2007

The Structure of Search Engine Law

James Grimmelmann

The Structure of Search Engine Law

James Grimmelmann[*]

* Associate Professor of Law, New York Law School. My thanks for their comments to Jack Balkin, Yochai Benkler, Shyam Balganesh, Aislinn Black, Michael Carroll, Eric Goldman, Anne Huang, Dan Hunter, David Johnson, Thomas Lee, Beth Noveck, Frank Pasquale, Guy Pessach, Chris Riley, Steven Wu, Tal Zarsky, and the participants in the workshops where I presented earlier versions of this Article. All otherwise-undated web sites in footnotes were last visited on August 28, 2007.

INTRODUCTION

Search engines are the new linchpins of the Internet.[1] A large and growing fraction of the Internet's immense volume of traffic flows through them. They are librarians, who bring order to the chaotic online accumulation of information. They are messengers, who bring writers and readers together. They are critics, who elevate content to prominence or consign it to obscurity. They are inventors, who devise new technologies and business models to remake the Internet. And they are spies, who are asked to carry out investigations with dispatch and discretion.[2]

Lawyers and the law have taken notice of search engines. Governments around the world are casting an increasingly skeptical eye on search engines, questioning whether their actions have always been in the interests of society. More and more parties are presenting themselves at the courthouse door with plausible stories of how they have been injured by search engines. Only a few foresighted legal scholars have recognized the growing importance of search engines.[3]

---

1. *See generally* JOHN BATTELLE, THE SEARCH: HOW GOOGLE AND ITS RIVALS REWROTE THE RULES OF BUSINESS AND TRANSFORMED OUR CULTURE (2005) (describing the history and significance of Internet search); JEAN-NOËL JEANNENEY, GOOGLE AND THE MYTH OF UNIVERSAL KNOWLEDGE (Teresa Lavender Fagan trans., 2007) (discussing the implications of search for European cultural heritage); DAVID A. VISE & MARK MALSEED, THE GOOGLE STORY (2005) (describing Google's history); IAN H. WITTEN ET AL., WEB DRAGONS: INSIDE THE MYTHS OF SEARCH ENGINE TECHNOLOGY (2007) (analyzing the role of search engines as the gatekeepers of information on the web).

2. *See, e.g.*, R. Scott Rappold, *Bumbling Bigg City Burglars Got $12K*, COLO. SPRINGS GAZETTE, July 10, 2007, *available at* http://www.gazette.com/articles/safes_24620___article.html/ackerman_google.html (describing criminals who Googled for "how to crack a safe" from a computer at the office they were burglarizing).

3. Many scholars have written about one legal controversy or another that involves a search engine. Fewer have linked the multiple, interconnected problems that search engines raise. The essential articles are Niva Elkin-Koren, *Let the Crawlers Crawl: On Virtual Gatekeepers and the Right to Exclude Indexing*, 26 DAYTON L. REV. 179 (2001); Urs Gasser, *Regulating Search Engines: Taking Stock and Looking Ahead*, 9 YALE J.L. & TECH. 201 (2006), *available at* http://www.yjolt.org/old/files/20052006Issue/Spring06-gasser.pdf; Eric Goldman, *Search Engine Bias and the Demise of Search Engine Utopianism*, 9 YALE J.L. & TECH. 188 (2006), *available at* http://www.yjolt.org/old/files/20052006Issue/spring06-goldman.pdf; Lucas Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169 (2000); Frank Pasquale, *Rankings, Reductionism, and Responsibility*, 54 CLEV. ST. L. REV. 115 (2006); and Frank Pasquale & Oren Bracha, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search* (2007), *available at* http://ssrn.com/abstract=1002453. Two unpublished theses also take a usefully broad view of search. *See* Michael Zimmer, The Quest for the Perfect Search Engine: Values, Technical Design, and the Flow of Personal Information in Spheres of Mobility (2007) (unpublished Ph.D. dissertation, New York University) (on file with the Iowa Law Review); Alejandro M. Diaz, Through the Google Goggles: Sociopolitical Bias in Search Engine Design (2005) (unpublished master's thesis), *available at* http://epl.scu.edu:16080/~stsvalues/readings /Diaz_thesis_final.pdf. Joshua A.T. Fairfield's *The Search Interest in Contract*, 92 IOWA L. REV. 1237 (2007), is, one may hope, the first of a wave of scholarly attention to the

This Article provides a roadmap to the legal issues posed by search. It indicates what questions we must consider when thinking about search engines, and it details the interconnections among those questions. It does not endorse any particular normative framework for search. Nor does it recommend who should regulate search.[4] Instead, it provides an analytic foundation to distinguish informed decisionmaking from random flailing.

The diverse questions of law it discusses form a coherent set because each affects the same few information flows. The essence of a search engine is that it combines its own knowledge of available content with user queries to provide recommendations to its users; the doctrines and policy values this Article discusses relate directly to this core process. Other law affects search engines—Google's well-publicized Initial Public Offering ("IPO"), for example, raised substantial issues of securities law,[5] and search engines have been sued for employment discrimination[6]—but these other issues can be resolved on their own merits, in isolation. In contrast, the concerns discussed in this Article must be balanced with one another because each relates to the same few information flows. Pushing on one affects the others.

Part I explains how modern search engines function and describes the business environment within which they operate. Search engine operations can be understood in terms of the information flows among four principal actors: (1) search engines themselves, (2) their users, (3) information providers, and (4) third parties with interests in particular content flows (such as copyright holders and censorious governments). There are, in turn, four significant information flows: (1) the indexing by which a search engine learns what content providers are making available, (2) user queries to the search engine for information about particular topics, (3) the results returned by the search engine to users, and finally, (4) the content that providers send to users who have found them through searching. Because so many major search engines are funded through advertising, this Part also

consequences of having omnipresent, powerful search tools available for legal questions that do not directly involve search.

Although not scholars, some bloggers have so deeply immersed themselves in the search world that they have a synoptic perspective on the field. A few blogs are indispensable reading for anyone interested in search. *See* John Battelle, SEARCHBLOG, http://battellemedia.com/; Matt Cutts, GADGETS, GOOGLE, AND SEO, http://www.mattcutts.com/blog/; Philipp Lenssen et al., GOOGLE BLOGOSCOPED, http://blogoscoped.com/; SEARCH ENGINE WATCH, http://blog.searchenginewatch.com/blog/; Danny Sullivan et al., SEARCH ENGINE LAND, http://searchengineland.com/.

4. *Cf.* Pasquale & Bracha, *supra* note 3 (comparing institutional forms for search regulation).

5. *See* Google, Inc., Registration Statement (Form S-1) (Apr. 29, 2004), *available at* http://www.sec.gov/Archives/edgar/data/1288776/000119312504073639/ds1.htm.

6. *See* Elwell v. Google, Inc., No. 05 Civ. 6487, 2006 U.S. Dist. LEXIS 3114, at *1 (S.D.N.Y. Jan. 31, 2006).

includes a brief survey of how search engine advertising works and the distinctive fraud problems confronting search engines and their advertisers.

Part II, the heart of the Article, presents a descriptive analysis of the legal struggles over search, showing how questions of search policy—many of which have long been latent in different fields of Internet law—are increasingly confronting lawyers, courts, and regulators. It describes those struggles in terms of the legitimate interests that each of these actors brings to debates over search. Users want high-quality results without too great a sacrifice of *privacy*. Content providers want favorable placement in search results without paying more than their fair share of the costs of supporting search and without facing unfair competition from search engines. Third parties want to prevent unauthorized distribution of copyrighted content, to preserve their own privacy, to protect their reputations, and to preserve what they see as "user virtue." And finally, search engines want to preserve their ability to innovate, to protect themselves from fraud, and to ensure that the search market remains open to competition. Each entry in this list of interests has its own associated legal theories; this systematic taxonomy allows us to recognize how any given legal theory affects the search ecology.

Part III then shows, with five examples, how taking a broad view of search yields otherwise-unavailable insights into pressing controversies. This is not to say that the end result must be a body of search-specific law,[7] but only that failing to consider the larger forces at work in search is antithetical to sensible policymaking. First, the broad, systematic view illustrates how various claims in search engine disputes can serve as functional substitutes for each other. Second, it shows that the degree of transparency of the search process is a highly contested variable, with some concerns pressing for greater transparency and others pressing for less. Third, it illustrates that user privacy is a deeply knotty problem and that preserving reasonable user expectations will involve difficult trade-offs with other interests—including some of the users' own. Fourth, it shows that we require a theory of search engine speech; the most sophisticated theory of search-engine-results-as-speech so far articulated by a court is too simplistic. And fifth, it illustrates the richness of debates over search engines' relationships to providers' trademarks.

Finally, a brief Conclusion takes note of some of the many open issues facing search engine law and scholarship.

7. *Cf.* Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207, 207 (arguing that there neither is, nor should be, a distinct body of cyberlaw). Lawrence Lessig has responded that "more than law alone enables legal values, and law alone cannot guarantee them," and has argued that cyberlaw (or "Internet law," depending on one's view of the subject) provides a broader view of law itself. Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501, 549 (1999). I do not make the same claims for search engine law. Search engines are more important in the consideration of what law should do than in the consideration of what law is.

I. SEARCH ENGINE TECHNOLOGY AND BUSINESS

Every major Internet application today is a search engine, contains a search engine, or depends on a search engine.[8] Because search is so increasingly indistinguishable from other applications, we shouldn't expect our definition of "search engine" to differentiate clearly between those things that are search services and those that aren't. Instead, let's start with a definition that accurately describes the core, paradigm cases: a search engine is a service that helps its users locate content on the Internet. That's what Google, Yahoo!, MSN Live, and Ask.com do: help people find stuff online. So do their smaller competitors, from AltaVista to Zoohoo.

As anyone who's played with Google's ever-expanding list of search services can attest, search engines help users find more than just web pages. Google Scholar searches journal articles; Yahoo! Local searches businesses near the user; the Internet Movie Database searches lists of film casts and crews; and Like.com searches online sales of clothes and jewelry by color, shape, and pattern. Thus, it's better to say that search engines help users find "content" than to say "pages" or "sites."[9]

All of these search engines help users find publicly accessible content, but others work with specialized sets of content not available to the public. Thus, LexisNexis, for a fee, allows users to search a large proprietary database of legal and news documents. Similarly, peer-to-peer file sharing systems such as Gnutella and Grokster allow users to search content that typically is accessible only through the peer-to-peer service itself.

Go far enough along these axes (away from the web and away from publicly accessible content) and you will reach things that are only marginally recognizable as search engines, according to the definition above. Google Desktop, for example, is one of several competing tools for users to search their own computers. Not every legal issue affecting search applies to these borderline cases. But some issues carry over even here: Google Desktop has raised privacy concerns resembling those that apply to plain-vanilla Google Web Search.[10] The point is that to the extent a technology resembles the paradigm case of public web search—and a great many technologies do—it raises many of the issues described below. Anyone working with that technology needs to think about how those issues fit

8. The description of search engine operations in this Part draws on John Battelle's THE SEARCH, *supra* note 1. Extended discussion of how search engines work and the business models of the search industry is available there and in David A. Vise and Mark Malseed's THE GOOGLE STORY, *supra* note 1, and WITTEN ET AL., *supra* note 1.

9. I say "content" rather than "information" to distinguish it from other significant information flows connected with search.

10. *See* Electronic Frontier Foundation, *Google Copies Your Hard Drive – Government Smiles in Anticipation*, DEEP LINKS, Feb. 9, 2006, http://www.eff.org/press/archives/2006/02/09 ("'[Google Desktop] highlights a key privacy problem in the digital age.'" (quoting Cindy Cohn, Legal Director, Electronic Frontier Foundation)).

together. The taxonomy that follows provides a framework for thinking both about search engines and about the large penumbra of things that resemble them.

