

WHAT GOOGLE KNOWS: PRIVACY AND INTERNET SEARCH ENGINES

Omer Tene*

Abstract

Search engines are the dominant actors on the Internet today and Google is undoubtedly, the undisputed king of search, evoking ambivalent feelings. It is adored for its ingenuity, simple, modest-looking interface, and superb services offered at no (evident) cost. Yet increasingly, it is feared by privacy advocates who view it as a private sector "big brother," posing what one commentator dubbed "the most difficult privacy problem in all of human history." Google is an informational gatekeeper, harboring previously unimaginable riches of personal data. Billions of search queries stream across Google servers each month, the aggregate thoughtstream of humankind online. Google compiles individual search query logs, containing information about each user's fears and expectations, interests and passions, and ripe with information that is financial, medical, sexual, political, in short, personal in nature.

The article begins with a technical and business analysis of search query logs. It utilizes Daniel Solove's taxonomy of privacy to analyze potential privacy harms inflicted by search engines discussing a range of responses to the search engine privacy problem, including technological, contractual, constitutional, statutory, and common law, emphasizing shortcomings of existing approaches and proposing solutions thereto. It advocates application of the breach of confidentiality tort to protect search users' privacy without impairing the ability of search engines to make use of the data collected.

* Assistant Professor, College of Management School of Law, Israel. LL.M., J.S.D (New York University); LL.B., LL.M. (Tel Aviv University); MBA (INSEAD). I would like to thank Michael Birnhack, Samuel Becher and Lee Bygrave for helpful comments. All web sites cited were last visited February 2008.

Introduction

I. Two types of search engine privacy

II. Use of data

- a) Use by search engine
- b) Use by third parties

III. Privacy problems

- a) Aggregation
- b) Distortion
- c) Exclusion
- d) Secondary use
- e) Chilling effect

IV. Privacy solutions

- a) Technological solutions
- b) Privacy policies and the limits of consent
- c) Constitutional protection—and the lack thereof
- d) Statutory protection—a cobweb full of holes
- e) Data retention v. data protection
- f) The law of confidentiality

Conclusion

"Don't be evil" Google¹

Introduction

Search engines are the central actors on the Internet today and Google is the undisputed king of search.² Google dominates the Internet,³ guiding users to the information they seek through an ocean of unrelated data with astonishing precision and speed. It is a powerful tool, evoking ambivalent feelings. On the one hand, we adore Google for its simple, modest-looking interface masking a hyper-complicated algorithm, which is the very essence of online ingenuity. We admire it for providing superb services at no (evident) cost, a practical miracle in today's market economy. On the other hand, we grow wary of Google's increasing clout as the ultimate arbiter of commercial success ("to exist is to be indexed by a search engine"⁴). And we fear potential abuse of its position as a central database for users' personal information, not only logging their search queries but also storing their e-mail (Gmail), calendars (Calendar), photos (Picasa), videos (YouTube), blogs (Blogger), documents (Docs & Spreadsheets), social networks (Orkut), news feeds (Reader), credit card information (Checkout)—in short, their entire digital lives.

¹ Google Code of Conduct, Preface, <http://investor.google.com/conduct.html>.

² For notable works in the growing body of literature on "search engine law," see Urs Gasser, *Regulating Search Engines: Taking Stock and Looking Ahead*, 8 YALE J. L. & TECH. 201 (2006); James Grimmelmann, *The Structure of Search Engine Law*, 93 IOWA L. REV. 3 (2007); Eric Goldman, *Search Engine Bias and the Demise of Search Engine Utopianism*, 8 YALE J. L. & TECH. 188 (2006); Roger Clarke, *Google's Gauntlets*, 22 COMP. L. & SEC. REP. 287 (2006); Heidi S. Padawer, *Google This: Search Engine Results Weave a Web for Trademark Infringement Actions on the Internet*, 81 WASH. U. L.Q. 1099 (2003); Lauren Troxclair, *Search Engines and Internet Advertisers: Just One Click Away from Trademark Infringement?*, 62 WASH. & LEE L. REV. 1365 (2005).

³ Google is estimated to account for nearly 60% of all Internet search queries in the United States—over six billion each month—more than double the next-largest search engine. See Press Release, Nilsen Netratings, Nielsen Online Announces December U.S. Search Share Rankings (Jan. 18, 2008), <http://biz.yahoo.com/iw/080118/0350483.html>. In 2006-07, Google is estimated to have received 76% of search revenue collected by the top three search engines. Yahoo, its top competitor, received 18%. See Business Wire, *Google Leads in Search Monetization as Yahoo's Market Share Stabilizes* (July 17, 2007), <http://tinyurl.com/yw7nhr>. Google accounts for 30% of online ad revenue in the U.S., compared with 16% for Yahoo and 7% for Microsoft. See Steve Lohr, *Yahoo Offer Is Strategy Shift for Microsoft*, N.Y. TIMES, Feb. 2, 2008, available at <http://www.nytimes.com/2008/02/02/technology/02soft.html?ref=technology>; see also Miguel Helft & Andrew Ross Sorkin, *Eyes on Google, Microsoft Bids \$44 Billion for Yahoo*, N.Y. TIMES, Feb. 2, 2008, available at <http://www.nytimes.com/2008/02/02/technology/02yahoo.html?ref=technology>.

⁴ Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INF. SOC. 169, 171 (2000).

Google's access to and storage of vast amounts of personally identifiable information create a serious privacy problem, one that Princeton computer scientist Edward Felten recently called "perhaps the most difficult privacy [problem] in all of human history."⁵ Every day, millions upon millions of users provide Google with unfettered access to their interests, needs, desires, fears, pleasures, and intentions. Many users do not realize that this information is logged and maintained in a form which can facilitate their identification. As John Battelle memorably put it, "[I]nk by link, click by click, search is building possibly the most lasting, ponderous, and significant cultural artifact in the history of humankind: the Database of Intentions."⁶ This "Database of Intentions," meaning "the aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result,"⁷ constitutes a honey pot for various actors. These range from the NSA and FBI, which expend billions of dollars on online surveillance⁸ and cannot overlook Google's information treasure trove, to hackers and identity thieves, who routinely overcome information security systems no matter how robust.

In April 2007, the Electronic Privacy Information Center (EPIC), a leading privacy group, filed a complaint with the Federal Trade Commission, arguing Google's contemplated 3.1 billion dollar merger with advertising powerhouse Doubleclick⁹ must be blocked on privacy grounds.¹⁰ Although the FTC approved the merger in

⁵ Economist Special Briefing, *Inside the Googleplex*, ECONOMIST, Aug. 30, 2007, available at http://www.economist.com/business/displaystory.cfm?story_id=9719610.

⁶ JOHN BATTELLE: HOW GOOGLE AND ITS RIVALS REWROTE THE RULES OF BUSINESS AND TRANSFORMED OUR CULTURE 6 (2005) [hereinafter BATTELLE, THE SEARCH].

⁷ John Battelle, *The Database of Intentions*, JOHN BATTELLE'S SEARCHBLOG, Nov. 13, 2003, <http://battellemedia.com/archives/000063.php>.

⁸ See, e.g., John Leyden, *US Warrantless Wiretapping Predates 9/11*, THE REGISTER, Dec. 18, 2007, http://www.theregister.co.uk/2007/12/18/warrantless_wiretapping_latest; Declan McCullagh, *Anger Grows over NSA Surveillance Report*, CNET NEWS.COM, May 11, 2006, <http://tinyurl.com/2r3abm>.

⁹ Elinor Mills, *Google Buys Ad Firm DoubleClick for \$3.1 billion*, CNET NEWS.COM, Apr. 13, 2007, http://www.news.com/Google+buys+ad+firm+DoubleClick+for+3.1+billion/2100-1024_3-6176079.html?tag=st.rn.

¹⁰ *In the Matter of Google and DoubleClick, Complaint and Request for Injunction, Request for Investigation and for Other Relief*, before the Federal Trade Commission (Apr. 20, 2007), available at http://www.epic.org/privacy/ftc/google/epic_complaint.pdf. DoubleClick is the leading provider of Internet-based advertising. It is a long-time nemesis of privacy advocates, who claim the company tracks users' behavior across cyberspace. In February 2000, EPIC filed a complaint with the FTC alleging that DoubleClick was unlawfully tracking users' online activities and combining surfing records with detailed personal profiles into a national marketing database. See *In the Matter of DoubleClick, Complaint and Request for Injunction, Request for Investigation and for Other Relief*, before the Federal Trade Commission (Feb. 10, 2000), available at

December 2007, it had done so largely sidestepping its privacy implications.¹¹ The transaction remains under review by European Union competition authorities, which intend to review not only antitrust but also privacy issues,¹² and in May 2007, European privacy regulators launched an investigation into Google's data retention and privacy practices¹³ that was quickly expanded to other search engines.¹⁴

A leading advocate for human rights, Privacy International, recently ranked Google's privacy practices as the worst out of a group of more than 20 leading Internet service providers, including Microsoft, Yahoo, Amazon, and eBay.¹⁵ Privacy International describes Google as "an endemic threat to privacy."¹⁶ It criticizes Google's "aggressive use of invasive or potentially invasive technologies and techniques" and claims the company "fails to follow generally accepted privacy practices such as the Organization for Economic Co-operation and Development (OECD) Privacy

http://www.epic.org/privacy/internet/ftc/DCLK_complaint.pdf. The case ended in a settlement, pursuant to which DoubleClick undertook a line of commitments to improve its data collection practices, increase transparency and provide users with opt out options. See Joel Winston, Acting Associate Dir., Div. of Fin. Practices, FTC, Letter to Christine Varney, Esq., Jan. 22, 2001, available at <http://www.ftc.gov/os/closings/staff/doubleclick.pdf>.

¹¹ The FTC approved the merger by a 4-1 decision. See Statement of Federal Trade Commission concerning Google/DoubleClick, FTC File No. 071-0170, Dec. 21, 2007, available at <http://ftc.gov/os/caselist/0710170/071220statement.pdf>, holding:

"Although such issues may present important policy questions for the Nation, the sole purpose of federal antitrust review of mergers and acquisitions is to identify and remedy transactions that harm competition ... [R]egulating the privacy requirements of just one company could itself pose a serious detriment to competition in this vast and rapidly evolving industry." *Id.* at 2. See also Dissenting Statement of Commissioner Pamela Jones Harbour, available at

<http://ftc.gov/os/caselist/0710170/071220harbour.pdf>. The FTC has nevertheless entered the fray recently, proposing a set of fair information principles for adoption through self regulation. See FTC, Online Behavioral Advertising—Moving the Discussion Forward to Possible Self-Regulatory Principles, Dec. 20, 2007, available at <http://www.ftc.gov/os/2007/12/P859900stmt.pdf>.

¹² See European Commission Directorate on Competition, Press Release, Mergers: Commission opens in-depth investigation into Google's proposed take over of DoubleClick, Nov. 13, 2007, available at <http://tinyurl.com/35hg4u>. See also Reuters, *Google Spars with European Lawmakers over Privacy*, CNET NEWS.COM, Jan. 21, 2008, http://www.news.com/Google-spars-with-European-lawmakers-over-privacy/2100-1030_3-6227031.html; Dawn Kawamoto & Elinor Mills, *Google-DoubleClick: Tough Sell in EU*, CNET NEWS.COM, Nov. 21, 2007, http://www.news.com/Google-DoubleClick-Tough-sell-in-EU/2100-1030_3-6219589.html?tag=html.alert.hed.

¹³ Article 29 Working Party Letter to Mr. Peter Fleischer, Global Privacy Counsel, Google (May 16, 2007), available at http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_google_16_05_07_en.pdf.

¹⁴ Article 29 Working Party Press Release (Dec. 5, 2007), available at

http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_05_12_07_en.pdf.

¹⁵ See Gemma Simpson, *Google Scores Lowest in Privacy Rankings*, ZDNET, June 12, 2007, <http://news.zdnet.co.uk/internet/0,1000000097,39287492,00.htm>.

¹⁶ Privacy International, *A Race to the Bottom—Privacy Ranking of Internet Service Companies, A Consultation Report*, June 9, 2007, available at <http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-553961>.

Guidelines and elements of EU data protection law."¹⁷ A recent report by research group One World Trust ranked Google as one of the least accountable and transparent organizations in the world.¹⁸ And in her dissenting opinion in the FTC decision upholding the Google/DoubleClick transaction, Commissioner Jones Harbour states that she is “uncomfortable accepting the merging parties’ nonbinding representations at face value. The truth is, we really do not know what Google/DoubleClick can or will do with its trove of information about consumers’ Internet habits. The merger creates a firm with vast knowledge of consumer preferences, subject to very little accountability.”¹⁹

How did Google evolve from being a benevolent giant seeking to “do no evil” into a potential privacy menace, depicted as a private sector “big brother” and reviled by human rights advocates worldwide?²⁰ What personally identifiable information should search engines be allowed to retain and for how long? What are the legal protections currently in place and are they sufficient to quell the emerging privacy crisis?

In Part I, I argue that since search query logs are typically traceable to an individual user, they create a serious privacy problem. In Part II, I show that such logs are used by search engines for various purposes, many of which are unknown to the average user. Moreover, the data in search query logs can be subpoenaed by government investigators or private litigants, who are thus afforded a peek into the private lives of unsuspecting users. More troubling yet, hackers, data thieves, and rogue employees may try to appropriate valuable personal information through illicit means. In Part III, I utilize Daniel Solove’s “Taxonomy of Privacy,” to analyze potential privacy

¹⁷ *Id.* See OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, Sept. 23, 1980, available at

http://www.oecd.org/document/20/0,3343,en_2649_201185_15589524_1_1_1_1,00.html.

¹⁸ See One World Trust, 2007 Global Accountability Report, Google Accountability Profile, Dec. 4, 2007, available at http://www.oneworldtrust.org/documents/Google_accountability_profile.pdf. See also John Oates, *Google Slightly Less Open than Interpol*, THE REGISTER, Dec. 4, 2007, http://www.theregister.co.uk/2007/12/04/google_privacy_transparency; cf. response of Google’s chief privacy counsel, Peter Fleischer, Transparency, Google and Privacy, PETER FLEISCHER BLOG, Dec. 5, 2007, <http://peterfleischer.blogspot.com/2007/12/transparency-google-and-privacy.html>.

¹⁹ Dissenting Statement of Commissioner Pamela Jones Harbour, *supra* note 11, at 9-10.

²⁰ See Leaders, *Who's Afraid of Google*, ECONOMIST, Aug. 30, 2007, available at http://www.economist.com/opinion/displaystory.cfm?story_id=9725272.

harms inflicted by search engines.²¹ This is the first systematic analysis in legal literature of search engines' privacy invasive activities.

In Part IV I discuss a range of solutions to search engine privacy problems, emphasizing the shortcomings of existing approaches, and proposing solutions thereto. I address six potential privacy responses. *First*, I describe technological solutions, such as cookie blocking, proxy servers, and anonymizing software. I argue that these tools, while useful, do not afford complete protection and are not readily available to the average user. *Second*, I address search engine privacy policies. These documents consist of self imposed, often opaque, contractual terms drafted by companies to protect their own interests as opposed to users' privacy. Moreover, user consent to such documents is implicit, uninformed, and partially coerced. *Third*, I present Fourth Amendment constitutional doctrine, under which a person has no "reasonable expectation to privacy" regarding information she turns over to a third party. I argue that in a day and age where third parties—such as financial institutions, telecommunication companies, and government agencies—maintain databases with massive amounts of personally identifiable information, U.S. constitutional doctrine has become obsolete. The European model, which applies a set of fair information principles to personally identifiable information in the hands of third parties, is more appropriate to deal with today's technological landscape. *Fourth*, I illustrate the Byzantine statutory scheme governing electronic communications stored by online service providers, which offers surprisingly weak privacy protection for search engine users. I argue that search query logs should be classified to secure stronger privacy protection such as that afforded to "contents of communications" as opposed to traffic data. The information in search query logs is highly revealing and cuts to the very core of a person's feelings and thoughts. Such information, while not the contents of a communication between a user and another person, is certainly the contents of a communication between a user and the search engine server. *Fifth*, I address the

²¹ See Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477 (2006) [hereinafter Solove, Taxonomy]; for notable previous attempts to organize the field, see Ken Gormley, *One Hundred Years of Privacy*, 1992 WIS. L. REV. 1335; Jerry Kang, *Information Privacy in Cyberspace Transactions*, 50 STAN. L. REV. 1193 (1998). The best known taxonomy is of course Prosser's, William L. Prosser, *Privacy*, 48 CAL. L. REV. 383 (1960). Other important contributions are ALAN F. WESTIN, *PRIVACY AND FREEDOM* (1967); Ruth Gavison, *Privacy*, 89 YALE L.J. 421 (1980). See also Robert C. Post, *The Social Foundations of Privacy*, 77 CAL. L. REV. 957 (1989); JULIE INNESS, *PRIVACY, INTIMACY, AND ISOLATION* (Oxford University Press, 1992); JUDITH W. DECEW, *IN PURSUIT OF PRIVACY: LAW, ETHICS AND THE RISE OF TECHNOLOGY* (1997); AMITAI ETZIONI, *THE LIMITS OF PRIVACY* (2000).

recent spate of national security inspired data retention legislation that not only permits, but actually mandates, the retention of users' search query logs. Such legislation raises the stakes for individual users, whose digital lives stand ready for summons from search engine servers by interested third parties. *Finally*, I review the law of confidentiality, which is well developed in the U.K., but much less so in the U.S.,²² and has largely been ignored in the debate over online privacy. I advocate application of the breach of confidentiality tort to protect search users' privacy without eliminating the ability of search engines to make use of the data they collect.

Throughout this article I use Google as a proxy for the entire search engine industry. While Google dominates search, it is by no means the only actor in the field, and, setting aside the Privacy International report discussed above, it is no worse than any of its major competitors.²³ I use Google for comfort of exposition and since, truth be said, I would not think of using another search engine myself.

I. Two types of search engine privacy

Search engine privacy comes in two flavors. *First*, there is the privacy interest of the search target.²⁴ The power of search has significantly reduced the transaction costs of compiling digital dossiers profiling a person's activities. Before the advent of search engines, we enjoyed a degree of "practical obscurity," protecting our privacy interest in issues such as litigation, asset ownership, past employment, and political opinion.²⁵ Although such information has always been in the public sphere, it was protected *de*

²² See Neil M. Richards & Daniel J. Solove, *Privacy's Other Path: Recovering the Law of Confidentiality*, 96 GEO. L. J. 123, 181 (2007), stating that "In contrast to the rather meagerly developed breach of confidentiality tort in America, the English tort is quite expansive and is enlarging its territory."

²³ Some search engines do provide a greater degree of privacy, competing with Google, Yahoo and Microsoft on precisely this issue. See, e.g., Jacqui Cheng, *Ask.com to Offer Anonymous Search with AskEraser*, ARSTECHNICA, July 20, 2007, <http://arstechnica.com/news.ars/post/20070720-ask-com-to-offer-anonymous-search-with-askeraser.html>. Yet the differences between the privacy practices of the major players are mundane and in some aspects Google has a better track record than its competitors. Google actively promotes sound privacy practices, recently calling for a globally coordinated approach to the problem. See Elinor Mills, *Google Proposes Global Privacy Standard*, CNET NEWS.COM, Sept. 13, 2007, http://www.news.com/Google-proposes-global-privacy-standard/2100-1030_3-6207927.html.

