate a "right[ ] equivalent to any of the exclusive rights within the general scope of copyright" as required before preemption may occur.[153]

However, to the extent that the effect of enforcement of the contractual provision would be to prohibit the entire public from engaging in the use of uncopyrightable information, it may be preempted not by the Copyright Act, but by the Constitution. Assuming that once on the site, a visitor is restricted in its use of information only by the copyright law, the prohibition on the use of spiders removes from the visitor one way of exercising its right to use the information. That other means, like indexing the site's content manually,[154] exist, would not necessarily save the term from preemption.[155]

---

151. J.H. Reichman & Jonathan A. Franklin, *Privately Legislated Intellectual Property Rights: Reconciling Freedom of Contract with Public Good Uses of Information*, 147 U. PA. L. REV. 875, 930 (1999).

152. *See* Ticketmaster Corp. v. Tickets.Com, Inc. ("Ticketmaster I"), 54 U.S.P.Q. 2d 1344, 1347 (C.D. Cal. 2000) (stating that because "the contract claim alleges adherence by Tickets to a contract not to use for commercial purposes," that claim is "possibly" not preempted). This court also noted that that the claim that the contract prohibits deep linking is not preempted. *See id.* These statements have limited value because the court issued a new order in August that superseded the prior cited one. *See* Ticketmaster Corp. v. Tickets.Com, Inc. ("Ticketmaster II"), No. CV99-7654, 2000 U.S. Dist. LEXIS 12987 (C.D. Cal. Aug. 10, 2000).

153. 17 U.S.C. § 301(a) (1994). The statute also requires that the rights at issue "come within the subject matter of copyright" for preemption to occur. *Id.* § 301(b)(1).

154. EBay would prohibit even manual efforts to index its site's contents. *See supra* note 149 and accompanying text.

155. *See, e.g.*, Bonito Boats, Inc. v. Thunder Craft Boats, Inc., 489 U.S. 141, 160, 167 (1989) (holding a Florida statute, prohibiting one way of duplicating boat hulls, was preempted by the federal Patent Act).

The preemption argument is not particularly strong, in part because at least until the DMCA, copyright law has never governed the permissible means by which one may obtain copyrighted information. Under current law, doctrines on contractual formation and enforcement (including unconscionability), will likely govern promises not to use spiders to access a web site. Contract law could benefit from a delineation of what factors should be relevant in determining whether, if an agreement is formed under traditional doctrine, the particular promise should be enforceable. Similar factors could also inform the question of the propriety of spidering in the absence of an agreement.

IV. ADDRESSING UNAUTHORIZED SPIDERING AND CONSIDERING THE NORMATIVE FRAMEWORK

A. Unauthorized Spidering

The policy review revealed that, in considering how to address claims that spiders should not be permitted to access and index the product and pricing information of a publicly available web site, the public benefit and private burden to the indexed site are relevant. None of the causes of action above, established in the physical world, really addresses these concerns in a systematic way.

As Professor Burk notes, because nuisance law allows a weighing of public and private interests, it, rather than trespass, may provide the best property analogy to use in evaluating not simply spiders, but all unwanted access of web sites. Unlike trespass, nuisance imposes liability on the unwanted visitor only if the costs of the intrusion outweigh its benefits.[156] This implies a balancing test. The key question in formulating such a test (regardless of whether or not one eventually places the label "nuisance" on it), is what benefits and costs should be legally cognizable in addressing claims of unauthorized access?

The *NBA* test provides a starting point. Courts applying that test consider whether "plaintiff generates or gathers information at a cost." Likewise, courts addressing claims of unauthorized access should consider: (i) the burden to the objecting site of processing unwanted requests because that imposition represents the direct costs of the unwanted access; (ii) whether the site loses revenue on which it depends for its existence as a result of the access; and (iii) the cost to the objecting site to gather the information the visitor is

156. *See* Burk, *supra* note 120, at 53.

taking and eventually using because this cost represents part of the free-riding benefit accruing to the visitor.[157]

Under the *NBA* test, courts also evaluate whether or not the plaintiff and defendant are in direct competition, and whether the particular free-riding at issue would so reduce incentives as to threaten the existence or quality of the product. Similarly, courts assessing claims of unauthorized access should consider the purpose for which the unauthorized visit is conducted, the nature of the information taken,[158] and whether the activity would threaten the site's existence or quality. Additionally, they should determine whether the visitor's eventual end use of that information is in direct competition with the visited site. Access to extract uncopyrighted data to market in direct competition with the visited site is less sympathetic than access and copying that results in a new product.