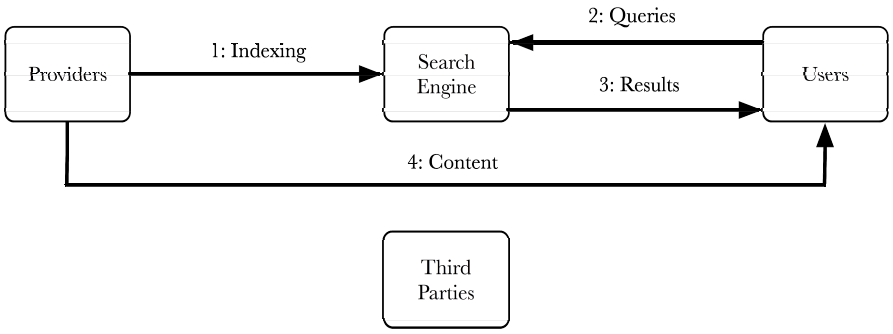
A. TECHNOLOGY

A search engine in isolation is useless. It becomes valuable only through interactions with content providers and with users. By aggregating information about providers' offerings and presenting it in a useful form, a search engine can match users with appropriate content providers, to the benefit of both. This matching, however, can antagonize third parties who would rather that certain connections not be made.[11] Visualizing the information flows between search engines and these three groups illustrates how search works.

Search involves four information flows, diagrammed in Figure 1: (1) the search engine gathers content; (2) a user queries the search engine; (3) the search engine provides the user with results; (4) the user obtains the content.

Let's take up these flows in order.

Figure 1. Information Flows in Search



1. Indexing

While users use a search engine to search for content, a search engine itself must search out the content it is to recommend to users. It must therefore work with content providers to learn what they have to offer. With web search, the process is normally driven by the search engine, which uses automated software agents—"robots," "spiders," or "crawlers"—to explore the web and find content. It generally does so in the same manner that a

11. As discussed in more detail below, such third parties include copyright holders, targets of libel, censorious governments, and others. *See infra* Part II.C.

user would, requesting web pages from content providers' servers and seeing what those pages contain.[12]

Other forms of search involve different forms of information gathering. Some search engines simply take existing collections of information and organize them more effectively—a nationwide phone-number search aggregates information that was, at one time, only available in a shelf of phone books. Others rely on content providers to come to them. Under "paid search inclusion," once practiced by a number of search engines, content providers pay a search engine and supply it with content; the engine promises to index any content for the appropriate fee.[13] Major search engines now offer tools for content providers to describe and organize their content in standardized formats.[14]

The line between search engine and content provider can be indistinct. Online merchants, such as Amazon.com, typically provide search engines for their own sites, as do most sites that aggregate user-supplied content (e.g., Wikipedia). One does not have to use the site-specific search engine to reach the content, but one can. Decentralized peer-to-peer systems use the same computers both to provide files and to index them.

2. Queries

Content is only one of two inputs to search. The other is the "search query," a request by a user for information on a particular topic. Most search engines use queries made up of a few keywords or short phrases. Some non-textual search engines allow users to issue queries in more exotic forms, such as hummed tunes[15] or pictures.[16] In whatever form, the user provides the search engine with some criteria to narrow the vast universe of possible results.

A query is typically only an approximation of the user's intentions. Common intentions include navigational queries (the user wishes to find a specific site or datum), informational queries (the user wishes to find information on a topic), and transactional queries (the user wishes to

12. *See* Google, How Google Crawls My Site, http://www.google.com/support/webmasters/bin/topic.py?topic=8843.

13. *See* Danny Sullivan, *The Evolution of Paid Inclusion*, SEARCH ENGINE WATCH, July 2, 2001, http://searchenginewatch.com/showPage.html?page=2163971 ("Paid inclusion programs mean that, in exchange for a payment, a search engine will guarantee to list pages from a web site.").

14. *See, e.g.*, Google, About Google Base, http://base.google.com/support/bin/answer.py?answer=59260&hl=en (describing Google Base as "a place where you can easily submit all types of online and offline content, which we'll make searchable on Google").

15. *See* Midomi Video Tour, http://www.midomi.com/index.php?action=main.video ("midomi lets you find music by singing or humming!").

16. *See* Bob Tedeschi, *Shopping Site Offers a Way to Raid a Celebrity's Closet*, N.Y. TIMES, Nov. 13, 2006, at C4; Like Visual Search, http://www.like.com ("Like.com is a visual search engine that lets you find items by color, shape, and pattern").

perform an activity, such as purchasing a good).[17] Because words have multiple meanings, the same query reasonably could be directed at many different possibilities, even within one of these categories. Further, the user may not have in her head a clear idea of what she is searching for. What she is searching for may not even exist. By way of example, consider a search for "apple." The user might have intended any of the following:

- To find the home page for Apple Computer (navigational);
- To learn about apples, the fruit (informational);
- To purchase an Apple MacBook computer (transactional);
- To purchase apples, the fruit (transactional, but different);
- To learn about oranges (informational, but confused);
- To find the home page for Apple Records, the Beatles' record label (navigational, but no such page exists[18]); or
- To test whether her connection to the Internet is working.

Along with the query itself may come various user information, such as geographic location, preferred types of results, operating system and browser, and preferences among search results revealed through clicks on past search results. Some search engines, such as Google and Yahoo!, keep extensive histories on the searches and preferences of registered users. Personalized search engines may customize their results, showing different results to users with different geographic locations or announced interests.

3. Results

In the defining step of search, the search engine returns to the user information about content relevant to her query. Web search engines typically show results on a web page that lists results ten at a time, starting with those results it thinks she would probably most like to see and following in descending order with results it thinks are probably less relevant. Each search result normally contains the name of the identified piece of content, its location, and a very short summary or excerpt that shows how the content relates to the query. As the user inspects a set of search results, she may choose to refine her query, choosing slightly different keywords in an attempt to better convey her intention to the search engine. The engine, in turn, will supply her a different set of results. This process is iterative, and

17. *See* Andrei Broder, *A Taxonomy of Web Search*, SIGIR FORUM, Fall 2002, *available at* http://www.sigir.org/forum/F2002/broder.pdf. Broder's taxonomy is obviously not exhaustive, nor are its three categories entirely distinct from one another, but it provides a good first approximation.

18. The closest substitutes are probably the home page for Apple Corps, Ltd., http://www.thebeatles.com/ (the conglomerate parent of Apple Records), and *The Complete Apple Records*, http://www.schomakers.com/ (an unaffiliated site with a complete Apple Records discography).

thus, even within a single search session, query and results flows may repeat any number of times.[19]

This is the step in which search engines really differentiate themselves: by using varied algorithms to summarize and organize the vast seas of available content. At one time, web search engines simply scanned the text of web pages to determine which topics the pages discussed. They then augmented this technique by analyzing additional information about pages (called "metadata"), such as their age, the number of links they contain, or the keywords used by their authors to describe them. More recently, powerful link-structure techniques involving study of which web pages link to which other pages have become the dominant web search paradigm.[20]

The biggest technical challenge in search today is integrating these computationally expensive analyses with a user's particular query. A user's query must be answered in a few seconds, but it can take a server farm of many thousands of computers weeks to build an index of the web. Search engines that deal with small domains can afford to gather information from content providers in response to a particular query, but most search engines preprocess information about available content to produce a specially optimized index that can be consulted quickly. The most extreme form of this optimization is to have pre-generated lists of results and merely allow users to choose, in effect, from a predetermined list of queries. Such search engines function like the index of a book, allowing a user to find quickly anything the search engine has specifically chosen to show.

Search technologists also distinguish between query-dependent and query-independent indexes. Google's PageRank, for example, is an example of the latter; PageRank estimates the general popularity of a web page. Google uses PageRank to help sort its lists of results relevant to a particular query and return more popular pages first, but PageRank does not by itself say anything about which pages are relevant. That is the job of query-dependent analysis, such as looking in the pages for query terms and their synonyms. Other search engines have their own algorithms, but this integration of query-dependent and query-independent information is a standard technique.

Search engines are also increasingly learning from the large volumes of query data they have accumulated. A user's history of queries can provide useful information about her probable intentions—for example, whether she tends toward navigational or transactional queries. Similarly, search

19. *See* Goldman, *supra* note 3, at 196. As Goldman explains, this refinement of results is a critical process by which users correct for ambiguities in their initial queries—and the interactivity of the process argues against trying to infer a single fixed meaning for a query. *Id.*

20. The best available reference on search algorithms is AMY N. LANGVILLE & CARL D. MEYER, GOOGLE'S PAGERANK AND BEYOND: THE SCIENCE OF SEARCH ENGINE RANKINGS (2006), which provides a clear presentation of Google's PageRank and related algorithms and contains an extensive bibliography of the technical literature on search ranking.

engines gain useful feedback into their own successes and failures by seeing which results users click on or by noticing long strings of searches on related terms, which may indicate that the user is having trouble finding what she's looking for.

4. Content

In the end, the user cares about the content. Most often, she takes the location information given her by the search engine and uses it to approach the appropriate content provider. For web pages, that typically means she reads the page and consumes the information on it. For other goods and services, she may purchase them (or decide not to, after considering her options). Typically, both the user and content provider value being matched up, since this sort of exchange is why the content provider is online and why the user turned to a search engine.

Search engines themselves provide some content to users. Simply telling users what the content is provides some information about it; often, a web page title, a short excerpt of its text, or a thumbnail image may be enough to meet the user's needs. Some search engines cache content, storing copies to make it easier for users to receive it quickly. Others archive content, enabling users to receive it even when the original provider cannot be reached.

B. BUSINESS

It is impossible to understand the legal controversies over search without some understanding of the most common search business models. The overwhelmingly predominant model for web search today is contextual advertising, in which the search engine shows its users advertisements alongside the search results. Most commonly, these ads are brief two- or three-line blocks of text containing a hyperlink to the web site being advertised and various visual indicators marking them as advertisements.[21] These ads are still often search "results" in the sense that the search engine presents particular ads based on the user's query.[22] Search engine

21. *But see* DEBORAH FALLOWS, PEW INTERNET & AM. LIFE PROJECT, SEARCH ENGINE USERS 17 (2005), http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf (finding that sixty-two percent of search engine users were unaware of the distinction between search results and advertisements).

22. *See* BATELLE, *supra* note 1, at 109–21. *See generally* Benjamin Edelman & Michael Ostrovsky, *Strategic Bidder Behavior in Sponsored Search Auctions*, 43 DECISION SUPPORT SYS. 192 (2007), *available at* http://www.benedelman.org/publications/cycling-060703.pdf (discussing ways to reduce strategic behavior). *See also* F. Gregory Lastowka, *Search Engines Under Siege: Do Paid Placement Listings Infringe Trademarks?*, 14 INTELL. PROP. & TECH. L.J. 1, 2 (2002). The process is closely related to paid inclusion; in each, a content provider has paid the search engine in the hope of being seen by users entering particular queries. *Id.* at 2. *But see* Thomas A. Weber & Zhiqiang (Eric) Zheng, *A Model of Search Intermediaries and Paid Referrals*, (Wharton

developers have created sophisticated bidding algorithms that balance advertisers' willingness to pay with the popularity of their ads in choosing which so-called "sponsored links" to show.[23] Thus, content providers can be found by users either by being listed as search results or by advertising on search engines.

Search engines use three common billing techniques to sell ads. An advertiser using pay-per-impression pays a given fee to the engine each time a user sees the ad. Under pay-per-click, the advertiser pays each time a user not only sees the ad but also clicks on it. Under pay-per-conversion (also known as pay-per-action or pay-per-performance), the advertiser pays only when the user makes a purchase or takes some similar action that indicates serious interest in the advertiser's site. Pay-per-conversion is the most closely correlated with actual sales, but it requires that advertisers turn over significant information to the search engine so that billing can be properly calculated. Pay-per-click currently strikes the most popular overall compromise between accuracy and convenience.[24]

A few web search engines do not show ads. Providers with site-specific search engines can monetize them directly because every result will be from their own site. Some search engines are maintained out of altruism—the search facility is provided ad-free as a public service. A few search engines have attempted to make money by analyzing user search patterns and selling aggregate information.[25]

Search advertising has generated its own distinctive forms of fraud. In click fraud, users upset at a particular advertiser or in competition with it

Sch. OPIM, Working Paper No. 02-12-01, 2006), *available at* http://ssrn.com/abstract=601903 (arguing that ranking based on provider bids decreases overall social welfare).

23. *See* BATTELLE, *supra* note 1, at 142.

