TOLLS ON THE INFORMATION SUPERHIGHWAY: ENTITLEMENT DEFAULTS FOR CLICKSTREAM DATA

Lee Kovarsky[*]

[*] J.D. University of Virginia, 2004. First, I would like to thank the Olin Foundation for the summer fellowship grant that made this Note possible. I owe special thanks to my fellowship advisor, Lillian BeVier, for her tireless reading and editing of different versions of this manuscript. I would also like to thank George Cohen, George Triantis, Dan Ortiz, Thomas Nachbar, Tim Wu, Charles Goetz, and Anne Coughlin for all of their thoughts and contributions to the concepts explored in this Note. Finally, I want to thank my family and Sarah largely for their love and support, but also because they are the people who listen to me complain.

INTRODUCTION

THE burgeoning online privacy debate is steeped in portentous rhetoric borrowing heavily from George Orwell's *Nineteen Eighty-Four*[1] and Jeremy Bentham's *Panopticon*,[2] but these literary and architectural metaphors egregiously misrepresent an already confusing controversy.[3] Commentators deploy these metaphors as descriptive shorthand, but in so doing they both misrepresent the contours of the problem and unfairly editorialize the description itself. Traditional privacy metaphors are generally ill-equipped to mediate our understanding of the way information changes hands in cyberspace.[4] In an attempt to capture the differences between the current state of informational privacy and that of Orwell's totalitarian dystopia, some commentators have sought modern variants on traditional privacy metaphors.

The concept of "little brother" is one such variation invoked by many to convey the private character of the "invasion."[5] Of equal

---

[1] George Orwell, Nineteen Eighty-Four (Secker & Warburg 1987) (1949) (describing a culture where all behavior is monitored by the state).

[2] Jeremy Bentham, The Panopticon Writings (Miran Bozovic ed., Verso 1995) (1791) (proposing plans for a progressive prison constructed such that all behavior could be monitored easily by a single agent).

[3] See, e.g., McVeigh v. Cohen, 983 F. Supp. 215, 220 (D.D.C. 1998) ("In these days of 'big brother,' where through technology and otherwise the privacy interests of individuals from all walks of life are being ignored or marginalized, it is imperative that statutes explicitly protecting these rights be strictly observed."); Paul M. Schwartz, Internet Privacy and the State, 32 Conn. L. Rev. 815, 853 (2000) ("The government's Herculean efforts to make Internet technology open for snooping resemble such an attempt to install the Panopticon."); Jonathan M. Winer, Regulating the Free Flow of Information: A Privacy Czar as the Ultimate Big Brother, 19 J. Marshall J. Computer & Info. L. 37, 68–70 (2000); William Branigin, Employment Database Proposal Raises Cries of 'Big Brother,' Wash. Post, Oct. 3, 1995, at A17; James Gleick, Big Brother is Us, N.Y. Times, Sept. 29, 1996, § 6 (Magazine), at 130. For an extensive discussion of the use of the Orwell metaphor, see generally Daniel J. Solove, Privacy and Power: Computer Databases and Metaphors for Information Privacy, 53 Stan. L. Rev. 1393 (2001) (arguing in favor of using Kafka's *The Trial* as a metaphor for mediating our understanding of online privacy).

[4] See Solove, supra note 3, at 1398–99.

[5] See, e.g., Justice Ben F. Overton & Katherine E. Giddings, The Right of Privacy in Florida in the Age of Technology and the Twenty-First Century: A Need for Protection from Private and Commercial Intrusion, 25 Fla. St. U. L. Rev. 25, 27 (1997) (ex-

importance as the difference in the institutional source of surveillance is the absence of what Frank Zappa calls a "central scrutinizer."[6] The Internet is not subject to monitoring by a single information collecting agent, and the overwhelming majority of those entities that do monitor online activity are private companies.[7] Of all the online information collected by these private agents, often referred to as "data miners," the vast majority consists of "cookie" or "clickstream data"—data collected on users as they go about the routine process of viewing web pages. To the extent that one may characterize the transmission of clickstream data from users to data miners as a market, the egregious information asymmetries between these two groups mean that, contrary to the implications of the traditional Orwell metaphor, by correcting these market imperfections the government may figure prominently in protecting the privacy of its citizens.

It is not the case that traditional privacy metaphors *always* distort the circumstances, but their expository appeal renders them a convenient camouflage for complexity. A solution to any online privacy problem must proceed from the understanding that an undesirable state of Internet privacy is not a monolithic species of problem at all, but rather a genus of related problems with a common denominator: Each concerns what different agents may do with information collected online. Some writers have a tendency to recite endless series of anecdotes, seemingly oblivious to the fact that the set of catalogued behavior shares only the property of having taken place at a computer.[8] From context to context the ob-

---

plaining that invasions of privacy no longer come from only the government, but also from various private commercial entities intent on making profits).

[6] The "central scrutinizer" is a character from one of Zappa's albums that serves to parody central censorship agents. Frank Zappa, Central Scrutinizer, *on* Joe's Garage (Rykodisc USA 1979).

[7] Over 1000 companies are engaged in compiling comprehensive online databases. See Mike Hatch, The Privatization of Big Brother: Protecting Sensitive Personal Information from Commercial Interests in the 21st Century, 27 Wm. Mitchell L. Rev. 1457, 1470–71 (2001).

[8] See, e.g., Ann Bartow, Our Data, Ourselves: Privacy, Propertization, and Gender, 34 U.S.F. L. Rev. 633, 645–46 (2000) (describing a scenario where computerized medical records are linked up to advertising data and email addresses); Jessica Litman, Information Privacy/Information Property, 52 Stan. L. Rev. 1283, 1306–08 (2000); Paul Rose, A Market Response to the European Union Directive on Privacy, 4 UCLA J. Int'l L. & Foreign Aff. 445, 455–65 (1999); Karl D. Belgum, Who Leads at

servers and the observed exhibit wildly different levels of technical sophistication, contractual expectations, and bargaining power.

This Note will address only one such context, the collection of clickstream data and the legal rules that should govern it. The focus is not arbitrary, however, as the clickstream data problem lends itself least to resolution by controversial philosophical principles and exhibits a commercial character rendering it particularly inviting for economic analysis. The Note will argue that the underlying objectives of any legal posture towards clickstream privacy are incompatible with a prophylactic standard of behavior and will propose a scheme that, using modern browser technology, assigns to each party (the user and the collecting agent) a certain set of default rights around which the two parties may bargain. The model would further protect user interests by incorporating, for the browser-negotiated privacy term, a user reservation price that would operate much like a liquid damages clause in the event of a breach. Although the model is presented as a case study of the clickstream privacy problem, many of the following arguments may be brought to bear on other modern contexts where privacy concerns coexist with important markets for information.

Part I will suggest a framework for analyzing privacy invasions and will clarify the scope of the activity to which the subsequent analysis applies. Part II will argue that while the concept of "information markets" remains inappropriate in some contexts, it is the best way to conceptualize and address clickstream data collection. Part III will explore the specific market failures in the clickstream context. Part IV will explain why existing privacy doctrines are unable to address the problem. Part V will argue that the plummeting transaction costs associated with developments in browser technology render a legal paradigm of default entitlements a viable option for mediating the clickstream data market. Part VI will contend that, in addition to specific default entitlements, clickstream data exchange should involve a default schedule of expectancy damages.

Halftime?: Three Conflicting Visions of Internet Privacy Policy, 6 Rich. J.L. & Tech. 1, ¶¶ 6–16 (2000), at http://www.law.richmond.edu/jolt/v6i1/belgum.html (on file with the Virginia Law Review Association).

I. COOKIES AND CLICKSTREAM DATA

A. A Framework for Analysis

At this point, it may be useful to conceptualize a state of informational privacy as a tripartite equilibrium among the observed, the collecting agent, and the searching agent.[9] Many times the collecting agent and the searching agent will be the same entity, and such a condition is what people usually think of as "surveillance."[10] Broadly speaking, informational privacy protects some right to control and condition the revelation of personal information to other entities.[11] Part II argues that this right does not imply an entitlement to control how one is treated or perceived by those to whom one divulges the information.[12] Almost all of the practices that implicate informational privacy (even offline) can be represented by some discrete combination of three attributes: (1) the ac-

[9] Professor Lawrence Lessig has extensively discussed the relationship between privacy and the costs associated with monitoring and search. See, e.g., Lawrence Lessig, Code and Other Laws of Cyberspace 18–19 (1999). To this author's knowledge, however, no one has used these concepts to construct an analytically precise means of classifying different privacy concerns. It has been suggested, however, that one could think about privacy in terms of the individuals about whom data is being gathered or in terms of who is doing the gathering. See A. Michael Froomkin, The Death of Privacy?, 52 Stan. L. Rev. 1461, 1468 (2000).

[10] Surveillance is not always described in these terms, but most situations we consider to be surveillance involve an entity collecting information for its own use.

[11] See Jeffrey Rosen, The Unwanted Gaze: The Destruction of Privacy in America 8 (2001) ("Privacy protects us from being misdefined and judged out of context in a world of short attention spans, a world in which information can easily be confused with knowledge."); Alan F. Westin, Privacy and Freedom 7 (1967) ("Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.").

[12] This conception of informational privacy would go further and suggest that the observed approve secondary exchanges between an entity that already "knows" something and another that seeks to learn it, even if the initial exchange between the observed and the first entity was not conditioned upon such subsequent consent. See, e.g., Eugene Volokh, Freedom of Speech and Information Privacy: The Troubling Implications of a Right to Stop People from Speaking About You, 52 Stan. L. Rev. 1049, 1050 (2000) (characterizing informational privacy as "[one person's] right to control [another person's] communication of personally identifiable information about [him or her]").

tivity in which the observed is engaged, (2) the identity of the collecting agent, and (3) the identity of the searching agent.[13]

The Video Privacy Protection Act[14] is nicknamed the "Bork Bill" because the legislation reflected the public outcry over the ease with which a reporter acquired the embattled jurist's video records.[15] The above framework would represent this practice as (1) renting video tapes (the activity), (2) the video store (the collecting agent), and (3) a reporter (the searching agent).

While connected to the Internet, the observed (an Internet user) may do any number of things that serve as a data point for collecting and searching agents: send an email, post messages in a chatroom, view a web page, download a piece of software, register for a service, or purchase a product. Collecting and searching agents may be employers, private companies, other people, or the government. An undesirable privacy state generally occurs when certain instances of one attribute coincide with certain instances of others. The term "attribute" refers to the variables (1), (2), and (3) themselves, and the term "instance" refers to a potential *value* for that variable. For example, an attribute is the property of having an agent of data collection itself and an instance of that attribute could be a website, an advertiser who places an ad on that page, an employer, or the government.

While in an Orwellian state the government is both the collecting and the searching agent, on the Internet its activity is generally limited to that of the latter.[16] This circumstance notwithstanding, people remain extremely concerned any time the government is the searching agent. Whether law restrains the government from acting as such generally depends on the circumstances. For example, prosecutors may use phone records to link names to Internet Protocol ("IP") addresses collected while users are on child por-

[13] This framework is merely general. Specific invasions would require variables that described, among other things, the time of the invasion and the exact identity (rather than the class) of persons both being observed and doing the collecting.

[14] 18 U.S.C. § 2710 (2000).

[15] See Elbert Lin, Prioritizing Privacy: A Constitutional Response to the Internet, 17 Berkeley Tech. L.J. 1085, 1148 n.430 (2002).

[16] There are certain exceptions. For example, the government surely collects data in .gov domains. Offline, the most obvious instance of the government as the collecting agent is when it acquires census data.

nography sites,[17] but the government does not enjoy unfettered access to the content of email transmissions.[18]

People are also suspicious of employers acting as searching agents, a circumstance that is the source of much discussion as companies routinely collect and search, among other things, the email and web-surfing histories of their employees.[19] That employers may monitor email and web surfing to promote productivity and protect against industrial espionage has become more of a fact of life than a controversy,[20] and employers would likely contract around any default rule to the contrary. Employers are rarely collecting agents without also being searching agents and there is a straight-forward rationale for why: Entities that are the former but not the latter generally subsist on selling data about the things they observe, and the profit derived from selling the web-surfing habits of its workforce is not worth the associated diminution in job appeal to potential employees. Regardless of its economic justification, many find employer surveillance objectionable because it represents an uninvited source of behavior modification.

People also become upset when they discover that they have been observed without notice. Several years ago Intel ignited a firestorm when the public learned that it embedded a unique identifier in each Pentium III chip, allowing the company to collect information about what its customers were doing on the Internet.[21] A

---

[17] See Daniel J. Solove, Digital Dossiers and the Dissipation of Fourth Amendment Privacy, 75 S. Cal. L. Rev. 1083, 1143–44 (2002).

[18] The Federal Wiretap Act of 1986, 18 U.S.C. §§ 2510–2522 (2000), restricts government access to the content of emails by requiring judicial issuance of a warrant in certain circumstances. See 18 U.S.C. §§ 2511(2)(ii)(A)–(B) (2000).

[19] See Lessig, supra note 9, at 145 (noting that the greatest invasion of privacy in cyberspace is employee-monitoring by companies); Rosen, supra note 11, at 160–62 (detailing the ease with which systems administrators can observe the online behavior of employees).

[20] See Julie E. Cohen, Privacy, Ideology, and Technology: A Response to Jeffrey Rosen, 89 Geo. L.J. 2029, 2031–32 (2001) [hereinafter Cohen, Privacy, Ideology, and Technology] (arguing that employers monitor to structure investment risk in their employees); Paul M. Schwartz, Beyond Lessig's *Code* for Internet Privacy: Cyberspace Filters, Privacy Control, and Fair Information Practices, 2000 Wis. L. Rev. 743, 770–71 (citing Jeffrey L. Seglin, As Office Snooping Grows, Who Watches the Watchers?, N.Y. Times, June 18, 2000, at BU4).

[21] See Shawn C. Helms, Translating Privacy Values with Technology, 7 B.U. J. Sci. & Tech. L. 288, 294, 298 (2001); Frank James, Intel Chip Fires up Privacy Debate, Chi. Trib., Jan. 22, 1999, at 1.

conceptually similar example of this type of invasion occurred when RealNetworks allowed surfers to download its RealJukebox audio product from its website.[22] RealNetworks, in turn, used a unique identifier embedded in the RealJukebox software to collect information about its customers' musical proclivities.[23]

Finally, people object not only to instance-attribute combinations involving specific searching agents but also to the manner in which these agents use the data. This is the "fourth attribute" violation. The "fourth attribute" refers to a representation of a privacy violation that includes, as one of its primary dimensions, what a searching agent does with the data. Properly understood, however, these types of objections do not relate to informational privacy at all. It borders on tautology to note that an entity compromises informational privacy when it reveals private information. When people object to receiving unsolicited emails or being dehumanized by advertising that treats them categorically,[24] however, their objection does not stem from the information revelation itself, but the *use* of that information. This activity, however, is closer to harassment than to an informational privacy violation. The public mistakenly analyzes the "fourth attribute" problem (in the clickstream context) as one of informational privacy for two reasons.

The first derives from faulty analogical reasoning. Many controversial uses of data result in users receiving unsolicited email. Because of the obvious analogy to real mail and real mailboxes, people tend to think of the receipt of unsolicited email in the same privacy terms as they think of unsolicited real mail. The analogy is surely warranted, and "junk mail" may well implicate a privacy interest. That privacy interest, however, is not informational. It is closer to the cluster of privacy interests that relate to being free from physical invasions or harassment at home.[25] The second

---

[22] See Lawrence Jenab, Will the Cookie Crumble?: An Analysis of Internet Privacy Regulatory Schemes Proposed in the 106th Congress, 49 U. Kan. L. Rev. 641, 662 (2001); Christopher D. Hunter, Recording the Architecture of Cyberspace Privacy: Why Self-Regulation and Technology Are Not Enough, at http://www.asc.upenn.edu /usr/chunter/net_privacy_architecture.html (Feb. 2000) (on file with the Virginia Law Review Association).

[23] See Hunter, supra note 22.

[24] See, e.g., Bartow, supra note 8, at 653–56.

[25] See infra note 57 and accompanying text.

source of confusion resides in the character of the economic cause-and-effect relationship. The ability to send junk mail and to deliver customized advertisements to consumers creates a significant demand for personal information. The market's response to this demand may compromise informational privacy, but the mailing and advertising themselves do not. Consequently, proscribing such behavior may be a means of vitiating the demand for informational privacy violations, but this does not mean the demand-inducing activity itself represents a violation.

Clickstream data collection superficially exhibits almost none of the qualities associated with typical claims about informational privacy and therefore represents a relatively novel problem. The government certainly acts as a searching agent, but for only a small fraction of the time, and lawmakers may easily proscribe such searches by enacting legislation. Employers monitor employee clickstreams,[26] but such monitoring is generally accepted as an economic necessity.[27] Moreover, most people are aware that their clickstream is collected,[28] and it is the absence of restrictions on such collection that animates much popular concern.[29] Finally, that searching agents use clickstream data for targeting is not a complaint about informational privacy at all. Further analysis of clickstream privacy requires a deeper understanding of the data exchange itself.

B. Characteristics of Clickstream Data

The most robust online databases store huge amounts of clickstream data collected using cookies to identify individual computers.[30] The vast majority of web interactions consist merely of a user downloading a web page, at which point a website operator (publisher) will write to that computer a tiny machine-readable

---

[26] See supra notes 19–20 and accompanying text.