²⁴ See, e.g., Herman T. Tavani, *Search Engines, Personal Information and the Problem of Privacy in Public*, 3 INT'L REV. INFO. ETHICS 39 (2005).

²⁵ See Chris Hoofnagle, *Search Engines and Individual Rights*, Pre-Conference Paper, "Regulating Search Conference," Yale Law School, Nov. 28, 2005, available at http://islandia.law.yale.edu/isp/search_papers/hoofnagle.pdf.

facto from all but skilled investigators or highly motivated researchers, due to the practical difficulty and costs involved in uncovering and compiling the data.²⁶ Today such information has become available instantly and costlessly through search engines such as Google. Generally, access to information is a good thing, of course. We all benefit from finding the best consumer goods at rock bottom prices. We greatly value the increased access to information for research, education, business, and pleasure.²⁷ Indeed, search engines create enormous social benefits. Yet these efficiency gains come at a cost to the search targets, whose private lives becomes accessible and searchable by current and prospective employers,²⁸ romantic prospects, nosy neighbors, press reporters, and even stalkers²⁹ and other criminals.³⁰ A balance must be struck between the efficiency benefits and the privacy costs of search engine activity.

Second, there is the privacy interest of the person conducting the search ("user"). In August 2005, as part of its longstanding effort to enforce the Child Online Protection Act ("COPA"),³¹ the U.S. government issued a subpoena to AOL, Google, Microsoft, and Yahoo, requesting the addresses of all web sites indexed by the search engines as

²⁶ As Battelle notes, "regardless of your prurient desire to know whether your new coworker has a messy divorce or a DUI in his otherwise well appointed closet, most of us will not spend an afternoon down in the basement of our county courthouse to find out." BATTLE, *THE SEARCH*, *supra* note 6, at 191.

²⁷ Richard Posner has written extensively on the informational and efficiency costs of the right to privacy. *See, e.g.*, RICHARD A. POSNER, *THE ECONOMICS OF JUSTICE* (1981); Richard A. Posner, *The Right to Privacy*, 12 GA. L. REV. 393 (1978); Richard A. Posner, *The Economics of Privacy*, 71 AM. ECON. REV. (1981); Richard A. Posner, Blackmail, Privacy, and Freedom of Contract, 141 U. PA. L. REV. 1817 (1993); Richard A. Posner, Privacy, Secrecy and Reputation, 28 BUFF. L. REV. 1 (1979). *See generally Symposium, The Law and Economics of Privacy*, 9 J. LEGAL STUD. 621 (1980).

²⁸ In *Mullins v. Department of Commerce*, 2007 WL 1302152 (Fed. Cir. 2007), David Mullins, a U.S. government employee, argued that he had been unlawfully dismissed due to a Google search by a supervisor, which revealed that he had been discharged from the Air Force.

²⁹ Consider the Amy Boyer "cyberstalking" case: Liam Youens, a former classmate of Ms. Boyer, who was obsessed with her since high school, obtained her personally identifiable information, including home and work address, from Docusearch.com, a self proclaimed "premier provider of on-line investigative solutions." Mr. Youens used the information to locate Ms. Boyer at her workplace, murder her, and commit suicide. *See* *Remsberg v. DocuSearch, Inc.*, 816 A.2d 1001 (N.H. 2003).

³⁰ *See* www.whosarat.com, a web site devoted to exposing the identities of witnesses cooperating with the government. The site posts police and FBI informants' names and mug shots, along with court documents detailing what they have agreed to do in exchange for lenient sentences. *See* Adam Liptak, *Web Sites Listing Informants Concern Justice Department*, N.Y. TIMES, May 22, 2007.

³¹ Pub. L. No. 105-277, 112 Stat. 2681 (1998) (codified as 47 U.S.C. § 231 (2000)). The law, intended to protect children from access to online pornography (not to confuse with child pornography), has repeatedly been challenged by the ACLU and struck down by the Supreme Court. *See* *Reno v. ACLU*, 521 U.S. 844 (1997) (invalidating COPA's predecessor, the Communications Decency Act of 1996, Pub. L. No. 104-104, 110 Stat. 133); *ACLU v. Ashcroft*, 322 F.3d 240 (3d Cir. 2003), *aff'd*, 124 S. Ct. 2783 (2004) (invalidating COPA).

well as every search term entered by search engine users during a period of two months. The government was seeking to refute the assertion that filtering devices may work as well as or better than criminal prosecutions in achieving the COPA's aims of keeping pornographic materials away from children. The government wanted to prove its point by showing what the average Internet user is searching for, surmising that many of the searches lead to material "harmful to minors."³² Of the four companies approached, only Google objected to the government subpoena, claiming that the request for information threatened its trade secrets and image as a protector of user privacy. A United States District Court ruled that the government was entitled to compel Google to provide a sample of URLs, but that Google would not have to disclose any of its users' search queries.³³

Most people who followed the story asked themselves not whether the government subpoena complied with the Federal Rules of Civil Procedure, but rather: "what? Google keeps a record of all of my online searches?" Surprisingly for users not rehearsed on Google's intricate privacy policy, the answer is simply "yes." Google records all search queries linked to a specific Internet Protocol (IP) address.³⁴ In its privacy policy, the company states:

[O]ur servers automatically record information that your browser sends whenever you visit a web site. These server logs may include information such as your web request, Internet Protocol address, browser type, browser language, the date and time of your request and one or more cookies that may uniquely identify your browser.³⁵

³² See, *Gonzales v. Google*, Trial Motion, Memorandum and Affidavit, Reply Memorandum in Support of the Motion to Compel Compliance With Subpoena Duces Tecum, 2006 WL 733758 (Feb. 24, 2006).

³³ *Gonzales v. Google, Inc.*, 234 F.R.D. 674 (N.D.Cal. 2006).

³⁴ For example, if a user enters a search for "kill neighbor" and "dispose of body," the URL for Google's reply, which will be logged by the search engine, is <http://www.google.com/search?hl=en&q=kill+neighbor+dispose+of+body>. A URL, or "Uniform Resource Locator," is the global address of documents and other resources on the World Wide Web. See URL, WEBOPEDIA, <http://www.webopedia.com/TERM/U/URL.html>.

³⁵ Google Privacy Policy, <http://www.google.com/intl/en/privacypolicy.html#information>. See also Google Privacy FAQ, http://www.google.com/intl/en/privacy_faq.html.

In addition, Google records the hyperlinks users click on after obtaining their search results.³⁶

Users' search query logs may contain highly revealing, personally identifiable information. We use search engines to explore job opportunities, financial investments, consumer goods, sexual interests, travel plans, friends and acquaintances, matchmaking services, political issues, religious beliefs, medical conditions, and more. One's search history eerily resembles a metaphorical X-ray photo of one's thoughts, beliefs, fears, and hopes. It is ripe with information that is financial, professional, political, sexual, and medical in nature. Data contained in search query logs may be far more embarrassing and privacy intrusive than that of the contents of e-mail correspondences or telephone calls. Consider the scrutiny you give to an e-mail message prior to clicking "send," compared to the utter carelessness before Googling a search query. Imagine an online dossier of yourself, residing on the servers of a multinational company, laden with terms such as "Britney nude," "growing marijuana," "impotence pills," "job search," "genital warts," "prozac side effects," "married gay men," etc.

A surprising peek into precisely such digital dossiers was provided courtesy of AOL in August 2006.³⁷ AOL posted on its "research" web site (research.aol.com), a list of 20 million search queries entered by 658,000 users over a period of three-months. After a few days, it rushed to take the data offline amid a maelstrom of public criticism. Yet much of the information had already been downloaded, reposted, and made searchable at a number of third party web sites. The privacy debacle formally ended when AOL issued a public apology and dismissed its chief technology officer.³⁸

³⁶ Google Privacy FAQ, *id.* at section 5.

³⁷ Saul Hansell, *AOL Removes Search Data on Vast Group of Web Users*, N.Y. TIMES, Aug. 8, 2006, available at <http://tinyurl.com/38rzpq>; J. Nicholas Hoover, *AOL Search-Term Data Was Anonymous, But Not Innocuous*, INFORMATIONWEEK, Aug. 14, 2006, <http://tinyurl.com/2wr2ue>.

³⁸ Elinor Mills & Anne Broache, *AOL Axes Staff Over Privacy Breach*, ZDNET, Aug. 22, 2006, <http://news.zdnet.co.uk/communications/0,1000000085,39281482,00.htm>.

The detailed search records revealed by AOL underscore how much users unintentionally reveal about themselves when they use search engines. Consider some of the search queries entered by user 1515830:

chai tea calories

calories in bananas

aftermath of incest

how to tell your family you're a victim of incest

surgical help for depression

oakland raiders comforter set

can you adopt after a suicide attempt

who is not allowed to adopt

i hate men

medication to enhance female desire

jobs in denver colorado

teaching positions in denver colorado

how long will the swelling last after my tummy tuck

divorce laws in ohio

free remote keyloggers

baked macaroni and cheese with sour cream

how to deal with anger

teaching jobs with the denver school system

marriage counseling tips

anti psychotic drugs

From just over a dozen search queries, it is easy to detect information concerning the user's health and mental condition, personal status, profession, geographical location, and even favorite sports team. Now imagine the wealth and depth of personal information contained in search query logs assembled over thousands and thousands of searches.

Queries entered by users such as number 17556639 appear to manifest criminal intent and may consequently be used at trial as evidence of wrongdoing:³⁹

how to kill your wife

pictures of dead people

photo of dead people

*car crash photo*⁴⁰

And while the AOL query data were purportedly anonymized and users assigned random serial numbers, the *New York Times* demonstrated how by a simple process of “reverse searching,” the identity of apparently anonymous users becomes easy to discern.⁴¹

Privacy concerns relate to personally identifiable information, that is, information which can be used to uniquely identify, contact, or locate a specific individual person. Information that cannot be linked to an individual person is less problematic from a

³⁹ See, e.g., U.S. v. Schuster, 467 F.3d 614 (7th Cir. 2006); see also Harriet Ryan, *Florida Man Convicted of Killing his Wife During Faked Mugging, Now Faces Death*, COURT TV NEWS, June 26, 2006, http://www.courttv.com/trials/barber/062406_verdict_ctv.html.

⁴⁰ Similarly, user 336865 searched for:

- sexy pregnant ladies naked
- child rape stories
- tamagotchi town.com
- preteen sex stories
- illegal child porn
- incest stories
- illegel anime porn

Other queries, such as those entered by user 100906, are less ominous but no less revealing:

- cinninati bell iwireless
- addicted to love
- women who love to much
- learning to be single
- should you call your ex
- when your ex goes out of his way to run into u
- slim upper thighs
- prophet mohamed life teaching
- missed period or light spotting
- birthcontrol for morning after pill
- l&n federal credit union
- hes just not that into u
- i dont have a career
- should i get back with my divorced husband
- questions about the bible
- do i quailfy for food stamps in Kentucky

Once again, there are “hints” concerning the user’s geographic location, marital status, ethnic and religious origin, medical history (and spelling skills).

⁴¹ Michael Barbaro & Tom Zeller, *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006, available at <http://www.nytimes.com/2006/08/09/technology/09aol.html>.

privacy standpoint. Imagine we have highly revealing data about AOL user 100906, but we do not know, nor can we find out, who the user *is*.⁴² It is akin to being told that John Doe is a heroin-addicted, schizophrenic Satan worshipper, who earns \$10,000 a month, half of which he spends on diet pills. Absent any indication as to the identity of John Doe, such information is not very meaningful from a privacy perspective.

Federal privacy legislation protects personally identifiable information in a number of contexts, such as health information,⁴³ financial data,⁴⁴ or credit reports.⁴⁵ Similarly, the European data protection framework applies to "personal data," defined as "any information relating to an identified or identifiable natural person"⁴⁶

Could specific individuals be identified according to the data in their search query logs? As noted above, search engines log users' search queries under their IP address. An IP address is a unique string of numbers assigned to a user's computer by her Internet Service Provider (ISP) in order to communicate with her computer on the network.⁴⁷ Simply put, it is the cyberspace equivalent of a real space street address or telephone number. An IP address may be dynamic, meaning a different address is assigned to a user each time she logs on to the network; or static, that is assigned to a computer by an ISP to be its permanent Internet address.

Does an IP address constitute "personally identifiable information"? This question is equivalent to asking whether "435 Fifth Avenue, New York, New York" or "+1(212)435-2170" constitutes personally identifiable information. The answer depends on whether the address might be linked to a specific individual through

⁴² Or, more precise, who the users *are*, since several users, such as family members or colleagues at work, might be using a single computer with a given IP address.

⁴³ See Health Insurance Portability and Accountability Act ("HIPAA") of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996) (codified as amended in various sections of 42 U.S.C. (2000)).

⁴⁴ See Gramm-Leach-Bliley Financial Modernization Act of 1999, Pub. L. No. 106-102, 113 Stat. 1338 (codified in scattered sections of 12 U.S.C. (2000) and 15 U.S.C. (2000)).

⁴⁵ See Fair Credit Reporting Act of 1970, Pub. L. No. 91-508, 84 Stat. 1114 (codified at 15 U.S.C. §§ 1681-1681t).

⁴⁶ Article 2(a) of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281) 31 [hereinafter EU Data Protection Directive].

⁴⁷ See IP Address, WIKIPEDIA, http://en.wikipedia.org/wiki/IP_address.

reasonable means.⁴⁸ In other words, the address of a 40-story apartment building in Manhattan does not constitute “personally identifiable information” absent a specific apartment number; whereas the address of an Upper East Side townhouse does. Clearly, a static address is more “personal” than a dynamic address. To use an analogy, it is easier to identify an individual based on her home (“static”) address than based on a string of hotel rooms (“dynamic”) she occupied on a business trip. However, even a dynamic address is personally identifiable in cyberspace, given the ability of a user’s ISP to link such an address to the individual (or company) that used it.⁴⁹ While such identification requires an additional step (requesting the information from an ISP), it is possible, meaning that an apparently anonymous string of numbers is not as anonymous as it seems.⁵⁰ Subsequently, European privacy regulators⁵¹ have recently opined that dynamic IP addresses constitute personally identifiable information, or “personal data” in European parlance.⁵²

To overcome the difficulty of profiling users who access search engines each time using a different dynamic IP address, search engines set “cookies”⁵³ which tag users’

⁴⁸ See LEE A. BYGRAVE, DATA PROTECTION LAW, APPROACHING ITS RATIONALE, LOGIC AND LIMITS 315-319 (2002); CHRISTOPHER KUNER, EUROPEAN DATA PRIVACY LAW AND ONLINE BUSINESS 91-95 (2nd ed. 2007); see also Article 29 Working Party Working Document, Privacy on the Internet—An Integrated EU Approach to On-line Data Protection, November 2000, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2000/wp37en.pdf.

⁴⁹ This is typically the case. In certain circumstances, such as a user logging on to the Internet anonymously in an Internet café, even the ISP cannot link the address to an individual user.

⁵⁰ For so called “John Doe” cases, where users’ personal details have been subpoenaed from ISPs, see *Doe v. Cahill*, 884 A.2d 451 (Del. Super. Ct. 2005); *Dendrite International, Inc. v. Doe No. 3*, 775 A.2d 756 (N.J. Super. Ct. App. Div. 2001); *In re Subpoena Duces Tecum to America Online, Inc.*, 2000 WL 1210372, 52 Va. Cir. 26 (2000), *rev’d* on other grounds, 542 S.E.2d 377 (2001); Michael S. Vogel, *Unmasking “John Doe” Defendants: The Case Against Excessive Hand-Wringing Over Legal Standards*, 83 OR. L. REV. 795 (2004); Ryan M. Martin, *Freezing the Net: Rejecting a One-Size-Fits-All Standard for Unmasking Anonymous Internet Speakers in Defamation Lawsuits*, 75 U. CIN. L. REV. 1217 (2007).

⁵¹ European privacy regulators, known as the “data protection commissioners,” meet periodically in a group created pursuant to Article 29 of the Data Protection Directive (the “Article 29 Working Party”). The group has an advisory status and its decisions are non-binding; yet they constitute a valuable interpretative tool, given that they reflect the views of the national regulators charged with enforcing the law. See generally Joel Reidenberg, *Resolving Conflicting International Data Privacy Rules in Cyberspace*, 52 STAN. L. REV. 1315, 1364-66 (2000).

⁵² See Article 29 Working Group Opinion 4/2007 on the concept of personal data, June 20, 2007, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf, at 17. See also Aoife White, *IP Addresses Are Personal Data*, E.U. Regulator Says, WASHINGTON POST, Jan. 22, 2008, available at <http://tinyurl.com/2umfpl>.

⁵³ See HTTP cookie, WIKIPEDIA, http://en.wikipedia.org/wiki/HTTP_cookie:

HTTP cookies, sometimes known as web cookies or just cookies, are parcels of text sent by a server to a web browser and then sent back unchanged by the browser each time it accesses that server. HTTP cookies are used for authenticating, tracking, and maintaining specific

browsers with unique identifying numbers.⁵⁴ Such cookies enable search engines to recognize a user as a recurring visitor and amass her search history, even if she connects to the Internet using different IP addresses. Google's cookie was initially programmed to expire in 2038. As a result of pressure by European privacy regulators, Google announced in 2007 that it would shorten the duration of its cookie to a period of two years after a user's last Google search.⁵⁵ The privacy benefits of such a move are doubtful, however, since as long as Google remains the Internet's leading search engine, users are bound to renew the two-year period on a frequent basis.⁵⁶

One of the major weaknesses of a cookie as a tracking device is the fact that it is accessible only by the web server that placed it on a user's computer. Google has overcome this weakness with its recent takeover of DoubleClick.⁵⁷ DoubleClick allegedly tracks users' behavior across multiple web sites, utilizing so called "third-party cookies"⁵⁸ as well as its "DART" (Dynamic, Advertising, Reporting, and Targeting) technology.⁵⁹ In her dissenting opinion in the FTC decision upholding the transaction, Commissioner Jones Harbour warns that:

information about users, such as site preferences or the contents of their electronic shopping carts..

⁵⁴ The Google privacy policy states:

We use cookies to improve the quality of our service by storing user preferences and tracking user trends, such as how people search. Most browsers are initially set up to accept cookies, but you can reset your browser to refuse all cookies or to indicate when a cookie is being sent. However, some Google features and services may not function properly if your cookies are disabled.

See Google Privacy Policy, *supra* note 35. As a matter of fact, few users change their browser's default settings to reject cookies. See Jessica J. Thill, *The Cookie Monster: From Sesame Street to Your Hard Drive*, 52 S.C. L. REV. 921 (2001).

⁵⁵ See Cookies Expiring Sooner to Improve Privacy, OFFICIAL GOOGLE BLOG, July 16, 2007, <http://googleblog.blogspot.com/2007/07/cookies-expiring-sooner-to-improve.html>.

⁵⁶ See, e.g., Ryan Singel, *Google Changes Cookie Policy but Privacy Effect is Small*, WIRED BLOG NETWORK, July 16, 2007, <http://blog.wired.com/27bstroke6/2007/07/google-changes-.html>.