24. A full discussion of the business issues involved in search engine advertising is beyond the scope of this paper. As Tal Zarsky has explained in correspondence, the different billing models have significantly different implications for the incentives of search engines and advertisers and for the privacy of search engine users. Thus, for example, under a pay-per-click system, search engines will try to favor not just those ads for which advertisers will pay the most per click but also those ads that will generate the most clicks. Advertisers, in turn, will therefore be able to purchase prominent advertising placement more cheaply if their ads are well designed to encourage clicks. The process of monitoring clicks, however, is both vulnerable to fraud and requires close observation of user behavior—two facts that implicate the general tension between openness and transparency in search advertising and search engine operations. E-mail from Dr. Tal Zarsky, Law Faculty, Haifa University, to James Grimmelmann (Dec. 31, 2006, 12:56) (on file with the Iowa Law Review); *see also* Ben Elgin, *The Vanishing Click-Fraud Case*, BUSINESSWEEK ONLINE, Dec. 4, 2006, *available at* http://www.businessweek.com/print/technology/content/dec2006/tc20061204_923336.htm.

25. The Alexa Toolbar uses this strategy. Once installed in a user's browser, it displays extra information about the pages that the user visits, including related links. In exchange, it can track which pages users visit. This aggregate user data is valuable even just as a Nielsen-type rating for web pages.

repeatedly view and click on its ads to run up the advertiser's bills.[26] When search engines use their ad-serving infrastructure to act as advertising brokers and place ads on other web sites, affiliate fraud becomes a possibility; there, web sites sign up as affiliates to have ads placed on their pages and then turn around and click on the ads themselves. They pocket some of the money, while the bill goes to the advertisers.[27]

The great demand for high placement[28]—when combined with the zero-sum nature of ranking decisions—leads to search engine optimization ("SEO"): the business of redesigning content (or creating it) to attract search engines and convince them to rank content highly.[29] Some "white hat" SEO techniques are generally considered desirable because they make the content easier for search engines and users to access. Other, "black hat" techniques involve mimicking the superficial features that search engines use as proxies for quality content. When search engines scanned page text and keywords, optimizers would hide popular keywords in invisible, tiny text on a page or show a search engine a different page (one larded with thousands of keywords) than the one shown to users. As search engines shifted to analyzing link structure, optimizers switched to creating "link farms": sets of thousands of sites and pages pointing to each other, mimicking a community of real users and hoping to trick search engines into treating them as authoritative, popular sources of information.[30] Search engines and black-hat SEOs are locked in a technical arms race that pits increasingly sophisticated algorithms that distinguish fraudulent from

26. *See* Charles C. Mann, *How Click-Fraud Could Swallow the Internet*, WIRED, Jan. 2006, *available at* http://www.wired.com/wired/archive/14.01/fraud_pr.html (discussing the ramifications of click fraud).

27. Affiliate advertising and affiliate fraud do not involve search as such. They are, however, a good example of the Protean nature of search technology. The algorithms that can determine whether a web page is relevant to a particular query are the same kind of algorithms that can determine whether a web page is a good place to show a particular ad.

28. *See* Nico Brooks, *The Atlas Rank Report: How Search Engine Rank Impacts Traffic*, ATLAS INST. (2004), http://www.atlassolutions.com/WorkArea/showcontent.aspx?id=1440&LangType=133. Users are far more likely to click on the first result in a list shown to them than on any other result; if they are shown a page of ten results, only a small fraction of them will click through to see even the second page. Results after the hundredth or so may as well not exist.

29. *See* David Kesmodel, '*Optimize' Rankings at Your Own Risk*, WALL ST. J. ONLINE, Sept. 23, 2005, http://www.startupjournal.com/ecommerce/20050923-kesmodel.html; Google, What's an SEO? Does Google Recommend Working with Companies That Offer to Make My Site Google-Friendly?, http://www.google.com/support/webmasters/bin/answer.py?answer=35291.

30. Link farms have cropped up in some surprising places. *See* Andy Baio, *Wordpress Website's Search Engine Spam*, WAXY.ORG, Mar. 30, 2005, http://www.waxy.org/archive/2005/03/30/wordpres.shtml (homepage for "popular open-source blogging software"); Philipp Lenssen, *Forbes Spam?*, GOOGLE BLOGOSCOPED, Apr. 17, 2007, http://blogoscoped.com/archive/2007-04-17-n53.html (Forbes magazine); Blake Ross, *Stanford Daily Link Spam Harms the Web and Students*, BLAKEROSS.COM, May 27, 2005, http://blakeross.com/index.php?p=136 (college newspapers).

authentic content against increasingly subtle forms of mimicry.[31] As might be expected, black-hat-SEO techniques are highly controversial, and the line between black- and white-hat techniques is both unclear and contested, as is the line between authentic and fraudulent content.

Some of the business practices of those who make their living in the search ecosystem have more broadly harmful consequences. On the one hand, search engines provide one of the main sources of traffic for various fraudulent schemes—unsuspecting users landing on pages they don't really want to see can be shown advertisements, tricked into downloading spyware,[32] or scammed out of their money.[33] In order to acquire that search traffic in the first place, black-hat SEOs use all sorts of techniques that make the Internet less usable. They have registered domains fraudulently,[34] posted link-filled comments to blogs and discussion boards,[35] created fake web sites and blogs,[36] hijacked popular domains,[37] and sent hyperlinks in e-mail, instant messages, and even in requests for web pages.[38] Major search engines provide a centralized target that makes such schemes profitable.

31. *See, e.g.*, Lee Gomes, *Our Columnist Creates Web 'Original Content' But Is in for a Surprise*, WALL ST. J., Mar. 1, 2006, at B1, *available at* http://online.wsj.com/public/article/SB114116587424585798-0qH9qUYuUug__vRSFKGvxIEwLGw_20070301.html?mod=blogs; Wikipedia, Wikipedia: Send in the Clones, http://en.wikipedia.org/wiki/Wikipedia:Send_in_the_clones (noting that much Wikipedia content is now available on other sites and discussing potential methods of competition). *Cf.* Stephen Baker, *Asbestos and the Art of Blogging for Money*, BUSINESSWEEK ONLINE, May 27, 2005, http://www.businessweek.com/the_thread/blogspotting/archives/2005/05/asbestos_and_th.html (describing an experiment in which Michael Buffington created a topical blog on asbestos specifically to capture some of the money that lawyers would pay per click on ads triggered by related keywords). There is no clear line between such experiments and "real" professional blogs.

32. *See* Ben Edelman & Hannah Rosenbaum, *The Safety of Internet Search Engines*, MCAFEE SITEADVISER, May 12, 2006, http://www.siteadvisor.com/studies/search_safety_may2006.html ("Despite search engines' efforts, we see too many sites trying to deceive unsuspecting users.").

33. *See* Ben Edelman, *False and Deceptive Pay-Per-Click Ads*, Oct. 10, 2006, http://www.benedelman.org/ppc-scams/ (describing ads that make false or misleading claims).

34. *See, e.g.*, Declan McCullagh, *Dotster Named in Massive Cybersquatting Suit*, CNET NEWS, June 2, 2006, http://news.com.com/2102-1032_3-6079567.html (discussing a lawsuit against a large domain-name registrar).

35. *See* Six Apart, Six Apart Guide to Combating Comment Spam, http://www.sixapart.com/pronet/comment_spam.html.

36. *See* Christopher Heun, *Spam Blogs Pollute Internet Search*, INFORMATIONWEEK, May 15, 2006, http://www.informationweek.com/showArticle.jhtml?articleID=187202310.

37. *See* Will Baude, *The Remains of Crescat*, CRESCAT SENTENTIA (NOV. 5, 2006), http://www.crescatsententia.net/archives/2006/11/05/#006824 (describing the loss of "crescatsententia.com," previous home of a popular legal blog, to an SEO domain-squatter).

38. *See* Referer Log Spam, METAFILTER (Oct. 24, 2002), http://www.metafilter.com/21063/ (discussing SEO attempts to fill a referer log with spammed URLs).

II. THE STRUCTURE OF SEARCH ENGINE LAW

The set of laws potentially applicable to search may seem bewilderingly large. It has, however, a recurring deep structure that becomes evident if we focus on four concepts: the actors involved, the information flows among them, the interests that they bring to search, and the legal theories that they use to vindicate their interests. We have already met the key actors: (1) search engines, (2) content providers, (3) users, and (4) concerned third parties. We have also discussed the relevant information flows: (1) indexing, (2) queries, (3) results, and (4) content. This Part details these actors' interests in these information flows and examines the legal theories associated with those interests.

Every dispute involving a search engine is, at base, a dispute among these constituencies. To see what is really at stake in any given legal battle over search, it is useful to ask what these groups stand to gain or lose.

On the one hand, search engines create enormous social benefits. They allow willing users and content providers to find each other, reducing transaction costs and enabling mutually beneficial exchanges. These benefits depend on the contributions of users, providers, and search engines in the form of queries, content, and ranking algorithms, respectively. Good search engine policy would therefore give each group appropriate incentives to maximize its productive contributions while deterring rent-seeking behavior.

On the other hand, search engines can also cause enormous harms to particular parties. By controlling the process matching users and content providers, they create winners and losers within these communities. Both users and providers entrust search engines with valuable information and may be upset at the terms on which search engines reveal that information. Third parties who would prefer that certain content not flow from providers to users also are injured when search engines enable such flows. Good search engine policy will prevent search engines from unreasonably inflicting serious harms on others.

Complicating matters, these two features of search are inextricably intertwined. Search engines do not generally cause harms out of inherent malice. They cause harms in the process of serving their other constituencies. It is precisely the fact that search engines create enormous value that gives them such power to cause enormous harm. Users value search engines because the search engines help them pick and choose among possible providers; third parties are most upset at new content flows precisely when users and providers value those flows the most. Attempts to remediate particular harms, then, almost invariably involve a contest among the interests of these other constituencies. The law's choice to intervene or not is a choice among their interests.

The central position that search engines occupy also creates problems even when the balance between their various constituencies seems appropriate. If the balance comes with too much deference to search

engines, the risk is that they will behave unaccountably and upset the balance by aligning themselves with one group against another. But if the balance comes with too many restrictions on the actions of search engines, the risk is that those restrictions will squander the innovative potential of search engines.

With these general principles in mind, let us take up the specific interests that each constituency brings to any discussion of search law:

Users are the most obvious group served by search engines. When they make search queries, they reveal some potentially private information; thus, they may be harmed if a search engine misuses or reveals that information. They also desire useful search results, so a search engine can harm them by providing low-quality or deliberately biased results.

Providers have three interlocking interests in search. On the one hand, they usually value prominent placement in search results because of the many users they can draw in. On the other hand, they'd prefer not to shoulder more than what they perceive as their fair share of the costs involved in providing search. That means limiting bandwidth and server time they burn to support indexing by search engines; it also means limiting the search advertising they buy. Content providers also fear that search engines will compete unfairly with them, either by opening up new ways to view content or sometimes simply by delivering content directly to users.

Third parties sometimes fear the content flows that search engines enable, for a variety of reasons. Some have a copyright interest in the material now flowing freely; some may be the subjects of that material and feel that it defames them or invades their privacy; some (most often governments) may simply wish to censor the flow of content they consider objectively harmful.

And finally, search engines themselves want to keep their business models viable. That means dealing with the operational threats of SEO and click fraud, preserving the incentives to innovate in developing new forms of search technology, and maintaining competitive freedom from the risk that other search engines will unfairly dominate the market for search.

These controversies are summarized in Table 1:

Table 1. Interests in Search

Constituency	**Interest**	**Information Flows**	**Sample Legal Theories**
Users	Privacy	Queries	Information Privacy, Contract
	Quality	Results	Consumer Protection
Providers	Costs	Indexing, content	Trespass, Contract
	Unfair Competition	Content	Copyright, Trademark, Contract
	Placement	Results	Trademark, Business torts
Third Parties	Ownership	Content	Copyright, Trademark
	Reputation	Content	Defamation
	Privacy	Content	Information Privacy
	User Virtue	Content	Direct Regulation
Search Engines	SEO	Indexing, Results	Fraud
	Click Fraud	Results	Contract
	Innovation	All	Intellectual Property
	Competition	All	Antitrust

The remainder of this Part takes up these issues in order.

A. USERS' INTERESTS

1. Query Privacy

Effective search requires that users disclose information about their interests and intentions. Whether they click on a topic heading in an index or craft a complex query with various exclusions and inclusions, the very fact that they are curious about something will be evident in their queries. If they are repeat users, the search engine may be able to construct an extensive history of their queries. It may also be able to correlate this curiosity with

users' actual behavior in obtaining content or with other information it has about them from their use of other applications and features it provides. In addition to collecting in-depth data on each user, a search engine also has broad access to information about many users.