[27] See supra notes 19–20 and accompanying text.

[28] In one survey, eighty-eight percent of respondents at least had some notion of what a cookie was. Lorrie Faith Cranor et al., Beyond Concern: Understanding Net Users' Attitudes About Online Privacy (AT&T Labs, Research Technical Report TR 99.4.3, 1999), at http://www.research.att.com/resources/trs/TRs/99/99.4/99.4.3/report.htm (Apr. 14, 1999) (on file with the Virginia Law Review Association).

[29] See id.

[30] See Solove, supra note 3, at 1411–12.

text file, called a cookie.[31] The cookie will contain some identifier (this can be anonymous, like an ID number, or non-anonymous, like a name) as well as other pieces of information that allow the publisher both to personalize treatment of that user and to collect information about that user's behavior on the site.[32] A cookie may therefore also contain a zip code, an IP address, a favorite color, or an indication of a preference for golf over tennis. On this author's computer, in the folder C:\WINDOWS\COOKIES, there are currently 953 such text files. Clickstream data is the information collecting agents store about the cookies that send requests to their servers.[33]

A page view is a transaction, and understanding why this is so is crucial to the process of identifying the privacy interests involved in the clickstream context.[34] When a user downloads a web page, she is effectively requesting content from a server.[35] When she types a URL address into her web browser, she is requesting content residing at that location (that is why it is called an address).[36] During the process of transmitting the information, in most instances the content server will either (1) read a cookie it has previously written to a specified directory in a user's computer, or (2) write a new cookie if one is not already there.[37] The cookie will contain information allowing the content server to customize user treatment based either on things that she has requested or on a profile of her cookie history.[38]

There are three important things to remember about cookies. First, cookies confer benefits upon users at the same time that they allow collecting agents to extract information from them. Cookies allow publishers to personalize user experience and to collect cer-

[31] For a general technical description of how collecting agents use cookies to track clickstreams, see Jenab, supra note 22, at 645–46.

[32] Id.

[33] See Solove, supra note 3, at 1411.

[34] See Jenab, supra note 22, at 643 ("The first and fundamental point to clarify is that all Internet activity is transactional.").

[35] Id. The distinction between "content" and "e-commerce" sites is not important here; a user always sees "content," in the sense of information sent from server to browser, even if the user is purchasing something.

[36] Id. at 643–44.

[37] See also id. at 645 (describing this process for third-party advertisers).

[38] See Solove, supra note 3, at 1411.

tain data about user behavior.[39] Second, for simplicity's sake this Note has thus far implied that only first-party content servers can place a cookie on a user's computer, but other third parties, generally advertisers or affiliates, may do so as well.[40] Third, this information may or may not be collected anonymously.[41] Generally cookies from third-party advertisers (entities that serve their clients' advertisements into designated spaces on publisher pages) will not contain personally identifiable information but will associate the information with a unique, but anonymous, identifier.[42]

The invisible yet highly technical character of the exchange often means that even if a user knows that a collecting agent is observing her, she may not know the identity of such agents, the identities of potential searching agents, or the potential uses to which the searches may be put. If she could condition her data transfer on adherence to certain expectations about these unknowns, she might consent to some of them but not to others. The seamlessness of the data collection allows collecting agents to take advantage of users that do not know that it is happening, and the presence of uncertainty regarding what may be done with the data overdeters Internet use by those who do know.[43]

---

[39] See Rita Heimes, Foreword to Susan Richey et al., Internet Privacy Law, Policy, and Practice: State, Federal, and International Perspectives, 54 Me. L. Rev. 95, 95 (2002); infra note 223 and accompanying text.

[40] See Jenab, supra note 22, at 645–46; Rachel K. Zimmerman, The Way the "Cookies" Crumble: Internet Privacy and Data Protection in the Twenty-First Century, 4 N.Y.U. J. Legis. & Pub. Pol'y 439, 444 (2000); Hilary Appelman, Ratings That Know What You're Looking at, and When, N.Y. Times, June 7, 2000, at H37; Peter H. Lewis, Battling Cookie Monsters, N.Y. Times, Feb. 24, 2000, at G1.

[41] Anonymity is a matter of degree, however, and there remain at least two ways anonymous cookies may allow collecting agents to identify users. First, cookies can be cross-referenced with other login information in the collecting agent's database to link behavior with personal information. Second, other information in cookies can be cross-referenced with data in other marketing databases. See Helms, supra note 21, at 297–98.

[42] An "anonymous" cookie is one that does not use any personal information, to determine which material to send to a user's browser. See, e.g., DoubleClick Privacy Policy, at http://www.doubleclick.com/us/corporate/privacy/privacy/default.asp?asp_object_1=& (last updated June 17, 2002) (on file with the Virginia Law Review Association) ("No personal information is used by DoubleClick to deliver Internet ads.").

[43] One may fairly object that in some instances it is better "not to know" or not to have to choose a course of action because the costs of acquiring information or of choosing respectively are so great. Parts III, V, and VI deal largely with this proposi-

Using this Note's syntax, the instances of the first two attributes in the clickstream context are (1) loading a web page and (2) first and third parties that serve information into that page. What is objectionable about clickstream data collection is not an undesirable combination of attributes, but instead the inability to create certainty in a transaction for information. The next Part attempts to place this concern, the inability to condition information exchange on certain behavior, in a broader philosophical context.

II. CLICKSTREAM DATA AND PRIVACY

A. Watching Clickstreams

The right of privacy is hardly a model of deontic precision.[44] Philosophers and legal scholars alike have spilled much ink attempting to distill a unitary privacy interest.[45] Such attempts at conceptualizing a privacy right, however, inevitably fail to capture adequately the full range of activity people tend to think of as protected by some sort of privacy interest.[46] The following Section of this Note argues that the way one conceptualizes the right has profound implications for clickstream privacy, as the only salient conceptualizations suggest conferring upon a user the right to determine what level of information privacy she is to enjoy. Before venturing into this more abstract territory, however, one should understand a little more about the ways in which collecting agents acquire, use, and sell clickstream data.

---

tion. Moreover, the desire not to make a decision does not automatically resolve the question about the appropriate legal regime in favor of the status quo.

[44] See Lessig, supra note 9, at 143–57 (arguing that this muddled conceptualization creates problems when new technical architectures implicate novel privacy interests). Privacy rights may, however, be classified broadly in three ways: (1) spatial privacy rights; (2) those rights concerned with autonomy of individual choice about significant decisions (e.g., abortion rights); and (3) the right to control flows of personal information. See Jerry Kang, Information Privacy in Cyberspace Transactions, 50 Stan. L. Rev. 1193, 1202–03 (1998); see also U.S. West, Inc. v. FCC, 182 F.3d 1224, 1234 (10th Cir. 1999) (citing Fred H. Cate, Privacy in the Information Age 19–22 (1997); Joseph I. Rosenbaum, Privacy on the Internet: Whose Information Is It Anyway?, 38 Jurimetrics J. 565, 566–67 (1998)).

[45] See, e.g., Rosen, supra note 11, at 9; Julie E. Cohen, A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace, 28 Conn. L. Rev. 981 (1996) [hereinafter Cohen, A Right to Read].

[46] Alternatively, such attempts might protect too much activity.

On the publisher's end, clickstream data may be (1) used to personalize user content, or (2) scrutinized for behavioral analysis of customers. A cookie is just an identification number, like a bar code, that a collecting agent writes to a user's computer. Each collecting agent sets its own cookie, so it is very difficult for any single agent to amass comprehensive web-wide profiles of user web-surfing patterns.[47] If Company *A* recognizes a browser as cookie #987654 and Company *B* recognizes that same browser as cookie #123456, Company *B* cannot realize substantial gains from purchasing Company *A*'s clickstream data because Company *B* has no way of knowing with which of its cookies it should associate Company *A* cookie profile #987654. For an entity to be able to marry two or more data sets, it needs what is called a "unique identifier"—some coded identification that is different for each web user and where that user has the same identification in each of the two data sets. Unless cookie data is packaged with some unique identifier allowing the buyer of the data to match it to its existing data,[48] there is no "dossier effect" whereby single collecting agents build highly specific surfing profiles of behavior on a number of sites.[49] Moreover, without a unique identifier associated with each browser, robust secondary markets for non-aggregated cookie data seem unlikely. Company *A* could buy clickstream data from Company *B*, but it would not know which of its cookies correlated with each of Company *B*'s.

These conditions change somewhat, however, when third parties are the collecting agents placing cookies on browsers. A third party, usually some sort of advertising agent,[50] is an entity that deploys code (including a cookie) to a web page using its own ad servers.[51] When a user downloads a page, the third party will either

[47] "A client does not serve up cookies simply to anyone who asks. In other words, not all servers have access to all cookies. Each cookie, when initially set, circumscribes the range of servers to whom the cookie may be subsequently given." Kang, supra note 44, at 1227–28.

[48] Id. at 1227–29.

[49] This "dossier effect" refers to the increasing comprehensiveness associated with profiles consisting of information mined while the user is on different websites. Jenab, supra note 22, at 646. See also Solove, supra note 3, at 1412 (noting that cookies have certain limits with respect to their monitoring capability).

[50] See Jenab, supra note 22, at 646.

[51] See id. at 645–46.

place a new cookie on, or read its existing cookie from, the user's browser. Third-party clickstream profiles are in some ways more comprehensive than those of web publishers because the third parties serve material on a number of *different websites*.[52] If Website *A*, Website *B*, and Website *C* all permit (or contract for) Third Party *Z* to serve the advertisements on each of their respective sites, Third Party *Z* can use the same cookie for each browser irrespective of which of the three sites the user is accessing. In other words, whereas Websites *A*, *B*, and *C* only know what users do while they are on their respective sites, Third Party *Z* can associate the behavior of a given user on any of these sites with that user's behavior on any of the others. Perhaps the most sophisticated of these third parties is DoubleClick, which by 2001 was serving over five billion targeted advertisements per week on over 11,500 websites.[53] Third parties are therefore more responsible than are other first-party collecting agents for the dossier phenomenon, but the scope of this "sinister" activity is subject to the same practical limitations as is that of first parties. Like first-party collecting agents, third parties cannot buy or sell any clickstream data they collect unless their data has a unique identifier that can be associated with a unique identifier in another data set.

B. Conceptualizing Privacy for Clickstream Data

The first step in articulating a more analytically rigorous understanding of privacy is to distinguish between the three broad privacy concepts. First, privacy may denote the existence of a "right to have sufficient moral freedom to exercise full individual autonomy."[54] Arguments in favor of the right to reproductive freedom fall into this category.[55] Second, privacy may require some sort of shield for activity one chooses to perform in seclusion rather than

[52] See id. at 646; Solove, supra note 3, at 1412. First parties, however, are much more likely to use personally identifiable information in cookies, so in that sense they have more detailed profiles within a much smaller portion of the Internet space.

[53] See Jenab, supra note 22, at 646.

[54] See, e.g., Paul v. Davis, 424 U.S. 693, 713 (1976) (noting that marriage, procreation, contraception, family relationships, child-rearing, and education are all contexts in which the individual has some sort of privacy protection).

[55] See, e.g., Roe v. Wade, 410 U.S. 113, 152–56 (1973) (outlining the scope of the abortion right); Griswold v. Connecticut, 381 U.S. 479, 484–86 (1965) (holding that contraceptive prohibition unconstitutionally encumbers the right to privacy).

in the public sphere.[56] Third, privacy may represent an individual's claim to control "the processing—i.e., the acquisition, disclosure, and use—of personal information."[57] The lines between these three are often blurred, and the last two in particular seem to protect similar interests. The frameworks discussed below, however, seek only to analyze the third category, informational privacy.

The collection of clickstream data makes more sense conceptualized as implicating a certain type of privacy interest than it does implicating others. This Note does not attempt to tackle the extraordinarily difficult question of which understanding of informational privacy is "correct." Instead, this Note argues that clickstream data collection implicates a limited set of privacy interests. These interests are most effectively captured when privacy is understood to represent the user's right to control the conditions under which she alienates information. This context-driven understanding of privacy is sometimes referred to as an "instrumental" understanding because it treats privacy as a second order right, a means to another deontic end. In other words, privacy's intrinsic value is regarded as secondary to its role in securing other liberties. In a recent article, Professor Daniel J. Solove wrote:

> Not all privacy problems are the same, and different conceptions of privacy work best in different contexts. Instead of trying to fit new problems into old conceptions, we should seek to understand the special circumstances of a particular problem. What practices are being disrupted? In what ways does the disruption resemble or differ from other forms of disruption? How does this disruption affect society and social structure? These are some of the questions that should be asked when grappling with privacy problems.[58]

Answering these questions for clickstream data reveals an informational privacy problem unlike any society has confronted before, and legal solutions should reflect that difference rather than stubbornly attempt to graft old doctrines onto new circumstances. The following discussion does not treat exhaustively all privacy

---

[56] See Kang, supra note 44, at 1202.
[57] Id. at 1243; see supra notes 11–12.
[58] Daniel J. Solove, Conceptualizing Privacy, 90 Cal. L. Rev. 1087, 1147 (2002).

conceptualizations, but is instead limited to those bearing a substantial relationship to the clickstream privacy debate.

The first conceptualization of privacy is as a right not to reveal information because such revelation imposes tangible burdens on the revealer.[59] Being forced to sign in at a law library is an example of a search that imposes such burdens. In other words, the right to privacy means the right to be free from tangible costs imposed by collecting and searching agents.[60] The problems with this conceptualization will not be discussed here because clickstream data collection poses no significant inconvenience for Internet users.[61]

The second conceptualization predicates the privacy right on the notion that monitoring and searching intrinsically offend human dignity,[62] and it therefore enjoys the moniker "dignity theory."[63] Under this conceptualization, privacy violations occur even if collecting agents impose tangible costs approaching zero. More importantly, this conceptualization identifies privacy violations even where the agent inflicting the harm may compensate the victim for the violation. Some would invoke this conceptualization, for example, to condemn profile-driven searches at airports.[64] This position is overstated, however, because the affront to dignity associated with such searches arises from its *discriminatory application*. The proposition that universally applied, costless searches could be what is protected by a privacy right is a problematic one. Looking at a person may be the paradigmatic example of a costless search

---

[59] See Lessig, supra note 9, at 146 (calling this conceptualization a "utility conception" that "seeks to minimize intrusion").

[60] See id.

[61] Clickstream data collection imposes no tangible burdens in the sense that the process whereby a server reads or writes a cookie does not inconvenience the user. This lack of tangible burdens should be distinguished from subsequent burdens imposed on the user because some third party imposes them as a result of obtaining access to the data point.

[62] See Lessig, supra note 9, at 147–48. People tend to frame these claims extravagantly, and it is often unclear exactly what they mean when they say that privacy offends dignity. See Volokh, supra note 12, at 1110–11.

[63] See, e.g., Joel R. Reidenberg, Resolving Conflicting International Data Privacy Rules in Cyberspace, 52 Stan. L. Rev. 1315, 1341 (2000).

[64] This example is imperfect because such searches certainly impose tangible costs on those who are searched. This condition, however, is unimportant to the subsequent point about discriminatory application. The point is merely that the tangible cost is usually very low, involving only a quick trip over to a security guard who brushes for explosives.

but, in most circumstances, privacy does not restrict the entitlement to see someone on the street. One may insult another by staring too long, but the glance is offensive because it is cast *selectively*. Ignoring further the dignity theory's obvious circularity problem,[65] the tangible costs imposed on web surfers through collecting clickstream data are close to zero,[66] and when most collecting agents "collect," they do so neither to identify nor to discriminate by race, national origin, sexual orientation, or by any other culturally sensitive variable.[67]

The third conceptualization of the right is the reciprocal of Victorian propriety. Under this theory, privacy protects the right to be naughty,[68] and privacy's role in defining the physical boundaries for adherence to propriety prompts many to term this paradigm "civility theory."[69] Civility theory comes in two substrains. First, privacy may represent a shield against punishment of illegal acts, so long as they are performed outside the public sphere.[70] The most troublesome issue for this substrain is deciding who determines which crimes the law should punish. Some may believe victimless deviance to be protected, but nobody believes murder, if only committed behind closed doors, is protected by a privacy interest. Moreover, if privacy protects the freedom to do otherwise-illegal acts at home, law could achieve these results much less ambiguously by building those contingencies into positive legislative enactments.[71] Although more clickstream privacy would increase protection for illegal activity,[72] this protection is not what animates clickstream

[65] The idea that privacy protects a sense of dignity that may only be defined in reference to what is properly private.

[66] See supra note 61.

[67] See, e.g., DoubleClick Privacy Policy, supra note 42 ("DoubleClick does not develop marketing scores that indicate a user's individual health condition, detailed financial information, sexual orientation or behavior, information that appears to relate to children under 13, racial and ethnic origin, political opinions, religious or philosophical opinions, and trade union membership."). DoubleClick is considered the collecting agent with the most comprehensive profiles.

[68] See Lessig, supra note 9, at 148.

[69] See, e.g., Reidenberg, supra note 63, at 1341.

[70] See Lessig, supra note 9, at 148.

[71] Id. (noting that other laws and amendments could restrict the scope of searches).