⁵⁷ See *supra* notes 9-12 and accompanying text.

⁵⁸ HTTP cookie, *supra* note 53:

While cookies are only sent to the server setting them or one in the same Internet domain, a web page may contain images or other components stored on servers in other domains. Cookies that are set during retrieval of these components are called third-party cookies. Advertising companies use third-party cookies to track a user across multiple sites. .

⁵⁹ In its complaint to the FTC, EPIC alleged that by purchasing Doubleclick, Google expanded its ability to pervasively monitor users not only on its web site but also on cyberspace as a whole. See Complaint and Request for Injunction, *supra* note 10; see also In the Matter of Google and DoubleClick, Supplemental Materials in Support of Pending Complaint and Request for Injunction, Request for Investigation and for Other Relief, (June 6, 2007), available at http://www.epic.org/privacy/ftc/google/supp_060607.pdf; Canadian Internet Policy and Public Interest Clinic, Section 9 Application for an Inquiry into the Proposed Merger of Google, Inc. and DoubleClick

post-merger, a user would visit one or more sites displaying DoubleClick ads, and also conduct one or more Google searches, during a time period when the IP address remained the same (a highly likely confluence of events, given each party's reach on the Internet). The merged firm would be able to use the common IP address to link the Google and DoubleClick cookies on that machine, and thereby cross-index that user among both databases.⁶⁰

One may argue that much like IP addresses, cookies do not constitute "personally identifiable information," since they identify a specific browser (typically, a computer) as opposed to an individual person. However, if a cookie and related search query log could be cross-referenced with an individual's name, the cookie itself would become personally identifiable. Google has this cross-referencing ability, since in addition to its search engine it provides users with a wide array of online services, many of which require registration using real name and e-mail address credentials. First and foremost is Gmail, the ubiquitous web based e-mail service launched in April 2004 as a private beta release by invitation only and opened to the public in February 2007.⁶¹ This article does not address the serious privacy issues raised by Gmail itself,⁶² but rather the synergetic privacy risk created by cross-referencing user search query logs with information collected by Gmail as part of the registration process. In other words, registration to Gmail or additional Google services such as

Inc. (Aug. 2, 2007) (addressed to Canadian Competition Bureau), *available at* http://www.cippic.ca/uploads/Google-DC_s.9_CompAct_complaint_FINAL.pdf.

⁶⁰ Dissenting Statement of Commissioner Pamela Jones Harbour, *supra* note 11, at 7, n.22. She further asserts that "the transaction will combine not only the two firms' products and services, but also their vast troves of data about consumer behavior on the Internet ... the merged firm will be capable of dominating the 'Database of Intentions.'" *Id.* at 4.

⁶¹ Gmail has been growing faster than any other e-mail site, nearly doubling its visitors to more than 20 million within the year 2007. It is the fourth largest web-mail provider, behind the much longer operating services of Yahoo, Microsoft and AOL. *See* Steve Lohr & Miguel Helft, Google Gets Ready to Rumble With Microsoft, N.Y. TIMES, Dec. 16, 2007, *available at* <http://tinyurl.com/ytysqo>.

⁶² Gmail gained its prominence (and notoriety) by providing a simple bargain for users (albeit not one that all users are aware of or understand): *get* an unprecedented amount of online storage space; *give* Google the opportunity to scan your e-mails' contents and add to them context-sensitive advertisements. The launch of Gmail turned out to be one of the most controversial product launches in the history of the Internet and placed Google at the center of a fierce privacy debate. *See* Matthew A. Goldberg, *The Googling of Online Privacy: Gmail, Search-Engine Histories and the New Frontier of Protecting Private Information on the Web*, 9 LEWIS & CLARK L. REV. 249, 250 (2005); Jason Isaac Miller, Note, "Don't Be Evil": Gmail's Relevant Text Advertisements Violate Google's Own Motto and Your E-Mail Privacy Rights, 33 HOFSTRA L. REV. 1607 (2005).

Google Talk, Google Reader, Google Calendar, or Google Checkout,⁶³ places the missing "name tag" on a user's search query log.⁶⁴

Finally, as demonstrated by the New York Times in the AOL case,⁶⁵ even apparently "pseudonymized" search query logs can be traced back to their originating user. This is done by a simple exercise of "reverse searching," combing search queries for personal identifiers, such as a social security numbers or credit card details. It becomes simpler yet by the tendency of users to run "ego searches" (also known as "vanity searches" or "egosurfing"), the practice of searching for one's own name on Google (once, twice, or many times per day).⁶⁶ In its effort to quash the government subpoena issued in *Gonzales v. Google*, Google itself posited that "search query contents can disclose identities and personally identifiable information such as user-initiated searches for their own social security or credit card numbers, or their mistakenly posted but revealing text."⁶⁷

To sum up, the contents of user search query logs are clearly personal in nature. They become privacy threatening if they can be traced back to a specific user. Google's ability to combine IP addresses, persistent cookies, and user registration information renders search query logs not only personal but also personally identifiable. Depending on their intended uses, search query logs may raise serious privacy problems.

III. Use of data

Why do search engines maintain search query logs? What is the information used for, and by whom? Who else may access the information and under what conditions? I

⁶³ There is also Google Web History, of course, which provides consenting users a personalized search experience linked to a personal account. Google Web History explicitly de-anonymizes one's search query log.

⁶⁴ Although users may register for services such as Gmail with a false or pseudonymous name, I suspect few do. I use Gmail as my main e-mail account due to its geographic and chronological versatility (you do not have to change e-mail addresses each time you relocate or switch jobs) and storage space. I use my real name, since I would not want colleagues or friends to receive e-mails from "Dr. No" or "Omer1970."

⁶⁵ See Barbaro & Zeller, *supra* note 41.

⁶⁶ Egosurfing, WIKIPEDIA, <http://en.wikipedia.org/wiki/Egosurfing>.

⁶⁷ See, *Gonzales v. Google*, Trial Motion, Memorandum and Affidavit, Google's Opposition to the Government's Motion to Compel, 2006 WL 728287 (Mar. 13, 2006).

show below that the answers to these questions affect the privacy analysis of user search query logs. This part distinguishes between use of information by search engines themselves and use by third parties. Use of search query logs by search engines may be anticipated by users and arguably agreed to as part of a transaction in which they are provided a service for free. However, use by third parties is more problematic and may be objectionable on the part of unsuspecting users.

a) Use by search engine

Google argues that the retention of search query logs is critical to its ability to operate and improve its services, and to provide adequate security for users.⁶⁸ Google faces the daunting task of having to guess what a user intends, essentially "read her mind," based on two or three words she enters as a search query. As Google co-founder Larry Page puts it, "[t]he perfect search engine would understand exactly what you mean and give back exactly what you want."⁶⁹

What complicates matters even more is that a single query may indicate different intentions depending on the context. For example, the words "Paris Hilton video" might be entered by a user searching for accommodation in the French capital, or (perhaps more likely) by one eager to follow the celebrity heiress' latest antics. Similarly, a "cheap apple" query might come from a user searching for fruit or for an iPhone. By analyzing search query logs, Google engineers can refine search quality and build new services, such as Google Spell Checker, which automatically looks at a query and checks whether the user entered the most common (and therefore, typically correct) version of a word's spelling. For example, if a user enters the words "Condoleza Rice," her search results would be preceded by the question: "Did you mean: Condoleezza Rice?"

Google also emphasizes the use of search query logs in preventing fraud and abuse and protecting the system from security attacks. To be sure, few if any users would

⁶⁸ Letter of Mr. Peter Fleischer, Global Privacy Counsel, Google, to Mr. Peter Schaar, Chairman, Article 29 Data Protection Working Party (June 10, 2007), *available at* http://www.epic.org/privacy/ftc/google/gres_a29_061007.pdf.

⁶⁹ See Google.com, Corporate Information, Our Philosophy, Never Settle for the Best, <http://www.google.com/corporate/tenthings.html>.

disapprove of optimizing search results and combating fraud. Yet Google also analyzes search query logs for revenue generating purposes, particularly for targeting and maximizing the effectiveness of advertisements. Google, after all, is an advertising company.⁷⁰ The predominant business model for search engines is contextual advertising, in which, alongside organic search results, users are displayed advertisements, most commonly textual, that are relevant to their search.⁷¹ The name of the game in online advertising, which is dominated by the pay-per-click (PPC) method of billing,⁷² is maximizing click-through rate (CTR), that is, the number of times users who visit a web page featuring an advertisement actually click the ad.⁷³ And in order to maximize CTR, search engines gauge user tastes, preferences, interests and needs. Google's chief executive officer, Eric Schmidt, stated: "If we target the right ad to the right person at the right time and they click it, we win."⁷⁴ Targeting "the right ad to the right person at the right time" requires knowing the users; and knowing the users means analyzing their search history.⁷⁵

Indeed, one might ask—why should search engines *not* retain user search query logs? Given the increasingly small costs of data warehousing,⁷⁶ relative dearth of regulation,⁷⁷ and potentially lucrative use of the information, search engines have little incentive to delete users' search query logs. This treasure trove of information is a "massive database of desires, needs, wants, and likes that can be discovered,

⁷⁰ Saul Hansell, *Google Wants to Dominate Madison Avenue, Too*, N.Y. TIMES, Oct. 30, 2005, available at <http://tinyurl.com/b3w6t>.

⁷¹ Grimmelmann, *supra* note 2, at 8; *see also* Statement of FTC, *supra* note 11.

⁷² "Pay per click (PPC) is an advertising model used on websites, advertising networks, and search engines where advertisers only pay when a user actually clicks on an ad to visit the advertiser's website." *See* Pay per click, WIKIPEDIA, http://en.wikipedia.org/wiki/Pay_per_click.

⁷³ Click-through rate, WIKIPEDIA, http://en.wikipedia.org/wiki/Click-through_rate.

⁷⁴ Hansell, *supra* note 70.

⁷⁵ No company evaluates user preferences as well as Google. Research shows that users click advertisements 50 percent to 100 percent more often on Google than they do on its main competitor, Yahoo. The cream of the crop in PPC advertising programs are Google's AdWords and AdSense programs, the company's main source of revenue. *See* Google AdWords, <http://adwords.google.com/select/Login>; Google AdSense, https://www.google.com/adsense/login/en_US/; *see also* FTC statement, *supra* note 11.

⁷⁶ Battelle notes that "the average cost per megabyte for storage has plummeted, and it will continue to drop until the point where it essentially reaches zero." BATTELLE, *THE SEARCH*, *supra* note 6, at 10. *See also* John Markoff, *Reshaping the Architecture of Memory*, N.Y. TIMES, Sept. 11, 2007, available at <http://www.nytimes.com/2007/09/11/technology/11storage.html?ref=technology>.

⁷⁷ *See* discussion *infra* notes 180-255 and accompanying text.

subpoenaed, archived, tracked, and exploited to all sorts of ends."⁷⁸ I now turn to discuss these additional potential uses.

b) Use by third parties

Google's Database of Intentions is an invaluable asset, a virtual honey pot for various third parties, ranging from national security and law enforcement officers to hackers and identity thieves. At present, search engines do not sell users' personally identifiable information to third parties,⁷⁹ yet they retain the ability to do so in the future.⁸⁰ Search engines do share user data with subsidiaries, affiliated companies, and other "trusted" business partners for the purpose of data processing and the provision of services.⁸¹ In addition, they retain the right to transfer data to a third party in case a merger or consolidation.⁸²

Certain third parties can—and in fact do—try to obtain user personally identifiable information from search engines through the legal process. First and foremost, the government may use search query logs for national security and law enforcement purposes, including the prevention, detection, and prosecution of crimes.⁸³ Clearly, a user searching for terms such as "illegal child pornography" or "prepare pipe bomb" warrants law enforcement intervention.⁸⁴ And indeed, governments tend to emphasize the most severe criminal activities, such as pedophilia, terrorism, and organized crime, when seeking authority to access user search query logs.⁸⁵ Few would dispute the imperative to provide government with all necessary tools to combat such heinous acts. Yet the picture becomes murkier when the government

⁷⁸ BATTELLE, *THE SEARCH*, *supra* note 6.

⁷⁹ See Google Privacy Policy, <http://www.google.com/intl/en/privacypolicy.html#information>; Yahoo! Privacy Policy, <http://info.yahoo.com/privacy/us/yahoo/details.html>; Microsoft Online Privacy Statement, <http://privacy.microsoft.com/en-us/fullnotice.aspx#use>.

⁸⁰ See discussion *infra* notes 170-174 and accompanying text.

⁸¹ See Privacy Policies, *supra* note 79. The term "trusted" is not defined in the Google and Yahoo privacy policies.

⁸² *Id.*

⁸³ See generally Michael D. Birnhack & Niva Elkin-Koren, *The Invisible Handshake: The Reemergence of the State in the Digital Environment*, 8 VA. J.L. & TECH. 6 (2003).

⁸⁴ Although even such ominous search queries might be entered, for example, by a researcher writing a paper on the subject.

⁸⁵ See, e.g., Declan McCullagh, *Terrorism Invoked in ISP Snooping Proposal*, CNET NEWS.COM, May 30, 2006, http://news.com.com/2100-1028_3-6078229.html; Prepared Remarks of Attorney General Alberto R. Gonzales at the National Center for Missing and Exploited Children (NCMEC) (Apr. 20, 2006), available at http://www.usdoj.gov/ag/speeches/2006/ag_speech_060420.html.

seeks to access search query logs of individuals who search for "how to cheat IRS." And a slippery slope may lead to the criminalization of search terms, such as "Falun Gong" or "democracy Tiananmen," in certain jurisdictions.⁸⁶

All major search engines declare in their privacy policies that they comply with legal process and government requests for information.⁸⁷ A full search warrant, supported by an affidavit showing probable cause, would in all cases enable law enforcement officers to access search engine data.⁸⁸ *The New York Times* recently reported that AOL alone responds to approximately 1,000 such criminal search warrants each month.⁸⁹ In most cases, however, much less than a full search warrant would suffice.⁹⁰ Search engines have been forthcoming in complying with government requests for users' personally identifiable information even when the consequences for identified users have been dire.⁹¹ Police is increasingly using search engine records as incriminating evidence in a variety of cases, ranging from homicide⁹² to wireless hacking.⁹³

Government access to user search query logs raises the risk of "function creep." Data intercepted in a search for terrorists may eventually be used by the government to prosecute tax offenders or collect debt. Surveillance tools, which may be accepted as necessary to combat serious crime or national security risks, appear disproportional when used for fiscal administration. Moreover, the evolving field of preventive law

⁸⁶ See Clive Thompson, *Google's China Problem (and China's Google Problem)*, N.Y. TIMESMAGAZINE, Apr. 23, 2006, available at <http://tinyurl.com/re2cn>.

⁸⁷ Google Privacy Policy, *supra* note 35.

⁸⁸ Grimmelmann, *supra* note 2, at 16.

⁸⁹ Saul Hansell, *Increasingly, Internet's Data Trail Leads to Court*, N.Y. TIMES, Feb. 4, 2006, available at <http://tinyurl.com/dpty3>; Adam Liptak, *In Case About Google's Secrets, Yours Are Safe*, N.Y. TIMES, Jan. 26, 2006, available at <http://tinyurl.com/cmnzg>.

⁹⁰ See discussion *infra* notes 243-255 and accompanying text; see also *Gonzales v. Google*, Amicus Brief of Center for Democracy & Technology in Support of Google's Opposition to the Motion to Compel of Attorney General Gonzales, 2006 WL 733757 (Mar. 13, 2006).

⁹¹ See, e.g., Jim Kerstetter, *Group says Yahoo helped jail Chinese journalist*, CNET NEWS.COM, Sept. 6, 2005, http://news.com.com/Group+says+Yahoo+helped+jail+Chinese+journalist/2100-1028_3-5851705.html; but see recently, AP, *Brazilian prosecutors say Google has not provided Orkut user information regarding crimes*, INT'L HERALD TRIB., Aug. 22, 2007, available at <http://www.ihf.com/articles/ap/2007/08/22/business/LA-FIN-Brazil-Google.php>.

⁹² See Lester Haines, *Alleged techie Killer Googled 'Neck Snap Break'*, THE REGISTER, Nov. 14, 2005, http://www.theregister.com/2005/11/14/techie_murder_evidence.

⁹³ U.S. v. Schuster, *supra* note 39.

enforcement tests the limits of legitimate government action in a democratic society.⁹⁴ Nabbing a terrorist before he realizes his plot to bomb a passenger jet is one thing.⁹⁵ It is quite another thing to arrest a teenager who runs Google searches for "kill guns," "prozac side effects," "brutal death metal bands," and "blood gore," and is therefore profiled by a data mining program as a potential "Columbine shooter." You might not want such a teenager to sit next to your daughter or son in class; but incarcerating him based on his Google searches—in essence applying guilt based on thoughts as opposed to deeds—is surely problematic.⁹⁶

In addition to criminal activity, search engine logs may be useful for litigants in civil cases, including copyright infringement, divorce, defamation, employment disputes, and shareholder actions.⁹⁷ The recording industry has been particularly aggressive in its attempts to identify online users who violate copyright law through service of subpoenas on online intermediaries, mainly ISPs.⁹⁸ While such cases have not yet been extended to search engines, the mega-lawsuit recently brought by Viacom against YouTube and its corporate parent Google for contributory and vicarious copyright infringement can have the effect of drawing search engines into the fray.⁹⁹

⁹⁴ See COLLEEN MCCUE, *DATA MINING AND PREDICTIVE ANALYSIS: INTELLIGENCE GATHERING AND CRIME ANALYSIS* (Butterworth-Heinemann 2006).

⁹⁵ See MARKLE FOUNDATION, *MOBILIZING INFORMATION TO PREVENT TERRORISM: THIRD REPORT OF THE MARKLE FOUNDATION TASK FORCE* (July 2006), available at http://www.markle.org/downloadable_assets/2006_nstf_report3.pdf.

⁹⁶ Consider the following exchange, from the film *Minority Report*:

"Knock, knock.

'Who's there?'

'FBI. You're under arrest.'

'But I haven't done anything.'

'You will if we don't arrest you,' replied Agent Smith of the Precrime Squad." *MINORITY REPORT* (20th Century Fox 2002), cited in K.A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 *COLUM. SCI. & TECH. L. REV.* 1 (2003); see also Clive Thompson, *Open Source Spying*, *N.Y. TIMES MAGAZINE*, Dec. 3, 2006, available at <http://tinyurl.com/2tewer>.

⁹⁷ Fred von Lohmann, *Could Future Subpoenas Tie You to 'Britney Spears Nude'?*, *LAW.COM*, Feb. 6, 2006, <http://www.law.com/jsp/article.jsp?id=1138961111185>.

⁹⁸ See *Recording Industry Association of America, Inc. v. Verizon Internet Services, Inc.*, 351 F.3d 1229 (D.C. Cir. 2003); see also *Sony Music Entertainment Inc. v. Does 1-40*, 326 F. Supp. 2d 556 (S.D.N.Y. 2004).

⁹⁹ See *Viacom v. YouTube and Google, Complaint for Declaratory and Injunctive Relief and Damages* (D.Ct. S.D.N.Y., Mar. 13, 2007), available at <http://news.com.com/pdf/ne/2007/ViacomYouTubeComplaint3-12-07.pdf>.