Much of this data is personal or sensitive. When AOL publicly released the search queries of some 650,000 search users, the logs included queries such as "can you adopt after a suicide attempt,"[39] "cocaine in urine,"[40] and "How to deal with mental abuse in a Christian marriage."[41] Even though the search logs identified users only by pseudonymous numbers, reporters showed by example that it was possible to take a list of searches and identify the searcher; people have a tendency to search on their own names, their addresses, and other personally identifying details.[42] Given the sensitivity of this information and the ease of linking it back to particular individuals, users have an evident privacy interest that their queries not be misused. Even where the information cannot be linked to an individual, it can still be used in ways that cause privacy harms.[43]

Case law dealing with collection of personal information online suggests that users will not enjoy significant legal recourse against search engines for the misuse of their queries. Multiple courts have held that users fail to state a claim when they allege that web advertising services (with the cooperation of the web sites on which the ads appear) have captured their browsing habits and tracked them through time.[44] Once in possession of the information, an engine is free to disclose it to others, unless it has undertaken not to.[45] A

39. Declan McCullagh, *AOL's Disturbing Glimpse into Users' Lives*, CNET NEWS, Aug. 8, 2006, http://news.com.com/2100-9588_22-6103098.html.

40. *Id.*

41. Lee Gomes, *What Are Web Surfers Seeking? Well, It's Just What You Think*, WALL ST. J., Aug. 16, 2006, at B1.

42. Michael Barbaro & Tom Zeller Jr., *A Face is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006, at A1.

43. In the terminology of Daniel Solove's *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 491–561 (2006), a search engine could harm a user through disclosure, breach of confidentiality, decisional interference, exposure, identification, secondary use, aggregation, and surveillance. For extensive discussion of the privacy implications of search, see Zimmer, *supra* note 3. Tal Zarsky, in *Mine Your Own Business!: Making the Case for the Implications of the Data Mining of Personal Information in the Form of Public Opinion*, 5 YALE J.L. & TECH. 4, 37–38 (2003), argues that the aggregation of information about consumer habits can lead to a harmful "autonomy trap." Julie Cohen's *A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace*, 28 CONN. L. REV. 981 (1996) explains clearly the dangerous chilling effect of surveillance of one's reading habits.

44. *See, e.g.*, Chance v. Ave. A, Inc., 165 F. Supp. 2d 1153, 1155 (W.D. Wash. 2001) (interpreting 18 U.S.C. §§ 1030, 2511(2)(a), 2701 (2000)); *In re* DoubleClick Inc. Privacy Litig., 154 F. Supp. 2d 497, 500 (S.D.N.Y. 2001) (interpreting 18 U.S.C. §§ 1030, 2510, 2701).

45. *See, e.g.*, *In re* JetBlue Airways Corp. Privacy Litig., 379 F. Supp. 2d 299, 330 (E.D.N.Y. 2005) (interpreting 18 U.S.C. § 2701). Legislation to change this baseline and require the deletion of stored queries was introduced in the 109th Congress, but it died in committee. *See*

search engine with carefully drafted terms of service, therefore, can largely immunize itself from user privacy suits. Even a violation of its own stated privacy policy may not expose it to significant legal risk.[46]

Even though courts may not obligate a search engine to keep queries private, is it permitted to do so against the legal demands of others? Answers vary. The Fourth Amendment and various statutes allow several procedures by which computers can be searched, balancing the government's showing of relevance to an investigation against user expectations of privacy.[47] A full search warrant, properly supported by an affidavit showing probable cause, will trump any expectation of privacy. Internationally, search engine operators (albeit not in their roles as search engines) have shown a willingness to comply with government demands for identifying data, even when the consequences for the identified users are severe.[48] Some have been proactive in cooperating with law enforcement.[49]

Where the demand comes from a private-sector third party, users may have more leverage. A traditional *subpoena duces tecum* issued to an intermediary, such as a search engine, affords the intermediary and the adversary in the underlying litigation an opportunity to object.[50] The court will balance the relevance of the information to the litigation against the burden on the recipient of the subpoena. Google recently resisted a Department of Justice subpoena for a random sample of user queries; it successfully argued that users' loss of trust in a search engine that releases their queries would constitute a burden.[51] Courts considering the use of subpoenas to learn the identity of particular users have developed tests that depend on the speech interests of the users,[52] although they have yet to

Eliminate Warehousing of Consumer Internet Data Act of 2006, H.R. 4731, 109th Cong. (introduced Feb. 8, 2006, referred to committee Feb. 17, 2006).

46. *See, e.g.*, *In re* Nw. Airlines Privacy Litig., No. 04-126, 2004 U.S. Dist. LEXIS 10580, at *15–17 (D. Minn. June 6, 2004) (dismissing a breach of contract claim); *In re* Geocities, 127 F.T.C. 94, 94 (1999) (entering an FTC consent order without fine or punishment).

47. *See generally* COMPUTER CRIME AND INTELLECTUAL PROP. SECTION, CRIMINAL DIV., U.S. DEP'T OF JUSTICE, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS (2002), *available at* http://www.cybersafe.gov/criminal/cybercrime/s&smanual2002.pdf.

48. *See, e.g.*, Anne Broache, *Google to Hand over Brazilian User Data*, CNET NEWS, Sept. 5, 2006, http://news.com.com/2061-10812_3-6112176.html; Philippe Naughton, *Yahoo Blamed for Jailing of Chinese Reporter*, TIMES (LONDON) ONLINE, Sept. 7, 2005, http://business.timesonline.co.uk/tol/business/industry_sectors/media/article563772.ece.

49. *See, e.g.*, Jonah Engle, *Buyer Beware: eBay Security Chief Turns Website into Arm of the Law*, THE NATION ONLINE, June 20, 2003, http://www.thenation.com/doc/20030707/engle.

50. *See* FED. R. CIV. P. 45(c).

51. Gonzales v. Google, 234 F.R.D. 674, 683–86 (N.D. Cal. 2006). AOL, Yahoo!, and Microsoft did not resist subpoenas served on them for similar data. *Id.* at 679.

52. *See, e.g.*, O'Grady v. Superior Court, 139 Cal. App. 4th 1423 (Cal. Ct. App. 2006) (granting reporters' motion to quash a subpoena served on an ISP to learn the identity of a confidential source); Sony Music Entm't Inc. v. Does 1–40, 326 F. Supp. 2d 556, 568 (S.D.N.Y. 2004) (requiring an ISP to disclose the identities of copyright-infringement defendants); Doe v.

determine what speech interests users have in anonymous search. Some courts have held that the statutory subpoena process of the Digital Millennium Copyright Act[53] does not apply to Internet Service Providers ("ISPs") that do not themselves host allegedly infringing files,[54] using reasoning that would appear to protect search engines as well.

One further possibility for protecting users' interest in privacy is self-help. Users might modify their searching behavior to prevent their queries from being correlated or linked back to them. Possible schemes involve anonymizing their connection to the search engine,[55] querying multiple search engines with partial queries and then correlating results,[56] and sending a cloud of off-topic queries to mask the true query of interest.[57] The practical effectiveness of these techniques remains questionable.[58]

2. Unbiased Results

Users turn to search engines to help them find useful, high-quality information. They therefore share search engines' interest in defeating SEO and promoting fair competition.[59] They want search engines to actually give them the highest-quality results that the engines are capable of delivering. Users are arguably harmed when a search engine gives some users higher-quality results than others or when it favors one content provider over another for reasons with which the users would disagree. Both could constitute bias, which Batya Friedman and Helen Nissenbaum have defined in this context as "systematic[] and unfair[] discriminat[ion] against certain individuals or groups of individuals in favor of others."[60]

2TheMart.com, Inc. 140 F. Supp. 2d 1088, 1097–98 (W.D. Wash. 2001) (requiring a showing that the need to learn the identity of a non-party is "directly and materially relevant to a core defense in the underlying . . . litigation").

53. Digital Millennium Copyright Act, 17 U.S.C. § 512(h) (2000).

54. *In re* Charter Commc'ns, Inc., Subpoena Enforcement Matter, 393 F.3d 771, 777–78 (8th Cir. 2005); Recording Indus. Ass'n of Am., Inc. v. Verizon Internet Servs., Inc., 351 F.3d 1229, 1238–39 (D.C. Cir. 2003).

55. *See* Ryan Singel, *How to Foil Search Engine Snoops*, WIRED.COM, Jan. 20, 2006, http://www.wired.com/news/technology/1,70051-0.html (recommending that users delete "cookies" each week or use a service that "masks the origins" of a query and evades filters).

56. *See generally* Benny Chor et al., *Private Information Retrieval*, 45 J. OF THE ACM 965 (1998), *available at* http://www.cs.technion.ac.il/~benny/PIR.pdf.

57. *See* Wakaha Ogata & Kaoru Kurosawa, *Oblivious Keyword Search*, 20 J. COMPLEXITY 356, 357 (2004); Daniel C. Howe & Helen Nissenbaum, TrackMeNot, http://mrl.nyu.edu/~dhowe/TrackMeNot/ (describing a program that obfuscates actual web searches).

58. *See, e.g.*, Bruce Schneier, *TrackMeNot*, SCHNEIER ON SECURITY (Aug. 23, 2006), http://www.schneier.com/blog/archives/2006/08/trackmenot_1.html ("Let's count the ways [TrackMeNot] doesn't work.").

59. *See infra* Part II.D.

60. Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFO. SYS. 330, 332 (1996) (emphasis omitted).

The first significant challenge here is to distinguish bias from simple failure by the search engine to do as well as it could have. Several problems make it difficult to set a proper baseline of "unbiased" results. As noted above, different users may have different needs and desires, a given query may reflect any number of different intentions, and even the user may not know what she is searching for when she queries a search engine. Some have even questioned whether users are the proper judges of what search results would be best for them.[61]

Even without a precise definition of "unbiased" results, some kinds of search decisions are troubling. Direct censorship—removing content from a search engine's index at the demand of a third party—is offensive to free-speech values.[62] In China, for example, at the demand of the government, major search engines try to prevent users from finding information on the banned Falun Gong movement.[63] A search engine could also bias its results by ranking favored content more highly and disfavored content more poorly; the disfavored content can still be found, but only if you're specifically looking for it. Both liberal and conservative groups have accused Google of bias toward the other in its advertising policies.[64] The concern is commercial as well as political: some have claimed (albeit without direct evidence) that search engines systematically favor their own advertisers and affiliated corporate providers.[65]

Technical design features of search engines can also introduce unconscious structural biases in their coverage and ranking of content.[66] John Hiler has suggested that search engines may rely too heavily on

61. *See* CASS SUNSTEIN, REPUBLIC.COM 8–10 (2001) (arguing that users will choose to see only content confirming their preexisting biases if they are given total control over their information inputs). *But see* NICHOLAS NEGROPONTE, BEING DIGITAL 163–65 (1995) (positing that users have tastes both for narrowly personalized information and for the common information seen by many others).

62. *See infra* Part II.C, and particularly Part II.C.4, for more on the issues raised by search censorship.

63. *See* Clive Thompson, *Google's China Problem (and China's Google Problem)*, N.Y. TIMES, Apr. 23, 2006, (Magazine), at 63, *available at* http://www.nytimes.com/2006/04/23/magazine/23google.html.

64. *See, e.g.*, *Google's Gag Order: An Internet Giant Threatens Free Speech*, PERRSPECTIVES, June 20, 2004, http://www.perrspectives.com/articles/art_gagorder01.htm (claiming apparent conservative bias); Rightmarch.com, Google says NO to Conservative Ads!, http://www.rightmarch.com/google.htm (claiming liberal bias). *But see* Eric Ulken, A Question of Balance: Are Google News Search Results Politically Biased? (May 5, 2005) (unpublished report, USC Annenberg School for Communication), *available at* http://ulken.com/thesis/googlenews-bias-study.pdf (claiming no bias toward either side in Google News selection of articles).

65. *See, e.g.*, Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, at App. A, http://infolab.stanford.edu/~backrub/google.html ("[W]e expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers."). *But see* Google, Why We Sell Advertising, Not Search Results, http://www.google.com/honestresults.html.