[72] See Richard Posner, The Right to Privacy, 12 Ga. L. Rev. 393, 397–401 (1978); see, e.g., Richard Posner, Overcoming Law 546–51 (1995) (discussing how increased privacy might enable blackmail).

privacy enthusiasts.[73] The illegal activities that this substrain seeks to protect tend to involve deviant sex acts, such as oral or anal penetration, that are physically impossible in light of the nature of the online medium.[74]

The deficiencies in the first substrain are not present in the second. According to this second theory, the right of privacy protects behavior that, while not illegal, is socially unpopular. Under a regime with strong privacy protection, those engaging in legal but socially deviant behavior may do so without fearing public exposure. This vision of privacy protects two types of values. First, there is intrinsic value in not forcing people to suppress their thoughts and feelings. People are allowed to "be themselves" behind closed doors, and in so doing are allowed to perform activities to which they are legally entitled, but from which they may refrain if forced to perform them publicly.[75] Second, it assures the socially advantageous survival of legal, but deviant, lifestyles and viewpoints. For example, if privacy enables anonymous speech, this anonymity allows people to advance unpopular political viewpoints without fearing the consequences of unpopularity.[76] Although privacy conceived of as protection against majoritarian social norms could explain the push for more anonymity in many web transactions, it does not bear strongly on the clickstream data controversy. This is because clickstream data transactions generally do not generate a personally identifiable *chronique scandaleuse* sufficient to chill deviant activity. Some scholars would counter that disclosure is not a necessary condition for chilling, as the mere act of surveillance is enough to cause people to modify their behavior. This assertion is plainly true, and one may say no more than that the magnitude of the modifications is proportionately related to the likelihood of disclosure. With respect to clickstream data, however, in almost no circumstance will the link between a name and deviant activity be

[73] Enthusiasts would hardly benefit from a perception that they wanted to protect criminals.

[74] For a discussion of rape in cyberspace, see Julian Dibbell, A Rape in Cyberspace, Village Voice, Dec. 21, 1993, at 36, *reprinted in* 1994 Ann. Surv. Am. L. 471 (1994).

[75] See Kang, supra note 44, at 1260 (noting that surveillance creates a threat of self-censorship); Hunter, supra note 22.

[76] See Belgum, supra note 8, ¶ 34.

disclosed to a third party unless that activity is both deviant *and* illegal.

Professor Julie Cohen is among the most persuasive exponents of the fourth conceptualization, which treats privacy as "constitutive" in the sense that it is an essential condition for one to function meaningfully in a democratic society.[77] Commentators generally refer to this position as "autonomy theory"[78] because it characterizes privacy as flowing directly from the notion that it is a necessary condition for individual autonomy.[79] While the quality of these arguments should not be understated, these arguments do not figure prominently in this discussion for at least two reasons. First, Cohen's position occupies highly contentious philosophical space. Powerful counterarguments cohabit this intellectual territory, and resolving the broader issues regarding the role of self-definition in political institutions lies beyond the scope of this Note.[80] Second, any attempt to apply this conceptualization of privacy to the clickstream context is almost impossible. Cohen's argument is so abstract that even if one accepted its philosophical propositions on-face, the application would remain subject to intense debate.[81]

Finally, some, including commentator Professor Jeffrey Rosen, conceive of informational privacy as a right to control how one is revealed to others.[82] To adopt the nomenclature favored by privacy

---

[77] See Cohen, A Right to Read Anonymously, supra note 45, at 1003–19; Julie E. Cohen, Examined Lives: Informational Privacy and the Subject as Object, 52 Stan. L. Rev. 1373, 1423–36 (2000) [hereinafter Cohen, Examined Lives] (arguing that there should be a constitutionally-created right against certain kinds of commercial collection); Cohen, Privacy, Ideology, and Technology, supra note 20, at 2039 ("Modern privacy advocates, including both [Jeffrey] Rosen and myself, conceive of privacy as a species of constitutive freedom and view that freedom as both intrinsically and instrumentally valuable."); see also Schwartz, supra note 3, at 834–43 (discussing various aspects and implications of constitutive privacy).

[78] See Reidenberg, supra note 63, at 1341.

[79] See Cohen, Examined Lives, supra note 77, at 1423–36.

[80] See, e.g., David Brin, The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom? 23 (1998) (arguing that more, not less, access to certain personal information would solve many privacy problems); Rosen, supra note 11, at 209 ("Not everyone is convinced by the social value of privacy.").

[81] For example, if one decided that there was a constitutive right to read anonymously, would web surfing count as reading? If web surfing were considered reading, how anonymous would the cookies have to be? Would such a paradigm allow pseudonymonous profiling?

[82] See, e.g., Rosen, supra note 11, at 10.

theorists, this idea fits under the umbrella of autonomy theory.[83] Distinctions between Rosen's theory and those of other autonomy theorists center in part on the acceptable conditions for alienating information.

In *The Unwanted Gaze*, Rosen uses the Monica Lewinsky debacle to illustrate this conceptualization.[84] During the course of the Paula Jones trial the court subpoenaed Lewinsky's purchase history from Amazon.com. Some of the purchases were embarrassing to Lewinsky, and Rosen argues that this violated privacy conceptualized as the right to selective revelation.[85] The government violated Lewinsky's rights because it impaired her ability to reveal herself selectively to others—she became "that type of person" who reads books about phone sex.[86] Lewinsky therefore became a personality subject to cultural stereotypes and uninformed generalizations.[87] Lewinsky ultimately elected to cede her purchase records to the Starr commission without a legal contest, so the issue of whether certain courts may subpoena this information remains unanswered.[88]

One can envision both a weak and a strong statement of this conceptualization. The weak version would say that people have the right to selectively reveal personal information to collecting agents. There are obvious exceptions—people must submit financial information to the IRS, for example—but most would not object to the gist of this assertion. The weak version, then, is more an obvious statement of moral consensus than anything else. Absent extenuating circumstances, people are free not to disclose personal information to collecting agents.

The strong version of the conceptualization, however, does resonate in the clickstream context.[89] This version suggests not only that

---

[83] See Cohen, Privacy, Ideology, and Technology, supra note 20, at 2030–31.

[84] Rosen, supra note 11, at 4–6.

[85] See id. at 229–30.

[86] See Daniel M. Filler, From Law to Content in the New Media Marketplace, 90 Cal. L. Rev. 1739, 1751–52 (2002) (book review) (discussing the type of person that reads a book about phone sex).

[87] See Rosen, supra note 11, at 9.

[88] See David Streitfeld, Lewinsky to Turn Over Book Purchase Information, Wash. Post, June 23, 1998, at A4.

[89] See Rosen, supra note 11, at 198 ("[P]eople don't want their browsing habits collected in personally identifiable dossiers, because those dossiers can be bought or

people may selectively reveal themselves to collecting agents, but that they may also control how those entities use the information and to whom they reveal it. Because Rosen's idea regarding the right to control and condition the revelation of personal information to other entities—to be judged in context—may be a bit ambiguous, the concept requires some elaboration. First, the right to be judged in context may mean the right to be judged only when the person doing the judging has all relevant information before making a judgment. Such a position is a conceptual impossibility, however, as every judgment one makes about another is based on an incomplete collection of information about that person.[90] Second, it may mean a right to be judged only when there is *sufficient* information to make an informed judgment. This threshold of sufficiency, however, is almost impossible to identify—how much information would be enough, and in which contexts? Third, and most plausibly, it may mean the right to have collecting agents abide by conditions of information transfer so as to give those about whom they are collecting information substantial control over data used to make probabilistic judgments in secondary contexts. A "secondary context" is some exchange of information involving the collecting agent after the subject of that information has transmitted it to that collecting agent. The following discussion proceeds from this third understanding.

The corollary of the notion that users may refuse to reveal information in the presence of certain conditions is the notion that they may disclose information. What most find objectionable in the clickstream context is the inability of users to condition revelation on a collecting agent's adherence to a set of mutually understood expectations. An Internet user will reveal information when the return on revealing it exceeds the cost. The "cost" of revealing information and, to a lesser extent, the return on revelation are highly subjective values, particularly in the clickstream context.

---

subpoenaed by employers, insurance companies, divorcing spouses, and others who have the ability to affect our lives in profound ways.").

[90] Such a claim is obviously predicated on the idea that any collection of information is an imperfect proxy for reality. See Cohen, Examined Lives, supra note 77, at 1404 ("Information theory, in contrast, recognizes that 'information' and 'reality' are different (though related) things, and that 'knowledge' forms an imperfect and culturally contingent bridge between them.").

The following Sections argue that these subjective valuations are difficult to reconcile with any prophylactic standard and that legal rules should complement market exchanges for this information.

III. INFORMATIONAL PRIVACY AND FAILING MARKETS

A. Defining the Market

Because many accept the proposition that a state of informational privacy is a by-product of the market for information,[91] a large subset of those who find that state undesirable may characterize that condition as one of "market failure."[92] Before devising legal regimes to correct market failures, however, one must understand the character of the failure itself.[93]

The first step in addressing market failure is identifying who is failing to buy or sell what at a competitive price from or to whom. The proposition that "the information market is failing" contemplates the market at an inappropriately high level of abstraction. The proposition that the "market for online information is failing" is similarly suspect (in that it is too abstract) for at least two reasons. First, collecting and searching agents obtain data in a variety of ways, and only a fraction of those exhibit properties whereby society can comfortably deploy market institutions to control the transfer of information. Second, even if one can characterize a standard practice as a market, it does not follow that all such practices that take place in the same medium (cyberspace) comprise the same market. Markets for email lists are quite distinct from markets for clickstream data. There is some overlap between the two because the same collecting and searching agents may be involved, but the transactional requirements governing information exchange in each context are radically different. Properly concep-

[91] See, e.g., Cohen, Privacy, Ideology and Technology, supra note 20, at 2031 ("Within the private sector, the impetus for the destruction of privacy is not prurience or prudishness, but core values that animate the rational marketplace behavior of profit-seeking entities.").

[92] See, e.g., Schwartz, supra note 3, at 833 ("Due to the pervasive failure in the privacy market in the United States, a subsidy is given to those data processing companies that exploit personal data.").

[93] See generally Charles Wolf, Jr., A Theory of Nonmarket Failure: Framework for Implementation Analysis, 22 J.L. & Econ. 107, 107–12 (1979) (discussing nonmarket failure and the four classic types of market failures).

tualized, the underlying exchange does not involve currency, but instead represents a sale of data where the marginal loss of informational privacy[94] associated with downloading a web page approaches the value a user gains by viewing it.[95]

The market for personal information collected during the course of online transactions, such as registrations or purchases, and the market for clickstream data may be failing for reasons that are, in the abstract, similar. Each market exhibits chronic information asymmetries, and, in each market, information-buyers exhibit suspect fidelity to expressed obligations.[96] The costs and returns on collecting data in each market, however, are radically different. Each market therefore requires a unique set of rules to correct its particular failures. The following is a discussion of the clickstream market.

B. Identifying Failure

The second step in addressing a market failure is determining which conditions constitute failure and whether the market in question exhibits them. In his illuminating 1979 article on *non*market failures, economist Charles Wolf, Jr. begins by summarizing "the essential points in the accepted theory" of market failures.[97] The four archetypal market failures occur in situations where there exist (1) externalities and public goods;[98] (2) increasing returns;[99] (3)

[94] In actuality, the marginal value could be equal to or exceed the value of the marginal privacy loss and whatever other marginal costs (e.g., time) are associated with the page view. I treat these values other than privacy loss as negligible.

[95] The notion that there may be a cognizable market for information is easier to accept once one understands that the act of viewing a web page is a transactional exchange. See supra notes 34–38 and accompanying text.

[96] See Schwartz, supra note 3, at 822–23.

[97] See Wolf, supra note 93, at 107–08. Wolf notes that the appropriate concept is actually closer to "market inadequacies" than to "market failures," since, strictly speaking, "market failures" refers exclusively to departures from Pareto-efficient outcomes and therefore ignores the distributional implications of different strategies. Id. at 108 n.5.

[98] Id. at 108.

[99] Id. at 109.

market imperfections relating to information asymmetries and transaction costs;[100] and (4) distributional inequities.[101]

Although there may be salient points to be made about the presence of (1) and (2) in the clickstream market,[102] most of the literature addresses (3) and (4).[103] Those failures associated with market imperfections are discussed extensively below, but conspicuously omitted is a similarly comprehensive discussion relating to the distributional implications of clickstream privacy. Many commentators lament the fact that consumers may be compelled to forgo their privacy because they cannot get the relevant goods or services elsewhere.[104] Because these choices can be made in a perfectly functioning market, however, the objection relating to the distribution of information value between collecting agents and users is one rooted in philosophy, not in economics. Moreover, such concerns are less pressing in the clickstream context because the good in question is often highly fungible reading material. To the extent that there will often exist many close substitutes for a given piece of nonessential material, a choice, made in a perfectly functioning market, to cede certain pieces of user information would not constitute a forced sale in the sense that the user sells because there are no substitute buyers.[105] If a user were unwilling to cede information under certain conditions, she would have the option of going to another website. The market for clickstream data therefore fails

---

[100] Id. at 110. Wolf does not actually identify "transaction costs," but such costs certainly fit under the umbrella of imperfect mobility characteristics of perfect markets. See id.

[101] Id. at 110–12.

[102] For example, transactions for personal data contain inevitable externalities. See Richard S. Murphy, Property Rights in Personal Information: An Economic Defense of Privacy, 84 Geo. L.J. 2381, 2401–02 (1996). With respect to the increasing returns scenario, there may well be certain instances where the marginal cost of collecting each data point is constant, but to the extent that this concern relates to how this constant figure manifests itself in monopoly markets, it is less salient in the clickstream context because of the absence of such monopolies.

[103] See, e.g., Cohen, Examined Lives, supra note 77, at 1397–98 (arguing that inadequate discounting distorts consumer privacy valuations); Solove, supra note 3, at 1450–51 (identifying several problems with the information market); Jeff Sovern, Protecting Privacy with Deceptive Trade Practices Legislation, 69 Fordham L. Rev. 1305, 1327–28 (2001) (addressing the character of information asymmetries).

[104] See, e.g., Cohen, Examined Lives, supra note 77, at 1397.

[105] For an example demonstrating that these conditions exist, see id. at 1397; infra Section VI.B.

primarily in the third sense, because there exist tremendous market imperfections relating to insufficient information and enforcement. Those problems are treated more comprehensively in Section C below.

Commentators almost pathologically appropriate survey results as evidence of widespread market failure.[106] These surveys, however, are rarely useful in helping lawmakers understand such failure. They provide little guidance for at least four reasons: (1) the wording of questions distorts results;[107] (2) attitudes about "privacy" reflect attitudes about activity that is conceptually distinct from that term's meaning;[108] (3) the surveys do not accurately represent tradeoffs between privacy and other consumer preferences;[109] and (4) they generally do not measure the concern for online privacy relative to other social desires.[110] Problems (1), (2), and (4) are all salient reasons for rejecting survey data as a cornerstone of online privacy discussion, but problem (3) is particularly compelling. To illustrate this point, assume a market comprised of only two cars, Mercedes and Subaru. If ninety-nine percent of survey respondents indicate a "strong preference" for Mercedes over Subaru, this fact does not mean that the market is in disequilibrium because only three percent of the population actually owns the German automobile. Naturally, most people prefer a state of privacy over one of no privacy, but such a statement says nothing about the strength of that preference. Surveys therefore distort the choice because in reality the decision is not between a state of privacy and that of no privacy, but between a state of more privacy

[106] See, e.g., Hatch, supra note 7, at 1477–81 (reciting a number of survey results); see Bartow, supra note 8, at 675 (citing Steve Lohr, Compressed Data: Survey Shows Few Trust Promises on Online Privacy, N.Y. Times, Apr. 17, 2000, at C4 (quoting a market research firm study that found that ninety-two percent of people don't trust companies to keep information confidential)); Jenab, supra note 22, at 651–52 (citing various poll results to suggest that Americans are concerned about privacy and that more legislation is necessary to protect it); Jim Harper & Solveig Singleton, With a Grain of Salt: What Consumer Privacy Surveys Don't Tell Us (June 2001) (unpublished manuscript, on file with the Virginia Law Review Association), available at Social Science Research Network, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=299930.

[107] Harper & Singleton, supra note 106, at 1.

[108] Id.

[109] Id.

[110] Id. at 2.

and other goods. Despite the ostensibly intense preference for privacy expressed in surveys, most consumers do not turn off their cookies.[111] In fact, one online study showed that less than one percent of users had their cookies disabled.[112] Why this tremendous disjunct between what consumers say and what they do? There are two reasons. First, as discussed above, consumers' actual privacy preferences are not nearly as strong as survey figures suggest.[113] Second, as will be discussed below, the current market systematically disadvantages consumers.

C. Causes of Market Failure

The third step in addressing market failure involves understanding why users are systematically failing to exact a competitive price from collecting agents. These failures are well chronicled in the legal literature,[114] but in the interest of simplicity they can be grouped into two broad categories. First, when consumers implicitly "consent" to observation by viewing a page, they generally do not know what collecting agents may do with that observation (if the consumers know that they are being observed at all).[115] Second, no effective enforcement mechanism guarantees compliance with the expectations generated if a user conditions her consent upon promises made by the collecting agent.[116] The two are closely related—one cannot enforce the conditions upon which a user grants consent without first discerning that a user has "consented"—but in order to understand how they relate, it is first necessary to discuss the two separately.

The first cause of clickstream market failure is an egregious information asymmetry between users and collecting agents because consumers lack information about what collecting agents may do

---

[111] Id. at 7.