If there are some benefits from the unauthorized access, courts should also consider whether any less intrusive but cost-effective means exist for the unwanted visitor to obtain the information it desires. This question ultimately reduces to deciding whether the particular technology the unwanted visitor is using is best developed by a finding of liability that effectively places control over it in the hands of the objecting site.

Finally, courts may creatively fashion remedies. When the public enjoys a substantial benefit from the unauthorized access, a court might appropriately allow it to continue but require the unwanted visitor to pay for the costs its access imposes. This would allow courts to safeguard the incentives of the visited site.

Under this test, a shopbot's unwanted spider would likely escape liability unless it so burdens the site as to threaten its very existence. The pro-competitive benefits that the spider's use of the uncopyrighted information offers would likely outweigh the cost to the system from processing its requests. There is, however, a question as to whether the shopbot could cost-effectively use less intru-

157. Copyright law would not consider this third cost in considering whether an infringement has occurred. *See* Feist Publications, Inc. v. Rural Tel. Servs., Co., 499 U.S. 340, 353-54 (1991) (rejecting the argument that "sweat of the brow" alone justifies granting copyright protection). As the question turns to permissible access, and the forum to the Internet, courts should consider these costs, in part because the relevant cause of action is more one sounding in misappropriation than in copyright.

158. Copyright law, including the DMCA, would govern the unwanted access and use of copyrighted information. Whether it has struck the balance appropriately is generally a question beyond the scope of this Article that deals only with access to copyrighted information as a step in extracting the uncopyrightable.

sive means to obtain that information. EBay, for example, argues that Bidder's Edge could equally effectively obtain the information it seeks by running real-time searches rather than using spiders. This fact (if true), while not determinative, should factor into the court's balance.

But some unwanted spiders may not escape liability. There are many ways that one site may seek to profit from the data of another. For example, a newsbot might allow a user to indicate that it would like the financial pages of the Wall Street Journal and the sports section of the New York Post. A newsbot could provide a user with such a personalized newspaper. But (assuming the newsbot presented only uncopyrighted, factual information), if the newspapers make money by advertising and users now never see the ads, the newsbot may seriously threaten the site's revenue model—and not because it offers a lower priced product of its own. In such a case, the balancing test may tip to holding the newsbot liable.

Similarly, a shopbot aggregator[159] that employs unauthorized spiders to build its database of comparison pricing information may not survive the test. The aggregator is in direct competition with the sites it indexes, providing the same information and seeking the same advertising dollars. It may very well threaten the existence of individual shopbots, even though its own aggregation cannot succeed without those shopbots.

If a site were to erect a higher barrier to entry than a robot exclusion to protect its uncopyrightable data—one that would be a "technological measure" under the DMCA—a court should still use the balancing test set forth above when that measure is effectively protecting against access to and use of uncopyrightable data. However, the burden of proof should be placed on the unwanted visitor to demonstrate that the test weighs in its favor. This recognizes the blocking site's interest in protecting its investment in blocking technology and, by extension, its heightened desire to keep out certain visitors.

B. The Normative Framework

The way in which one answers the essential question of who should have what rights and privileges in accessing a web site reveals much about the normative framework to which one subscribes. The choices are essentially three: (i) a system in which us-

159. Recall that a shopbot aggregator is one that gathers the data of a number of pure shopbots. *See supra* notos 11 and 19 and accompanying toxt.

ers have unfettered access to sites; (ii) a strong property rights regime premised on a real property analogy that allows a site owner to exclude whomever it wants for whatever reason; and (iii) a regime of divided entitlements in which web site owners have certain property rights but users also have certain privileges. The nuisance-like regime suggested above subscribes to the third.

The notion that the web should be openly accessible to all has some appeal. Public funds helped to establish the standards that allow the web to function and paid for much of its infrastructure. The raison d'etre of the web is to connect many users, and there is something intuitively disturbing about a web site owner who fragments the network by restricting access to some.

But if web sites are required by law to allow the public to enter using whatever means of access is available with copyright law providing the only limiting principle, the web will lose some of its flexibility. For example, a site could not use a subscription based revenue model if it were not legally entitled to keep out those who have not paid. Giving sites at least some ability to regulate access would allow the Internet to accommodate a variety of different types of web sites.