66. *See* Introna & Nissenbaum, *supra* note 3, at 175–76.

weblogs because weblogs tend to be frequently updated and contain many hyperlinks.[67] Studies of relative traffic and links to web sites have also caused some to discern a "Googlearchy," in which the most popular content receives more attention from users and therefore becomes even more popular, creating a popularity-feedback loop that new providers can never hope to break into.[68] Mathematical models and empirical studies both support and undercut this theory.[69] Separating cause from effect in the wildly uneven popularity of content has proven difficult, as have attempts to show whether these differences are undesirable or not.[70]

If users or regulators are concerned about bias, what might they do? Users are not usually in a position to evaluate the search engine's performance directly.[71] Users making navigational queries can often tell whether their intended destination is among the search results; users making transactional queries, however, may be less able to tell whether the engine has really directed them to the best sources. Some of this concern may be alleviated by users' ability to compare results from different search engines, although other search engines may suffer from the same systematic biases (think of the Chinese censorship and Googlearchy problems). Users might respond by demanding additional information from a search engine

67. John Hiler, *Google ♥ Blogs—How Weblogs Influence a Billion Google Searches a Week*, MICROCONTENT NEWS, Feb. 26, 2002, http://www.microcontentnews.com/articles/googleblogs.htm, *available at* http://web.archive.org/web/20030205011440/http://www.microcontentnews.com/articles/googleblogs.htm; *see also* Ulken, *supra* note 64 (claiming inclusion of "non-traditional news sources" in Google News creates increased prominence for extreme viewpoints). *Cf.* Anil Dash, *Nigritude Ultramarine*, DASHES.COM, June 4, 2004, http://www.dashes.com/anil/2004/06/04/nigritude_ultra (describing how collaboration among blog authors to link Dash's entry using keywords "nigritude ultramarine" nearly won an SEO contest).

68. Matthew Hindman et al., "Googlearchy": How a Few Heavily-Linked Sites Dominate Politics on the Web 1 (July 28, 2003) (unpublished manuscript), *available at* http://www.cs.princeton.edu/~kt/mpsa03.pdf.

69. *Compare* Junghoo Cho & Sourashis Roy, Impact of Search Engines on Page Popularity 1 (2004) (unpublished manuscript, *available at* http://oak.cs.ucla.edu/~cho/papers/cho-bias.pdf) ("New and valuable pages are ignored just because they have not been given a chance to be noticed by people."), *with* Santo Fortunato et al., *Topical Interests and the Mitigation of Search Engine Bias*, 103 PROC. NAT. ACAD. SCI. 12,684, Aug. 22, 2006, *available at* http://cxnets.googlepages.com/PNAS_engine.pdf ("Yet, despite the rich-get-richer dynamics implicit in the link analysis used to rank results, the use of search engines appears to mitigate the average traffic attraction of high-degree pages.").

70. *See* Clay Shirky, *Power Laws, Weblogs, and Inequality*, CLAY SHIRKY'S WRITINGS ABOUT THE INTERNET (Feb. 8, 2003), http://www.shirky.com/writings/powerlaw_weblog.html ("Inequality . . . is a reliable property that emerges from the normal functioning of the system."). Notably, even where scholars agree on the overall distribution of attention, they disagree on its implications. *Compare* Hindman et al., *supra* note 68 (x^{-1} powerlaw distribution is a sign of lack of diversity), *with* CHRIS ANDERSON, THE LONG TAIL 126–27 (2006) (x^{-1} powerlaw distribution is a sign of diversity).

71. *See, e.g.*, Diaz, *supra* note 3, at 147 ("The complexity and opacity of search technology makes it almost impossible for users to notice what is 'missing' from their search results.").

about its ranking algorithms, this time to understand why it has made the choices it has.[72]

Legally, search engines have a strong first line of defense against user suits for bias in their browsewrap terms of service.[73] Further, few business-tort theories provide users with enforceable rights. One exception may be the Federal Trade Commission's ("FTC") jurisdiction over misleading business practices. Although it has not taken direct action against any search engines, the FTC has communicated to search engines its belief that any paid placement results should be clearly disclosed and distinguished from organic search results.[74] In cases of conscious, human-directed manipulation of results, search engines may also have opened themselves to claims of fraud on consumers by emphasizing the mechanical and supposedly "objective" nature of their algorithms.[75]

Those who are concerned about systematic biases have also proposed various forms of forced ranking or inclusions. One proposal would require search engines to randomly intermix new content that has not yet had the time to establish itself with older and already-popular content.[76] Others would require search engines to show users more diverse content to break down their biases toward the familiar.[77] There is a strong counterargument, however, that regulators would be even more biased, as well as grossly incompetent, at the task of dictating search results.[78]

---

72. *But see* James Grimmelmann, Note, *Regulation by Software*, 114 YALE L.J. 1719, 1734–38 (2005) (observing that some algorithms are so complicated that it may not be possible to say "why" a computer made a particular decision).

73. *See, e.g.*, Hubbert v. Dell Corp., 359 Ill. App. 3d 976, 983–84 (Ill. App. Ct. 2005) (holding enforceable the arbitration clause in the "Terms and Conditions" on Dell's web site).

74. *See* Letter from Heather Hippsley, Acting Associate Director, F.T.C. Division of Advertising Practices, to [Search Engine Company] (June 27, 2002), *available at* http://www.ftc.gov/os/closings/staff/commercialalertattatch.htm.

75. *See* Rebecca Tushnet, *KinderStart: The Return*, 43(B)LOG (Sept. 20, 2006), http://tushnet.blogspot.com/2006/09/kinderstart-return.html ("If Google continues to tell searchers one thing about how search results are generated and tell webmasters another, it might behoove the FTC—the only entity with a realistic chance of affecting Google—to look into the matter."). *But see* Search King, Inc. v. Google Tech., Inc., No. CIV-02-1457-M, 2003 U.S. Dist. LEXIS 27193, at *11 (D. Okla. May 27, 2003) ("Google and Page's statements as to the purported objectivity of the PageRank system cannot transform a subjective representation into an objectively verifiable fact.").

76. *See* Sandeep Pandey et al., *Shuffling a Stacked Deck: The Case for Partially Randomized Ranking of Search Engine Results*, PROC. OF 31ST INT'L CONF. ON VERY LARGE DATABASES (VLDB) (2005), *available at* http://oak.cs.ucla.edu/~cho/papers/cho-shuffle.pdf.

77. *See* SUNSTEIN, *supra* note 61, at 182–89 (arguing that "must carry" rules could be an effective means of diversifying individual experiences); Susan L. Gerhart: *Do Web Search Engines Suppress Controversy?*, FIRST MONDAY, Jan. 2004, http://firstmonday.org/issues/issue9_1/gerhart/index.html (stating that a normal surfer would be "pulled" toward dominant sites instead of more diverse ones).

78. *See* Eric Goldman, *A Coasean Analysis of Marketing*, 2006 WIS. L. REV. 1151.

B. PROVIDERS' INTERESTS

1. Minimizing Costs

Turning to providers, we start with the first step in search: indexing. Every time a search engine asks to index content from a provider, the provider must use a little server time and network bandwidth to respond. Every time a user requests content from the provider, the provider must similarly expend a little server time and bandwidth. Multiply by a large collection of content, many search engines, frequent indexing, and many users—and many a mickle makes a muckle.[79]

Most of the time, providers willingly cooperate: getting lots of users from search engines is usually a good thing, well worth the technical costs. Still, some providers object. Their reasons vary. Providers may be bothered by inefficient search engines that index too often or with unwelcome burstiness. They may want to be available online to users but not to be searchable.[80] They may find that flows of user traffic from search engines don't include the sorts of users for whom they're looking.[81] Some of them may simply welcome any leverage they can employ to force search engines to pay for the privilege of indexing.[82]

Turning to the law, providers have attempted to defend their servers from unnecessary indexing burdens on three principal grounds. First, they have used the common-law tort of trespass to chattels. Despite some successes for this theory,[83] the 2003 California Supreme Court's heavily publicized holding in *Intel v. Hamidi* [84] (that the tort would not lie without a showing of harm to the chattel or the owner's ability to use it) would largely

79. In 2006, forty-six percent of all requests for pages from this author's web site came from the Yahoo! robot.

80. *See generally* danah boyd, *Facebook's "Privacy Trainwreck": Exposure, Invasion, and Drama*, Sept. 8, 2006, http://www.danah.org/papers/FacebookAndPrivacy.html (arguing that individuals may be more comfortable revealing personal information when they do not expect it to be easily searchable); *see also infra* Part II.B.2 (discussing unfair-competition reasons why providers may not like searchability).

81. In the so-called "Slashdot effect," being recommended by a popular web site produces a sudden and huge influx of traffic. Not only does the linked site face a potentially huge bandwidth bill, but it also runs the risk of seeing its servers crash under the load, meaning that it reaches none of the users it is trying to. For more on the Slashdot effect, see generally Jason Kottke, *Digg vs. Slashdot (or, Traffic vs. Influence)*, KOTTKE.ORG, Jan. 12, 2006, http://www.kottke.org/06/01/digg-vs-slashdot.

82. *See* Danny Sullivan, *Google's Belgium Fight: Show Me the Money, Not the Opt-Out, Say Publishers*, SEARCH ENGINE WATCH (Sept. 20, 2006), http://blog.searchenginewatch.com/blog/060920-152314 (analyzing the motivations of publishers suing Google to prevent indexing).

83. *See generally* Register.com, Inc. v. Verio, Inc., 356 F.3d 393 (2d Cir. 2004); Oyster Software, Inc. v. Forms Processing, Inc., No. C-00-0724, 2001 WL 1736382 (N.D. Cal. Dec. 6, 2001); eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058 (N.D. Cal. 2000).

84. Intel Corp. v. Hamidi, 71 P.3d 296, 308–09 (Cal. 2003).

preclude a trespass-to-chattels claim, unless the search engine was egregiously burdensome.[85]

Providers have had more luck under *sui generis* state and federal computer-intrusion statutes, which generally prohibit access to a computer system without authorization. Here, courts have been willing to say that any use of a server—including spidering—that the owner does not condone is *ipso facto* unauthorized. Despite academic criticism[86] and at least one court's discomfort with the theory,[87] courts considering claims against search engines under the federal Computer Fraud and Abuse Act's civil provisions[88] have held for the providers.[89]

Third, providers have at times alleged the existence of a contract prohibiting a search engine from spidering their content. In one notable 2004 case, the Second Circuit found that displaying a notice on information returned by a server forbidding certain uses of that information was sufficient to bind even a company accessing the information purely through a spidering program.[90] In reasoning directly applicable to search engines, the court held that the repeated access meant that knowledge of the purported contractual terms should be imputed to the spider's operator.[91] The leading case on consumer interactions with contracts presented on the web, *Specht v. Netscape Communications Corp.*, is in accord. The consumer in that case was not bound by the contract but only because the contract was not clearly displayed on the web page in a way that forced the user to see it or acknowledge it before clicking on a download link. The court left little doubt that a provider with sufficient willingness could craft terms and an interface for displaying them that would bind users who clicked through.[92]

None of these theories has been settled definitively. Some of the analyses are cursory at best, and none has found liability for a general-purpose search engine, rather than a specialized service that could be characterized as a direct competitor to the aggrieved provider. Nonetheless,

85. *See also* Ticketmaster Corp. v. Tickets.com, Inc., No. 99-7654, 2003 WL 21406289, at *3 (C.D. Cal. Mar. 7, 2003) (requiring a showing of actual harm "pending appellate guidance").

86. *See generally* Orin Kerr, *Cybercrime's Scope: Interpreting "Access" and "Authorization" in Computer Misuse Statutes*, 78 N.Y.U. L. REV. 1596 (2003).

87. *See* Lockheed Martin Corp. v. Speed, No. 6:05.cv-1580-Orl-31KRS, 2006 U.S. Dist. LEXIS 53108, at *18–19 (M.D. Fla. Aug. 1, 2006).

88. 18 U.S.C. § 1030(g) (2000).

89. *See* EF Cultural Travel BV v. Explorica, Inc., 274 F.3d 577, 585 (1st Cir. 2001); SW. Airlines Co. v. Farechase, Inc., 318 F. Supp. 2d 435, 438–41 (N.D. Tex. 2004).

90. *See* Register.com, Inc. v. Verio, Inc., 356 F.3d 393, 393 (2d Cir. 2004).

91. In one remarkable case, a court held that a plaintiff's allegations of a binding browsewrap contract could survive a motion to dismiss, even when the contract's terms were not displayed in a manner that would signal to a spider's creator that the provider was attempting to form a contract. *See* Internet Archive v. Shell, No. 06-cv-01726-LTB-CBS, 2007 U.S. Dist. LEXIS 10239, at *18–19 (D. Colo. Feb. 13, 2007).