[112] Id. (citing Privacy: Caught With the Cookie Jar, Nat'l J. Tech. Daily, A.M. Edition, Apr. 4, 2001, at http://nationaljournal.com/pubs/techdaily/ ("Web analysis service Web Side Story found in a review of more than 1 billion page views cookies were disabled just .68 percent of the time.")).

[113] See supra notes 106–12 and accompanying text.

[114] See Wolf, supra note 93, at 107.

[115] See Solove, supra note 3, at 1451; Hunter, supra note 22.

[116] See Solove, supra note 3, at 1451–52; Robert Thibadeau, A Critique of P3P: Privacy on the Web, at http://dollar.ecom.cmu.edu/p3pcritique/ (Aug. 23, 2000) (on file with the Virginia Law Review Association).

with their data.[117] It almost goes without saying that, as opposed to users, the collecting agents themselves know which data they are collecting and with some exceptions, what they will do with it.[118] In the absence of legal enforcement, the market could nonetheless function efficiently if consumers could roughly calibrate their surfing patterns to reflect the amount of information they wished to divulge.[119] In such a model, a user would effectively "vote with her mouse," visiting only those pages for which the value of the "view" exceeded the marginal loss associated with revealing the data point (and other costs associated with looking at the page).[120]

Consumers must acquire knowledge about what collecting agents do with data, but the problem is that they struggle to do so when the agent does not conveniently provide such knowledge. The relational information defaults of the consumer-collecting agent transaction (in the clickstream context) give buyers no incentive to inform the sellers. The absence of meaningful consumer awareness, coupled with the even-more-conspicuous absence of legal rules for enforcing the promises contained in online privacy policies, means that the collecting agent has little incentive to facilitate an explicit understanding of user expectations.

Many collecting agents now have hyperlinks to privacy policies housed on a company website.[121] Anyone who has tried to read them, however, knows that discerning their clear meaning is an exercise in frustration. The policies are often difficult to locate.[122] The language of the policies is often unclear and difficult for laypeople

---

[117] See Solove, supra note 3, at 1431; Hunter, supra note 22.

[118] Collecting agents can never anticipate all the ways they may be able to use data. New uses of data spring up all the time, and part of the value collecting agents attribute to a certain datum is surely some residual possibility of an unexpected use.

[119] This is simply the standard economic argument that, *ceteris paribus*, consumers would gravitate toward those services that exacted privacy of some value less than that acquired through the page view.

[120] See supra note 94.

[121] See, e.g., Walt Disney Internet Group (WDIG) Privacy Policy, at http://disney.go.com/legal/privacy_policy.html (last visited Mar. 27, 2003) (containing the privacy policy for, among other websites, ESPN.com); see also Schwartz, supra note 3, at 823–24 (noting that the Georgetown Internet Privacy Policy Survey, sponsored by the Federal Trade Commission, "found that Web sites with the most passenger traffic were increasingly offering click-on 'Privacy Notices'").

[122] These policies are increasingly used and available, but they are rarely featured prominently on the homepage.

to understand.[123] Although it is rarely clear to the user what activity constitutes consent to which terms, "sticking around" is increasingly treated as granting consent to use data collected as users navigate the website.[124] Websites often reserve the right to change their privacy policy at any time.[125] Finally, the policies rarely clarify the information practices of third parties with material on the website.[126]

Some commentators resist legal solutions to this conundrum because they maintain that, given time, industry norms will adjust to rectify the situation.[127] They point to industry initiatives to control data mining and the advent of certificatory companies like TRUSTe as evidence of the success of these developing norms.[128] These responses, however, have quite noticeably failed to force collecting agents to reveal a substantial amount of information.[129] Generally speaking, these certificatory companies merely verify the existence of privacy policies rather than examine their accuracy.[130] Even if these certificatory companies succeed in compelling collecting agents to post privacy policies, such success would not

---

[123] See Schwartz, supra note 3, at 824.

[124] See id. (noting that the rule of thumb seems to be that if an Internet user sticks around, then that person has effectively consented to all terms, even if those terms change over time).

[125] See, e.g., The Kroger Co. Privacy Policy, Kroger, at http://kroger.com/privacy-policy.htm (last visited May 28, 2003) (on file with the Virginia Law Review Association) ("We reserve the right to change our privacy policy at any time.").

[126] See Hunter, supra note 22.

[127] See, e.g., Electronic Commerce: The Current Status of Privacy Protections for Online Consumers: Hearing Before the Subcomm. on Telecomm., Trade, and Consumer Prot. of the House Comm. on Commerce, 106th Cong. 89–94 (1999) (statement of Solveig Singleton, Director of Information Studies, The CATO Institute), available at www.cato.org/testimony/ct-ss071399.html.

[128] See Froomkin, supra note 9, at 1525 (stating that TRUSTe.com may be "the most visible and successful self-regulatory initiative"). Companies like TRUSTe provide seals authenticating the integrity of a site's professed data practices. Id.

[129] See Heimes, supra note 39, at 97; Jenab, supra note 22, at 660 ("Market participants are not, then, rushing to enact meaningful privacy policies."). Jenab further cites a study indicating "that fewer than 10% of the web's busiest sites adequately inform consumers of how their personal information is being processed," while 92.9% of them collect personal data. Id. (citing Comments on the Georgetown Internet Privacy Policy Survey, Center for Democracy and Technology, at http://www.msb.edu/faculty/culnanm/GIPPS/cdt.pdf (last visited Feb. 19, 2001)).

[130] See Froomkin, supra note 9, at 1525 (discussing the procedures of the TRUSTe certificatory company).

address problems involving what constitutes consent to such policies. Finally, collecting agents are not consistently re-evaluated to ensure that they are complying with the policy as originally conveyed to the certificatory company.[131] These certificatory companies possess an economic incentive not to be too harsh on clients.[132] Substantial evidence shows that these privacy policies do not figure prominently in users' decisions to allow collecting agents to exact information.[133] Consequently, the risk of net losses associated with aggressive enforcement through suspension of trustmarks (icons indicating that a given collecting agent meets certain standards) would render these programs sufficiently unappealing to potential clients that these certificatory companies would be enforcing themselves right out of the market.[134]

Although the relevant literature extensively discusses the failure of legal rules to force collecting agents to disclose information regarding data practices,[135] it largely ignores their failure to force consumers to reveal certain useful pieces of information about themselves: their privacy preferences. The effect is to promote a regime where buyers have to reveal what they do without sellers having to reveal what they want. Such an oversight is understandable because, up until quite recently,[136] there had been no cost-efficient way for users to transmit this information to collecting agents. Even if websites had been willing to provide a form to determine the clickstream privacy preferences of their users, the number of people completing the form would have been so small that the economic benefit to a website differentiating its service offering based on those preferences would be minimal.[137] A drop in the costs of revealing preferences would increase gains from trade by giving

---

[131] See, e.g., id. at 1526.
[132] Id. at 1526–27.
[133] See Cranor et al., supra note 28.
[134] See Froomkin, supra note 9, at 1526–27.
[135] See supra notes 118–34 and accompanying text.
[136] See infra Section V.B.
[137] Completion of the form would surely be voluntary, and the return on time spent filling out the form would in almost every instance fail to give sufficient incentives for the user to fill it out.

collecting agents valuable information about the potential consumer response to changes in privacy policies.[138]

The second broad cause of clickstream market failure is the glaring absence of a meaningful legal enforcement mechanism. Even if a website's privacy policy is abundantly clear, no legal force guarantees the expectations it generates.[139] Granting a civil cause of action to the person whose expectations are disappointed does not make sense if the cost of bringing it exceeds the value lost by the collecting agent's violation.

The absence of any enforcement mechanism means that sites do not have to honor their own privacy policies and that they are free to use whatever language they choose in crafting potentially misleading guarantees. Collecting agents have an incentive to avoid misusing consumer information only to the extent that such use risks negative publicity. The violation of consumer expectations must be enormous before it inflicts sufficiently concentrated losses to generate a response. DoubleClick's acquisition of Abacas Direct, an offline data miner, was one such public incident.[140] Collecting agents are free to commit smaller violations because in such

---

[138] In a market with accurate information about data uses, a collecting agent would stand to gain from knowledge of the privacy preferences of its audience/user base. Based on that information, the agent may be able to diversify its services so as to satisfy each user's privacy preference. At very high costs of acquiring information about these preferences, however, the return on understanding the privacy preferences of its audience/user base is not worth the costs of figuring them out.

[139] See Froomkin, supra note 9, at 1527 (stating that the United States might be unique in having a self-regulatory system without any means of enforcing it); James Goodale et al., Panel I: The Conflict Between Commercial Speech and Legislation Governing the Commercialization of Public Sector Data, 11 Fordham Intell. Prop. Media & Ent. L.J. 21, 22–23 (2000) (identifying the need for enforcement as a critical issue in online privacy discourse); Solove, supra note 3, at 1451 ("Most privacy policies have no way to prevent changes in policy or a binding enforcement mechanism."); Jeff Sovern, Opting In, Opting Out, or No Options at All: The Fight for Control of Personal Information, 74 Wash. L. Rev. 1033, 1100 (1999) ("[S]ome enforcement apparatus would be necessary to ensure that businesses were living up to their obligations under the regulation."); Belgum, supra note 8, ¶ 46; Thibadeau, supra note 116.

[140] See Rosen, supra note 11, at 164. DoubleClick, the largest advertising network on the Internet, purchased Abacus Direct, which possessed an enormous marketing database profiling offline consumer behavior. DoubleClick declared its intention to cross-reference its online data with that of the offline marketer, igniting a firestorm ultimately causing its stock to plummet and prompting it to alter its plans for the data integration. See id.; see also Zimmerman, supra note 40, at 445 (describing DoubleClick's acquisition and the subsequent public reaction).

situations they do not incur the losses associated with public relations disasters. One should not overstate the latitude collecting agents currently enjoy, however, because these disasters carry a sufficiently severe penalty that data miners already have an incentive to avoid the most egregious practices implicating informational privacy.

The failure of consumers to track and report back on misuse of their clickstream data represents the classic collective-action problem.[141] A single consumer cannot appropriate all the information gains a market would accrue through such unilateral audits. Only for the most extreme privacy sociopath is the return on monitoring sufficiently large that it exceeds its cost, so only such user anomalies would bother to see what happens to their data. In the absence of a legal entitlement to preclude misuse, news about a violation does nothing but provide consumers with information about the likelihood of such an occurrence in the future. Without a cause of action, the user performing such meticulous tracking gets only this predictive information in return for her exhausting audit.

Norms also fail to remedy this aspect of market failure. If users could successfully organize to fund a single auditing agent, then information about misuse, even absent a cause of action, could be worth the pro-rata cost of contribution. Certificatory companies such as TRUSTe, however, do not do any such back-end inquiries into whether collecting agents adhere to the obligations expressed in their privacy policies.[142] The absence of such unitary auditing entities is likely a testament to the fact that, without a cause of action, consumers are not willing to pay for them.

The absence of any back-end check on whether collecting agents are fulfilling their obligations has disastrous effects on the market. Unenforceability renders the cost of exchanging information extremely uncertain for consumers. As adherence to expressed obligations becomes more uncertain, and because users are risk-averse, the valuation a user places on a privacy term, V_{User}, falls to a level lower than the expected likelihood of adherence multiplied by V_{User} where all obligations are honored. This uncertainty destroys poten-

[141] See Schwartz, supra note 3, at 822. For a textbook explanation of the collective-action problem, see Charles J. Goetz, Cases and Materials on Law and Economics 24–29 (1984).

[142] See Froomkin, supra note 9, at 1525.

tial gains from trade with risk-averse users who are aware that they are being observed. Some would doubtlessly respond to this claim by noting that the explosion of Internet surfing over the last several years belies uncertainty's value-destroying effects,[143] but such uncertainty is only one of many variables that govern the decision about whether to surf the Internet. The argument is not that uncertainty grinds web surfing to a halt, but instead that, *ceteris paribus*, it retards growth relative to a market without it.

D. Two Objectives for a Solution

The final step in addressing market failure involves identifying the objectives that should drive measures to fix it. If one accepts that under appropriate conditions certain information should be freely alienable, then it is relatively uncontroversial that some mix of the following two objectives should animate the market for clickstream data: (1) Consumer expectations should be protected (the "compensatory objective") and (2) information should flow to the highest valued use (the "utility objective"). Unfortunately, these objectives are not always complementary. The relationship between these objectives may be illustrated by a mathematical relationship. If costs of information, transaction, and enforcement were all zero and V_{User}[144] was known by all parties, then any promotion of the compensatory objective would always promote the utility objective so long as the user received V_{User} upon any single unauthorized use of the data, DU_{Unauth}, where the value of such use, DUV_{Unauth}, exceeds V_{User}. Then, by definition, collecting agents would always and only violate expectations when the value of the unauthorized use exceeded the user's reservation price for parting with her privacy.[145] If collecting agents always and only violate consumer expectations when $DUV_{Unauth} > V_{User}$, and if the aforementioned assumptions hold, a legal regime facilitating such decisionmaking would satisfy both objectives.

---

[143] See, e.g., Volokh, supra note 12, at 1118 (arguing that the absence of such enforcement mechanisms demonstrates that uncertainty about privacy does not chill Internet use).

[144] Again, V_{User} is the value the user places on adherence to the privacy term—*not* the value the user places on the page view itself.

[145] This conclusion is true assuming a user's reservation price is generally equivalent to V_{User} (an assumption relaxed in subsequent analysis).

The fact that none of these assumptions holds in the real world means that, at some level of fidelity to privacy expectations, increased enforcement will diminish gains from trade. In a situation where there existed no fidelity to privacy expectations, increased enforcement would almost certainly *bolster* gains from trade by keeping the Internet attractive to consumers.[146] A state of informational privacy where collecting agents exhibited no fidelity to user expectations would generate uncertainty that would almost certainly deter web surfing, making it more difficult for collecting agents to acquire data in the first place. At some higher level of enforcement, however, the associated costs may begin to constrain gains from trade as they would either decrease the net benefit of data collection for collecting agents or decrease the reservation price paid to the user in the event of a violation.[147]

Nevertheless, holding the level of privacy protection constant, one legal regime is preferable to another if it can be implemented with fewer sacrifices in the form of gains from trade in information. Ideally, a legal regime would seek to encourage unauthorized data uses so long as (1) the collecting agent pays the user at least V_{User} (satisfying the compensatory objective), and (2) $DUV_{Unauth} > V_{User}$ (satisfying the utility objective). Economists would consider these secondary data uses to be "Pareto superior" because all parties would be at least as well off as they would be prior to the violation of the privacy term.[148]

The practical workability of a legal regime grounded in such thinking turns on the availability of its technological predicates. It would seem obtuse to speak of a user's reservation price if there

---

[146] This would happen because the increased certainty would come without any implementation or enforcement costs. For a general discussion of the ways in which fair information practices relate to network use, see Joel R. Reidenberg & Francoise Gamet-Pol, The Fundamental Role of Privacy and Confidence in the Network, 30 Wake Forest L. Rev. 105, 123–25 (1995). This analysis assumes people know that collecting agents were acquiring information on them.

[147] In other words, because implementation and enforcement do cost something, at some point the marginal expenditure on enforcement will not be worth the marginal increase in those willing to use the Internet. This happens because either the user or the collecting agent will bear much of the implementation and enforcement cost, effectively diminishing the return on transactions for information.

[148] The user is better off because she gets paid an amount equivalent to her valuation on the privacy expectation, V_{User}, and the collecting agent is better off because it gets value equivalent to $DUV_{Unauth} - V_{User}$.

were no realistic way of determining it. It would be similarly wrongheaded to premise rules on the ability to return these values to consumers in the event that collecting agents violate expectations if there is no realistic way either of identifying the violation or of inexpensively compensating the "victim."

Parts V and VI advance some ideas about how a regime might solve these problems, but there admittedly remains a salient objection regarding the difficulty of enforcing expectations at a sufficiently low transaction cost so as to preserve gains from trade. Following Part IV's discussion of current privacy doctrine is an attempt to outline a statutory regime that achieves both compensatory and utility objectives by minimizing the transaction cost of monitoring and punishing clickstream privacy violations.

IV. THE CURRENT STATE OF LEGAL PROTECTION

Existing legal doctrine is not equipped to deal with the novel demand for information that the World Wide Web enables.[149] Statutory remedies, on at least three occasions, have proven to be equally incapable of redressing alleged injury from abuse of cookie data.[150] There exists no comprehensive U.S. privacy legislation, and courts have, for various reasons, refused to subject clickstream pro-

[149] See Lessig, supra note 9, at 152–53 (arguing that the advanced modern capacity to monitor and archive data exacerbates privacy concerns); Murphy, supra note 102, at 2402 ("The plummeting cost of data storage and dissemination and the expanding uses of particularized information have conspired to drive up the value of such information to third parties, and, hence, to the merchant who collects the information.").