Third party subpoenas (*subpoena duces tecum*) are issued as a matter of course in civil litigation based on the relevance of evidence held by the intermediary.¹⁰⁰ Employers seek to summon an employee's search query logs to prove the employee used his computer for private purposes. A couple engaged in divorce proceedings subpoena each other's search query logs; the husband to prove his wife planned a secret vacation getaway; the wife to prove her husband sought homosexual escort services. Shareholders subpoena corporate insiders' search queries to prove that they engaged in insider trading.

Under the Federal Rules of Civil Procedure, an overbroad subpoena seeking irrelevant information may be quashed or modified if it subjects a non-litigant to an undue burden.¹⁰¹ In *Gonzales v. Google*, Google argued that the government subpoena of search query logs constituted an undue burden, based on the time and resources required to gather the requested information, as well as the risk to Google's trade secrets and confidential commercial information.¹⁰² Tellingly, users' privacy rights were raised by neither the government nor Google in their arguments in the case. In fact, the court explicitly stated it "raises, *sua sponte*, its concerns about the privacy of Google's users apart from Google's business goodwill argument."¹⁰³

In addition to *government* and *private actors* serving legal process, Google's information goldmine is bound to attract *hackers* and *data thieves*. Valuable databases get infiltrated all the time, regardless of the robustness of security measures. Security breaches abound even in highly guarded industries such as financial services, health services, and telecommunications. Rogue employees sell data to criminals; negligent employees lose laptops; computers are stolen and back up tapes lost;

¹⁰⁰ Fed. R. Civ. Proc. 26(b), 45. Online intermediaries seeking to resist third party subpoenas have occasionally relied on users' fundamental right of free speech (but not privacy). This line of cases is based on the Supreme Court ruling in *McIntyre v. Ohio Elections Comm'n*, 514 U.S. 334, 115 S.Ct. 1511, 131 L.Ed.2d 426 (1995), establishing the right to anonymous speech. See, e.g., *In re Subpoena Duces Tecum to America Online*, *supra* note 50; *but see Doe v. Cahill*, *supra* note 50. See also *2TheMart.Com*, 140 F.Supp.2d at 1090 (granting motion to quash subpoena seeking identities of anonymous ISP subscribers in shareholder derivative suit). Courts have yet to determine what speech interests, if any, users have in anonymous *search*. See *Grimmelmann*, *supra* note 2, at 16.

¹⁰¹ Fed. R. Civ. Proc. 45(c)(3)(A)(iv). See *Mattel Inc. v. Walking Mountain Prods.*, 353 F.3d 792 (9th Cir. 2003).

¹⁰² Google further claimed that the information requested by the government imposed on Google the risk of responding to inadequate process based on the Electronic Communications Privacy Act. See Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended at 18 U.S.C. §§ 2510-2520, 2701-2711, 3121-3127 (2000)) [hereinafter ECPA].

¹⁰³ *Gonzales v. Google*, *supra* note 33, at 687.

passwords are compromised and firewalls lowered. California's Security Breach Information Act ("SB 1386") of 2003,¹⁰⁴ which was followed by a spate of state legislation across the U.S.,¹⁰⁵ has led to the disclosure of security breaches in companies such as Citigroup, Bank of America, CardSystems, Merrill Lynch, T-Mobile, LexisNexis, Choicepoint, and Time Warner, as well as in dozens of colleges and universities, hospitals, and federal, state, and municipal government departments.¹⁰⁶ Overseas, the United Kingdom government admitted in November 2007 to having lost two HM Revenue & Customs CDs containing the personal and financial details of 25 million citizens, which could be worth more than \$3 billion to criminals.¹⁰⁷ The number of people whose personal data have been affected by security breaches through January 2008 is estimated at more than 218 million.¹⁰⁸ The point is that no matter what security measures are in place, data stored will eventually be data breached. The best method to secure data, and consequently guard individuals' privacy, is not to store them in the first place.¹⁰⁹

To sum up, far from being restricted to use by search engines themselves, search query logs may haunt users in future government investigations or private litigation and can be illicitly accessed by hackers and data thieves.

IV. Privacy problems

¹⁰⁴ Cal. Civ. Code § § 1798.29, .82, .84 (West Supp. 2006).

¹⁰⁵ See generally John Kennedy, *Slouching Towards Security Standards: The Legacy Of California's SB 1386*, 865 PLI/PAT 91 (2006) (reviewing legislation); see also Gramm-Leach-Bliley Act § 5, 15 U.S.C. § 6801, 6805 (2000), and the Interagency Guidance issued pursuant thereto: INTERAGENCY GUIDANCE ON RESPONSE PROGRAMS FOR UNAUTHORIZED ACCESS TO CUSTOMER INFORMATION AND CUSTOMER NOTICE, 70 Fed. Reg. 15,736, 15,743 (Mar. 29, 2005); INTERAGENCY GUIDANCE ESTABLISHING INFORMATION SECURITY STANDARDS, 69 Fed. Reg. 77,610 (Dec. 28, 2004).

¹⁰⁶ See Privacy Rights Clearinghouse, A Chronology of Data Breaches, <http://www.privacyrights.org/ar/ChronDataBreaches.htm>.

¹⁰⁷ See Andy McCue, U.K. Government's Lost Data 'Worth Billions to Criminals', CNET NEWS.COM, Nov. 29, 2007, <http://tinyurl.com/288ywo>.

¹⁰⁸ These include, for example, 40 million Visa and MasterCard accounts compromised by a hacking incident at data processor CardSystems Solutions; 28 million veterans whose names, Social Security numbers, dates of birth, phone numbers and addresses were stored on a laptop computer stolen from a government employee's home; and 3.9 million accountholders whose data have been compromised by Citigroup when it lost a shipment of computer backup tapes sent via UPS. See Chronology of Data Breaches, *supra* note 106.

¹⁰⁹ See generally, Lynn M. LoPucki, *Human Identification Theory and the Identity Theft Problem*, 80 TEX. L. REV. 89 (2001).

Any discussion of the right to privacy ultimately rests on the most basic of questions, namely "what does privacy *mean*?" Why exactly do I argue that the collection and use of search query logs may be privacy invasive? Numerous attempts have been made to define privacy and many are no doubt forthcoming.¹¹⁰ For the purposes of this article, I use Daniel Solove's, "Taxonomy of Privacy," which is a comprehensive, topical, and robust classification of *privacy invasive activities*.¹¹¹ In the first systematic analysis in legal literature of search engines' privacy invasive activities,¹¹² I show in this section that collection, aggregation, storage, use, and transfer of search query logs inflict many of the privacy harms surveyed by Solove.

Search engines raise the problem of *aggregation*, because intimate and comprehensive user profiles are assembled from bits of information revealed over time; *distortion*, because information in search query logs may be highly misleading with potentially harsh results for users; *exclusion*, because search engine users are not granted access to their files; and *secondary use*, because search engines use data collected from users for one purpose (search) to different ends (commercial, security, law enforcement, litigation). Finally, I discuss the *chilling effect* that search engines' privacy practices could have for search and online activity generally.

a) Aggregation

Solove defines aggregation as the "gathering together of information about a person."¹¹³ He explains that "combining information creates synergies. When analyzed, aggregated information can reveal new facts about a person that she did not expect would be known about her when the original, isolated data was collected."¹¹⁴ User search query logs aggregate vast amounts of data from tiny bits of information revealed by users gradually over time. Entering a search query for "French

¹¹⁰ Some of the notable works are Ruth Gavison, *Privacy*, 89 YALE L.J. 421 (1980); ALAN F. WESTIN, *PRIVACY AND FREEDOM* (1967); Charles Fried, *Privacy*, 77 YALE L.J. 475 (1968); Prosser, *supra* note 21; Solove, *Taxonomy*, *supra* note 21;

¹¹¹ Solove, *Taxonomy*, *supra* note 21. Neil Richards characterized Solove's work as part of the "Information Privacy Law Project," a group of scholars seeking to establish information privacy law as a field of study distinct from traditional U.S. constitutional privacy. See Neil M. Richards, *The Information Privacy Law Project*, 94 GEO. L. J. 10087 (2006).

¹¹² I focus on search engine activities that infringe on the legal right to privacy, as opposed to privacy as a social, economic or psychological concept.

¹¹³ Solove, *Taxonomy*. *supra* note 21, at 507.

¹¹⁴ *Id.*

mountains," may not give much away; "French mountains" and "ski vacation" is more telling; add to that "Christmas deals," "gift to grandchild," "NY Paris flights," "category D car rentals," "five star hotels," and "disabled access" and a lucid picture begins to emerge. Search by search, click by click, the profile and identity of a user becomes clearer.¹¹⁵ And if this is evident after half a dozen searches, consider the wealth and depth of information collected in search query logs that contain thousands and thousands of searches aggregated over a period of months or years. Even the few users who are aware of search engines' data compilation practices probably underestimate the impact of search query logs on their privacy, effectively making their lives "transparent" over time.¹¹⁶

What complicates matters even more is the highly concentrated nature of the search engine industry.¹¹⁷ With search, you not only know that voluminous data are being compiled, but also who is compiling them. Government, private litigants, and hackers alike know that Google and, to a lesser extent, Yahoo and MSN harbor this personal information.¹¹⁸

b) Distortion

Information in search query logs may be highly misleading, with potentially troubling results for users. A user searching for "assassinate U.S. president" is not necessarily a terrorist or criminal; instead, she might be a student conducting research for a history class. Similarly, a user searching for "growing marijuana" is not necessarily considering an agricultural endeavor; she may be a parent concerned with growing drug use in schools.

A real-life example of the elusive distinction between fact and fiction in search query logs was presented by *The New York Times* reporters who exposed Thelma Arnold as

¹¹⁵ Barbaro & Zeller, *supra* note 41.

¹¹⁶ DAVID BRIN, *THE TRANSPARENT SOCIETY – WILL TECHNOLOGY FORCE US TO CHOOSE BETWEEN PRIVACY AND FREEDOM?* (1998).

¹¹⁷ See Nielsen Netratings, *supra* note 3; see also Tair-Rong Sheu & Kathleen Carley, *Monopoly Power on the Web: A Preliminary Investigation of Search Engines*, available at <http://arxiv.org/ftp/cs/papers/0109/0109054.pdf>.

¹¹⁸ See, e.g., BATTELLE, *THE SEARCH*, *supra* note 6, at 6, noting that "the Database of Intentions ... lives in many places, but three or four places in particular—AOL, Google, MSN, Yahoo (...)"

the face behind the randomly assigned "AOL Searcher No. 4417749."¹¹⁹ Although the reporters were able to glean Ms. Arnold's identity from her search query log, they were also led astray by many of her search queries, such as "hand tremors," "nicotine effects on the body," "dry mouth," and even "bipolar," which appear to imply a wide range of ailments (or fear thereof). Ms. Arnold explained that "she routinely researched medical conditions for her friends to assuage their anxieties."¹²⁰ Ms. Arnold, who is a 62-year-old widow, also searched for the terms "dances by laura," "dances by lori," "single dances" and "single dances in Atlanta." She explained these entries as follows: "A woman was in a [public] bathroom crying. She was going through a divorce. I thought there was a place called 'Dances by Lori' for singles."¹²¹ Hence, in user search query logs, what you see is not always what you get.

Solove defines distortion as "the manipulation of the way a person is perceived and judged by others, and involves the victim being inaccurately exposed to the public."¹²² Recognizing the potentially harmful effects of inaccurate information, the EU Data Protection Directive provides that personally identifiable information must be "accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified."¹²³ In addition, individuals in the EU enjoy the right to access their personally identifiable information without delay, and to rectify, erase, or block data that are inaccurate or incomplete.¹²⁴ The combination of inaccurate and misleading data, ease of government access, and lack of transparency and accountability to users, makes user search query logs highly problematic from a privacy perspective.

c) Exclusion

The prohibition against secret databases is one of the doctrinal foundations of European privacy law, gleaned following decades of totalitarian regimes that used information in secret databases to police and terrorize citizens into conformity and

¹¹⁹ Barbaro & Zeller, *supra* note 41.

¹²⁰ *Id.*

¹²¹ *Id.*

¹²² Solove, Taxonomy, *supra* note 21, at 550.

¹²³ Article 6(1)(d) of the Data Protection Directive.

¹²⁴ Article 12 of the Data Protection Directive.

submission.¹²⁵ A corollary of the basic prohibition on secret databases is the right of individuals in Europe to be notified which data are collected about them, by whom, and for what purposes.¹²⁶ Individuals are entitled to access their personally identifiable information and, if necessary, correct or amend them.¹²⁷ Solove refers to "the failure to provide individuals with notice and input about their records as exclusion."¹²⁸ He explains that "exclusion creates a sense of vulnerability and uncertainty in individuals . . . [I]n a world where personal information is increasingly used to make important decisions about our lives, powerlessness in this arena can be significantly troublesome."¹²⁹

Public awareness to the extent of data retention by search engines is minimal. A survey held pursuant to the government's request for Google search records reveals that "89% of respondents believe that their web searches are kept private, and 77% believe that Google web searches do not reveal their personal identities."¹³⁰ To a great extent, then, Google's collection of search queries is a *de facto* "secret database."

In its complaint to the FTC concerning the Google/DoubleClick merger, EPIC points out that a user must click on four links from Google's ubiquitous homepage¹³¹ in order to obtain information concerning the company's data collection practices.¹³²

¹²⁵ For a good recent exposé see the German film *Das Leben der Anderen* (The Lives of Others) (Bayerischer Rundfunk, Germany 2006) (documenting the activities of the omniscient East German Stasi). MICHEL FOUCAULT, *DISCIPLINE AND PUNISH* (Vintage Books, 2d ed. 1995); Spiros Simitis, *Reviewing Privacy in an Information Society*, 135 U. PA. L. REV. 707 (1987). The classical work is of course GEORGE ORWELL, 1984 (1948).

¹²⁶ Articles 10-11 of the EU Data Protection Directive.

¹²⁷ Article 12 of the EU Data Protection Directive.

¹²⁸ Solove, *Taxonomy*, *supra* note 21, at 523.

¹²⁹ *Id.* at 523-24.

¹³⁰ Linda Rosencrance, *Survey Finds Solid Opposition to Release of Google Data to Feds*, COMPUTERWORLD, Jan. 24, 2006,

<http://www.computerworld.com/securitytopics/security/privacy/story/0,10801,107993,00.html>. This article will (hopefully) be read by students and lawyers interested in privacy or cyberlaw; yet I urge you to consider when *you* first became aware of search engines' data aggregation practices. Given that I assume the answer will be "not too long ago" (if that), consider the lack of knowledge by the general public.

¹³¹ Recently asked why the Google homepage does not contain a link to the company's privacy policy, Mr. Fleischer explained: "Google has a very sparse homepage. It's one of the things that we're very proud about. It's kind of clean and Zen-like ... We don't believe in pushing things into people's face." Matthew Magee, *Google privacy chief talks*, OUT-LAW RADIO, July 5, 2007, <http://www.out-law.com/page-8285>.

¹³² EPIC complaint, *supra* note 10, at 7. *First*, on the Google homepage, a user must click on "About Google." *Second*, the user must click on "Privacy Policy," which displays the "Google Privacy Policy

Moreover, even the full privacy policy fails to explain clearly what Google *does* with information in search query logs. In addition, it is not clear whether and to what extent users have access to their search query logs.¹³³

User access to search query logs is now provided as part of the Google Web History service.¹³⁴ Users of Google Web History may access their search query logs and edit or delete items therein. Yet such access comes at a significant privacy cost, because Google stores not only the *search queries* of Web History users, but also the *web pages* visited. Moreover, Google users who do not subscribe to Google Web History, ostensibly due to that very "cost," are arguably already paying a similar privacy price, given Google's retention of their search query logs. Finally, counter to Web History users, Google search users are not provided with the opportunity to edit or delete search query logs (at least not by simple means).

d) Secondary use

One of the fundamental principles of privacy law embodied in international instruments ranging from the OECD privacy guidelines to the EU Data Protection Directive is the principle of purpose specification.¹³⁵ Under the purpose specification principle, personally identifiable information obtained for one purpose must not be used or made available for another purpose without the affected individual's prior informed consent. Solove explains that secondary use of personally identifiable

Highlights" page. *Third*, the user has to click on the link to Google's full Privacy Policy, which outlines the information Google collects and how it uses it. Included in this list is the term "log information," which is described in text that contains the hyperlinked term "server logs." A fourth click on the term "server logs" leads the user to a FAQ entry for "What are server logs?" It is only there that the user can learn that Google retains her IP address in connection with her search queries. See Google Privacy Policy Highlights, <http://www.google.com/intl/en/privacy.html>; Google's full Privacy Policy, <http://www.google.com/intl/en/privacypolicy.html>; Google's explanation for the term "search query logs," Google Privacy FAQ, http://www.google.com/intl/en/privacy_faq.html#serverlogs.

¹³³ See Google's full Privacy Policy, *supra* note 35.

¹³⁴ Google Web History, www.google.com/psearch; see Margaret Kane, *Your Web History, Courtesy of Google*, CNET NEWS.COM, Apr. 20, 2007, http://news.com.com/8301-10784_3-9710855-7.html; Tom Espiner, *Google Launches Web History Tool in U.K.*, CNET NEWS.COM, Aug. 3, 2007, http://news.com.com/2100-1030_3-6200619.html.

¹³⁵ Section 9 of the OECD Guidelines; Article 5(b) of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Council of Europe Treaties No. 108 (Jan 28, 1981), available at <http://conventions.coe.int/Treaty/en/Treaties/html/108.htm>; Article 6(1)(B) of the EU Data Protection Directive, which provides that personal data must be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes."

information "creates a dignitary harm . . . emerging from denying people control over the future use of their data, which can be used in ways that have significant effects on their lives."¹³⁶ He points out that "secondary use resembles breach of confidentiality, in that there is a betrayal of the person's expectations when giving out information."¹³⁷

The case of user search query logs is instructive. My intuition (without having conducted empirical research) is that when you enter a search term in Google, you expect that information to be used to respond to your query—and no more. You do not (knowingly, necessarily) expect Google to aggregate your current query with all of your past searches and mine the data in order to improve its service. You probably do not expect Google to make use of this information to target you with effective advertising or analyze your ad viewing behavior.¹³⁸ And you most certainly do not expect Google to disburse this information to the government or private parties engaged in litigation against you. When Google uses the information in your search query log for purposes diverging from those you reasonably envisaged, it breaches your trust—your "reasonable expectation of privacy"¹³⁹—as well as the purpose specification principle.

A possible retort is that you do indeed consent, implicitly at least, to all of these uses, since they are specified in Google's privacy policy. However, the implicit consent argument is tenuous at best. *First*, consent is based in this case on a browse-wrap agreement,¹⁴⁰ which must be assembled from several distinct web pages¹⁴¹ and is hard

¹³⁶ Article 6(1)(B) of the EU Data Protection Directive, *supra* note 135, at 521-22.

¹³⁷ *Id.* at 522. For discussion of the tort of breach of confidentiality, *see infra* notes 277-295 and accompanying text.

¹³⁸ *See* Grimmelmann, *supra* note 2, at 15, noting that "even the display of advertising keyed to a user's query is arguably a purpose other than that intended by the user."