92. Specht v. Netscape Comm. Corp., 306 F.3d 17, 35 (2d Cir. 2002).

the trend seems to be that a search engine could be prohibited from indexing an unwilling provider, at least on contractual grounds.

On the web, the matter appears to have settled into a rough equilibrium, with most providers using robot exclusion protocols to inform the public which robots are allowed to spider which portions of their content. Most major search engines respect these requests.[93] This informal compromise does not directly bind users, however, because the norms of robot exclusion protocols apply only to the operators of indexing robots.[94]

So much for indexing. Compare the legal regime governing excessive user attention. It is well-established that deliberately orchestrated denial-of-service attacks are both crimes and torts,[95] and normal principles of co-conspirator liability suggest that recruiting thousands of others to act in concert to attack a computer system might make the recruiter (and quite possibly each user) liable for all resulting harm.[96] The more interesting cases involve intermediaries (such as search engines) that overwhelm a site by directing to it users who don't know they're part of an online mob.

At least one case has held that a person who deliberately manipulates the inputs to an information-location tool so that a provider is incorrectly listed may be liable in trespass to chattels.[97] Similarly, a search engine that misleads users as to what they will find may be indirectly intermeddling and, thus, also liable in trespass to chattels.[98] A substantial gray area remains. On the one hand, any result returned by a search engine involves an intent to cause the user to access the site (and thus to use the chattel of the server); on the other, even an intention by a search engine to cause harm to a site may not be accompanied by any deception of users.[99]

93. *See* Patricia Bellia, *Defending Cyberproperty*, 79 N.Y.U. L. REV. 2164, 2250–51 (2004) (arguing that this result is, for the most part, efficient and just).

94. *See* Eric J. Feigin, *Architectures of Consent: Internet Protocols and Their Legal Implications*, 56 STAN. L. REV. 901, 916–17 (2004) (arguing that Internet norms of access and restriction embedded in technical protocols are entitled to legal respect).

95. *See* Computer Fraud and Abuse Act, 18 U.S.C. § 1030(c)(2)(B)(ii) (2000); United States v. Ancheta, No. CR 05-1060 (C.D. Cal. Feb. 2005) (indictment), *available at* http://fl1.findlaw.com/news.findlaw.com/hdocs/docs/cyberlaw/usanchetaind.pdf.

96. *See* U.S. SENTENCING GUIDELINES MANUAL § 1B1.3(a)(1)(B) (2006) (stating that the base offense level for jointly undertaken activity is based on "all reasonably foreseeable acts and omissions of others in furtherance of the jointly undertaken criminal activity").

97. Sch. of Visual Arts v. Kuprewicz, 771 N.Y.S.2d 804, 808 (N.Y. Sup. Ct. 2003) (holding that one who subscribed another to high-volume e-mail lists could be liable in trespass to chattels).

98. *See* RESTATEMENT (SECOND) OF TORTS § 217 cmt. e (1965) ("[A] trespass may be committed by causing a third person through duress or *fraud* to intermeddle with another's chattel." (emphasis added)).

99. On the difficulties involved in assessing intent and effects, compare Universal Tube & Rollform Equip. Corp. v. YouTube, Inc., No. 3:06CV02628, 2007 U.S. Dist. LEXIS 40395, at *19–20 (N.D. Ohio June 4, 2007) (holding that choosing the "youtube.com" domain name does not involve "duress or fraud" in causing users to visit "utube.com"), with Dave Plonka, *Flawed Routers Flood University of Wisconsin Internet Time Server* (Aug. 21, 2003), http://www.cs.wisc.edu/

One final cost requires discussion: Money spent on search engine advertising could have been spent on some other useful cause. Search advertising campaigns are increasingly managed like any other advertising campaign, often as part of a comprehensive marketing strategy—and with the same advertiser pressure for lower rates.[100] Since users could also be directed to a provider through organic search results, search advertising is a partial substitute for search rankings. There have been occasional accusations that search engines churn their rankings or deliberately demote some providers to spur them to purchase search advertising.[101]

2. Avoiding Unfair Competition

Search engines can help users to acquire content in new ways. Some of those ways disrupt providers' preferred relationships with users. Does the fact that the content originally came from providers give them a right to demand that search engines only present it to users in ways that providers approve of? Answering such a question is the domain of unfair-competition law, which defines the limits of businesses' rights to prevent others from diverting customers. Search engines raise questions that blur the boundaries between copyright, trademark, and common-law theories of misappropriation.[102] This Section analyzes unfair-competition arguments as they apply to the different information flows in search, but the area is sufficiently unstable that this taxonomy is a matter of convenience rather than settled law.

Start with indexing. A search engine's spidering processes require making at least one initial copy of any content the engine wishes to index; providers have complained that this initial copy is unauthorized and hence

~plonka/netgear-sntp/ (discussing a company that shipped thousands of home networking devices that accessed a university server for its intended purpose but in far greater numbers than appropriate).

100. *See generally* Search Engine Marketing Professional Organization ("SEMPO"), About SEMPO, http://www.sempo.org/about.

101. Evaluating such charges is difficult, given the secrecy surrounding search ranking algorithms. Independent search analyst Danny Sullivan argues plausibly that the churn in rankings is a side effect of anti-SEO efforts, rather than a deliberate strategy. *See* Danny Sullivan, *What Happened to My Site on Google?*, SEARCH ENGINE WATCH, Dec. 7, 2003, http://searchenginewatch.com/showPage.html?page=3286101 ("[T]here are far easier ways that Google could boost ad revenue uptake without doing sneaky, behind-the-scene actions . . . .").

102. The difference between the unfair-competition concerns discussed in this Section, Part II.B.2, and the unwanted-access concerns discussed above, *supra* Part II.B.1, is that unfair competition deals with a loss of *human* attention from users, while unwanted access has to do with *technical* burdens. The difference between unfair-competition concerns and the placement concerns discussed below, *infra* Part II.B.3, is that the unfair-competition concerns are about competition from *search engines*, while the placement concerns are about competition with *other providers*. In the former, the search engine itself steals users; in the latter, it shows undue favoritism among providers.

infringes the provider's copyright. This is the principal doctrinal peg on which the Authors' Guild hangs its copyright suit against the Google Book indexing project.[103] Because indexing by itself doesn't involve users and thus harms only an author's abstract interest in controlling her work, courts have tended to find it fair if they approve of the subsequent purposes to which the engine puts the copies—and unfair if they don't.[104] Providers, however, have begun to argue that even copies retained but not shown to users are troublesome because of the risk of a security breach exposing the archives to bulk copying.[105]

Next, consider the case in which the search engine directly provides users with content. We could describe this provision in copyright terms; the search engine is arguably making and publicly distributing unauthorized copies.[106] Unauthorized caching suggests infringement; providers who sell the content or show it with advertising see their business models disrupted by a search engine that serves up the content.[107] Major search engines generally honor requests not to cache, but they have forced providers to use standard technical measures to make those requests.[108] Giving users content

103. *See* Complaint at 2, Authors' Guild v. Google, Inc., No. 05-CV-8136 (S.D.N.Y. Sept. 20, 2005), *available at* http://pub.bna.com/eclr/05cv8136comp.pdf (calling indexing "massive copyright infringement").

104. *See* Kelly v. Arriba Soft Corp., 336 F.3d 811, 822 (9th Cir. 2003) (finding fair use). *But see* UMG Recordings, Inc. v. MP3.com, Inc., 92 F. Supp. 2d 349, 352–53 (S.D.N.Y. 2000) (finding no fair use). In these cases, the line between providers and third parties is at its least distinct. *See, e.g.*, Sony Computer Entm't, Inc. v. Connectix Corp., 203 F.3d 596, 608 (9th Cir. 2002) (finding fair use in initial copying for purposes of reverse engineering).

105. *See Fair Use: Its Effects on Consumers and Industry: Hearing Before the H. Subcomm. on Commerce, Trade, and Consumer Protection of the H. Comm. on Energy and Commerce*, 109th Cong. 66 (2005) (testimony of Paul Aiken, Executive Director, Authors Guild) ("Since there's no license needed, in Google's view, Google doesn't have to give rightsholders contractual assurances of the security of their database.").

106. *See* 17 U.S.C. § 106(1), (3) (2000) (providing certain exclusive rights to copyright owners, including reproduction and distribution).

107. *Cf.* Twentieth Century Fox Film Corp. v. iCraveTV, No. 00-121, 2000 U.S. Dist. LEXIS 1013, at *2 (W.D. Pa. Jan. 28, 2000) (granting an injunction against Internet sites framing TV stations' live feeds with advertisements and rebroadcasting them on the Internet).

108. *See* Field v. Google, Inc., 412 F. Supp. 2d 1106, 1122–23 (D. Nev. 2006); Internet Archive, Removing Documents from the Wayback Machine, http://www.archive.org/about/exclude.php. *But see* Rebecca Bolin, *Locking Down the Library*, 29 HASTINGS COMM. & ENT. L.J. 1, 9–19 (2006) (arguing that copyright-related removal from online archives threatens preservation of human memory). Providers and search engines have both been developing technical measures to improve the precision of signals in this area. *See* Jeffrey Goldfarb, *Publishers Aim for Some Control of Search Results*, INT'L BUS. TIMES, Sept. 22, 2006, http://ibtimes.com/articles/20060922/global-publishers-google-search-engines-copyright.htm (detailing an attempt by publishers to develop an automated system to grant content-use permission); The Web Robots Pages, The Web Robots FAQ, http://www.robotstxt.org/wc/faq.html. *But see* Niva Elkin-Koren, *What Contracts Cannot Do: The Limits of Private Ordering in Facilitating a Creative Commons*, 74 FORDHAM L. REV. 375, 407–17 (2005) (arguing that increased precision of permissions increases the perception that permission is required for any reuse).

directly may also interfere with proper attribution by severing the link between source and content. This objection appeals to trademark policies,[109] but trademark causes of action, ironically, may be hemmed in by copyright.[110]

The issue becomes murkier in the case of thumbnailing and other practices that give users only excerpts or summaries of content. Some authority holds that the thumbnails are protected fair use;[111] this holding is not clearly established, however, and drawing the line between thumbnail and full copy (or derivative work) may require significant case-by-case analysis.[112] Some providers, moreover, have argued that thumbnailing—or even offering search itself—involves the exploitation of value produced by content and properly attributable to that content.[113] Precisely because this related market is valuable, goes the argument, the content owners should have the exclusive right to exploit it.[114]

Even when users obtain content directly from providers, the search engine may change *how* they obtain it, through techniques such as framing, deep linking, and inlining. If the engine provides the user with specific technical instructions for obtaining content, the user may experience the content in a context not intended by providers, even though the provider itself supplied the content to the user upon request. Providers have been particularly upset when search engines cause users to bypass advertising on the provider's site.[115] United States courts have not been able to agree either

109. *See* 15 U.S.C. § 1125(a)(1) & (a)(1)(A) (2000) (making actionable a "false designation of origin . . . likely to cause confusion").

110. *See generally* Dastar Corp. v. Twentieth Century Fox Film Corp., 539 U.S. 23 (2003) (holding that authorship is not "origin" for purposes of Section 1125(a)). *But see* F. Gregory Lastowka, *The Trademark Function of Authorship*, 85 B.U. L. REV. 1171, 1172 (2005) ("[T]he Dastar approach is misguided.").

111. *See* Kelly v. Arriba Soft Corp., 336 F.3d 811, 822 (9th Cir. 2003).

112. *See* Perfect 10 v. Amazon.com, 487 F.3d 701 (9th Cir. 2007) (holding thumbnailing likely to be a fair use as against a third-party copyright holder).

113. *See, e.g.*, Nick Taylor, . . . *But Not at Writers' Expense*, WASH. POST, Oct. 21, 2005, at A18 ("The alphabet ought to be free, most certainly, but the people who painstakingly arrange it into books deserve to be paid for their work."). *But see* Emily Anne Proskine, Note, *Google's Technicolor Dreamcoat: A Copyright Analysis of the Google Book Search Library Project*, 21 BERKELEY TECH. L.J. 213, 231–32 (2006) ("None of copyright's exclusive rights suggest that publishers or authors should possess a monopoly over the indexing and searching of their works.").

114. *But see* Rochelle Dreyfus, *Expressive Genericity: Trademarks as Language in the Pepsi Generation*, 65 NOTRE DAME L. REV. 397, 405 (1990) ("Furthermore, fallacies in the fundamental assumptions made by courts that have approved this 'if value, then right' theory mean that the right lacks a coherent limit.").