[150] See Chance v. Ave. A, Inc., 165 F. Supp. 2d 1153 (W.D. Wash. 2001) (dismissing claims against a third-party ad server arising under the Stored Communications Act and the Federal Wiretap Act); In re DoubleClick, 154 F. Supp. 2d 497 (S.D.N.Y. 2001) (dismissing claims against a third-party ad network arising under the Electronic Communications Privacy Act, the Federal Wiretap Act, and the Computer Fraud and Abuse Act); In re Intuit, 138 F. Supp. 2d 1272 (C.D. Cal. 2001) (dismissing claims against a first-party website operator arising under the Computer Fraud and Abuse Act and the Federal Wiretap Act). The reasoning for the dismissals varies from case to case. Even in the event that courts were to apply a statutory prohibition to cookie-assignment, such a prohibition would doubtlessly be prophylactic in character and therefore subject to many of the same criticisms leveled in notes 177–83 infra and accompanying text. There has been one case that has held cookie collection to be in violation of the Electronics Communications Privacy Act, but in that case, the violation stemmed from the fact that the third party collecting personally identifiable information on users did not secure adequate consent from its first-party website business partners to do so. See In re Pharmatrak, Inc., 439 F.3d 9, 20 (1st Cir. 2003).

tocols to the Stored Communications Act,[151] the Federal Wiretap Act,[152] and the Computer Fraud and Abuse Act.[153] Fourth Amendment law and privacy torts are the primary nonstatutory vessels available to those who would invoke the law to control entities that buy and sell clickstream data. Unfortunately, applying privacy tort law would require courts to completely rework existing doctrine, and applying the Fourth Amendment would require the same to be done to the Constitution.[154]

Under the state action doctrine, the Constitution grants individuals protection against only state actors.[155] Most of the entities that systematically collect and distribute information about Internet users, however, are private.[156] Constitutional protection surrounding the exchange of personal information is therefore limited.[157]

Most of what people find objectionable in the clickstream context has little or nothing to do with government. Private companies collect cookie data and other private companies buy it. In 1890, Professor Samuel Warren and then-Professor Louis Brandeis first advanced the modern conception of privacy in their famous article

---

[151] Stored Communications Act, 18 U.S.C. §§ 2701–2710 (2000); see supra note 150. The scope of statutory regulations is quite small. The most substantial privacy law is the Stored Communications Act, 18 U.S.C. §§ 2701–2710. The online scope of the act is limited to civil and criminal interception, use, and disclosure rules for email content. See David L. Sobel, The Process that "John Doe" is Due: Addressing the Legal Challenge to Internet Anonymity, 5 Va. J.L. & Tech. 3, ¶¶ 9–10 (2000), at http://www.vjolt.net/vol5/symposium/v5ila3-Sobel.html (on file with the Virginia Law Review Association).

[152] Federal Wiretap Act, 18 U.S.C. §§ 2510–2522 (2000); see supra note 150.

[153] Computer Fraud and Abuse Act, 18 U.S.C. § 1030 (2000); see supra note 150.

[154] The Supreme Court's holding in *Whalen v. Roe*, 429 U.S. 589 (1977), suggests that some constitutional protection for informational privacy may exist where the state is the collecting agent. See id. at 606 (Brennan, J. concurring).

[155] See, e.g., United States v. Jacobsen, 466 U.S. 109, 113 (1984); Spetalieri v. Kavanaugh, 36 F. Supp. 2d 92, 107 (N.D.N.Y 1998).

[156] See supra note 5.

[157] Professor Lawrence Lessig has argued that, because the government can control the development of Internet architecture, the state action doctrine is not as inapposite in the online context as one might think. See Lessig, supra note 9, at 217. This position seems to ignore the fact that the government can *almost* always control the terms by which private parties exchange information, but such control does not render the underlying private activity subject to constitutional restriction. Additionally, there is no Fourteenth Amendment privacy interest in less sensitive affairs. *Spetalieri*, 36 F. Supp. 2d at 109.

in the *Harvard Law Review*.[158] The market for information has undergone dramatic changes over the last century, and, as a result, privacy torts cannot serve as a doctrinal foundation for regulating online data exchange.

For many years, tort law has protected a cluster of privacy interests centered largely around the need for celebrities and other notable public figures to insulate themselves from overzealous public scrutiny.[159] In an episode of *The Simpsons* guest starring Alec Baldwin and Kim Basinger, Homer quips, "Look, all I'm saying is, if these big stars didn't want people going through their garbage and saying they're gay, then they shouldn't have tried to express themselves creatively."[160] Famous people have always had to guard their privacy closely because there has always been money to be made in celebrity gossip, which is really nothing more than valuable personal information. With the advent of the World Wide Web, however, it can be lucrative to gather information about anybody. Costs of collection, storage, and search have fallen to such a degree that now more than just celebrities are concerned about protocols for managing personal information.[161] The threat to which law must respond is no longer merely that of people like Homer Simpson slandering celebrities and pawing through their trash, but that of highly organized companies systematically collecting and selling data about ordinary people. The market for information has changed in ways that render existing privacy torts a particularly clumsy doctrinal vehicle for addressing many modern privacy concerns.

The Second Restatement of Torts delineates four distinct actions for invasion of privacy: unreasonable intrusion into another's seclusion, appropriation of another's name or likeness, unreasonable publicity given to another's private life, and publicity that unrea-

---

[158] Samuel D. Warren & Louis D. Brandeis, The Right to Privacy, 4 Harv. L. Rev. 193 (1890).

[159] The exceptions to this generalization primarily involve the "Intrusion Upon Seclusion" tort. See Restatement (Second) of Torts § 652C (1977).

[160] The Simpsons: When You Dish Upon a Star (Fox television broadcast, Nov. 8, 1998).

[161] "'Is freedom inversely related to the efficiency of the available means of surveillance? If so, we have much to fear.'" Lessig, supra note 9, at 18 (quoting James Boyle, Shamans, Software, and Spleens: Law and the Construction of the Information Society 4 (1996)).

sonably places another in a false light before the public.[162] None of these actions would be useful in addressing clickstream privacy concerns.[163]

The false light tort plainly cannot apply because the relevant data does not represent false information.[164] The publicity tort cannot apply because it generally requires that the information be made available for public consumption. Intrusion upon seclusion and misappropriation are therefore the only conceivable tort actions available for protecting clickstream privacy interests.

The intrusion upon seclusion tort sets an extremely high bar for liability. The intrusion must (1) be into either a person's personal seclusion or her "private affairs," and (2) be "highly offensive to a reasonable person."[165] Courts appear unlikely to consider online intrusions as exhibiting either of these characteristics.[166] Clickstream data represents a record of which browsers downloaded which pages. For the action to lie, either the information contained in the cookie or the fact that this information can be associated with viewing certain content must be a "private matter." The data in the cookie itself may be personal, but it is rarely private.[167] One would also have a difficult time characterizing as an "intrusion" the fact that the collecting agent knows characteristics about the person to whom it is delivering content—particularly when offline commercial transactions generate this sort of information all the time. Of equal importance is that courts seem unwilling to characterize

---

[162] Restatement (Second) of Torts §§ 652A–652E (1977).

[163] See Jenab, supra note 22, at 655–56; Schwartz, supra note 20, at 778–79; Belgum, supra note 8, ¶¶ 19–23.

[164] See Belgum, supra note 8, ¶ 20.

[165] Restatement (Second) of Torts § 625B (1977).

[166] See Dwyer v. Am. Express Co., 652 N.E.2d 1351, 1353 (Ill. App. Ct. 1995) (rejecting intrusion claim against charge card company for selling card holder names, scored by purchase patterns, for remarketing). Moreover, most would agree that clickstream data is less "personal" than a profile of purchase habits associated with a name and address or phone number as in *Dwyer*. Finally, courts tend to reject such claims when the information is ceded "voluntarily," and the fact that most browsers can be configured to reject cookies would likely be sufficient to satisfy this "voluntariness" requirement. See Helms, supra note 21, at 310.

[167] The cookie may contain a zip code or an indication that the user prefers golf over tennis, but that information is not generally considered private. See Belgum, supra note 8, ¶ 23 (explaining that personal data is not synonymous with private data).

these practices as "highly offensive to a reasonable person,"[169] the second element of an intrusion claim.

Examination of misappropriation tort cases reveals that liability attaches only when a defendant uses a recognizable person's image or likeness to endorse a particular thing.[170] Although the misappropriation tort protects people from certain unauthorized uses of identities, this protection does not extend to the exchange of facts about ordinary people. In *Cox v. Hatch*, several workers objected to their appearance in a campaign photo with Orrin Hatch.[171] The court rejected the misappropriation claim in part because their identities did not have "intrinsic value."[172] The word "intrinsic" is inappropriate because even if an identity can be said to possess "value," that value is not in any sense primordial or inherent. The value of someone's identity or likeness is what the market is willing to pay for it. To say that celebrities' identities have worth is to say that people would be willing to pay for their likeness. The workers' identities did not lack "intrinsic value" because there is no such thing. The proper way of understanding why the workers were not deprived of economic value is that their identities' market value was sufficiently close to zero that the unauthorized use did not deprive them of any economic gain. Paul McCartney's likeness is worth something because (1) people respond positively to a perceived association with Paul McCartney, and (2) there are not many Paul McCartneys. Conversely, John Doe's likeness is not worth much because (1) an endorsement from John Doe is not uniquely compelling, and (2) there are many substitute John Does. Unauthorized use of John Doe's identity does not deprive him of an economic gain because no lucrative possibilities existed for him

---

[168] See Helms, supra note 21, at 311 ("[I]t is not likely that even the most detailed profile would rise to the level of a 'highly offensive' disclosure of a private fact.").

[169] See Helms, supra note 21, at 311 ("[I]t is not likely that even the most detailed profile would rise to the level of a 'highly offensive' disclosure of a private fact.").

[170] See, e.g., Shibley v. Time, Inc., 341 N.E.2d 337, 339 (Ohio Ct. App. 1975) ("It is clear from a reading of the authorities dealing with invasion of privacy that the 'appropriation or exploitation of one's personality' referred to . . . refers to those situations where the plaintiff's name or likeness is displayed to the public to indicate that the plaintiff indorses the defendant's product or business."). Cases where courts have relied on the misappropriation tort in relation to facts are limited to "hot-news." See, e.g., NBA v. Motorola, Inc., 105 F.3d 841, 853 (2d Cir. 1997).

[171] 761 P.2d 556, 558 (Utah 1988).

[172] Id. at 564.

to sell his identity in the first place. The distinction between the court's characterization of economic value and that presented here would be of little consequence in *Cox*, but would be extremely important in analyzing clickstream data collection precisely because in such a context an ordinary person's identity *does* have market value.

The analytic clarification in the preceding paragraph pushes the misappropriation tort further towards cognizability in the clickstream context because, under that understanding, the presense or absence of value does not turn on a user's status as ordinary. Despite this clarification, the misappropriation action would still fail to mediate online data collection because the objection to unauthorized identity use is not the revelation of the identity itself, but rather the use of that entity to *endorse* something. In *Tropeano v. Atlantic Monthly Co.*, for example, a court rejected a misappropriation claim against *The Atlantic Monthly* when a woman objected to her appearance in a photo associated with a story about changing sexual mores.[173] The court invoked the "incidental use" doctrine in the process of rejecting the claim.[174]

The magazine did not, the court reasoned, appropriate the woman's specific likeness for pecuniary gain—the photo was just illustrative. The author merely used the picture to describe the change in mores. Comment (d) of the Restatement suggests, and the *Tropeano* court seems to accept, that if one appropriates an identity for a description rather than for an endorsement then courts do not consider that identity to be misappropriated.[175]

---

[173] 400 N.E.2d 847, 848 (Mass. 1980).

[174] Id. at 850–51. The incidental use doctrine states:

> The value of the plaintiff's name is not appropriated by mere mention of it, or by reference to it in connection with legitimate mention of his public activities; nor is the value of his likeness appropriated when it is published for purposes other than taking advantage of his reputation, prestige, or other value associated with him, for purposes of publicity. No one has the right to object merely because his name or his appearance is brought before the public, since neither is in any way a private matter and both are open to public observation. It is only when the publicity is given for the purpose of appropriating to the defendant's benefit the commercial or other values associated with the name or the likeness that the right of privacy is invaded.

Restatement (Second) of Torts § 652C cmt. d (1977).

[175] See *Tropeano*, 400 N.E.2d at 850; Restatement (Second) of Torts § 652C cmt. d (1977).

Cookie data is used for describing and predicting behavior, not as a means of using one person's identity or visage to manipulate the behavior of another. Less abstractly, a successful misappropriation claim seems to require that the plaintiff's identity be used in some sort of publication. Clickstream data, however, is almost never published and in the rare instances when it is, the data is almost always anonymous and aggregated.[176]

Even if privacy torts could be strained to the point of relevance in the clickstream context, applying them would be economically undesirable. First, the inevitable reference to some objective standard of reasonableness introduces substantial uncertainty into the status of legal liability.[177] Because both websites and users would, *ceteris paribus*, disfavor greater risk, such legal ambiguity would chill both data collection and web surfing.[178]

Second, even if the standard were clear, few would sue. Tortious damage would generally be both small and dispersed amongst a tremendous number of users. Although class action lawsuits are designed to remedy precisely these conditions,[179] in the clickstream context they are sufficiently extreme that even the transaction cost of participating in a class action could dwarf the return on litigation. Even more importantly, the character of the injury is such that it imposes transaction costs associated with identifying the appropriate class of plaintiffs.

Third, even if people sued, calibrating damages so as to encourage socially optimal data collection would be impossible. Ideally, a

---

[176] In other words, when clickstream data is published, it is not published in such a way that indicates that Person *1* did *A*, Person *2* did *B*, etc., but instead the information is aggregated into broad generalizations like "the class of people with these characteristics tended to behave in the following way." See, e.g., John Chandler-Pepelnjak & Jaymee Johnson, New Industry Benchmarks: Purchase and Registration Drop-off Rates, Atlas Institute, at http://atlasdmt.com/media/pdfs/insights/DMIDropeview.pdf (2001) (on file with the Virginia Law Review Association) (detailing consumer "drop-off" rates aggregated by industry).

[177] For a discussion of the implications of juror-imposed uncertainty, see Goetz, supra note 141, at 299–302.

[178] For a discussion of why uncertainty can diminish participant utility relative to a certain outcome with the same expected value, see id. at 75–82.

[179] See Owen M. Fiss, The Political Theory of the Class Action, 53 Wash. & Lee L. Rev. 21, 24 (1996) ("In short, the class action could be viewed as a device to fund the private attorney general and is able to play that role because of the aggregation of the claims of a large number of persons who have similar or identical claims, none of which—standing alone—would justify the suit.").

tort regime would seek to assess liability only where the value of the unauthorized use (DUV_{Unauth}) is less than V_{User}. Private actors are likely to pursue this socially optimal strategy only where their marginal private cost equals the marginal social cost. The only way courts can force socially optimal information practices is therefore to calibrate the damages to reflect the marginal social cost. A number of violations will doubtlessly go unreported, so the damage measure for those that are successfully litigated would also have to impose a penalty in disproportion to the injury incurred by the user. The coefficient by which to scale the damage measure up or down, however, would be pure guesswork.[180]

Finally, even a perfectly administered liability rule would destroy potentially beneficial gains from trade. People have highly differentiated privacy preferences,[181] meaning that some strongly value their privacy while others do not. If downloading a web page is viewed as a transaction (content for information), some people would be willing to make this exchange with few restrictions, while others would make the exchange only after securing the most stringent protection. A prophylactic standard[182] eliminates possible gains from trade associated with such highly differentiated preferences.[183] A single, perfectly administered standard imputes to each user the preference of the average. A rule set at this average would therefore (1) fail to protect those whose preference is above the average and (2) foreclose the possibility of mutually beneficial transactions with those parties whose preference is below it.

---

[180] An appropriate coefficient would have to incorporate a figure representing the percentage of cognizable claims that, for one reason or another, are not filed. The very nature of these claims, however, makes measuring the frequency with which they occur almost impossible.

[181] See Cranor et al., supra note 28.

[182] It may strike one as odd to refer to a tort standard as prophylactic because most tort cases involve ex post assessments. One may justify such a proposition, however, because the context in which courts gauge the reasonableness of behavior is unlikely to change. What is reasonable to do with clickstream data collected on one website, for example, is likely very similar to what is reasonable to do with data collected on another. Any variation in the standard, then, would more likely be attributable to the subjective variation of juries rather than objective differences in context.

[183] Cranor et al., supra note 28 ("[I]t seems unlikely that a one-size-fits-all approach to online privacy is likely to succeed.").

In sum, existing privacy law is neither doctrinally appropriate nor economically desirable as a vehicle for addressing clickstream privacy concerns.

V. TOWARDS A REGIME OF DEFAULT ENTITLEMENTS

A. Oversight of Default Models

Promulgating a scheme of statutory default entitlements around which users and collecting agents could bargain could be, under certain conditions, a viable means of both protecting user expectations and preserving gains from trade. As intimated in Part III, the most important such condition is low bargaining costs. Default entitlement proposals (many times in the form of majoritarian contract default rules) for regulating personal data exchange are nothing new,[184] although to this author's knowledge no scholars have advocated such solutions for clickstream data. Default models are most desirable where transaction and enforcement costs do not dwarf the value of the underlying good being traded. Only recently has the advent of new browser technology and web surfing protocols rendered such a default system feasible.

Some opponents critical of more prophylactic tort standards have advocated propertizing online data.[185] This property paradigm, theorists reason, would avoid the opportunity costs of applying a prophylactic standard to a privacy context with highly differentiated preferences and would preserve users' autonomy to control how their data would be used. Intellectual property law, however,

---

[184] See, e.g., Kang, supra note 44, at 1246–59 (arguing for default rules vesting rights in consumers for anything other than "functionally necessary" information); Murphy, supra note 102, at 2407–17 (endorsing contractual privacy). I call this a "pseudo-contract" scheme because it would actually have to be a statutory scheme that imposed contract-like obligations on the basis of parameters determined by interaction between a user agent and the website.