Certainly, one may make economic arguments that sites should have a relatively unfettered right to exclude analogous to that of owners of real property.[160] For example, the law could grant sites the entitlement to ban access for the purpose of indexing. If a site opts out of allowing indexing altogether or permits it only selectively, this sends a market signal for consumers to interpret. Because a site generally has an incentive to attract the maximum number of customers, its policy toward indexing must reflect that which is efficient both for it and its customers. Moreover, if a site is entitled to determine who indexes it, it can provide better service because its systems would never be overburdened by unauthorized spiders.

Further, a rule like this would not impose an insuperable obstacle to metasites. Transaction costs are low enough on the Internet to facilitate metasites' contacting and negotiating with the tar-

160. *See generally* Trotter Hardy, *Property (and Copyright) in Cyberspace*, 1996 U. CHI. LEGAL F. 217, 259-60 (1996) (arguing, *inter alia*, that lower transaction costs and the ease of defining boundaries in cyberspace justify a strong property rights regime there); I. Trotter Hardy, *The Ancient Doctrine of Trespass to Web Sites*, 1996 J. ONLINE L., (last visited Sep. 18, 2000) http://www.wm.edu/law/publications/jol/hardy/html (reviewing various theories of property and concluding that they support both treating web sites as property, and applying trespass te them).

get sites for permission to index before doing so. Additionally, by requiring metasites to seek permission first, the law would force them to internalize the costs that they impose.[161] Moreover, the site owner would not have to incur the costs of erecting a fence that the law might or might not respect.

The problem, of course, is that it is quite difficult to know at this point whether or not these economic assertions are true in practice. The earlier analysis revealed that sites may be more concerned with limiting competition than enhancing aggregate efficiency. This renders questionable the assumption that a site's indexing policy is necessarily efficient. Additionally, transaction costs might be much higher than anticipated.[162] For example, as auction sites proliferate, it may become quite costly for Bidder's Edge to obtain permission from each one that it seeks to index. Finally, the public interest in access to information likely exceeds its interest in accessing another's tangible real or personal property. The question really is on which side of the coin—strong or weak property rights—should the law err at the beginning stages of a new technology? The answer is a close one.

Because we know so little about the direction in which technology will develop and because incentives are not always aligned optimally to ensure efficient licensing, a cautious approach to property rights may be appropriate. Effectively giving control over the development of technology to one or a few firms through the grant of strong exclusionary rights may impede its efficient exploitation. The high cost of litigation generally and its strategic direction may hamper the emergence of competitors.

Throughout history, many countries, including the United States, have found weak property rights useful particularly in the early stages of a new industry. As that industry develops, the pressure for stronger rights to protect the new base of knowledge grows. While the analogy is not direct because the Internet may be the ultimate cross-jurisdictional industrial tool, it lends support to the idea that weaker rights may help a more vibrant network to develop.

As more information becomes available about how Internet markets work and how firms and consumers behave there, the law

---

161. *See* Burk, *supra* note 120, at 49 ("Typically, external costs imposed upon a common resource may be internalized by creating private property interests that give users an incentive to consider the full cost of their usage, or seek permission to impose usage costs upon the property of another.").

162. *See* Dan L. Burk, *Muddy Rules for Cyberspace*, 21 CARDOZO L. REV. 121, 152-58 (1999).

can make more reasoned choices about how it wishes to provide the appropriate incentive structure for both innovation and competition on the Internet. This may eventually result in stronger and more certain property rights. But to implement them now may effectively place costly restrictions on a developing marketplace that hamper both innovation and competition.

CONCLUSION

The Internet and its technology offer the potential to transform the way in which retailing is conducted and to shift some bargaining power from producers to consumers. The law therefore must carefully scrutinize claims by sites that seek to restrict access to software tools that would help this shift to occur. It must separate sites' "legitimate" motivations from their anti-competitive ones and consider how best to balance legally cognizable private harms against the public benefit.

Merely applying existing law by analogy is unsatisfactory because it lacks an appropriate sensitivity to the unique characteristics and opportunities of the Internet. By understanding what motivates sites, the law can formulate a cause of action that addresses questions of unauthorized access on the Internet more effectively. As time and technology march forward, the real question will be how to implement flexible legal rules that can deal with the myriad of ways in which information is accessed and used on the Internet. This Article starts the debate with a proposed balancing test to address unauthorized access to obtain uncopyrighted information, but its main contribution is to remind Internet users of all types that the law has a major role to play in shaping competition in this still relatively new medium.