115. *See, e.g.*, Complaint at ¶36, Wash. Post Co. v. TotalNews, Inc., No. 97 Civ. 1190 (S.D.N.Y. Feb. 20, 1997), *available at* http://legal.web.aol.com/decisions/dlip/washcomp.html ("Yet an advertisement on one of Plaintiffs' sites, when seen through the totalnews.com window, is reduced in size, may even be totally obscured by the totalnews frame, and is forced to compete for the user's attention with the visual clutter of the totalnews.com frame, including other advertising – possibly including advertising for directly competitive products.").

on the copyright[116] or trademark[117] implications of these techniques, but European courts have been willing to find that deep linking to news violates the European Database Directive.[118] An additional set of complications arises when the relevant content is uncopyrightable (e.g., because it consists of unprotectable fact, rather than protectable expression). Here, providers may wish to complain both of direct provision and altered presentation, but they must steer clear of copyright preemption.[119]

Finally, these questions cannot be addressed without considering users' rights. Not only do users enjoy significant statutory and fair-use rights to make copies for their own caching and archival purposes,[120] but there are strong arguments that many of the transformations to which search engines subject content would be wholly legal if carried out by users directly. The role of the search engine as an intermediary carrying out those activities for them raises deeply contentious issues of intellectual-property policy.[121]

116. *See* Kelly v. Arriba Soft Corp., 280 F.3d 934, 944–47 (9th Cir. 2002), *withdrawn by*, 336 F.3d 811 (9th Cir. 2003) (finding that deep linking and framing infringe on the exclusive public-display right).

117. *See* Digital Equip. Corp. v. AltaVista Tech., 960 F. Supp. 456, 478–79 (D. Mass. 1997) (enjoining the defendant from linking to the plaintiff and creating a false impression of affiliation); Knight-McConnell v. Cummins, No. 03 Civ. 5035 (NRB), 2004 U.S. Dist. LEXIS 14746, at *7 (S.D.N.Y. July 29, 2004) ("The mere appearance on a website of a hyperlink to another site will not lead a web-user to conclude that the owner of the site he is visiting is associated with the owner of the linked site.").

118. *See* Copiepresse v. Google, Inc., Tribunal de Premiere Instances de Bruxelles [Court of First Instance] Brussels, 13 février 2007, J.B.C. 7964 (Belg.) (prohibiting deep linking), *available at* http://www.copiepresse.be/13-02-07-jugement-en.pdf; Press Release, Newsbooster.com, Newsbooster Keeps on Fighting (July 12, 2002), *available at* http://www.chillingeffects.org/news.cgi?NewsID=199 (discussing Danish Newspaper Publishers' Association v. Newsbooster.com (2002) (Denmark), which prohibited Newsbooster.com's deep linking). *But see* Zoekallehuizen.nl/NVM, Rechtbank Arnhem [Court Arnhem], 16 maart 2006, LJN AV5236 (Neth.) (allowing deep linking), *available at* http://zoeken.rechtspraak.nl/resultpage.aspx?snelzoeken=true&searchtype=ljn&ljn= AV5236&u_ljn=AV5236. For a discussion of international cases relating to deep linking, see Linking Cases—Deep Links/Search Engines, http://www.linksandlaw.com/linkingcases-deeplinks-3.htm (summarizing and linking to Home v. Ofir (2006) (Denmark) (allowing deep linking), Zoekallehuizen.nl v. NVM (2006) (Neth.) (same), and Bixee.com v. Nahuri.com (2006) (Ind.) (prohibiting deep linking)).

119. *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., Inc., 499 U.S. 340, 364 (1991) (holding that a phone book was not original enough to be protected by copyright); Ticketmaster Corp. v. Tickets.com, Inc., No. CV 99-7654 HLH(BQRX), 2000 WL 525390, at *2, *4 (C.D. Cal. March 27, 2000) (applying *Feist*).

120. *See* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 454–55 (1984) (holding that a VCR owner's taping of television shows is fair use). *See generally* Jessica Litman, *Lawful Personal Use*, 85 TEX. L. REV. 1871 (2007) (discussing personal uses of copyrighted material, such as private home copying).

121. *See, e.g.*, Marvel Enters. v. NCSoft Corp., No CV 04-9253-RGK (PLAx), 2005 U.S. Dist. LEXIS 8448, at *11–17 (C.D. Cal. March 9, 2005) (considering copyright and trademark implications of actions by users that would have been lawful private uses offline); Paramount v.

3. Prominent Placement in Results

Prominent search-result placements carry immense value. Users are more likely to click on the first result than the second, the second than the third, and so on. If you don't appear on the first few pages of results, you may as well not exist. Accordingly, providers seek the highest possible placement. Some turn to SEO. Others turn to the law.

When a search engine returns result R in response to query Q, it effectively asserts that "R is a good source for information about Q," or perhaps that "R is a good place to acquire Q." Such assertions implicate the policies of trademark and advertising law, which seek to keep consumers from being confused by misleading claims about the relationship between businesses and the goods they offer.

Trademark law has taken a strong interest in Internet information-location tools. Trademark holders have been able to appeal to ordinary principles of trademark law, as well as to various *sui generis* legal regimes dealing with domain names,[122] against those who use confusing domain names to draw away consumers.[123] Courts have split on whether purchasing search advertisements tied to queries on a competitor's trademark can constitute infringement.[124] In a decision that has attracted substantial criticism,[125] the Ninth Circuit held that the use of trademarks in hidden metatags (an early, albeit probably ineffective, form of SEO) could be actionable as causing "initial interest confusion."[126]

Three years later, the Ninth Circuit extended the doctrine so that liability could run against search engines themselves.[127] The court reasoned

ReplayTV, 298 F. Supp. 2d 921, 923 (C.D. Cal. 2004) (objecting to a feature of a personal video recorder allowing consumers to skip commercials automatically).

122. *See* Internet Corporation for Assigned Names and Numbers, Uniform Domain Name Dispute Resolution Policy ("UDRP") (1999), http://www.icann.org/udrp/udrp-policy-24oct99.htm; *see also* Anti-Cybersquatting Consumer Protection Act ("ACPA"), 15 U.S.C. § 1125(d) (2000); *id.* § 1117. *See generally* Lamparello v. Falwell, 420 F.3d 309 (4th Cir. 2005).

123. *See generally, e.g.*, People for the Ethical Treatment of Animals v. Doughney, 263 F.3d 359 (4th Cir. 2001). *But see Lamparello*, 420 F.3d at 316–17 (distinguishing *Doughney*).

124. *Compare* Edina Realty, Inc. v. TheMLSonline.com, No. Civ. 04-4371, 2006 WL 737064, at *7 (D. Minn. Mar. 20, 2006) (yes), *and* Buying for the Home, LLC v. Humble Abode, LLC, 459 F. Supp. 2d 310, 332 (D.N.J. 2006) (yes), *and* J.G. Wentworth, S.S.C. v. Settlement Funding LLC, No. 06-0597, 2007 U.S. Dist. LEXIS 288, at *8 (E.D. Pa. Jan. 4, 2007) (yes), *with* Site Pro-1 v. Better Metal, LLC, No. 06-6508, 2007 U.S. Dist. LEXIS 34107, at *5–6 (E.D.N.Y. May 9, 2007) (no), *and* Merck & Co., Inc. v. Mediplan Health Consulting, Inc., 425 F. Supp. 2d 402, 415–16 (S.D.N.Y. 2006) (no), *and* Rescuecom Corp. v. Google, Inc., 456 F. Supp. 2d 393, 403–04 (N.D.N.Y. 2006) (no).

125. *See* Eric Goldman, *Deregulating Relevancy in Internet Trademark Law*, 54 EMORY L.J. 507, 565 (2005); *see also* Playboy Enters., Inc. v. Netscape Commc'ns Corp., 354 F.3d 1020, 1034–36 (9th Cir. 2004) (Berzon, J., concurring).

126. Brookfield Commc'ns, Inc., v. W. Coast Entm't Corp., 174 F.3d 1036, 1066 (9th Cir. 1999).

127. *Playboy*, 354 F.3d at 1024–29.

that a search engine's use of a trademarked term as an advertising keyword was a use of the trademark capable of causing confusion, whether or not the keywords were used in the advertisements themselves.[128] By way of contrast, other courts have held that adware vendors, who sell software that displays advertisements based on the web pages visited by users, do not violate the trademark rights of the providers of the underlying web pages.[129]

These precedents provide conflicting guidance on the obligations of search engines that trigger advertisements based on search queries containing trademarked terms, and search engines have vacillated in their policies on selling such advertisements.[130] Google, which has been the most vigorous of the major search engines in contesting suits by trademark holders, has both won and lost motions to dismiss on nearly identical facts.[131] Utah has amended its trademark law to prohibit many keyword-triggered advertisements,[132] although the prohibition raises substantial constitutional and preemption issues.[133] So far, no search engine has been sued for returning a competitor's web page as an organic search result in a search on a trademarked term, but such suits cannot be ruled out. Some

128. *Id.* at 1025–26.

129. 1-800-Contacts, Inc. v. WhenU.com, Inc., 414 F.3d 400, 410–12 (2d Cir. 2005); Wells Fargo & Co. v. WhenU.com, Inc., 293 F. Supp. 2d 734, 772 (E.D. Mich. 2003); U-Haul Int'l, Inc. v. WhenU.com, Inc., 279 F. Supp. 2d 723, 728–29 (E.D. Va. 2003). These cases also raised copyright theories tied to the allegedly altered display of the web pages.

130. *Compare* Google, AdWorlds Trademark Complaint Procedure, http://www.google.co.uk/tm_complaint_adwords.html (forbidding only the use of the trademark in ad text), *with* Yahoo! Search Marketing, Trademarks, http://searchmarketing.yahoo.com/legal/trademarks.php (forbidding bidding on trademarked terms except under stated conditions).

131. *See generally* Google, Inc. v. Am. Blind & Wallpaper Factory, Inc., No. C 03-5340 JF CRSJ, 2007 U.S. Dist. LEXIS 32450 (N.D. Cal. Apr. 18, 2007) (allowing a theory of infringement for use in keywords to proceed on motion for summary judgment); Gov't Employees Ins. Co. v. Google, Inc., No. 01:04cv507, 2005 U.S. Dist. LEXIS 18642 (E.D. Va. Aug. 8, 2005) (finding for the search engine in a bench trial); Google, Inc. v. Am. Blind & Wallpaper Factory, Inc., No. 03-05340, 2005 U.S. Dist. LEXIS 6228 (N.D. Cal. Mar. 30, 2005) (allowing a theory of infringement for use in keywords to proceed on motion to dismiss); Gov't Employees Ins. Co. v. Google, Inc., 330 F. Supp. 2d 700 (E.D. Va. 2004) (dismissing a theory of trademark infringement for use in advertising keywords but allowing a theory of infringement for use in advertising text to proceed). *See also* 800-JR-Cigar, Inc. v. GoTo.com, 437 F. Supp. 2d 273, 297 (D.N.J. 2006) (allowing a keyword-use theory to proceed); Elinor Mills, *Google Loses French Trademark Lawsuit*, CNET NEWS, June 28, 2006, http://www.news.com/2100-1030_3-6089307.html (reporting on similar French decisions in favor of trademark holders).

132. Trademark Protection Act, § 70-3a-402(1)(c), S.B. 236, 2007 Gen. Sess. (Utah 2007), *available at* http://le.utah.gov/~2007/bills/sbillamd/sb0236.htm.

133. *See* Eric Goldman, *Utah Bans Keyword Advertising*, TECH. & MKTG. L. BLOG (Apr. 3, 2007), http://blog.ericgoldman.org/archives/2007/04/utah_bans_keywo.htm (stating that the law raises Dormant Commerce Clause and First Amendment concerns).

commentators have proposed that a similar result be required by regulation.[134]

Providers aggrieved by their poor placement have also sued search engines for business libel and related theories.[135] The providers generally have argued either that (1) their content was highly relevant by any objective standard and that a poor ranking is, in effect, a lie, or (2) the search engine reduced the providers' rankings out of malice.[136] These state-law claims have not fared well in court. It may not have helped some providers' legal cases that their prelitigation actions looked a lot like black-hat SEO.[137]

C. THIRD PARTIES' INTERESTS

We turn now to third parties' interests in suppressing certain content flows. Bringing these third parties into the picture complicates the politics. The flow of content from providers to search engines to users will simply stop if any of those three groups pulls out entirely. They may argue over the division of spoils and jockey for relative advantage, but they are united in wanting to use search to make information more usefully accessible. Third parties, on the other hand, object to the content flow itself. This creates an adversarial dynamic: third party versus everyone else—with some interesting fault lines within the "everyone else" coalition.