[185] See, e.g., Kenneth C. Laudon, Extensions to the Theory of Markets and Privacy: Mechanics of Pricing Information, *in* Privacy and Self-Regulation in the Information Age 41, 42–43 (U.S. Dep't of Commerce ed., 1997), available at http://www.ntia.doc.gov/reports/privacy/selfreg1.htm (advocating a property regime). Professor Alan F. Westin first proposed the idea of propertizing personal information in 1967. Westin, supra note 11, at 324–25. For a discussion of the contexts in which government institutions have recognized property interests in facts, see Hatch, supra note 7, at 1467–68.

has stubbornly refused to propertize facts,[186] and it is doubtful whether those advocating property rights in information could advance a compelling economic argument for clickstream data's being excepted. Aside from broader concerns regarding the social value of freely flowing information, the costs of enforcing property rights over facts almost surely exceed the costs of ensuring exclusive use protectively.[187] A relationship to a piece of information mediated by property law is a right against the world,[188] whereas a relationship to that same piece of information mediated by contract is merely a right against the entity with whom the disclosing agent is transacting. First, property rules fail to promote the utility objective because data buyers (collecting agents and secondary purchasers) would have to secure from the user the right to buy and use the data, imposing substantial transaction costs that would make almost any secondary data use impossible. Second, property rules also fail to protect users in that they may be forced to make an all-or-nothing decision without being able to attach conditions to data collection. The reason for this problem is that the First Amendment protects non-contractually restricted speech (data disclosure) about information a speaker has legally acquired.[189] Courts, however, have consistently held that bargained-for restrictions on the right to speak trump a speaker's First Amendment right to disclose information.[190]

---

[186] See Feist Publ'ns v. Rural Tel. Serv. Co., 499 U.S. 340, 345–47 (1991).

[187] For example, it costs a man less to control information exchange regarding his receding hairline by wearing a hat than it would cost society to control this sort of exchange by enforcing some sort of property right in this fact.

[188] See Robert P. Merges, The End of Friction? Property Rights and Contract in the "Newtonian" World of On-line Commerce, 12 Berkeley Tech. L.J. 115, 121–22 (1997) ("The difference is that a bilateral agreement applies only between contracting parties, whereas a restraint on alienation 'runs with the property,' and hence interferes with all potential future parties as well.").

[189] Courts will generally enforce contracts not to speak, but not property rights against those seeking to disclose information that the speaker has legally acquired. See Ellen Alderman & Caroline Kennedy, The Right to Privacy 329 (1995) (noting that property ownership of personal information is at odds with the First Amendment). But see L.A. Police Dep't v. United Reporting Publ'g Corp., 528 U.S. 32, 34 (1999) (rejecting a facial challenge to a statutory amendment limiting commercial users' access to arrestee addresses). For a general discussion of the tension between privacy and the First Amendment, see Goodale et al., supra note 139.

[190] See, e.g., Cohen v. Cowles Media Co., 501 U.S. 663, 665 (1991) (holding that the First Amendment does not prohibit one from recovering promissory estoppel dam-

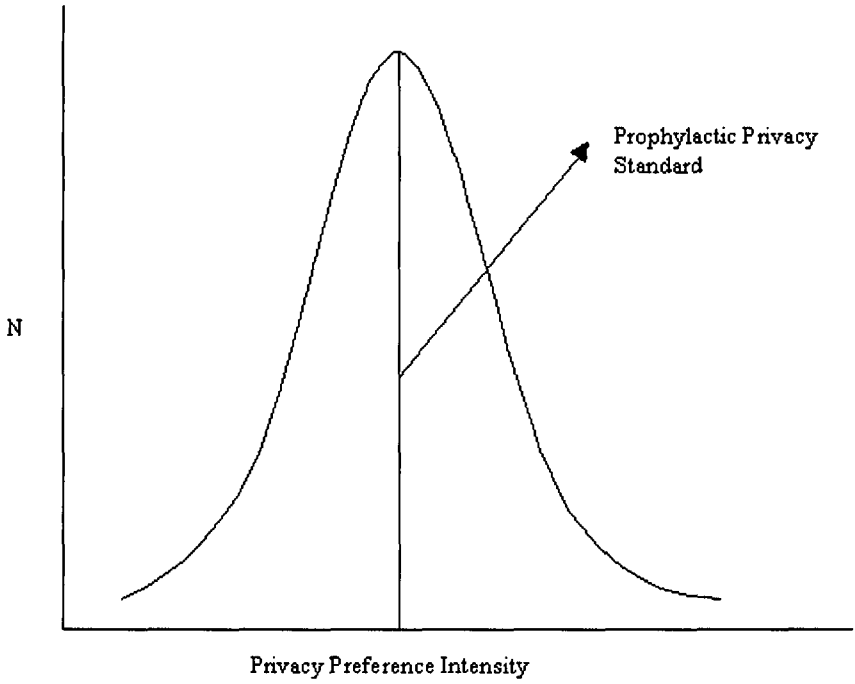

Figure 1

A model that allocates default entitlements between the user and the collecting agent and then allows them to bargain around the baseline enjoys several additional advantages over models premised on more prophylactic standards. First, the concept of default rules can impute to an exchange certain implied conditions when the parties fail to contract for express conditions. These defaults play a critical role in forcing valuable information from both the user and from the collecting agent. Second, a contract model can accommodate the highly differentiated set of expectations—an invaluable feature in the clickstream context because, as Figure 1

ages for a publisher's breach of a confidentiality promise that was given in exchange for information); see also Volokh, supra note 12, at 1057–61 (arguing that courts may even recognize implied contractual privacy terms).

illustrates, V_{User} can vary significantly from user to user.[191] If users' privacy preferences are highly differentiated, then although a standard set at the "reasonable expectation" may well enforce the average expectation, that average will rarely reflect V_{User} for any given consumer. Any prophylactic damage measure failing to enforce an actual valuation either overdeters secondary data use (where V_{User} is lower than that protected by the standard) or undercompensates the user (where V_{User} is higher) for her violated expectation. Such a damage measure, then, would simultaneously constrain gains from trade (undermining the utility objective) and underprotect user expectations (undermining the compensatory objective). The statistical corollary of these propositions is that, from a pure utility standpoint, the desirability of a prophylactic privacy standard is inversely related to the standard deviation of V_{User}.[192]

The requisite assumptions of such a bargaining model were, until recently, so preposterous that, although thinkers posed contractual defaults for other informational privacy quagmires, they rightfully ignored such possibilities in the clickstream context. Clickstream default proposals have escaped academic attention for several reasons. First, the high transaction costs of bargaining rendered making explicit the obligations of clickstream data exchange not worth the benefit of exchanging it. Second, no compelling mechanism existed for discerning individual reservation prices in a context where user valuations were highly differentiated. Third, enforcement of an expectation damage measure (in this scheme the reservation price, V_{User}) would have been so expensive that it would have suffocated gains from trade, rendering otherwise mutually beneficial transactions economically undesirable for either the user or the collecting agent. These conditions have led most to call for either a common law or statutory remedy predicated on prophylactic rules regarding what collecting agents may do with clickstream data when they acquire it from users. If, however, one could engineer a

[191] Figure 1 represents a V_{User} distribution with a small mean. Although the absolute value of the distribution around the mean is likely small (because the mean itself is so infinitesimal), the relative size of the dispersion remains crucial to the analysis.

[192] Because a larger standard deviation implies greater dispersion around the mean, the greater the standard deviation, the greater the dispersal around the mean, and the greater the degree to which a prophylactic standard would either undercompensate or overdeter.

market environment that inexpensively created, stored, and enforced expectations, then one could only justify such inflexible rules on paternalistic grounds (namely, that people should want more privacy than they do). Ultimately, the bargaining model would be "Pareto superior" because unlike the inflexible rules, it would improve the position of all those involved.

B. The Platform for Privacy Preferences ("P3P")

The Platform for Privacy Preferences ("P3P") project has enjoyed intense intellectual scrutiny for several years.[193] P3P is a product of the World Wide Web Consortium ("W3C") and represents a set of recommended protocols governing information exchange over the Internet.[194] The debate surrounding the ultimate viability of P3P remains extremely volatile, with the protocols receiving treatment from commentators that ranges from unabashed enthusiasm to scathing derision.[195] This Note's discussion, however, involves only the protocols for clickstream data and may therefore be exempt from some of the more extreme characterizations of P3P as a whole. Although P3P may indeed be an invaluable tool for mediating certain informational privacy contexts, people should remain skeptical of claims of its talismanic significance in others.

Microsoft's most recent browser product, Internet Explorer 6.0, uses P3P as its cookie-handling protocol.[196] Some earlier browser versions have a more inchoate protocol that either rejects or accepts all cookies.[197] Some commentators have reported version 6.0

---

[193] See, e.g., Lessig, supra note 9, at 160–63 (discussing the possibility of using P3P in conjunction with a legal remedy); Rosen, supra note 11, at 172.

[194] The full P3P specification can be found at World Wide Web Consortium, at http://www.w3.org/tr/p3p/ (last visited July 5, 2003) (on file with the Virginia Law Review Association).

[195] See, e.g., William McGeveran, Programmed Privacy Promises: P3P and Web Privacy Law, 76 N.Y.U. L. Rev. 1812, 1815 (2001) (espousing P3P as a means of creating a privacy market); Marc Rotenberg, Fair Information Practices and the Architecture of Privacy (What Larry Doesn't Get), 2001 Stan. Tech. L. Rev. 1, ¶¶ 75–89 (2001) (leveling numerous criticisms at P3P).

[196] See Aaron Goldfeder & Lisa Leibfried, Privacy in Internet Explorer 6, at http://msdn.microsoft.com/library/en-us/dnpriv/html/ie6privacyfeature.asp?frame=true (last updated Oct. 16, 2001) (on file with the Virginia Law Review Association).

[197] See Jeffrey Benner, MS Gets Privacy-Happy With New IE, Wired News, at http://www.wired.com/news/privacy/0,1848,43686,00.html (May 15, 2001) (on file with the Virginia Law Review Association).

market penetration to be as high as 30%,[198] and that figure will surely increase. As collecting agents increasingly deploy the protocol, one can expect the next generation of browsers to be equipped with some sort of cookie-handling feature and that companies like Microsoft will set the defaults.[199]

Under the "Internet Options" tab of the Microsoft Internet Explorer software 6.0 there is a tab for "Privacy."[200] Although there are a number of online privacy issues,[201] these settings deal exclusively with cookies. There are six gradations of cookie-handling from which the consumer may choose: accept all cookies, low, medium, medium-high, high, and block all cookies.[202] The default setting is medium, which, among other things, blocks any cookies from servers without compact privacy policies.[203] Under the P3P cookie-handling functionality deployed by Explorer, when a user first types the URL address into the browser, the browser requests a privacy policy reference file from all collecting agents (domains) serving cookie content into that web page.[204] These agents generally consist of the website and any affiliates or advertisers also serving code into that page. If the site has what is called a "compact privacy policy," the collecting agent will return the reference file detailing the location

---

[198] Scarlet Pruitt, Internet Explorer 6.0 Leaves Netscape Behind, IDG News Service, at http://www.pcworld.com/news/article/0,aid,91483,00.asp (Mar. 27, 2002) (on file with the Virginia Law Review Association).

[199] See James Tierney, Remarks at the Technology Law Center at the University of Maine School of Law (June 2001), *in* Internet Privacy Law, Policy, and Practice: State, Federal, and International Perspectives, 54 Me. L. Rev. 95, 111 (2002).

> However, the default levels will, of course, be set by the person in the company who owns your operating system. So, I want you to go out and look all across the country and decide which operating system you all have. Ninety-six percent of us have one from one company and it will decide the level of default that you will have, and nobody will ever change that default.

Id.

[200] See Goldfeder & Leibfried, supra note 196.

[201] See supra notes 16–29 and accompanying text.

[202] See How to Manage Cookies in Internet Explorer 6, Microsoft Knowledge Base Article, at http://support.microsoft.com/default.aspx?scid=kb;en-us;283185 (last visited Feb. 21, 2003) [hereinafter How to Manage Cookies] (on file with the Virginia Law Review Association).

[203] A user may actually input customized settings, but this functionality is beyond the scope of this discussion.

[204] These cookies could be from first or third parties. How to Manage Cookies, supra note 202.

of that policy to the browser.[205] The browser will then request the XML- (Extensible Markup Language) encoded compact privacy policy, which it then compares to the user's preset privacy preferences.[206] If the collecting agent's treatment of cookies conforms to the user-determined parameters, it may set the cookie in the manner stipulated by the browser setting.[207] All of this (aside from the user setting her preferences) occurs almost instantaneously. The user does not, as in some earlier generations of browsers,[208] receive a prompt every time a collecting agent tries to set a cookie, thereby incurring significant transaction costs in attempting to protect her privacy. If a collecting agent attempts to set a cookie that is either blocked or downgraded by the browser, a little red icon signaling that event appears in the bottom right corner of the screen.[209] If the user wishes to investigate the specifics of the rejection or modification,[210] she can click on that icon for a report.

C. Default Rules

The concept of default rules has certainly received attention in the online context.[211] In 1989, Professors Ian Ayres and Robert Gertner wrote an enormously influential article on the appropriate criteria for default rule selection.[212] Professor Jerry Kang has systematically incorporated those criteria into a contract-based pro-

---

[205] Laurie Cranor, P3P: A More Detailed Look, The O'Reilly Network, at http://www.oreillynet.com/lpt/a/1554 (last visited Feb. 21, 2003) (on file with the Virginia Law Review Association).

[206] Id.

[207] Usually this will either be a "block" or an "accept," but certain settings provide that certain types of cookies be either "leashed" or "restricted." A leashed cookie is one that can be read only by the agent that issued it. A restricted cookie is treated like a "session cookie" and is deleted when the user closes her browser. See Goldfeder & Leibfried, supra note 196.

[208] In Internet Explorer 5.5, the user can either reject all cookies, accept all cookies, or be prompted every time a collecting agent attempts to set a cookie. See Benner, supra note 197.

[209] See Goldfeder & Leibfried, supra note 196.

[210] Upon clicking on the report she will see a pop-up window detailing all of the cookies that collecting agents with material on that page seek to assign to her browser. The window will explain how each cookie was treated—whether it was accepted, blocked, leashed, or converted to a session cookie. Id.

[211] See Kang, supra note 44, at 1246–59.

[212] Ian Ayres & Robert Gertner, Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules, 99 Yale L.J. 87 (1989).

posal for the collection of non-clickstream personal information (such as that collected during purchase or registration).[213] Kang likely failed to treat clickstream data for largely the same reasons as did other commentators—the transaction costs associated with making explicit the conditions of data exchange would dwarf the value of the data point itself.[214] The integration of new browser technology with P3P protocols depresses the transaction costs involved in securing express obligations from collecting agents and, in so doing, crowds out fewer data exchanges that, in the face of larger contracting costs, would not be worthwhile for the user nor for the collecting agent.

Ayres and Gertner's seminal paper takes two major positions. First, they argue that penalty defaults, or default rules that force one of the transacting parties to reveal certain critical pieces of information to the other, may be justified where information asymmetries would otherwise allow one party to protect its "share of the pie" by strategically withholding information at the expense of the collective "size of the pie."[215] Second, Professors Ayres and Gertner attack arguments for majoritarian defaults that employ criteria resembling rules that most people would choose.[216] This condition, they argue, holds only when (1) the losses associated with not being able to "flip out" of each of the two default rules are equivalent,[217] and (2) the transaction costs of flipping, for each default rule, are also equivalent.[218] It may in some cases make sense to have

---

[213] He concluded that the default should vest the right only to "functionally necessary" information with the collecting agents. Kang, supra note 44, at 1201.

[214] This is not to say that clickstream data is not valuable, but simply that the value of a single clickstream data point and that of the page view for which it is traded are sufficiently small that even a minor inconvenience would probably be enough to cause the user to view another page. A collecting agent may be more willing than a user to engage in this sort of bargaining, but only because it could have millions of such interactions and could therefore capture returns from scale on bargaining for these data points.

[215] This statement is a more colloquial way of saying that this posture is justified where otherwise one of the parties, in the interest of capturing a greater fraction of the surplus from the bargain, would selectively conceal information the revelation of which would have increased the overall surplus to be shared by both parties to the bargain. See Ayres & Gertner, supra note 212, at 94.

[216] Id. at 89–91.

[217] To "flip out" means to switch from the default term to the other option. "Flipping out" is the terminology used by Jerry Kang. See Kang, supra note 44, at 1257.

[218] See Ayres & Gertner, supra note 212, at 114–15.

a minority default where flipping out is less expensive and/or failure to flip entails a less significant loss for one of the parties.

In part because the transaction costs of a user flip would likely be the same without regard to how the default itself is set,[219] this discussion focuses primarily (1) on the loss associated with users getting "stuck" with an undesirable term, and (2) on the first part of the argument, the information-forcing function of default rules. The following is a discussion of the desired informational defaults for P3P cookie protocols.