¹³⁹ In the U.S., the predominant test for a legally protected right to privacy is the "reasonable expectation of privacy" test established in *Katz v. United States*, 389 U.S. 347, 360-61; 88 S. Ct. 507; 19 L. Ed. 2d 576 (1967) (Harlan, J., concurring) [hereinafter *Katz*].

¹⁴⁰ A browse-wrap agreement is typically presented at the bottom of a web site; acceptance is based on "use" of the site. Hence, there is no affirmative signal of the user's assent to the contract's terms. Browse-wrap agreements are distinguished from, and obviously more problematic than "click-through agreements," which require an offeree to click on an acceptance icon, manifesting assent to be bound. *See Specht v. Netscape Communications Corp.*, 306 F.3d 17 (2d Cir. 2002). *See also* Mark A. Lemley, *Terms of Use*, 91 MINN. L. REV. 359 (2006); Ian Rambarran & Robert Hunt, *Are Browse-Wrap Agreements All They Are Wrapped Up To Be?*, 9 TUL. J. TECH. & INTEL. PROP. 173 (2007); Terry Haridi, *Mass Licensing—Part I: Shrinkwraps, Clickwraps & Browsewraps*, 831 PLI/PAT 251 (2005); Christina Kunz et al., *Browse-Wrap Agreements: Validity of Implied Assent in Electronic Form Agreements*, 59 BUS. LAW. 279 (2003).

¹⁴¹ *See supra* notes 131-132 and accompanying text.

to comprehend. *Second*, Google's privacy policy remains constructively opaque concerning the *primary* use of search query logs, rendering secondary use all the more difficult to accept.

Google's use of search data for secondary purposes and the privacy issues it raises, expose a broad rift between U.S. and European privacy law. The purpose specification principle, so deeply ingrained in EU law,¹⁴² is not at all evident in the U.S., where the underlying assumption has traditionally been that as between any individual and a company collecting her personally identifiable information, it is the company that owns the data and may use, reuse, or sell it to third parties at will.¹⁴³

e) Chilling effect

While not a privacy problem under Solove's taxonomy, Google's data retention and use may have a chilling effect on online search. I have argued that most users are probably not aware of Google's privacy practices. Increased public awareness may mean decreased use of search engines, or, at least, self-censored search. Google itself made this point in its response to the government's subpoena of search queries, arguing that "the production of the requested data will result in a chilling effect on Google's business and user trust."¹⁴⁴ Search engine users in China and other totalitarian regimes must of course think long and hard before looking for information about unpopular opposition groups or historic events.¹⁴⁵ A user entering a search query such as "free Taiwan" in China or "Islamic Jihad" in Egypt may pay a dear

¹⁴² In the EU, the purpose specification principle is based on the underlying belief that personal data must be controlled by the "data subject" and may be collected, used and transferred (collectively, "processed") by the user of the data (in the EU, "data controller"), strictly for those purposes consented to by the data subject or prescribed by law. See Article 8 of the Charter of Fundamental Rights of the European Union, Official Journal 2007/C 303/1 (Dec. 14, 2007) [hereinafter Charter], providing: "(1) Everyone has the right to the protection of personal data concerning him or her. (2) Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law."

¹⁴³ See Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373 (2000). Certain exceptions exist in specific legislation, which incorporates the purpose limitation principle. See, e.g., The Privacy Act of 1974, Pub. L. No. 93-579, 88 Stat. 1897 (Dec. 31, 1974).

¹⁴⁴ Google's Opposition, *supra* note 67. According to Google, "[i]f users believe that the text of their search queries into Google's search engine may become public knowledge, it only logically follows that they will be less likely to use the service . . . this chilling effect on Google's business is potentially severe."

¹⁴⁵ See, e.g., Lester Haines, *Egyptian Blogger Jailed for Four Years*, THE REGISTER, Feb. 22, 2007, http://www.theregister.co.uk/2007/02/22/egyptian_blogger_jailed. See also Thompson, *supra* note 86.

price for her curiosity. Yet self-censorship will afflict not only citizens of dictatorships. In Western democracies as well, in order to avoid potential embarrassment and remain above suspicion, users may refrain from intimate or potentially unpopular search queries.¹⁴⁶ As Julie Cohen thoughtfully observes, "[p]ervasive monitoring of every first move or false start will, at the margin, incline choices toward the bland and the mainstream."¹⁴⁷

V. Privacy solutions

This part outlines the main solutions to the search query logs privacy problem. As I illustrate below, current approaches are flawed and afford inadequate protection to search engine users. *First*, technological solutions permit users to mask their identity and browse anonymously, yet are complicated to implement and not entirely foolproof. *Second*, privacy policies are drafted by lawyers to protect search engines from liability, not users' privacy, and are based on user consent that is neither informed nor freely given. *Third*, constitutional doctrine in the U.S. is flawed insofar as it affords no protection for information held by third parties. At the same time, statutory provisions are difficult to decipher and provide a surprisingly low level of protection for the contents of communications. Moreover, emerging data retention requirements advanced by national security and law enforcement agencies further restrict user privacy by compelling service providers to maintain traffic data for extended periods of time. After reviewing these approaches and their shortcomings I suggest that a return to the law of confidentiality may reinforce user privacy without eliminating the ability of search engines themselves to make use of the data they collect.

a) Technological solutions

Technological problems often have technological solutions and search privacy is no exception. Privacy invasive technologies are countered by an array of privacy enhancing technologies (PETs) that enable users achieve a degree of (though rarely

¹⁴⁶ See Susan W. Brenner & Leo L. Clarke, *Fourth Amendment Protection for Shared Privacy Rights in Stored Transactional Data*, 14 J.L. & POL'Y 211, 265 (2006).

¹⁴⁷ Cohen, *supra* note 143, at 1426.

complete) online anonymity.¹⁴⁸ PETs cover a range of different technologies, including encryption tools, cookie management, Internet browser settings, and anonymization schemes.¹⁴⁹ Unfortunately, the vast majority of search users remain oblivious to PETs.

Within the context of search engines, users also may implement various technological measures, ranging from simple steps providing partial protection to more complicated procedures providing greater safeguards.¹⁵⁰ To begin with, search users may avoid logging in to their search engine or any related services, or using their ISP's search tool. As long as users manage to separate registration information from search query logs, it is difficult to link their identity to their search history.¹⁵¹ This, however, will not suffice to protect users from the retention of search query logs based on persistent cookies.

To combat this problem, users may set their browsers to block cookies from search engines or allow only session cookies, *i.e.*, cookies that will be erased each time the browser shuts down. More sophisticated users will use anonymous proxy servers and anonymizing software. A proxy server is a buffer between a user's computer and the Internet.¹⁵² A proxy server that removes identifying information from user requests for the purpose of anonymity is called an anonymizing server, or simply an anonymizer. Anonymizers effectively hide from third parties any information regarding a user and her search and browsing habits.¹⁵³ However, the anonymizer

¹⁴⁸ For technology as a regulatory mechanism, see LAWRENCE LESSIG, *CODE AND OTHER LAWS OF CYBERSPACE* (Basic Books, 1999); see also Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553 (1998).

¹⁴⁹ The government, in turn, is diligent in devising "responsive" surveillance technologies to counter PETs such as encryption and anonymization tools, which might be put to use by organized crime or terrorists. See, e.g., Ric Simmons, *Why 2007 is Not Like 1984: A Broader Perspective on Technology's Effect on Privacy and Fourth Amendment Jurisprudence*, 97 J. CRIM. L. & CRIMINOLOGY 531 (2007); Orin S. Kerr, *The Fourth Amendment in Cyberspace: Can Encryption Create a "Reasonable Expectation of Privacy?"*, 33 CONN. L. REV. 503 (2001).

¹⁵⁰ For a review of options and practical tips, see EFF, *Six Tips to Protect Your Online Search Privacy*, <http://www.eff.org/Privacy/search/searchtips.php>; Ethan Zuckerman, *A Technical Guide to Anonymous Blogging: Security Measures for Hiding Your Identity Online*, TECHSOUP, Dec. 15, 2006, <http://www.techsoup.org/learningcenter/internet/page6042.cfm>.

¹⁵¹ See discussion *supra* notes 61-64 and accompanying text.

¹⁵² See Proxy server, WIKIPEDIA, http://en.wikipedia.org/wiki/Proxy_server.

¹⁵³ For a variety of anonymous browsing options, see Anonymizer web site, <http://www.anonymizer.com/>. Another popular option is Privoxy, which strips out hidden identifying

itself may collect information concerning a user, and there have been instances of malicious proxy servers recording sensitive information, including users' unencrypted logins and passwords. Another anonymizing option is Tor, also known as the "Onion Router," a software product that first encrypts users' Internet traffic and then sends it through a series of randomly selected computers, thus obscuring the source and route of the data request.¹⁵⁴ Yet Tor, too, is not foolproof,¹⁵⁵ and it slows down browsing rendering it far less attractive for users.

While anonymizers and cookie management may be used to make traffic faceless across a broad range of Internet activities, TrackMeNot, a lightweight (41K) browser extension, has been invented by NYU law professor Helen Nissenbaum and researcher Daniel C. Howe specifically to address search engine privacy.¹⁵⁶ TrackMeNot periodically issues randomized search queries to leading search engines, thereby hiding users' actual search trails in a cloud of "ghost" queries.

The main problem with PETs is that few people use them. The reason might be the so called "blinking 12:00 syndrome,"¹⁵⁷ even apparently simple products are too complex for users who are not technologically savvy; or the fact that PETs slow down or complicate the browsing experience¹⁵⁸ and rarely provide complete protection. In any event, it appears that the technological arms race between PETs and privacy invasive technologies is markedly tilted toward the latter, necessitating legal intervention to protect users' rights.

b) Privacy policies and the limits of consent

information from Internet traffic, blocks advertisements and can be configured to manage cookies. See Privoxy, <http://www.privoxy.org/>.

¹⁵⁴ TOR, <http://tor.eff.org/>. See Tor (anonymity network), WIKIPEDIA, [http://en.wikipedia.org/wiki/Tor_\(anonymity_network\)](http://en.wikipedia.org/wiki/Tor_(anonymity_network)).

¹⁵⁵ See, e.g., Ryan Naraine, *Hacker Builds Tracking System to Nab Tor Pedophiles*, ZDNET, Mar. 6, 2007, <http://blogs.zdnet.com/security/?p=114>; Steven J. Murdoch & George Danezis, *Low-Cost Traffic Analysis of Tor*, paper presented at 2005 IEEE Symposium on Security and Privacy, Oakland CA, May 2005, available at <http://www.cl.cam.ac.uk/~sjm217/papers/oakland05torta.pdf>.

¹⁵⁶ TrackMeNot, <http://mrl.nyu.edu/~dhowe/trackmenot/>.

¹⁵⁷ Millions of VCRs across the world blink "12:00," because very few people can figure out how to program them.

¹⁵⁸ Instead of Google "remembering" your search queries, you have to enter and re-enter them over and over again.

In the absence of federal law governing the collection, retention, and use of search query logs, it has fallen upon search engines to craft their own privacy policies.¹⁵⁹ Google's privacy policy declares that "privacy is important" and promises to protect users' personally identifiable information.¹⁶⁰ Privacy policies are incorporated by reference into search engines' terms of use, which are service agreements implicitly agreed upon by use of the companies' services (*i.e.*, browse-wrap agreements).¹⁶¹

Reliance on industry self-regulation and user consent is ill advised in the context of search engine privacy.¹⁶² EPIC, for example, opposes the architecture of Google's privacy policy, which places information concerning user search query logs at a distance of four links from the company's homepage.¹⁶³ In addition, certain terms in Google's privacy policy may be interpreted in various ways.¹⁶⁴ For example, Google's Privacy Policy Highlights state: "We may also share information with third parties in limited circumstances, including when complying with legal process"¹⁶⁵ "Limited circumstances" is certainly a broad enough term to encompass a host of data transfers that are detrimental to user privacy. And what does "legal process" mean in this context? In *Gonzales v. Google*,¹⁶⁶ Yahoo, Microsoft, and AOL complied with the government's request for user search queries without requiring a search warrant. The term "legal process" has vastly different privacy implications depending on whether the standard is "probable cause" (Fourth Amendment standard for search warrants),¹⁶⁷ "specific and articulate facts giving reason to believe" (Stored Communications Act

¹⁵⁹ See Privacy Policies, *supra* note 79. See, e.g., Peter Fleischer, Eric Schmidt on Global Privacy Standards, PETER FLEISCHER BLOG, Sept. 19, 2007, <http://peterfleischer.blogspot.com/2007/09/eric-schmidt-on-global-privacy.html>, quoting Google CEO Eric Schmidt, who states that "it's important we develop new privacy rules to govern the increasingly transparent world which is emerging online today – and by new rules I don't automatically mean new laws. In my experience self regulation often works better than legislation (...)"

¹⁶⁰ Google Privacy Policy, *supra* note 35.

¹⁶¹ See Google Terms of Service (Apr. 16, 2007), <http://www.google.com/accounts/TOS>; Yahoo! Terms of Service, <http://info.yahoo.com/legal/us/yahoo/utos/utos-173.html>; Microsoft Service Agreement (May 2007), <http://tou.live.com/en-us/default.aspx>.

¹⁶² *But see* FTC, Online Behavioral Advertising, *supra* note 11.

¹⁶³ See discussion *supra* notes 131-132 and accompanying text.

¹⁶⁴ See analysis in Roger Clarke, Evaluation of Google's Privacy Statement against the Privacy Statement Template of 19 December 2005 (Xamax Consultancy Pty Ltd, 2005), available at <http://www.anu.edu.au/people/Roger.Clarke/DV/PST-Google.html>.

¹⁶⁵ See Google Privacy Policy Highlights, *supra* note 132.

¹⁶⁶ *Gonzales v. Google*, *supra* note 33.

¹⁶⁷ Wiretaps require a higher standard sometimes referred to as "probable cause plus." See Jayni Foley, *Are Google Searches Private? An Originalist Interpretation of the Fourth Amendment in Online Communication Cases*, 22 BERKELEY TECH. L.J. 447, 454 (2007); James X. Dempsey, *Digital Search & Seizure: Updating Privacy Protections to Keep Pace with Technology*, 865 PLI/PAT 505 (2006).

standard for access to certain stored records),¹⁶⁸ or simply "relevance" to an investigation (ECPA standard for pen registers).¹⁶⁹

Even if a privacy policy appears satisfactory, users should be aware that search engines typically reserve the right to modify and amend it unilaterally, at any time and without notice.¹⁷⁰ Although Google warrants that it "will not reduce your rights under this [Privacy] Policy without your explicit consent,"¹⁷¹ it may be difficult to decide whether a given policy modification "reduces" a user's right. And in any case, Google is unique in this respect among other leading search engines, which do not restrict their right to modify privacy policies.¹⁷² Moreover, the ability to modify privacy practices to reduce user rights may be concealed in apparently innocuous language. For example, Google claims it does not correlate users' e-mail and search records,¹⁷³ yet Google's Privacy Policy Highlights provide that: "Google collects personal information when you register for a Google service or otherwise voluntarily provide such information. We may combine personal information collected from you with information from other Google services or third parties to provide a better user experience, including customizing contents for you."¹⁷⁴ Thus, Google reserves the right to correlate users' e-mail and search data and it may do so under the current privacy policy without modifying its terms to "reduce" users' rights.

The fleeting nature of privacy protections under self imposed (and generally self serving) privacy policies, as well as companies' retention of the right to unilaterally modify their agreements, raise broader contractual issues related to browse-wrap agreements.¹⁷⁵ In a growing number of cases, customers have challenged the

¹⁶⁸ 18 U.S.C. § 2703(d). Title II of ECPA, the Stored Communications Act (SCA), is codified at 18 U.S.C. §§ 2701-2711.

¹⁶⁹ See federal wiretap law, codified as 18 U.S.C. § 2510-2522 (enacted as Title III of the Omnibus Crime Control & Safe Streets Act of 1968, Pub. L. 90-351, 82 Stat. 212).

¹⁷⁰ Google's chief privacy officer has recently announced the company would change its privacy policy to reflect "the additional data gathered through third party ad serving" (as a result of the DoubleClick transaction). See Peter Fleischer, Online Advertising: Privacy Issues are Important, but They Don't Belong in Merger Reviews, PETER FLEISCHER BLOG, Oct. 23, 2007, <http://peterfleischer.blogspot.com/2007/10/online-advertising-privacy-issues-are.html>.

¹⁷¹ Google Privacy Policy, *supra* note 132.

¹⁷² See, e.g., Microsoft Online Privacy Statement, <http://privacy.microsoft.com/en-us/fullnotice.aspx>.

¹⁷³ Goldberg, *supra* note 62, at 254.

¹⁷⁴ Google Privacy Policy Highlights, *supra* note 132.

¹⁷⁵ See generally Lemley, *supra* note 140; Rambarran & Hunt, *supra* note 140; Kunz et al., *supra* note 140; Sharon K. Sandeen, *The Sense and Nonsense of Web Site Terms of Use Agreements*, 26 HAMLINE L. REV. 499 (2003); Robert A. Hillman & Jeffery J. Rachlinski, *Standard-Form Contracting in the*

enforceability of browse-wrap agreements, based on insufficient notice, lack of consent, or unconscionable terms.¹⁷⁶ The fact that Google's privacy policy and terms of use do not appear on the search engine's homepage arguably casts a shadow over their enforceability.¹⁷⁷

Perhaps the greatest shortcoming of privacy policies is their grounding in user consent. After all, if users agree to their search queries being logged, retained, analyzed, and possibly disclosed, who is to complain? Yet too much is made of consent in this context. To be meaningful, consent must be informed and freely given. However, most users are probably not aware that their transactions with Google leave a personally identifiable, permanent track record, much less agree to such a result. Thus, user consent is not well informed, nor is it freely given.¹⁷⁸ Freely given consent assumes voluntary *choice*. However, since Google and its main competitors implement similar privacy practices,¹⁷⁹ search engine users do not have any real choice. The choice between using search engines under current policies and forgoing use altogether is no choice at all. Not using Google means not participating in today's information society. It is tantamount to never using a telephone, not riding a car, or residing in a secluded cabin in the woods. Google has become ubiquitous—practically a public utility. "Consent" is illusory where it is given (implicitly) by a captive audience, agreeing to contractual terms few users have ever read, which were unilaterally drafted to serve corporate interests. A privacy protection regime based on such consent provides no privacy protection at all.

Electronic Age, 77 N.Y.U. L. REV. 429 (2002); Comment, *Into Contract's Undiscovered Country: A Defense of Browse-Wrap Licenses*, 39 SAN DIEGO L. REV. 1363 (2002).

¹⁷⁶ *Specht v. Netscape Communications*, *supra* note 140; *see also* *Brazil v. Dell Inc.*, 2007 WL 2255296 (N.D.Cal. 2007); *Ticketmaster Corp. v. Tickets.com, Inc.*, 2003 WL 21406289 (C.D. Cal., 2003); *Mary E. DeFontes, et al. v. Dell Computers Corporation, et al.*, 2004 R.I. Super, Lexis 32 (Sup. Ct. R.I. 2004); *cf.* *Register.com, Inc. v. Verio, Inc.*, 356 F.3d 393 (2d Cir. 2004).

¹⁷⁷ *See* Goldberg, *supra* note 62, at n.33.

¹⁷⁸ *See, e.g.*, definition of "the data subject's consent" in Article 2(h) of the EU Data Protection Directive: "any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed."

¹⁷⁹ *See* Declan McCullagh & Elinor Mills, *How Search Engines Rate on Privacy*, CNET NEWS.COM, Aug. 13, 2007, http://www.news.com/How+search+engines+rate+on+privacy/2100-1029_3-6202068.html. There are conspicuous exceptions: Search engine Ask.com has chosen to differentiate itself from competitors by providing users with choice as to whether or not their search histories would be maintained. *See* Ryan Paul, *Ask.com Adds New "AskEraser" Search Privacy Feature*, ARSTECHNICA, Dec. 11, 2007, <http://arstechnica.com/news.ars/post/20071211-ask-com-adds-new-search-privacy-feature.html>.

c) Constitutional protection—and the lack thereof

Given that the government is a significant “client” of user data, the constitutional right to privacy is implicated in the collection and use by Google of search query logs. I argue below that constitutional doctrine for privacy protection in the U.S. is overly narrow and outdated, particularly in light of the market and technological developments of the past three decades.

The Fourth Amendment provides that “[t]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause”¹⁸⁰ In its landmark 1967 decision in *Katz*,¹⁸¹ the Supreme Court established a two part test to measure whether a person has a “reasonable expectation of privacy” that is entitled to constitutional protection. In his famous concurring opinion, Justice Harlan held that the appropriate inquiry is composed of a subjective prong, checking whether “a person [has] exhibited an actual (subjective) expectation of privacy” and an objective prong, verifying whether “the expectation [is] one that society is prepared to recognize as ‘reasonable.’”¹⁸²

The Supreme Court's decision in *Katz* became a fortress of privacy protection over the past 40 years. However, two Supreme Court decisions dating from the late 1970's destabilized one of the fortress foundations, eroding privacy protection in situations personally identifiable information is held by a third party, such as Google.¹⁸³ In the first case, *United States v. Miller*,¹⁸⁴ the Supreme Court held in 1976 that bank customers had no “reasonable expectation of privacy” in financial records held by their bank. The Court reasoned that a customer who voluntarily reveals her financial data to a third party (the bank) “assumes the risk” that the third party would pass the information on to the government.¹⁸⁵ The Court reached its conclusion notwithstanding the fact that “the information is revealed on the assumption that it

¹⁸⁰ U.S. Const. amend. IV.

¹⁸¹ *Katz v. United States*, *supra* note 139.

¹⁸² *Id.* at 361 (Harlan, J., concurring).

¹⁸³ See generally Orin S. Kerr, *The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution*, 102 MICH. L. REV. 801 (2004).

¹⁸⁴ *United States v. Miller*, 425 U.S. 435 (1976) [hereinafter *Miller*].

¹⁸⁵ *Id.* at 443.

will be used only for a limited purpose and the confidence placed in the third party will not be betrayed."¹⁸⁶ The Court's rationale follows the proverb attributed to Benjamin Franklin, "three may keep a secret, if two of them are dead."¹⁸⁷ Once the "secret" is out, even if revealed in confidence as part of a banker-customer relationship, the customer can expect no privacy.¹⁸⁸

Miller's assumption of risk analysis was extended in 1979 in *Smith v. Maryland*, which held that telephone users lack a reasonable expectation of privacy in the numbers they dial.¹⁸⁹ Once again, the Court held that "a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties."¹⁹⁰ Distinguishing *Katz*, the Court held that the pen registers at issue in *Smith*, which capture numbers dialed, "do not acquire the 'contents' of communications."¹⁹¹ Hence, Fourth Amendment protection continues to apply insofar as personally identifiable information held by a third party includes the "contents" of a communication. Constitutional protection is lost in situations no contents are involved.¹⁹²

Courts have extended the *Miller* and *Smith* "assumption of risk" paradigm to a wide variety of circumstances involving the disclosure of personally identifiable information to trusted third parties, who then proceed to transfer the data to the government.¹⁹³ In a string of cases, courts authorized warrantless government access to ISP customer records, including names, screen names, addresses, birthdates, and

¹⁸⁶ *Id.*

¹⁸⁷ Benjamin Franklin, WIKIQUOTE, http://en.wikiquote.org/wiki/Benjamin_Franklin.

¹⁸⁸ See Patricia L. Bellia, *Surveillance Law Through Cyberlaw's Lens*, 72 GEO. WASH. L. REV. 1375, 1402 (2004). This approach is rejected outright by the English tort of breach of confidentiality. See discussion *infra* notes 277-295 and accompanying text.

¹⁸⁹ *Smith v. Maryland*, 442 U.S. 735, 743-44 (1979) [hereinafter *Smith*].

¹⁹⁰ *Id.*

¹⁹¹ *Id.* at 747-48.

¹⁹² Justice Stewart, dissenting in *Smith*, questions the sharp contents-non contents distinction. *Id.* at 748. Analogizing electronic communications to postal mail, Orin Kerr refers to the distinction as one between "contents" (constitutionally protected) and "envelope" (not constitutionally protected). Orin S. Kerr, *Internet Surveillance Law After the USA Patriot Act: The Big Brother That Isn't*, 97 NW. U. L. REV. 607, 611-16 (2003) [hereinafter *Kerr, Patriot Act*]; cf. Daniel Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1264, 1288 (2004).

¹⁹³ See, e.g., *United States v. Jacobsen*, 466 U.S. 109 (1984) (a package of drugs sent via Federal Express); *SEC v. Jerry T. O'Brien, Inc.*, 467 U.S. 735 (1984) (financial records held by broker-dealer); *California v. Greenwood*, 486 U.S. 35 (1988) (garbage bags left at the curb); *United States v. Phibbs*, 999 F.2d 1053 (6th Cir. 1993) (credit card statements and phone records).

passwords.¹⁹⁴ The Ninth Circuit Court of Appeals recently applied the *Miller* doctrine to a government request for ISP subscriber information, including not only registration details but also to/from addresses of e-mail messages, IP addresses of websites visited, and the total amount of data transmitted to or from an account.¹⁹⁵ The Court did set aside discussion of government access to a list of URLs visited by ISP subscribers, noting that "[s]urveillance techniques that enable the government to determine not only the IP addresses that a person accesses but also the uniform resource locators ('URL') of the pages visited might be more constitutionally problematic."¹⁹⁶ Hence, *Smith*, with its "assumption of risk" analysis, applies to government access to non-contents information, whereas *Katz* continues to hold for communication contents.

Are user search query logs entitled to Fourth Amendment protection?¹⁹⁷ Under the "assumption of risk" doctrine, users may be held to have relinquished any reasonable expectation of privacy in search queries once they have typed them into a Google search screen. Such users have "voluntarily turned over information to a third party" and are therefore arguably not entitled to Fourth Amendment protection. Alternatively, search queries may be characterized as the *contents* of a communication, reasserting constitutional protection under the *Smith* exception.¹⁹⁸

Numerous commentators have criticized the *Miller* and *Smith* "assumption of risk" doctrine.¹⁹⁹ One basic problem emanates from the *Katz* two-pronged test itself, since

¹⁹⁴ See *Guest v. Leis*, 255 F.3d 325 (6th Cir. 2001); *United States v. Kennedy*, 81 F. Supp. 2d 1103 (D. Kan. 2000); *United States v. Hambrick*, 55 F.Supp.2d 504 (4th Cir. 2000); *United States v. Cox*, 190 F. Supp.2d 330, 332 (N.D.N.Y. 2002); *United States v. Bach*, 310 F.3d 1063 (8th Cir. 2002).

¹⁹⁵ *United States v. Forrester*, ___ F.3d ___, 2007 WL 2120271 (9th Cir. 2007). *But see* Steven Warshak v. *United States*, ___ F. 3d. ___, 2007 WL 1730094 (6th Cir. 2007), holding Fourth Amendment protection does apply to the contents of e-mail stored on an ISP's server. *See also* *United States v. Charbonneau*, 979 F. Supp. 1177, 1184 (S.D. Ohio 1997); *Smyth v. Pillsbury Co.*, 914 F. Supp. 97, 101 (E.D. Pa. 1996).

¹⁹⁶ *United States v. Forrester*, *supra* note 194, at *6 n.6.

¹⁹⁷ See Matthew D. Lawless, *The Third Party Doctrine Redux: Internet Search Records and the Case for a "Crazy Quilt" of Fourth Amendment Protection*, 2007 UCLA J. L. & TECH. 1.

¹⁹⁸ The question of search queries as contents of communications is addressed *infra* notes 237-242 and accompanying text. I concentrate here on the shortcomings of the constitutional doctrine.

¹⁹⁹ See, e.g., Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1135 (2002); Bellia, *supra* note 188, at 1397-1412; Gavin Skok, *Establishing a Legitimate Expectation of Privacy in Clickstream Data*, 6 MICH. TELECOMM. TECH. L. REV. 61, 78 (2000); Ric Simmons, *From Katz to Kyllo: A Blueprint for Adapting the Fourth Amendment to Twenty-First Century Technologies*, 53 HASTINGS L.J. 1303 (2002); Raymond Shih Ray Ku, *The Founders'*

the greater the expectation one has of being subject to surveillance, the less constitutional protection one has. The Court in *Smith* was well aware of this shortcoming, stating that "if the Government were suddenly to announce on nationwide television that all homes henceforth would be subject to warrantless entry, individuals thereafter might not in fact entertain any actual expectation of privacy regarding their homes, papers, and effects."²⁰⁰ Hence, the *Katz* test as applied in *Miller* and *Smith* becomes a self-fulfilling paranoid prophecy, where one's suspicion of government surveillance strips one of constitutional protection. In other words, what you expect is what you get, and you are probably right to expect the worst. It is a constitutional race to the bottom, where the least protective expectation sets the standard for citizens' privacy rights.²⁰¹

Moreover, in his dissent in *Smith*, Justice Marshall states that it is idle to speak of voluntary "assumption of risk" where, as a practical matter, individuals have no realistic choice. Justice Marshall observes that "unless a person is prepared to forgo use of what for many has become a personal or professional necessity, he cannot help but accept the risk of surveillance."²⁰² This observation reverberates in the search engine context. Google users have no plausible alternative to using the leading Internet search engine or one of its competitors which also apply similar privacy policies. "Assumption of risk" analysis is misleading in this context. Users do not convey personally identifiable information to Google because they have chosen to do so after careful deliberation and cost-benefit analysis. They do so simply because they have to.

An additional problem concerns the scope of constitutional protection. The Fourth Amendment protects individuals from *government* search and seizure. It curtails the investigatory power of government officials. It does not apply to the private sector and therefore does not limit Google from collecting, using, retaining, or transferring

Privacy: The Fourth Amendment and the Power of Technological Surveillance, 86 MINN. L. REV. 1325 (2002).

²⁰⁰ *Smith*, *supra* note 189, at 740 n.5.

²⁰¹ Commentators note that the *Miller* and *Smith* application of the *Katz* formula is fatally flawed, because it treats the objective prong of the *Katz* test as a positive rather than normative question. See Brenner & Clarke, *supra* note 146, at 247-50; Lawless, *supra* note 197 (advocating an "operational realities test").

²⁰² *Smith*, *supra* note 189, at 749-50.

data to corporate third parties.²⁰³ The private sector, so the theory goes, will self regulate to reach an efficient equilibrium based on consumers' privacy preferences and companies' information needs.²⁰⁴

Yet commentators question both the fairness and efficiency of a market based solution in the context of privacy.²⁰⁵ They point out that privacy invasions typically cause many small, individualized injuries that might be difficult to vindicate through litigation.²⁰⁶ They argue that in information transactions, consumers are hampered by psychological limitations, which Michael Froomkin dubbed "privacy myopia," causing them to "sell their privacy bit by bit for frequent flyer miles."²⁰⁷ In her dissenting opinion in the FTC decision upholding the Google/DoubleClick transaction, Commissioner Jones Harbour expresses her skepticism concerning the market equilibrium, asserting that "Congress ultimately will need to decide whether a market-oriented solution is workable, given the disconnect between the financial incentives of advertisers and publishers (*i.e.*, to exploit data) and the privacy incentives of some consumers (*i.e.*, to protect data)."²⁰⁸

In contrast to the narrow scope of constitutional privacy protection in the U.S., European constitutional law has recognized privacy as a fundamental right in instruments ranging from the 1950 ECHR²⁰⁹ to the 2007 Charter of Fundamental Rights of the European Union.²¹⁰ In Europe, not only privacy but also data protection

²⁰³ See *United States v. Jacobsen*, *supra* note 193, at 113, holding that "[the Fourth Amendment] is wholly inapplicable to a search or seizure, even an unreasonable one, effected by a private individual".

²⁰⁴ See Peter P. Swire, *Markets, Self-Regulation and Government Enforcement in the Protection of Personal Information*, in U.S. DEP'T OF COMMERCE, *PRIVACY AND SELF-REGULATION IN THE INFORMATION AGE* (1997), available at <http://www.ntia.doc.gov/reports/privacy/selfreg1.htm>. The classic law and economics analyses of privacy are Posner, *The Right of Privacy*, *supra* note 27 ; and George J. Stigler, *An Introduction to Privacy in Economics and Politics*, 9 J. LEGAL STUD. 623 (1980). But see Richard S. Murphy, *Property Rights in Personal Information: An Economic Defense of Privacy*, 84 GEO. L.J. 2381 (1996); and Cohen, *supra* note 143. See generally *supra* note 27.

²⁰⁵ See Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2055, 2076-94 (2004).

²⁰⁶ Richards, *supra* note 111, at 1099.

²⁰⁷ A. Michael Froomkin, *The Death of Privacy*, 52 STAN. L. REV. 1461, 1502 (2000).

²⁰⁸ Dissenting Statement of Commissioner Pamela Jones Harbour, *supra* note 11, at 11-12.

²⁰⁹ Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (1950), 213 UN Treaty Ser. 221 (1955) ("ECHR") provides: "Everyone has the right to respect for his private and family life, his home and his correspondence."

²¹⁰ Article 7 of the Charter provides: "Everyone has the right to respect for his or her private and family life, home and communications." *Supra* note 142. See also Article II-67 of the Treaty establishing a Constitution for Europe, Official Journal 2004/C 310/01 (Dec. 16, 2004) (the "Constitutional Treaty").

is a constitutional right.²¹¹ Counter to American constitutional privacy, which is based on *liberty* and individual freedom from *government* intervention²¹² (particularly in one's home²¹³ or family),²¹⁴ the European constitutional approach is grounded on the underlying value of *human dignity*.²¹⁵ The fundamental value of human dignity is not restricted to interaction with the government; rather it applies in equal force to the private sector. Dignitary harms, such as unlawful discrimination or invasion of privacy, may be inflicted not only by the government but also by individuals and businesses.²¹⁶

Nowhere is the difference between the U.S. and European constitutional frameworks starker than in the context of the *Miller* and *Smith* assumption of risk doctrine. Under *Miller* and *Smith*, constitutional analysis ends if personal information is voluntarily turned over to a third party. Conversely, in Europe, this is specifically the point constitutional analysis begins. Indeed, the whole thrust of European data protection law, which affords individuals control over their personally identifiable information, pertains to the fair and lawful use of information by *third parties*, including government and private entities alike. The EU Data Protection Directive requires Member States to implement an intricate statutory framework governing all aspects of collection, use and transfer of personally identifiable information, and create independent regulatory authorities to enforce the law.

²¹¹ Article 8 of the Charter provides: "Everyone has the right to the protection of personal data concerning him or her." *Id.*; see also Article II-68(1) of the Constitutional Treaty.

²¹² See JEFFREY ROSEN, *THE UNWANTED GAZE: THE DESTRUCTION OF PRIVACY IN AMERICA* 5 (2000); THOMAS PAINE, *COMMON SENSE* 65 (Penguin Classics, Penguin Books 1986) (1776).

²¹³ See *Boyd v. United States*, 116 U.S. 616 (1886). See also Note, *The Right to Privacy in Nineteenth Century America*, 94 HARV. L. REV. 1892 (1981).

²¹⁴ See, e.g., *Griswold v. Connecticut*, 381 U.S. 479 (1965); *Roe v. Wade*, 410 U.S. 113 (1973); *Lawrence v. Texas*, 539 U.S. 558 (2003). Justice Kennedy begins his decision in *Lawrence* writing that "[l]iberty protects the person from unwarranted government intrusions into a dwelling or other private places. In our tradition the State is not omnipresent in the home." *Id.* at 562 (emphasis added). See Jacob Strahilevitz, *Consent, Aesthetics, and the Boundaries of Sexual Privacy after Lawrence v. Texas*, 54 DEPAUL L. REV. 671 (2005); Laurence H. Tribe, *Lawrence v. Texas: The "Fundamental Right" that Dare Not Speak Its Name*, 117 HARV. L. REV. 1893 (2004).

²¹⁵ For a fascinating account see James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151 (2004); see also Robert C. Post, *Three Concepts of Privacy*, 89 GEO. L.J. 2087 (2001); Matthew W. Finkin, *Menschenbild: The Conception of the Employee as a Person in Western Law*, 23 COMP. LAB. L. & POL'Y REV. 577 (2002); EDWARD J. EBERLE, *DIGNITY AND LIBERTY: CONSTITUTIONAL VISIONS IN GERMANY AND THE UNITED STATES* (2002).

²¹⁶ See, e.g., Paul M. Schwartz, *German and U.S. Telecommunications Privacy Law: Legal Regulation of Domestic Law Enforcement Surveillance*, 54 HASTINGS L.J. 751 (2003).

Consequently, while the collection, use, and retention of search query logs by search engines do not raise a constitutional issue in the U.S., at least insofar as the government is not involved, they fall squarely within the ambit of European privacy law. In this context too, search engine users' privacy is inadequately protected in the U.S.²¹⁷

d) Statutory protection—a cobweb full of holes

Enacted in 1986 as an amendment to the federal wiretap statute,²¹⁸ the ECPA²¹⁹ is a highly complex piece of legislation.²²⁰ Originally intended to adapt federal privacy protections to new and emerging technologies, ECPA has itself become technologically outdated.²²¹ Most troubling, it provides a surprisingly low level of protection for contents and non-contents of communications so long as these are not intercepted by the government in mid-traffic.²²²

ECPA consists of three statutes, the Wiretap Act,²²³ the Pen Register Act,²²⁴ and the Stored Communications Act (SCA).²²⁵ The SCA, which applies to communications

²¹⁷ See Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531 (2006), stating "[c]ommon wisdom teaches that the Fourth Amendment exists to protect privacy – and that it does a miserable job of it."

²¹⁸ Title III of the Omnibus Crime Control and Safe Streets Act of 1968.

²¹⁹ *Supra* note 102.