What is unique, indeed unprecedented, about clickstream defaults is that they can literally be encoded for every exchange of clickstream data. They are not there as "gap-fillers," but are technologically embedded as conditions for each page view. P3P cookie-handling devices epitomize the "code" that Professor Lawrence Lessig sees as the linchpin of cyber-governance.[220] If the transacting parties achieve a proxied agreement on expectations based upon either default privacy preferences or those input by the user, the default settings effectively *are* the default legal rules. As long as a legal regime enforces the expectation as expressed by the browser menu, legal defaults deviating from their technical analogues would be inefficient because, in every instance, one would have to override the other. Moreover, the default settings exhibit a tremendous amount of inertia because (1) many do not know about them, and (2) if they do, they may not be comfortable changing them. The default browser settings should therefore be subjected to intense scrutiny. Except for the first point, the following discussion deals largely with default "rules" insofar as the encoded browser settings reify them.

First, the default rules should apply only when the user has a browser that would enable her to express her expectations using the browser's privacy menu. It may be possible to bargain over terms without the browser features, but such dickering is more expensive (in terms of effort) than is an automated exchange between browser and server. To the extent that the fixed cost of acquiring the browser technology is small (one can download

[219] Irrespective of the default settings, in order to change them the user would still have to go to the same privacy menu. The difference resides merely in the button she would click to do it.

[220] Lessig, supra note 9, at 160.

Explorer 6.0 for free), a default term construed against users—in other words, one that locates most of the "ownership" of a clickstream data point with the collecting agent—in the event that they fail to use browsers with cookie-handling functionality encourages them to procure browsers with this feature.[221]

Second, the default browser settings should (and currently do) block cookies from collecting agents that have no compact, machine-readable privacy policy. This setting forces information about data usage from collecting agents because it penalizes those that do not provide a clear and instantaneously readable privacy policy. This default setting eliminates problems associated with web operators concealing and manipulating current privacy policies so as to protect their "piece of the pie."

Third, the defaults should and do "leash" all first-party cookies, meaning that only the issuing collecting agent may read them. Most people appreciate the benefits of personalization,[222] although many may not realize that cookies and customization are related. Keeping a virtual "shopping cart" and having news items that are of idiosyncratic interest placed prominently on news sites are two examples of the advantages cookies confer upon users.[223] Many people, however, would not consent to a cookie placed by one site that may be read by other, unidentified sites. Consumers should be responsible for understanding that websites do set cookies, but they should not be responsible for knowing, without being told, which third parties may read them. A default that accepted cookies readable by third parties would chill web surfing by introducing more uncertainty into the clickstream collection context.[224] Because consumers do appreciate the gains from personalization associated with a leashed first-party cookie, a blanket default set to reject all first-party cookies linking clickstream data to personal information would either (1) stick a number of people with a rule that, absent

[221] This setting is particularly important because otherwise users would lack an incentive to get the browser that allows users and collecting agents to make explicit the exact privacy terms.

[222] In one survey, seventy-eight percent of respondents indicated that "they would definitely or probably agree to Web sites using persistent identifiers . . . to provide a customized service." Cranor et al., supra note 28.

[223] Cookies are used to link a user with material that that user has placed in the shopping cart. That information is stored in the collecting agent's database.

[224] See supra notes 177–80 and accompanying text.

the costs of contracting, they would rather contract around or (2) force those who indeed elect to contract around it to incur gratuitously the costs of doing so. Devising a default rule leashing first-party cookies therefore avoids the transaction costs associated with a rule that blocks them entirely and eliminates uncertainty by forcing websites to explicitly and specifically identify any additional entities that may enjoy access to the cookies.

Finally, the defaults should accept anonymous, third-party cookies. The presence of third-party cookies evokes the strongest reaction from users because these files allow collecting agents to compile user profiles across a number of sites.[225] This concern represents a reaction to the "dossier effect" because a single collecting agent can monitor and compile behavior across a variety of contexts. That being said, such dossier effects are less objectionable to the extent that they are anonymous, meaning that they cannot be tied to any personally identifiable information. The difference between personally identifiable dossiers and anonymous ones is roughly the difference between a pharmacy knowing the identity of a person buying a certain mix of medication and the mere fact that a person is buying it. A cookie is "anonymous" if the third party does not collect it to associate it with, or set it to contain, a name, an email address, a real-space address, a social security number, or any other information that could be used to identify the behavior of the associated person in the real world. Without certainty regarding uses to which the data may be put offline, the average informed consumer would attach a highly negative valuation to the third-party use of personally identifiable clickstream data.

Some would go even further and maintain that the default should block third-party cookies entirely. Such a setting would be a terrible misstep, as it is little exaggeration to say that targeted advertising is the lifeblood of free publications.[226] Those desiring a default that does not accept third-party cookies would respond by arguing (1) that disallowing third-party cookies would not represent a death knell for online targeted advertising, and (2) that those us-

[225] In one survey, only forty-four percent indicated that they would agree to allow an identifier to provide advertising across a number of websites. Cranor et al., supra note 28.

[226] See Dorothy Kabakeris, How to Find Investment Information on the Web, 3 No. 21 Law. J. 8, 8 (2001).

ers desiring third-party cookies could just adjust the privacy menu from the default to reflect such preferences.

With respect to the first point, targeted advertising would persist, but with the same limitations as exist offline. Certain publications cater to targeted audiences, and third-party advertisers would retain the ability to talk to that audience generally, but they would not be able to differentiate *within* that audience. The ability to differentiate within the broader audience of a targeted publication is the advantage of advertising online versus off. Advertisers are willing to pay a premium for online advertising precisely because when third parties serve an advertisement to a user, they can target the material pursuant to that user's cookie history.[227] If default rules effectively disable this targeting capability then many web publishers that subsist on advertising revenue would be forced either to go under or to switch to other revenue models. Consumers would find themselves either without the same selection of content or having to pay more for it.

With respect to the second point, the question is not quite fair. If a user never had to internalize the costs of a decision to refuse third-party cookies then it is quite possible that she would elect to block them. A user internalizes the effects of her decision to block *first*-party cookies by sacrificing the personalization that the cookie provides. Unless websites can efficiently calibrate access or service to reflect the aggregate advertising loss inflicted by a single user's decision to block a *third*-party cookie, however, the user will not have to internalize the cost of her decision.

Again, the point may be expressed as a mathematical relationship. The user captures all the return on her decision to refuse the cookie, R_{Block_3}, but shares the aggregate burden in the form of sacrificed advertising opportunity, $L_{Advertising}$, with the rest of the users on the website. Forcing users to internalize the costs of the decision to block third-party cookies would necessitate the Herculean technical undertaking of segmenting a site's web service along another dimension—whether the user accepts third-party cookies. If websites cannot segment their service offering so as to force users to incur 100% of $L_{Advertising}$, then each user on a website with an audi-

[227] By "premium" I mean to suggest that advertisers pay more than they otherwise would if they could not differentiate within a vendor audience.

ence of size N will reject third-party cookies as long as $R_{Block\_3} > L_{Advertising}/N$.[228] Consequently, users systematically undervalue third-party cookies. In sum, the argument that most consumers would elect to block third-party cookies is misleading because their decision to do so rarely reflects the true cost of blocking them.[229] Moreover, when advertisers use targeted ads, many times their intention is to use knowledge about the cookie to control for different testing variables, not to ruthlessly exploit knowledge of web-surfing habits.[230]

VI. DAMAGES AND ENFORCEMENT

A. Inappropriate Damage Measures

Reserving private causes of action exclusively for individual users fails to implement an appropriate litigious check on collecting agents. The following Section argues that a regime seeking to adequately address both compensatory and utility objectives should enforce expectations not through the private suits of those injured by violation of the privacy term, but instead through a statutory scheme of regulatory fines or penalties for violating default or bargained-for entitlements. While contract law represents an effective blueprint for structuring exchanges and for identifying the scope of the appropriate damages, inherent limitations involving the costs-to-return ratio of litigation render the private cause of action, relative to a statutory penalty, a far inferior means of protecting clickstream privacy.

---

[228] There exist some simplifying assumptions for this assertion. First, the number of users on a given website would have to be static. Second, all users would have to use the site with the same frequency. That these assertions are plainly untrue, however, does not seriously undermine the point. The fraction of the loss borne by a single user may not be exactly $1/N$, but will nevertheless be substantially less than one hundred percent.

[229] One might recognize that setting the default to accept third-party cookies uses one type of market imperfection (the friction of flipping out of the default rule) to correct another type of market imperfection (externalities).

[230] See, e.g., Digital Marketing Insights, Atlas DMT, at http://atlasdmt.com/insights/dm.asp (last visited Feb. 21, 2003) (on file with the Virginia Law Review Association) (chronicling a number of different tests using cookie data).

The Achilles heel of almost every proposed privacy solution is the mechanism for determining and enforcing damages.[231] Should the cause of action be private or public? What should the damage measure be? In solutions where the cause of action resides with the user, the damage award must be high enough to induce that user to bring the cause of action.[232] From an ex ante perspective, collecting agents will not violate the privacy term if the return on the unauthorized use (recall that this is DUV_{Unauth}) is less than the expected cost (to the violator) of violating the privacy term, $C_{Violate}$. Forgetting for a moment that litigation costs the collecting agent money and further assuming that all initiated suits are successful, then $C_{Violate}$ is equal to the expected probability that a user brings a suit, P_E, times the magnitude of the damages, D_E. (That is, $C_{violate} = P_E * D_E$.) Assuming that law should at least deter breaches when the breach yields less value to the breacher than adherence yields for the potential victim (if $DUV_{Unauth} < V_{User}$), then D_E must be large enough to make the expected penalty greater than the value of the breach (or, mathematically speaking, large enough to satisfy the condition $P_E * D_E > DUV_{Unauth}$). When a cause of action resides with an individual user, D_E would need to be extraordinarily high (because P_E is so low) in order to deter sufficiently breach of the privacy term.

There are two major problems with setting D_E high enough to correctly calibrate $C_{violate}$. First, it fails as a compensatory measure because D_E would far exceed V_{User}, the actual loss incurred by the user.[233] Second, if the purpose of calibrating $C_{violate}$ is to encourage data uses where $DUV_{Unauth} > V_{User}$, then $C_{violate}$ should approximate V_{User}.[234] At high levels of D_E, however, even slight errors in estimating P_E would render $C_{violate}$ a gross misapproximation for V_{User}, and these errors are almost unavoidable.

Some have argued that class-action lawsuits can effectively increase P_E by allowing users to pool claims for small damages into

---

[231] See Froomkin, supra note 9, at 1527 (noting that the United States may be unique in having self-regulation without any enforcement mechanism).

[232] See supra note 180 and accompanying text.

[233] The relationship between D_E and P_E is actually far more complicated, as neither of the two variables is independent. Increasing D_E would effectively increase the return on litigation, pulling P_E upward.

[234] This way, the decision to make unauthorized use of data despite the penalty, occurs only when that use garners a return higher than the user's privacy valuation.

one suit.[235] Transaction costs, however, still inhere in class-action suits, and the damages are not likely large enough to induce people to incur even a fraction of the costs of participating. Particularly with respect to clickstream data, where the loss incurred by breach of the privacy term is low relative to the costs of litigating, class action is not a viable mechanism for preserving a private cause of action. Advocates of class-action-related solutions may respond by noting that there are many cases where a potential participant's individual stake may be very small, even several cents, but that response mistakenly assumes that the costs of identifying the class are analogous. Identifying the class of persons whose privacy terms are violated is probably a far more expensive proposition than is that exercise for other groups of plaintiffs.[236]

Other proposals would grant a cause of action to the government for violation of a prophylactic standard. Such proposals, however, would enforce this standard at the expense of both compensatory and utility objectives. A prophylactic statutory regime would fail to protect differentiated expectations because the damages would not be repaid to the users who incurred the loss (and even if they were, they would not reflect V_{User}).[237] Moreover, the inability to correctly calibrate D_E so that collecting agents make unauthorized use of data only where $DUV_{Unauth} > V_{User}$ leaves such strategies vulnerable to the same problems as are detailed at the end of Part IV.

B. Clickstream Nanocontracting[238]

The discussion surrounding the user reservation price wastes ink if there exists no feasible mechanism for identifying V_{User}. Prophylactic rules labor under a crude approximation for the average V_{User}, so a statutory scheme predicated on a low cost mechanism for identifying the subjective V_{User} for each cookie, akin to the P3P mechanism for identifying the scope of the privacy term itself, would be

---

[235] See supra notes 179–80 and accompanying text.

[236] It would be more expensive in part because of the costs associated with executing subpoenas to force violators to provide the list of users.

[237] They would not be repaid unless the government remitted payments to those users whose privacy terms were violated.

[238] As used in this Note, the term "contract" does not invoke the idea of contractual remedies. A "nanocontract" is merely a convenient way of describing the very small agreement.

quite attractive. The most sensible point in the process of clickstream data exchange to identify V_{User} would reside at a technical bottleneck—some attendant digital exchange that must occur between every user and every collecting agent. Browsers are the most obvious technical bottleneck for such clickstream exchanges because in order for a user to view a page there must be some communication between the browser and the collecting agent's server. It is precisely this bottleneck property that rendered the browser the ripest vehicle for communicating the privacy term itself. Just as the newest generation of browsers can effectively encode a privacy term, so too can they serve as a mechanism for identifying V_{User}. This valuation could be represented by a reservation price, P_R,[239] that conveys V_{User} and, like the privacy term, may be set as a *default* according to a schedule. The schedule would vary with respect to the level of privacy preference itself, with stricter privacy terms defaulting to lower reservation prices.[240] The language of this clause would be to the effect of, "I would be willing to allow this collecting agent to breach these expectations for $[P_R]." Although courts could hypothetically enforce the reservation price as a liquidated damages clause, legislatures could bypass many of the idiosyncrasies of contract doctrine by promulgating an alternative statutory enforcement scheme. This scheme could assess severe penalties for failure to return P_R to the user independent of any common-law contract doctrine. Section VI.C discusses this possibility in more detail.

The reasons why default values for P_R (it may at times be useful to conceptualize this value as a liquidated damage) should be inversely related to the stringency of the privacy term require a little elaboration. Consider the following seven variables, several of which are used in earlier analysis: P_R, representing the reservation price that approximates V_{User}; DUV_{Unauth}, representing the value of a single unauthorized data use (unauthorized means that the use breaches the privacy term); DUV_{Unauth_profit}, representing the value of a single profitable, unauthorized data use; $DUV_{Unauth\_unprofit}$, representing the value of a single *un*profitable, unauthorized data use;

[239] One would expect P_R to approach V_{User} because one is generally willing to accept value at least equivalent to the value she places on the term.

[240] See infra this Section.

SDU_{Unauth}, representing the size of the set of unauthorized uses (both profitable and unprofitable); SDU_{Unauth_profit} (a subset of SDU_{Unauth}), representing the size of the set of profitable, unauthorized data uses; $SDU_{Unauth\_unprofit}$ (also a subset of SDU_{Unauth}), representing the size of the set of *un*profitable, unauthorized data uses. A collecting agent will breach a privacy term if the potential use is profitable, that is, where $DUV_{Unauth} > P_R$. One should conceptualize the privacy term as a line dividing the total set of potential data uses into those that are authorized by the privacy term and those that are not (SDU_{Unauth}), with the latter being further subdivided into SDU_{Unauth_profit} and $SDU_{Unauth\_unprofit}$.

Two further assumptions are required. First, the number of unauthorized data positively correlates with the number of profitable such uses ($SDU_{Unauth} \sim SDU_{Unauth\_profit}$). Since a stricter privacy policy implies more unauthorized uses, it therefore follows that a more stringent term implies more *profitable* such uses. Second, assume an inverse relationship between the value of the unauthorized data uses and the number of such uses. The downward sloping line in Figure 2 captures this relationship. There are therefore fewer unauthorized uses that are *very profitable* than there are unauthorized uses that are *slightly profitable.*

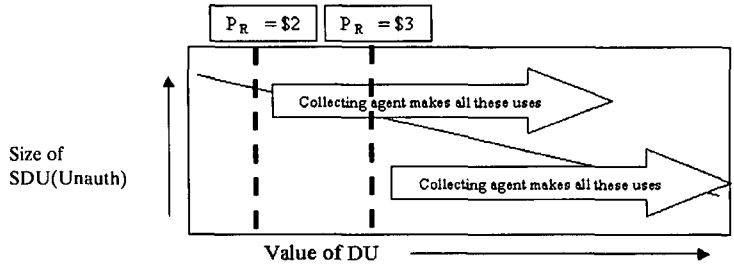

Figure 2

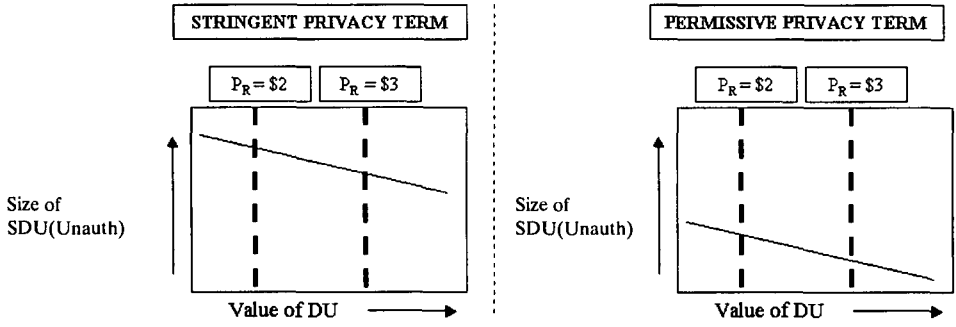

Figure 3

Figure 2 illustrates conditions with two different reservation prices attached to the same privacy term. The set of unauthorized uses where the value of the unauthorized use exceeds P_R is considerably larger when P_R is two dollars than when it is three dollars. The area between the hash marks under the curve represents the loss to a collecting agent when,[241] for a given privacy term, P_R is in-

[241] Some sort of marginal gain in terms of credibility probably results from a collecting agent's agreement to pay an increased reservation price, but this analysis does not incorporate this phenomenon because what ultimately matters here is whether this marginal increase varies with respect to the stringency of the privacy term. This analysis assumes that because an extra dollar of P_R should generally inspire roughly the same amount of incremental credibility irrespective of how stringently the user has set

creased from two to three dollars.[242] Figure 3 draws these same curves for two different privacy terms, one that is stringent and one that is permissive. Recall that because the privacy term divides data uses into those that are authorized and those that are unauthorized, more permissive terms imply a smaller set of unauthorized uses as depicted in the right side of Figure 3. Because the size of SDU_{Unauth_profit} is generally proportional to the size of SDU_{Unauth}, a more permissive term implies not only a smaller SDU_{Unauth}, but also a smaller SDU_{Unauth_profit}. The area between the hash marks under the curve for the permissive term is therefore much smaller than the analogous area for the stringent term. The loss associated with formerly profitable data uses being rendered unprofitable with an increased P_R is larger with stringent than with permissive privacy terms. Similarly, for a fixed P_R the *total* number of unprofitable unauthorized data uses is larger with stringent than with permissive privacy terms. The upshot of this discussion is that a default P_R schedule that varies with respect to the stringency of the privacy term should assign lower default P_R values to stricter privacy terms (keeping a greater fraction of potential data uses profitable) and higher P_R values to more permissive ones.[243]

---

the cookie-handling privacy term, the discussion regarding which way the scale of defaults should slide with respect to privacy term stringency may ignore this dynamic.

[242] The lost value derives from two distinct phenomena. First, an additional dollar of P_R would render some formerly profitable data uses unprofitable. Second, the additional dollar of P_R would diminish the profit (by a dollar) on each data use that remains profitable.

[243] One might object that projected collecting agent valuations of privacy terms should not drive the default settings because such methodology ignores the pattern of user valuations for that term. Suppose, the argument would go, that users would prefer that lower levels of P_R be associated with permissive, not stringent policies. The answer to this objection resides in the extreme dispersal of user valuations around the mean V_{User}. A default P_R reflecting the mean *user* valuation is, for any single user, more likely to deviate significantly from that *user's* actual valuation than is likely, for a given collecting agent, a default P_R reflecting the mean *collecting agent* valuation to deviate from that *collecting agent's* actual valuation. This condition means that setting the default P_R relative to collecting agent preferences, rather than user preferences, minimizes the likelihood that one of these parties need incur the transaction cost of adjusting the P_R from the default. P_R will rarely reflect the actual valuation of a given user at any level to which it is set, but if it is set in accordance with collecting agent valuations, at least that class need not request that the P_R be adjusted upon each page view. This is an assumption and would require empirical verification. Given, however, that business valuations are driven by economic need rather than subjective preference, they are probably less dispersed than their consumer counterparts.

A regime incorporating a default P_R schedule would be desirable for a number of reasons.[244] First, making explicit a proxy for V_{User} in the form of P_R would make the threshold for identifying value-added data uses clearer. When $DUV_{Unauth} > P_R$, the collecting agent can put data points to net-beneficial use without incurring the risk of excessive penalty so long as it pays out the reservation price. Data can therefore flow inexpensively to its highest valued use.

Second, default valuations, set according to a particular schedule, could serve important information-forcing functions. They would force users with idiosyncratically high privacy valuations to reveal their preferences. More importantly, they would also force collecting agents that have particularly high values of SDU_{Unauth_profit} (meaning they have some special reason to believe they will breach the term) to reveal this condition by explicitly requesting a lower P_R from the user.

Third, the valuations would diminish the uncertainty associated with both ceding data and breaching the privacy term. Users would know how much they would get in the event of a breach and collecting agents would know how much they would have to pay.

Finally, using valuations would effectively check the incentives for consumers simply to demand as much privacy as they want without respect to relative value. In certain instances, collecting agents may refuse to transact (serve a web page) with those browsers demanding both unreasonably stringent privacy terms and exorbitant values for P_R. A market with so many technological predicates will no doubt function imperfectly, but it will create incentives for consumers to relinquish data at their actual reservation prices.

Some commentators would surely object to this solution as one enabling the forced sale of clickstream data, and Section III.B has addressed some of these concerns already. The strict notion of a forced sale is one generally employed in the context of compulsory licensing—a context in which legislatures identify some price at which buyers must be able to buy a property right from sellers.[245] This condition, however, leaves a market dependent on "property

[244] Determining the absolute value of the default reservation price, in dollar terms, is beyond the scope of this Note.

[245] See Robert P. Merges, Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations, 84 Cal. L. Rev. 1293, 1295 (1996).

rules" in name only because the compulsory license effectively sets a standard liability threshold beyond which the buyer may acquire an entitlement from the seller. The flaws in the analogy should be obvious. First, the privacy term and the reservation prices would merely be defaults, whereas the compulsory license rules are mandatory. Users could unconditionally preclude collecting agents from acquiring their data in at least three ways: (1) They could change P_R from the default setting to one indicating that there may be no breach of the term (basically, at infinity); (2) they could disable their cookies; or (3) they could stay off the Internet entirely. The last strategy only sounds harsh if one refuses to acknowledge that the sacrifice in privacy *is* the toll users pay to operate cyberspace. To say that a user would never have to relinquish her data involuntarily is not to say that she would never be expected to transmit this information in order to secure some measure of reciprocal value. Instead, she would always have a choice as to whether her privacy is worth whatever she could be getting in exchange. Moreover, should a user decide that what she is getting in exchange for her information is not worth her sacrifice in privacy, she could opt for a provider of fungible content that may well compete in the market by offering more stringent privacy terms. Most users, of course, would never choose such a radical holdout strategy because, even with perfect information, they would be more than happy to trade conditional access to their web-surfing habits for something of value.

C. Enforcing Nanocontracts

The ability to structure these types of nanotransactions for each piece of clickstream data is not beyond our technological capacity. Browsers are already equipped with privacy preference menus, and adding a default valuation functionality requires only the addition of another field. The primary hurdles to creating this type of market for clickstream data are enforcement-related. What follows are several related ideas about how enforcement of these nanocontracts might work. The enforcement costs of such a regime would be substantial, but the same could be said about the costs of administering any legal restrictions.

Collecting agents collect data points on users and store them as entries in their databases. Each entry has a value for any number of

fields. A field is a dimension in a database that stores values for a given variable. As discussed in Part I, the core of clickstream data fields usually consists of at least a cookie number, an IP address, a timestamp (time that the page view took place), and something identifying the page that the user viewed. These cookie "log files" will generally contain a number of other fields, but each data collector's particular mix depends on the specific business needs of the collecting agent.

Collecting agents could easily earmark another field for storing the privacy contract configuration.[246] That field would consist of (1) the privacy term negotiated between the user's browser and the collecting agent's server and (2) P_R. The costs of writing and storing this data are not insignificant, but they would not represent an inordinate expenditure.[247] Moreover, collecting agents destroy data when it is no longer useful, and the usefulness of the privacy configuration expires at the same time as does the usefulness of the associated data point.

The difficult part of this scheme, as with any other, is detecting breaches since each page view generates a nanocontract. Luckily, market constraints make the task a little less daunting. Business transactions in cookie data are not concerned with one data point. The market value of a clickstream data point is infinitesimal, so collecting agents make money by using or transacting for millions of these points together. One could therefore use sampling methodology to minimize auditing costs. By auditing only a fraction of the data, an auditor could determine (with a substantial degree of certainty) whether a collecting agent has violated a privacy term.

Ideally, database owners should have to open their databases to a central auditor—one that is either public or that is commissioned by the government—in order to assure that, in the event they breach the privacy term, they comply with the statutory mechanism for returning to the user P_R. If the privacy term constrains the use

[246] See Thibadeau, supra note 116.

[247] Costs of storing data are substantial, and for larger data miners, adding two fields may not be the optimal strategy for linking privacy configurations with data points. An alternate strategy would be to store the privacy configurations of every user in a *different* table, and to query that table anytime the data is audited. That way, if a collecting agent sees the same cookie fifty times, it need not store that cookie's configuration fifty times as well. Obviously date parameters are important, as users may change their preferences over time, but the author will ignore that here.

to which a collecting agent may put a piece of data, such as forbidding it to link a cookie number with personal information, auditors could easily detect that violation by looking at existing tables and recorded searches.[248] Moreover, if the privacy term prohibits sale to other searching agents, that condition may be detected simply by footing the name of the agent being audited with the name of the original data collector, which data miners can easily embed in the privacy configuration.[249] In this context, the Orwell metaphor does online privacy a disservice, since central auditing of clickstream data exchange may be far less expensive than competition among auditing agents.[250]

This Note has at times referred to the privacy agreement between a user's browser and the collecting agent's server as a nanocontract. Although the legal paradigm seeks to enforce the agreement by requiring compensation equivalent to that of an expectation measure familiar to anyone who has taken first-year contracts, the scheme may actually work most efficiently with a system of regulatory penalties or fines triggered when the collecting agent fails to return to the user her reservation price. The penalties could be calibrated to accomplish the objectives set forth in Part III without requiring that injured users pursue individual contract claims. Instead of punishing the breach itself, a regulatory agent, perhaps the FTC,[251] should sanction *the failure to return to the user* or a relevant substitute (discussed below) *the reservation price embedded in the nanocontract.* A sufficiently substantial criminal penalty could diminish the necessity for exhaustive auditing. The penalty, conceptually equivalent to D_E in the preceding analysis, need only be set high enough such that the expected loss from violating

---

[248] Expert computer science auditors can quickly audit vast quantities of data. See, e.g., In re Pharmatrak, Inc., 329 F.2d 9, 17 (1st Cir. 2003) (describing how the plaintiff's expert witness was able to conduct a thorough audit of the defendant's log files in a matter of hours).

[249] See Thibadeau, supra note 116.

[250] An argument for the superiority of a central auditing agent in the context of content ratings is presented in Thomas B. Nachbar, Paradox and Structure: Relying on Government Regulation to Preserve the Internet's Unregulated Character, 85 Minn. L. Rev. 215, 270–87 (2000).

[251] See Volokh, supra note 12, at 1060 ("Though breach of contract has traditionally been seen as a purely private wrong, to be remedied through a private lawsuit, it's similar enough . . . to fraud or false advertising that there's nothing startling about . . . the Federal Trade Commission prosecuting some such breaches itself.").

the privacy term and not paying P_R to the user (the probability of getting caught times the magnitude of the sanction) is greater than the value of P_R itself. Mathematically speaking, D_E need only be substantial enough such that those damages times the likelihood of getting caught, $P_{Detection}$, renders a decision to breach the privacy term without paying to the user P_R an inferior strategy to paying P_R voluntarily. In this context, the precise value of D_E is not as important as it was in a prophylactic scheme as, again, it need only be large enough so that $P_{Detection} * D_E > P_R$. The margin by which the left side of the equation exceeds the right matters less than in a regime that penalized violation of the privacy term itself. Penalizing the failure to pay rather than a breach of the term could significantly diminish auditing costs because the enforcing agency could afford to implement a regulatory structure with a very low $P_{Detection}$ so long as D_E would be sufficiently high to make compensating, in the event of a violation, the more profitable strategy (as compared to not compensating).

After a collecting agent violates user expectations, some value must be exacted from the violator and some value must be remitted to the consumer. Were there no transaction costs these values should be equal both to each other and to P_R (and, for that matter, V_{User}). The most significant transaction costs in this regulatory framework would involve returning money to the injured user. On the one hand, if the violator bears too much transaction cost, the scheme would fail to achieve the utility objective because collecting agents will not be able to make unauthorized uses of data with values that exceed P_R. On the other hand, if users bear too much transaction cost, the scheme would fail to achieve the compensatory objective because users would receive too much less than V_{User}. If each violation required a separate check to be issued directly to the user, the transaction costs could stifle, rather than stimulate, Internet use.

The problem involved in inexpensively returning P_R to consumers is something of a paradox because the very subject of the transaction, privacy, can prevent violators from knowing the identity of the person from whom they collected the data.[252] There may be a

[252] See Thibadeau, supra note 116 (identifying problems that inhere in a system of identifying privacy violations involving data that is supposed to be anonymous).

way around this unique problem. Two of the fields that collecting agents routinely store are a user's IP address and a timestamp. Local Area Networks ("LAN"s) and Internet Service Providers ("ISP"s) can generally identify the account associated with values for these two fields, using a procedure called a "reverse IP lookup."[253] Violators should have to transfer the sum of all reservation prices for violated nanocontracts to the ISP or LAN that owns the IP address. Violators could also transfer timestamps of the compromised data so the LAN or ISP can attribute the violation to a particular account. ISPs and LANs generally retain information about which account was using a particular IP address at a particular time. If the failure of these entities to retain this information becomes a problem, then the regulatory framework could simply require them to do so.

Orchestrating a system that can identify the users whose terms are violated, however, is only half of the enforcement solution. The other half involves returning to them some value approximating P_R without incurring excessive check-cutting costs. Fortunately, users have some established relationship to their ISPs or LANs. Most users that maintain accounts with ISPs pay these service providers periodically and most users that are on LANs either have similar accounts or are employed by the entity that deploys the LAN. In any of these circumstances, the user and the ISP/LAN owner exchange money periodically. The periodic monetary exchange could account for the money the ISP/LAN received from the violator without requiring any additional bill or check—P_R may simply be embedded as a line-item in a statement or a paycheck. Realistically, there are transaction costs associated with processing, but this method may represent one way of getting the "check cost" to a sufficiently low level that costs of administration do not cannibalize any benefits associated with increased fidelity to privacy terms and gains from socially desirable data usage.[254]

---

[253] "Reverse IP lookup" is a fancy name for a simple procedure. The LANs and ISPs can just associate the timestamp and the IP address recorded in the clickstream data with the account their records show using that IP address at that time.

[254] Processing costs associated with paying LANs/ISPs would be smaller than those associated with paying users directly because there are, by many orders of magnitude, more users than there are LANs and ISPs.

One objection may be that this seems to vest ISPs and LANs with exactly the type of control over personal information that the scheme seeks to wrest from the collecting agents. This is not a problem, however, because ISPs would not know what the user viewed, but only that the collecting agent violated the associated nanocontract. Knowledge regarding nothing more than whether a violation occurs is not particularly sensitive information.

Another objection may involve imperfect administration. Many different users, for example, may use the same computer on a LAN. The circumstances creating these enforcement difficulties, however, are the same ones that render unauthorized clickstream data use utterly inoffensive. Using the multiple-user example, there is little privacy violation if the "dossier" associated with the cookie reflects the surfing habits of a hundred different undergraduates at a public terminal. There may be other circumstances where, although a violation can be associated with a particular computer, one may not be able to associate a relationship between that box and a person. As the example illustrates, if one cannot associate a cookie profile (which is a property of a computer) with a person, then there exists little invasion of privacy at all.

In sum, any viable legal solution to the clickstream privacy controversy must involve some sort of centrally organized auditing. Otherwise, only in rare circumstances will violations inflict sufficiently large and concentrated user losses to induce people to bring claims. Absent transaction costs, one would almost always want a legal regime to protect, where possible, a highly differentiated set of preferences for privacy without suffocating socially desirable transactions for the underlying information. Although user valuations are highly differentiated, they are small enough that even relatively minor administrative costs can render the underlying transactions not worth the trouble. The advent of new browser technology, however, allows users to communicate almost costlessly their expectations to collecting agents and could just as easily be configured to communicate their privacy valuations as well. Leveraging current contractual relationships between LANs/ISPs and users may provide a channel for returning to the consumer her reservation price at a sufficiently low administrative expense so as to make such a scheme more desirable than enforcing prophylactic standards.

CONCLUSION

Companies will continue to invent new ways to use data, and, as a consequence, they will inevitably invent new ways to collect it. The stakes are simply too high to treat every informational privacy concern as a facsimile of the concerns coming before it. It is perhaps a collective social skepticism towards "newfangled technology" that militates against piecemeal, non-holistic approaches to online privacy reform. Perhaps this same sentiment fuels the desire to treat modern privacy concerns as though the law may easily address them using existing ways of thinking about informational privacy.

Although the term "privacy" may assume a number of deontic meanings, given the character of clickstream data collection it may well be the case that none of these conceptualizations adequately captures the appropriate way of thinking about information exchange. The intensity with which a user values privacy varies significantly within the Internet population, and a scheme that seeks to impose a uniform, prophylactic privacy standard therefore paradoxically protects clickstream informational privacy too much and, at the same time, not enough. Lawmakers can leverage existing browser technology to force users to reveal the intensity of their privacy preference, and they can leverage users' existing relationships with LANs and ISPs to compensate those whose preferences have been violated. Up until now, the law has struggled to generate possibilities other than prophylactic standards for clickstream data regulation. The law now stands on sufficiently sound technical footing to implement a solution acknowledging different preference intensity without incurring transaction costs suffocating all social gains associated with a market for clickstream data.