²²⁰ Orin Kerr promises that "[a]lthough the rules found in § 2702 and § 2703 can seem maddeningly complicated at first, they prove surprisingly straightforward in practice." Orin S. Kerr, *A User's Guide to the Stored Communications Act, and a Legislator's Guide to Amending it*, 72 GEO. WASH. L. REV. 1208, 1222 (2004) [hereinafter, Kerr, SCA]. Surveillance powers under the ECPA were expanded by the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act of 2001, Pub. L. No. 107-56, 115 Stat. 272. See generally Marc Rotenberg, *Foreword: Privacy and Secrecy after September 11*, 86 MINN. L. REV. 1115 (2002); Sharon H. Rackow, *Comment, How the USA PATRIOT Act Will Permit Governmental Infringement upon the Privacy of Americans in the Name of "Intelligence" Investigations*, 150 U. PA. L. REV. 1651 (2002) (criticizing the impact of the Patriot Act on civil rights); cf. Note, *The Patriot Act's Impact on the Government's Ability to Conduct Electronic Surveillance of Ongoing Domestic Communications*, 52 DUKE L.J. 179 (2002) (arguing that the Patriot Act will have less impact on privacy than feared). For a discussion of state surveillance statutes, see Charles H. Kennedy & Peter P. Swire, *State Wiretaps and Electronic Surveillance After September 11*, 54 HASTINGS L.J. 971 (2003).

²²¹ See Dempsey, *supra* note 167, at 521; Bellia, *supra* note 188, at 1423-24, stating that "under the government's approach, seemingly trivial choices by a subscriber among different technical options a service provider offers have tremendous legal consequences." Kerr, SCA, *supra* note 220, at 1216-17.

²²² See discussion *infra* notes 243-255 and accompanying text.

²²³ 18 U.S.C. §§ 2510-2522 (2000 & Supp. II 2002).

²²⁴ 18 U.S.C. §§ 3121-3127 (2000 & Supp. II 2002).

²²⁵ 18 U.S.C. §§ 2701-2711 (2000 & Supp. II 2002).

stored by third parties, is most relevant to search engine users' privacy.²²⁶ The level of privacy protection set forth by the SCA depends on whether we are dealing with (a) voluntary or compelled disclosure of information; (b) by an "electronic communication service" or a "remote computing service"; (c) that offers services "to the public" or not; (d) of the "contents of a communication" or of non-contents; and (e) of communications that are in "electronic storage" or in transit.

The SCA applies to two types of communications service providers: providers of "electronic communication service" (ECS) and providers of "remote computing service" (RCS). An ECS means "any service which provides to users thereof the ability to send or receive wire or electronic communications."²²⁷ An RCS means "the provision to the public of computer storage or processing services by means of an electronic communications system."²²⁸ The RCS provisions were originally conceived to cover outsourced data processing services,²²⁹ yet are currently applicable to online search engines.²³⁰ Much like in traditional data processing, a user transmits data (a search query) to Google via an electronic communication; Google processes the data according to its proprietary algorithm and sends the result (a list of hyperlinks) back to the user. Substantially, Google maintains a log of its communications with users, which is precisely the aspect of RCS that SCA drafters were concerned with.

A fundamental distinction in the SCA is that between voluntary disclosure: a service provider chooses to disclose information to the government or a third party,²³¹ and compelled disclosure: the government uses the law to force disclosure.²³² The rules concerning voluntary disclosure revolve around the distinction between contents and non-contents information and between government and non-government recipients.²³³

²²⁶ For a good exposition see Kerr, SCA, *supra* note 220; Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1701 (2004).

²²⁷ 18 U.S.C. § 2510(15).

²²⁸ 18 U.S.C. § 2711(2).

²²⁹ See S.Rep. No. 99-541 (1986), reprinted in 1986 U.S.C.C.A.N. 3555, 3564.

²³⁰ As a provider of webmail services, Google also appears to fit the definition of ECS. See Goldberg, *supra* note 62, at 267-69.

²³¹ 18 U.S.C. § 2702.

²³² 18 U.S.C. § 2703.

²³³ Another critical distinction in this context is between providers that offer services to the public and those that do not. The SCA's voluntary disclosure limitations apply strictly to providers that offer services to the public. 18 U.S.C. § 2702. Google clearly belongs to this category.

Voluntary disclosure of communication contents is prohibited, whether the information is disclosed to a government or non-government entity, subject to a list of exceptions specified in Section 2702(b).²³⁴ Service providers are free to disclose non-contents information to non-government entities,²³⁵ whereas disclosure to a government entity, even of non-contents, is banned.²³⁶ Determining whether a transferee is a government or non-government entity is straightforward. I therefore turn to the question of whether the data disclosed, in our case user search queries, constitute contents or non-contents information.

The definition of "contents" applicable throughout the SCA appears in the Wiretap Act.²³⁷ Section 2510(8) provides: "'contents', when used with respect to any wire, oral, or electronic communication, includes any information concerning the substance, purport, or meaning of that communication."²³⁸ Kerr simplifies this rather cryptic definition, explaining that the contents of a communication consist of information that a person wishes to share with or communicate to another person, whereas non-contents (sometimes referred to as "envelope" information) is information about the communication that the network uses to deliver and process the contents.²³⁹ In simple terms, contents are what you write in a letter and non-contents are what you write on an envelope.

However, in the online context, the distinction becomes blurred.²⁴⁰ Does a search query constitute "contents of an electronic communication"? As discussed above, this determination is critical not only for SCA analysis but also for Fourth Amendment purposes.²⁴¹ No court has yet addressed the question squarely.²⁴² On the one hand,

²³⁴ 18 U.S.C. § 2702(a).

²³⁵ 18 U.S.C. § 2702(c)(6).

²³⁶ 18 U.S.C. § 2702(a).

²³⁷ 18 U.S.C. § 2711(1), providing that "the terms defined in section 2510 of this title have, respectively, the definitions given such terms in that section."

²³⁸ 18 U.S.C. § 2510(8). Non-contents information is labeled by the SCA "a record or other information pertaining to a subscriber to or customer of such service (not including the contents of communications)." 18 U.S.C. § 2703(c)(1).

²³⁹ See Kerr, Patriot Act, *supra* note 192, at 611-16.

²⁴⁰ Kerr argues that "the legal categories of 'contents' and 'addressing information' are straightforward in the case of human-to-human communications, but can be quite murky when considering human-to-computer communications." Kerr, Patriot Act, *supra* note 192, at 645-46; see also *United States v. Forrester*, *supra* note 195.

²⁴¹ See discussion *supra* notes 243-255 and accompanying text.

search queries do appear to contain "substance, purport, or meaning," because they convey a person's interests, passions, needs or fears—information that goes well beyond routing or addressing data. On the other hand, a search query may be regarded as a signpost pointing the search engine in the direction of the required contents. Under this line of reasoning, the contents are the hyperlinks listed in Google's search result; whereas the search query is merely a non-contents tool used to obtain the contents.

In my opinion, search queries constitute "contents of communications." The information conveyed in search query logs is far more revealing than typical "envelope" addressing data, such as telephone numbers or e-mail to/from fields. It cuts to the very core of a person's thoughts and feelings, telling much about what she wants to buy or sell; where she plans to go on vacation; what kind of job, husband, music, or shoes she might be interested in; whom she adores and which diseases she abhors; what her political opinions are; and which religious faith she subscribes to. Such information, while not the contents of a communication between a user and another person, is most certainly the contents of a communication between a user and the Google server. And if the 1980's featured extension of federal wiretapping laws to electronic communication networks, the natural progression for the new millennium is to extend protection of communication contents to the contents of communications between man and machine.

Assuming that search queries constitute contents of a communication and that Google is an RCS provider, voluntary disclosure by Google of user search queries is prohibited, regardless of whether such disclosure is made to a government or non-government entity. Section 2702(b) sets forth seven exceptions to this rule.²⁴³ Most pertinent to the disclosure of search query logs by Google are the exceptions in

²⁴² *But see* United States v. Forrester, *supra* note 195, at *6 n.6 (discussing the proper classification of a list of URLs). *See also In re* United States for an Order Authorizing the Use of a Pen Register & Trap, 396 F.Supp.2d 45, 49 (D. Mass. 2005), holding that:

There is the issue of search terms. A user may visit the Google site. Presumably the pen register would capture the IP address for that site. However, if the user then enters a search phrase, that search phrase would appear in the URL after the first forward slash. This would reveal contents . . . The "substance" and "meaning" of the communication is that the user is conducting a search for information on a particular topic.

²⁴³ Section 2702(c) sets similar exceptions for disclosure of non-contents information, the major difference being that non-contents can be disclosed to non-government entities without restriction. 18 U.S.C. § 2702(c)(6).

Sections 2702(b)(2) and 2702(b)(3).²⁴⁴ Under Section 2702(b)(3), a service provider may divulge the contents of a communication to a government or non-government entity "with the lawful consent of the . . . subscriber."²⁴⁵ Google may rely on user consent to its privacy policy to justify voluntary disclosure under Section 2702(b)(3). I argued above, however, that user consent is neither informed nor freely given, and is at best tenuously inferred from use of the Google site.²⁴⁶ It is therefore an unacceptable basis for disclosure of communication contents under the SCA.

Section 2702(b)(2) sanctions disclosure of information to the government "as otherwise authorized in Section 2517, 2511(2)(a), or 2703 of this title."²⁴⁷ The relevant provision in our case is Section 2703, which governs compelled disclosure of communications data to the government.²⁴⁸ Thus, the standards for government compelled disclosures become intertwined with those applicable to voluntary disclosure of communication contents to a government entity. The main distinction drawn by Section 2703 is between disclosure of contents (Section 2703(a) and (b)) and non-contents (Section 2703(c)). For disclosure of contents, Section 2703 further distinguishes between contents in "electronic storage" for 180 days or less (Section 2703(a)); in "electronic storage" for more than 180 days (Section 2703(b)); or permanently held by an RCS provider (Section 2703(b)). Under Section 2703, a full search warrant is required only to access un-retrieved and unopened e-mail messages and other temporarily stored files held *pending transmission* for 180 days or less.²⁴⁹

Section 2703(b) establishes the requirements the government must meet to compel disclosure of the contents of communications (such as user search query logs) held by an RCS provider (such as Google). Under Section 2703(b), the government may

²⁴⁴ 18 U.S.C. §§ 2702(b)(2)–(b)(3).

²⁴⁵ 18 U.S.C. § 2702(b)(3).

²⁴⁶ See discussion *supra* notes 178-179 and accompanying text.

²⁴⁷ 18 U.S.C. § 2702(b)(3).

²⁴⁸ 18 U.S.C. § 2703.

²⁴⁹ The statute defines "electronic storage" as "any temporary, intermediate storage of a wire or electronic communication incidental to the electronic transmission thereof," and "any storage of such communication by an electronic communication service for purposes of backup protection of such communication." 18 U.S.C. § 2510. The prevailing government approach is that only messages that have not yet been opened or retrieved by a customer are in "electronic storage." Once a message is opened, its storage is no longer "incidental to the electronic transmission thereof." Such a message is therefore "exiled" from the rather strict privacy protections of Section 2703(a) to the looser standards of Section 2703(b). See COMPUTER CRIME AND INTELLECTUAL PROPERTY SECTION, U.S. DEP'T OF JUSTICE, MANUAL ON SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS (2001). Cf. Theofel v. Farey-Jones, 359 F.3d 1066 (9th Cir. 2004).

compel an RCS provider to disclose the contents of a communication using one of five tools: (a) a criminal search warrant; (b) an administrative subpoena; (c) a grand jury subpoena; (d) a trial subpoena; or (e) a court order issued under Section 2703(d). A court order issued under Section 2703(d) is not equivalent to a search warrant, which requires a showing of "probable cause." Instead, a court may issue a Section 2703(d) order if the government offers "specific and articulable facts showing reasonable grounds to believe" that the communications sought are "relevant and material" to an ongoing criminal investigation.²⁵⁰

By settling for a subpoena or Section 2703(d) order rather than a full search warrant, SCA drafters presume that a user retains no "reasonable expectation of privacy" in the contents of communications stored by an RCS provider.²⁵¹ Consequently, Kerr notes that "[t]he most obvious problem with the current version of the SCA is the surprisingly weak protection the statute affords to compelled contents of communications under the traditional understanding of ECS and RCS."²⁵² This is evident particularly in the case of subpoenas, which are issued with no prior judicial approval and are enforced on a mere showing of relevance. Worse yet, when a subpoena is served on the affected individual herself, she at least has notice and an opportunity to file a motion to quash or modify.²⁵³ But where a subpoena is served on a disinterested third party, such as Google, that third party typically has little or no reason to object.²⁵⁴ Consequently, statutory safeguards under the ECPA mirror the weak constitutional protection and perpetuate the vulnerability of search engine users' privacy rights.²⁵⁵

To sum up, I argue that users' search query logs should be classified as "contents of communications" and therefore afforded statutory protection against voluntary

²⁵⁰ 18 U.S.C. § 2703(d).

²⁵¹ Bellia, *supra* note 188, at 1422.

²⁵² Kerr, SCA, *supra* note 220, at 1233; *see also* Note, *Email Privacy after United States v. Councilman: Legislative Options for Amending ECPA*, 21 BERKELEY TECH. L.J. 499 (2006).

²⁵³ *See* Christopher Slobogin, *Subpoenas and Privacy*, 54 DEPAUL L. REV. 805 (2005).

²⁵⁴ In addition, notice to the person whose privacy is affected may be deferred for long periods of time under Section 2705. 18 U.S.C. § 2705.

²⁵⁵ For the inadequacy of statutory protection, *see also* Peter Swire, *Katz Is Dead. Long Live Katz*, 102 MICH. L. REV. 904 (2004); Deirdre K. Mulligan, *Reasonable Expectations in Electronic Communications: A Critical Perspective on the Electronic Communications Privacy Act*, 72 GEO. WASH. L. REV. 1557 (2004).

disclosure to non-government third parties. Even then, however, the statutory protection under the dense provisions of the SCA leaves a lot to be desired.

e) Data retention v. data protection

Online privacy is increasingly clashing with a formidable enemy in the shape of data retention requirements promoted by national security and law enforcement agencies worldwide. Ostensibly, the best privacy solution for user search query logs would be their immediate deletion. The mere existence of the so called Database of Intentions constitutes a magnet for government investigators, private litigants, data thieves, and commercial parties.²⁵⁶ But are search engines even *allowed* to delete users' search query logs?

Governments increasingly impose data retention requirements to make online activity traceable by law enforcement agencies.²⁵⁷ Data retention laws compel telecom companies and ISPs to collect and store customer data. Typically, retention is restricted to non-contents data, such as subscriber registration information, traffic, and location data.

In the EU, the legal and technical differences between data retention standards across Member States posed difficult dilemmas for service providers. In addition, data retention requirements apparently conflicted with data privacy laws.²⁵⁸ The need for European-wide harmonization and clarification of the interplay between privacy and data retention has led to the adoption of a Data Retention Directive in March 2006.²⁵⁹

²⁵⁶ See 28th International Data Protection and Privacy Commissioners' Conference, London, UK, Resolution on Privacy Protection and Search Engines (Nov. 2-3, 2006).

²⁵⁷ See, e.g., Regulation of Investigatory Powers Act 2000, 2000 Chap. 23, Part I, Chapter II; Anti-terrorism, Crime and Security Act 2001, 2001 Chap. 24, Part 11 (UK); see also Home Office, Retention of Communications Data Under Part 11: Anti-Terrorism, Crime & Security Act 2001, Voluntary Code of Practice, available at <http://www.opsi.gov.uk/si/si2003/draft/5b.pdf>. Cf. EDRI-gram, *Italy Decrees Data Retention until 31 December 2007* (Aug. 10, 2005), <http://www.edri.org/edrigram/number3.16/Italy>; EDRI-gram, *Telecom Data to be Retained for One Year in France*, <http://www.edri.org/edrigram/number4.6/franceretantion>.

²⁵⁸ Article 6(e) of the EU Data Protection Directive.

²⁵⁹ Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, L 105/54, Official Journal (April 13, 2006) [hereinafter Data Retention Directive]. See generally Francesca Bignami, *Privacy and Law Enforcement in the European Union: The Data Retention Directive*, 8 CHL J. INT'L L. 233 (2007).

Under the Data Retention Directive, providers of "electronic communications services" are required to store traffic data related to telephone calls, e-mails, and online activity for a period of six months to two years, depending on the law in each Member State. Traffic data include the identities of a customer's correspondents; the date, time, and duration of phone calls, VoIP calls,²⁶⁰ or e-mail messages; and the location of the device used for a communication; but not the contents of a communication.²⁶¹

Although the U.S. has not yet followed the European lead, adoption of data retention legislation has been advocated by prominent politicians, including former Attorney General Alberto Gonzales.²⁶² Unlike European companies that are hemmed-in by data protection laws, American services providers usually retain users' traffic data for commercial reasons even without being required to do so. Moreover, the Sarbanes-Oxley Act,²⁶³ tax laws, and accounting regulations reflect mounting data retention requirements applicable to U.S. companies even without a special data retention statute. Finally, U.S. authorities benefit from a related, if less sweeping law enforcement tool, known as "data preservation."²⁶⁴ Data preservation is set forth in the Electronic Communication Transactional Records Act of 1996,²⁶⁵ which requires ISPs to retain any "record" in their possession for 90 days "upon the request of a governmental entity." However, counter to European data retention, which applies across the board, American data preservation is targeted at the traffic data of a specific individual already under investigation.²⁶⁶

²⁶⁰ Voice over IP, WIKIPEDIA, <http://en.wikipedia.org/wiki/Voip>.

²⁶¹ Articles 2, 5 of the Data Retention Directive.

²⁶² See Declan McCullagh, *GOP revives ISP-tracking legislation*, CNET NEWS.COM, May 30, 2006, http://news.com.com/GOP+revives+ISP-tracking+legislation/2100-1028_3-6156948.html; Declan McCullagh, *Terrorism Invoked in ISP Snooping Proposal*, CNET NEWS.COM, May 30, 2006, http://news.com.com/2100-1028_3-6078229.html.

²⁶³ Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (codified in scattered sections of 11, 15, 18, 28, and 29 U.S.C.).

²⁶⁴ The concept of data preservation exists in Europe. See Articles 16-17 of Council of Europe Convention on Cybercrime, Europ. T.S. No. 185 (Nov. 23, 2001), available at <http://conventions.coe.int/Treaty/EN/Treaties/Html/185.htm>.

²⁶⁵ 18 U.S.C. §§ 2701-2712.

²⁶⁶ See Catherine Crump, *Data Retention: Privacy, Anonymity, and Accountability Online*, 56 STAN. L. REV. 191, 194 (2003).

A U.S. District Court in California recently implemented an expansive approach to data preservation in a case involving TorrentSpy,²⁶⁷ a BitTorrent²⁶⁸ indexing web site. The Motion Picture Association of America (MPAA) sued TorrentSpy in February 2006, accusing the web site of facilitating illegal downloads of copyrighted materials.²⁶⁹ As part of the discovery process, the MPAA filed a motion to compel TorrentSpy to preserve and produce server log data, including IP addresses of users seeking "dot-torrent" files. TorrentSpy, which operated its web site from servers based in the Netherlands, pointed out that it had never retained server logs because the information was not necessary for its business, and since data retention was restricted by Dutch privacy law. TorrentSpy claimed that requiring it to log user data would force it to act in a manner contrary to its privacy policy. The court granted the MPAA's motion, holding that since the data sought by the MPAA were at least temporarily available, they were covered by the rules of evidence and must therefore be logged and turned over to the plaintiff.²⁷⁰ The Court's ruling is much broader than a preservation order, because it is not targeted at a specific suspect and goes so far as to require the web site to store data not ordinarily kept on its servers.²⁷¹

The tension between privacy and data retention requirements featured prominently in Google's latest investigation by European regulators.²⁷² European regulators claimed that the 18 to 24 month period during which Google retains search query logs is excessive.²⁷³ Google responded by shortening the storage period to 18 months, but pointed out that if EU Member States implement the Data Retention Directive, mandating a 24 month retention period, it would have to adjust its policy to comply. Thus, on the one hand, Google is slapped on the wrist by privacy regulators for its

²⁶⁷ See TorrentSpy, <http://www.torrentspy.com/>.

²⁶⁸ BitTorrent is a peer-to-peer (P2P) file sharing communications protocol, allowing for broad distribution of large amounts of data without the distributor having to incur the costs of hardware, hosting and bandwidth. See BitTorrent, WIKIPEDIA, <http://en.wikipedia.org/wiki/BitTorrent>.

²⁶⁹ See John Borland, *MPAA Sues Newsgroup, P2P Search Sites*, CNET NEWS.COM, Feb. 23, 2006, <http://tinyurl.com/36zj9n>.

²⁷⁰ Columbia Pictures Inds. V. Bunneli, ___ F.Supp.3d ___ No. CV 06-1093 FMC(JCx) (C.D. Cal. May 29, 2007).

²⁷¹ The decision has been stayed pending appeal. TorrentSpy announced it would block all search queries from U.S. users rather than log user queries in contravention of its privacy policy. See Jacqui Cheng, *TorrentSpy to MPAA: Log this! Site blocks US searches*, ARSTECHNICA, Aug 27, 2007, <http://arstechnica.com/news.ars/post/20070827-torrentspy-to-mpaa-log-this-site-blocks-us-searches.html>.

²⁷² Article 29 Working Party Letter, *supra* note 13.

²⁷³ *Id.* Google has consequently shortened its data retention period to 18 months and reset its cookie to expire two years after a user's last search query. See Fleischer Letter, *supra* note 68.

retention policies. On the other hand, Google is mandated by data retention requirements to store user data for lengthy time periods.²⁷⁴

The solution to Google's quandary requires finding the golden path between privacy and data retention requirements. To be sure, massive data retention is privacy intrusive and risks turning service providers into data warehouses for government investigations.²⁷⁵ It sets the stage for pervasive surveillance of ordinary citizens whose personally identifiable information may be mined and analyzed in grand "fishing expeditions" by security and law enforcement agencies. Nevertheless, prohibiting data retention would constitute a boon for terrorists, pedophiles, organized crime and hackers, and put law enforcement agencies at a disadvantage against an increasingly sophisticated opponent.

The balance between law enforcement and privacy rights should permit limited data retention for narrowly tailored purposes, such as fighting terrorism and organized crime. In addition, mechanisms must be put in place to ensure that there be no further use of retained data for unrelated purposes; that prevention of terrorism not include large-scale data mining schemes; that access to data be duly authorized on a case by case basis by a judicial authority; and that systems for storage of data for law enforcement purposes be separated from systems used for business purposes.²⁷⁶

To sum up, far from requiring search engines to purge search query logs shortly after users complete their sessions, governments are increasingly *mandating* data retention. Such legislation raises the stakes for users, whose personal information piles up on search engine servers, ready for use by interested government or private third parties.

²⁷⁴ In his response to European privacy regulators, Google's chief privacy officer suggests "[a] public discussion is needed between officials working in data protection and law enforcement." Fleischer Letter, *id.*

²⁷⁵ See, e.g., EUROISPA and US ISPA Position on the Impact of Data Retention Laws on the Fight Against Cybercrime (Sept. 30, 2002), available at http://www.euroispa.org/docs/020930eurousispa_dretent.pdf; Home Office Voluntary Code of Practice, *supra* note 257, at s. 23-24.

²⁷⁶ See Article 29 Working Party Opinion 4/2005, 21 Oct., 2005, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2005/wp113_en.pdf, at 8-10; see also Opinion 3/2006, Mar. 25, 2006, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2006/wp119_en.pdf; Opinion 9/2004, Nov. 9, 2004, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2004/wp99_en.pdf.

f) The law of confidentiality

Rather than protecting information that is held in secrecy, the law of confidentiality protects information that is shared with others based on a relationship of trust and an expectation of privacy.²⁷⁷ I argue that breach of confidentiality offers a more satisfactory remedy than U.S. constitutional or statutory law for search engine users whose expectations of privacy are trumped.

Ever since Warren and Brandeis "reinvented" the right of privacy in their seminal article in 1890, privacy has been closely intertwined with the law of confidentiality.²⁷⁸ English courts to this day hesitate to acknowledge an independent right of privacy, preferring to seek the comfort of traditional breach of confidence law instead.²⁷⁹ They do so, even at a price of "stretching" the confidentiality doctrine to account for practically nonexistent relations between the parties.²⁸⁰

In the U.S., the law of confidentiality has been slow to develop compared to the tort of public disclosure of private facts, which was classified as a privacy cause of action in William Prosser's classic taxonomy.²⁸¹ Comparing the two causes of action, Solove explains that "both involve the revelation of secrets about a person, but breaches of confidentiality also violate the trust in a specific relationship. In this way, the tort emerges from the concept of a fiduciary relationship."²⁸² Hence, "the harm from a breach of confidence . . . is not simply that information has been disclosed, but that the victim has been betrayed."²⁸³ In other words, the fundamental rationale of

²⁷⁷ Solove, Taxonomy, *supra* note 21, at 526-29.

²⁷⁸ Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890).

Warren and Brandeis cite a line of 19th Century English cases, the most well known of which is *Prince Albert v. Strange* [1849] 2 De G & Sm 293, 1 Mac & G 25, which imposed liability on disclosures of information protected under an implied contract or a trust or confidence. Richards and Solove recently argued that Warren and Brandeis' groundbreaking article, as well as William Prosser's work, have had the (unfortunate) effect of "divorcing" U.S. privacy law from the law of confidentiality. The latter continued to evolve as a robust branch of privacy law in the U.K. See Richards & Solove, *supra* note 22.

²⁷⁹ See *Wainwright v. Home Office*, [2003] UKHL 53; *Kaye v. Robertson*, [1991] F.S.R. 62, 66 (C.A.) (U.K.).

²⁸⁰ See *Campbell v. MGN Ltd*, [2004] 2 A.C. 457; *Douglas v. Hello!, Ltd* [2001] QB 967.

²⁸¹ William L. Prosser, *Privacy*, 48 CAL. L. REV. 383 (1960); see also Restatement (Second) of Torts § 652D (1976); WILLIAM L. PROSSER & W. PAGE KEETON, *LAW OF TORTS* 856-63 (1984 & Supp.1988). Richards and Solove state that "[i]n the United States, the breach of confidentiality tort has grown up in the shadow of the Warren and Brandeis torts." Richards & Solove, *supra* note 22, at 156.

²⁸² Solove, Taxonomy, *supra* note 21, at 526-27.

²⁸³ *Id.* at 527.

confidentiality law is not the protection of information but, rather, the protection of a *relationship* of confidence.²⁸⁴

Are Google users "betrayed" by the company when it makes secondary use of their personally identifiable information or divulges it to third parties? Courts have traditionally applied the confidentiality paradigm to professionals in fiduciary roles,²⁸⁵ such as lawyers,²⁸⁶ doctors,²⁸⁷ therapists,²⁸⁸ and banks.²⁸⁹ Yet English law has gradually expanded the confidentiality doctrine to non-fiduciaries, including the press.²⁹⁰ Ironically, this paradigm shift, which was influenced by European legal instruments,²⁹¹ has had the effect of bringing the English concept of "confidentiality" closer to the U.S. notion of privacy, captured in Justice Harlan's celebrated "reasonable expectation of privacy" test.²⁹² As Lord Nicholls holds in *Campbell v. MGN*, "essentially the touchstone of private life is whether in respect of the disclosed facts the person in question had a reasonable expectation of privacy."²⁹³

²⁸⁴ See John D. McCamus, *Celebrity Newsgathering and Privacy: The Transformation of Breach of Confidence in English Law*, 39 AKRON L. REV. 1191, 1209 (2006); *Hammonds v. Aetna Cas. & Sur. Co.*, 243 F. Supp. 793, 801 (N.D. Ohio 1965).

²⁸⁵ See generally Susan M. Gilles, *Promises Betrayed: Breach of Confidence as a Remedy for Invasions of Privacy*, 43 BUFF. L. REV. 1 (1995); G. Michael Harvey, *Comment, Confidentiality: A Measured Response To the Failure of Privacy*, 140 U. PA. L. REV. 2385 (1992); Alan B. Vickery, *Note, Breach of Confidence: An Emerging Tort*, 82 COLUM. L. REV. 1426 (1982).

²⁸⁶ See Lee A. Pizzimenti, *The Lawyer's Duty to Warn Clients About Limits on Confidentiality*, 39 CATH. U. L. REV. 441, 463-71 (1990).

²⁸⁷ See, e.g., *South Carolina State Board of Medical Examiners v. Hedgepath*, 325 S.C. 166, 480 S.E.2d 724 (1997); *McCormick v. England*, 494 S.E.2d 431, 432 (S.C. Ct. App. 1997); *Hammonds v. Aetna Casualty & Sur. Co.*, 243 F. Supp. 793 (N.D. Ohio 1965); *Mull v. String*, 448 So. 2d 952 (Ala. 1984); see also Joseph White, *Physicians' Liability for Breach of Confidentiality: Beyond the Limitations of the Privacy Tort*, 49 S.C. L. REV. 1271 (1998).

²⁸⁸ See, e.g., *Doe v. Roe*, 93 Misc.2d 201, 400 N.Y.S.2d 668 (NY Sup. 1977); *MacDonald v. Clinger*, 446 N.Y.S.2d 801, 805 (App. Div. 1982).

²⁸⁹ See, e.g., *Young v. U.S. Dept. of Justice*, 882 F.2d 633 (2nd Cir. 1989); *Rush v. Maine*, 387 A.2d 1127 (Me. 1978); *Peterson v. Idaho First Nat'l Bank*, 367 P.2d 284, 290 (Idaho 1961). See generally Edward L. Raymond, *Annotation, Bank's Liability Under State Law For Disclosing Financial Information Concerning Depositor or Customer*, 81 A.L.R. 4th 377 (1992).

²⁹⁰ See *Attorney General v. Guardian Newspapers Ltd (No 2)* [1990] 1 AC 109. As Lord Nicholls observes in the *Naomi Campbell* case, "[t]his cause of action has now firmly shaken off the limiting constraint of the need for an initial confidential relationship . . . Now the law imposes a 'duty of confidence' whenever a person receives information he knows or ought to know is fairly and reasonably to be regarded as confidential." *Campbell v. MGN*, *supra* note 280, at 464-65.

²⁹¹ British courts have broadened their interpretation of the right to privacy pursuant to the enactment of the Human Rights Act, 1998, which incorporates into British law the provisions of the ECHR. As discussed above, these include Article 8, which protects the right to privacy. See *supra* note 209 and accompanying text.

²⁹² *Katz*, *supra* note 139; see discussion in H. Tomas Gomez-Arostegui, *Defining Private Life Under the European Convention on Human Rights by Referring to Reasonable Expectations*, 35 CAL. W. INT'L L. J. 153 (2005).

²⁹³ *Campbell v. MGN*, *supra* note 280, at 466. Baroness Hale too refers to "the 'reasonable expectation of privacy' [as] a threshold test." *Id.* at 496. See also *A. v. B. Plc.*, [2003] Q.B. 195 (U.K.), where

Viewing the disclosure of information by users to search engines as a disclosure to the public, which trumps such users' reasonable expectation of privacy, may fit Fourth Amendment doctrine but is out of sync with the actual beliefs and expectations of users. As Solove puts it, "when people establish a relationship with banks, Internet service providers, phone companies, and other businesses, they are not disclosing their information to the world. They are giving it to a party with implicit (and often explicit) promises that the information will not be disseminated."²⁹⁴

When you enter a search query in Google you simply do not expect it to haunt you in criminal or civil proceedings, nor do you expect it to be transferred to third party businesses for marketing or data mining purposes. Information revealed to search engines may be highly sensitive and similar to data covered by confidential relationships with physicians (consider the query "hypertension impotence treatment"), psychotherapists ("Zyprexa side effects"), lawyers ("income tax misrepresentation"), or priests ("Jesus savior baptizing"). Indeed, users likely consult search engines for issues that they would hesitate to address to any of the traditional professionals, based on a misguided belief that such queries remain between them and their computer.

Does the relationship between search engines and their users espouse trust, a reasonable expectation of privacy, and therefore a duty of confidentiality? Well, Google itself seems to think so.²⁹⁵ Whether based on an implied term of contract between Google and its users or on the private nature of the information itself, Google

Lord Woolf holds: "A duty of confidence will arise whenever the party subject to the duty is in a situation where he either knows or ought to know that the other person can reasonably expect his privacy to be protected."

²⁹⁴ Solove, *Taxonomy*, *supra* note 21, at 529; *see also* Richards & Solove, *supra* note 278, at 175, stating that "confidentiality stands directly at odds with the notion that when people share information with others they necessarily assume the risk of betrayal. The very purpose of confidentiality law is to recognize and enforce expectations of trust."

²⁹⁵ *See* Ken Auletta, *The Search Party*, *THE NEW YORKER*, Jan. 14, 2008, *available at* http://www.newyorker.com/reporting/2008/01/14/080114fa_fact_auletta; "As for risks to personal privacy, Eric Schmidt says that Google would never cross that boundary; if it violated user trust, 'we'll really be hosed.'" *See also* Nicole Wong, *Responsible Use and Protection of Information in Online Advertising* (presentation at FTC Town Hall entitled "Ehavioral Advertising: Tracking, Targeting, and Technology"), Nov. 1, 2007, *available at* <http://www.ftc.gov/bcp/workshops/ehavioral/presentations/4nwong.pdf>, stating "our users' trust and their privacy are critical to our business."

should account to users in case of disclosure of information to third parties. Users who confide in Google their fears and wishes, needs and interests, may perhaps expect their search query logs to be used by the company for improvement of its services or prevention of fraud. But they do not expect their personally identifiable information to be transferred to third parties. Subsequently, they must be compensated if Google breaches their trust.

The law of confidentiality solves the problem of trust between patient and physician, customer and banker, and other additional fiduciary relationships. It should be applied equally to the transaction between search engines and users. A physician divulging personal information to a third party breaches her fiduciary duty of confidentiality, commits a tort, and violates professional codes of ethics. The same is true when a search engine discloses user queries to a government or non-government entity.

Conclusion

The *Gonzales v. Google* case and the AOL privacy debacle were not isolated or exceptional occurrences. They are, but the tip of the iceberg of an emerging privacy problem on a grand scale, featuring Internet search engines as informational gatekeepers harboring previously unimaginable riches of personally identifiable information. Billions of search queries stream across Google's servers each month, "the aggregate thoughtstream of humankind, online."²⁹⁶ Google compiles individual search query logs, containing users' fears and expectations, interests and passions, and ripe with information that is financial, medical, sexual, political, in short—personal in nature. Google puts these data to secondary uses, such as improving its search service, ensuring network security, and targeting ads. Users may stomach such use of their personally identifiable information as part of their transaction with a company that offers an amazing service for no apparent cost. Yet they are less inclined to appreciate the sharing of their data with third parties, be they commercial, governmental, or, of course, criminal.

²⁹⁶ BATTELLE, THE SEARCH, *supra* note 6, at 6.

As this article demonstrates, the collection, retention, and use of personally identifiable information by search engines raise a host of privacy problems, including *aggregation, distortion, exclusion, and secondary use*. These problems and the public realization that they exist may have a *chilling effect* on search and online activity. Search engine users who become aware that the government may be privy to their communications—or more accurately in the context of search, to their thought process—may be cowed into submission to mainstream views and opinions.

Users may counter privacy invasive technologies with PETs in a perpetual game of "hide and seek." Yet they are often unwilling to expend the time and effort, or simply not technology-savvy enough, to fight for what many believe is a lost cause. Privacy policies—the one-sided browse-wrap agreements typically not read by anyone, save the lawyers who draft them—cannot be relied upon to protect users' privacy. As contractual documents, they are based on user consent, which is inferred from mere use of a web site, uninformed, and not truly voluntary. Having exhausted technological and contractual privacy protections, the fall back for users is the constitutional and statutory scheme provided by the state. Users are bound to be disappointed, as current doctrine is ill-suited to protect their interests.

In a world of pervasive surveillance and rapidly evolving data mining technologies, the U.S. doctrine, which denies protection to personal information that has been transferred to third parties, has become outdated. In this day and age, third parties—such as financial institutions, insurance companies, online service providers, and government agencies—maintain databases with massive amounts of personally identifiable information, including in certain cases information not known to the individuals themselves. The line dividing protected and unprotected data must be redrawn, since under current doctrine individuals have no rights whatsoever in these vast data pools. The EU data protection framework, with its set of fair information principles, provides better protection for personally identifiable information held by third parties, but has been criticized as cumbersome and overly bureaucratic.²⁹⁷

²⁹⁷ Peter Fleischer, The Need for Global Privacy Standards, PETER FLEISCHER BLOG, Sept. 14, 2007, <http://peterfleischer.blogspot.com/2007/09/need-for-global-privacy-standards.html>.

Statutory protection for search query logs is also fundamentally flawed. Privacy in electronic communications is protected by a Byzantine statutory framework dating to the 1980's, when the Internet was in its infancy and search but a distant dream.²⁹⁸ It is not clear whether search queries constitute “contents of communications” entitled to protection under the statutory scheme. And even if they do, protection under the SCA is surprisingly weak, permitting access to the contents of communications pursuant to a mere administrative subpoena. In updating the ECPA for the new millennium, lawmakers should clarify the classification of search queries as contents and require a full search warrant for their disclosure.

Information privacy is about to receive a severe blow with the advent of data retention legislation. Such laws not only permit service providers to retain personally identifiable information but actually compel them to do so. They are advanced by national security and law enforcement agencies possessing far greater political clout than privacy advocates. In the public debate about combating terrorism and serious crime, the voice of privacy advocates is often muted. A reasonable balance must be struck between the needs of law enforcement and the democratic imperative of not casting a light of suspicion on all law abiding citizens.

The law of confidentiality may offer a partial solution. Search engines owe a duty of confidentiality to users, whether by contract or due to the inherently private nature of search data. The law of confidentiality remedies the upsetting results of existing constitutional doctrine, under which people assume the risk of betrayal when they share secrets with others. Customers reveal sensitive personal information to professionals, such as physicians, psychotherapists, lawyers, and bankers, based on trust and an expectation that confidential information will not be disclosed. A similar duty of confidentiality must apply to search engines with respect to user search query logs that they compile and store.

²⁹⁸ Battelle deems Archie, created in 1990 by a group of McGill University students, the first Internet search engine. See BATTLE, THE SEARCH, *supra* note 6, at 39-40; see also Archie search engine, WIKIPEDIA, http://en.wikipedia.org/wiki/Archie_search_engine.