1. Ownership

We begin with intellectual-property interests. Here, providers and users have a common interest in the flow of infringing content. Search engines have a business interest in serving them, counterbalanced by an interest in avoiding implication in the infringement. Note that there is some doctrinal

134. *See, e.g.*, Pasquale, *supra* note 3, at 136 (proposing that trademark holders be allowed to place an asterisk next to unsanctioned search results); James A. Rossi, *Protection for Trademark Owners: The Ultimate System of Regulating Search Engine Results*, 42 SANTA CLARA L. REV. 295, 347–54 (proposing that search engines should always offer the option of returning results that the holder of the trademarked search term selects).

135. *See generally* KinderStart.com, LLC v. Google, Inc., No. C 06-2057 JF CRSJ, 2007 U.S. Dist. LEXIS 22637, at *3 (N.D. Cal. Mar. 16, 2007); Langdon v. Google, Inc., 474 F. Supp. 2d 622, 626 (D. Del. 2006); Roberts v. Google, Inc., No. 1-06-CV-063047 (Cal. Sup. Ct. May 12, 2006) (voluntary dismissal); Search King, Inc., v. Google Tech., Inc., No. CIV-02-1457-M, 2003 U.S. Dist. LEXIS 27193, at *3 (W.D. Okla. May 27, 2003); Complaint, Datner v. Yahoo! Inc., No. BC355217 (Cal. Sup. Ct. July 11, 2006); Complaint, CLRB Hanson Industries, LLC v. Google, Inc., No. 1-05-CV-046409 (Cal. Sup. Ct. Aug. 3, 2005).

136. *Search King, Inc.*, 2003 U.S. Dist. LEXIS 27193, at *8–9.

137. *See* Dahlia Lithwick, *Google-Opoly: The Game No One but Google Can Play*, SLATE, Jan. 29, 2003, http://www.slate.com/id/2077875/ ("SearchKing in effect has its clients collude to trick Google into boosting everyone's ratings."). *Cf.* David Kesmodel, *Blogger Faces Lawsuit Over Comments Posted by Readers*, WALL ST. J. ONLINE, Aug. 31, 2005, http://online.wsj.com/pubic/article/SB112541909221726743-_vX2YpePQV7AOIl2Jeebz4FAfS4.20060831.html?mod=blogs (describing a defamation lawsuit against a blogger whose commenters criticized the plaintiff's SEO tools).

overlap between these claims by third parties and the unfair-competition claims providers might bring against search engines.[138]

Starting, as above, with indexing, the initial copies a search engine makes while spidering seem to be unambiguous fair uses.[139] So far, search engines have also been doing well at the other end of the chain: courts have held that neither linking to infringing content nor framing it constitutes direct infringement and that thumbnailing is a fair use.[140] Caching and archiving are riskier; most major search engines behave cautiously, removing allegedly infringing content from their caches.[141]

Secondary copyright liability is murkier. *Grokster* teaches that the makers of infringement-facilitating technologies must both pass the *Sony* staple-article-of-commerce test and steer clear of purposeful, culpable inducement of infringement.[142] Given that Grokster and its brethren consisted of a search application fused with a file-transfer application, this holding applies directly to search engines.[143] The principal web search engines easily pass the "capable of substantial noninfringing uses" prong of the *Sony* test, but more specialized search engines may not. As for purposeful, culpable inducement, the application of this language to many technologies, not just search engines, remains unclear.[144]

The immunities and subpoena processes detailed in Section 512 of the Digital Millennium Copyright Act ("DMCA")[145] also will be significant in search engine copyright litigation. ISP litigation has clarified some of the issues, but the courts have not yet extensively glossed Section 512(d), the immunity for "Information Location Tools." That section incorporates by

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

139. *See* Kelly v. Arriba Soft Corp., 280 F.3d 934 (9th Cir. 2002), *withdrawn by* 336 F.3d 811, 822 (9th Cir. 2003) (finding that reproduction of thumbnail images is a fair use).

140. *See* Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701 (9th Cir. 2007).

141. *See, e.g.*, MSN.com, Site Owner Help: Control Which of Your Pages are Indexed, http://search.msn.com.sg/docs/siteowner.aspx?t=SEARCH_WEBMASTER_REF_RestrictAccessToSite.htm. ("Prevent MSNBot from caching a page"). *But see* 17 U.S.C. § 512(b) (2000) (providing a safe harbor for "system caching"); Parker v. Google, Inc., 422 F. Supp. 2d 492 (E.D. Pa. 2006), *aff'd*, No. 06-3074, 2007 WL 1989660, at *3 (3d Cir. July 10, 2007) (per curiam) (holding that archiving is not "volitional" and, therefore, is not copyright infringement).

142. MGM Studios, Inc. v. Grokster, Ltd., 545 U.S. 913, 935 (2005). *Sony* held that one who manufactures and distributes a technology will not be liable for infringements committed by its users as long as the technology has "substantial noninfringing uses." Sony Corp. v. Universal City Studios, Inc., 464 U.S. 417, 456 (1984). *Grokster* added to that test a requirement that the maker must not have acted with the intent of inducing its users to infringe. *Grokster*, 545 U.S. at 935.

143. *See Grokster*, 545 U.S. at 921 (describing Grokster's search function).

144. *Compare id.* at 919 ("[O]ne who distributes a device with the object of promoting its use to infringe copyright, as shown by *clear expression or other affirmative steps taken to foster infringement*, is liable" (emphasis added)), *with* MGM Studios, Inc. v. Grokster, Ltd., 454 F. Supp. 2d 966, 985 (C.D. Cal. 2006) ("Plaintiffs need prove only that StreamCast distributed the product with the intent to encourage infringement.").

145. Digital Millennium Copyright Act, 17 U.S.C. § 512 (2000).

reference the ISP-focused notice-and-takedown procedure for "Information Residing on Systems or Networks at Direction of Users" described in Section 512(c), and search engines have been diligent about removing links for which they receive notices alleging copyright infringement.[146] But the parallel is not exact. On the one hand, because search engines do not have the direct relationship with users that hosting services have, providers are more vulnerable to abuse of the notice-and-takedown process. On the other hand, because the search engine's role in such cases is generally only to link to information, and because the notices must specify the "reference or link, to material or activity claimed to be infringing," search engines can generally undermine a takedown notice by displaying the notice itself.[147] It also remains unclear how other provisions of Section 512, such as the termination-of-repeat-infringers requirement of Section 512(i) and the subpoena-to-identify-infringers provision of Section 512(h), apply to search engines.[148]

Cutting across all of these copyright issues are the general problems of what it means for a search engine to have "knowledge" of infringement and the extent to which a search engine profits from infringing activity or can control infringers.[149] Napster was charged with knowledge of particular infringing MP3 files based on notifications from copyright holders.[150] Identification of infringing providers is also a difficult issue, given that such identifications may effectively be ex parte—in many cases, there's no good way for a content provider to learn of or respond to an accusation of infringement. Also, the assessment of vicarious liability will involve some close scrutiny into search business models; it appears that Google's affiliate-

146. Some providers, both in the ISP and search engine contexts, have used such notices to enforce non-copyright-related desires to suppress the material. *See* Online Policy Group v. Diebold, Inc., 337 F. Supp. 2d 1195, 1198–1204 (N.D. Cal. 2004) (finding "material misrepresentation" of infringement where posted "embarrassing" content was "not subject to copyright protection").

147. Joshua Urist, *Who's Feeling Lucky? Skewed Incentives, Lack of Transparency, and Manipulation of Google Search Results Under the DMCA*, 1 BROOK. J. CORP. FIN. &. COMM. L. 209, 219 (2006) (arguing that such Section 512(d) takedown requests should be publicly archived). *See* Google, Digital Millennium Copyright Act, http://www.google.com/dmca.html ("Please note that in addition to being forwarded to the person who provided the allegedly infringing content, a copy of this legal notice may be sent to a third-party partner for publication and annotation.").

148. Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102, 1109–15, 1116–17 (9th Cir. 2007) (discussing Section 512(d) and Section 512(i)).

149. The question of knowledge is relevant both in the standard for contributory infringement and directly under Section 512; the ability to control infringers is relevant in vicarious infringement and directly under Section 512.

150. *See* A&M Records v. Napster, Inc., 284 F.3d 1091, 1098–99 (9th Cir. 2002).

network advertising might be decisive in finding vicarious liability under the right circumstances.[151]

Third-party trademark rights in content (as distinguished from trademark rights in queries) are generally less dangerous for search engines. Direct liability in trademark law for trademark use on providers' pages seems unlikely, given the trend in the keyword-advertising cases. Secondary liability based on tests paralleling those in copyright might be argued by analogy to the offline "swap meet" cases.[152] eBay, the most like a swap meet of the major search engines, has a rigorous trademark-protection policy and will take down an auction based on a complaint from a trademark holder.[153]

2. Reputation

How different things are when the content flows are defamatory, rather than infringing! Here, search engines are protected by Section 230 of the Communications Decency Act, which gives any "provider . . . of an interactive computer service" blanket immunity from being treated as the "speaker of any information provided by another information content provider or user."[154] Moreover, while search engines need not filter such material, they are also immunized if they voluntarily, "in good faith," remove material that they believe to be "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable."[155] Except in exceptional circumstances, a party has no recourse against a search engine that facilitates the distribution of personally harmful or defamatory content.[156]

---

151. *See* Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 711–12, 729–32 (9th Cir. 2007) (noting that Google placed affiliate advertisements on sites allegedly hosting infringing images).

152. *See, e.g.*, Hard Rock Cafe Licensing Corp. v. Concession Serv., Inc., 955 F.2d 1143, 1148–50 (7th Cir. 1992) (discussing contributory and vicarious trademark liability for the operator of a flea market where counterfeit goods are sold).

153. *See* eBay, How eBay Protects Intellectual Property (VeRO), http://pages.ebay.com/help/tp/programs-vero-ov.html [hereinafter VeRO]. *But see* Complaint, Tiffany, Inc. v. eBay, Inc., No. 04 CV 4607 (S.D.N.Y. June 18, 2004), *available at* http://cyberlaw.stanford.edu/attachments/Tiffany%20eBay%20complaint.pdf (claiming that eBay is secondarily liable for sales of trademarked goods). eBay is potentially more liable than most stand-alone search engines under a vicarious-liability theory, since it has the ability to disable any auction on its site. *See* VeRO, *supra*. Contributory trademark liability might be more broadly generalizeable beyond eBay.

154. 47 U.S.C. § 230(c)(1)(2000).

155. *Id.* at § 230(c)(2)(A).

156. *See* Chicago Lawyers' Comm. for Civil Rights Under the Law, Inc. v. Craigslist, Inc., 461 F. Supp. 2d 681, 698–99 (N.D. Ill. 2006) (applying Section 230 to preempt a claim under the Fair Housing Act against an online classified-ads site and search engine for the discriminatory housing ads posted by its users). For a remarkable example of the broadness of this immunity, consider *Sturm v. eBay*, Inc., No. 1-06-CV-057926 (Cal. Super. Ct. July 27, 2006), in which eBay was not required to remove defamatory feedback about a user, even when the user and the defamer had entered into a settlement stipulating that the feedback was defamatory and both had written eBay asking that it be removed. *See* Elise Ackerman, *EBay Lawsuit Reveals Foibles of*

Those exceptional circumstances might arise when the engine itself has taken sufficient steps so that it could be identified as the provider of the content. First, if it has encouraged the creation of the content and directed its creation, it might be identified with the provider for liability purposes.[157] This scenario is not a significant concern for a pure search engine, but the growing integration of search engines with other applications raises concerns, particularly for search engines associated with creative communities.[158] Second, to the extent that a search engine is viewed as a speaker—something search engines are eager to encourage in the context of defending themselves in suits over rankings[159]—its recommendations of content potentially become endorsements of that content's message. Applicable precedents hold that services have substantial leeway to choose which messages to pass along, but if the search engine itself adds some content of its own to the recommendation (even, for example, a sentence describing the linked-to page[160]), that additional content might fall outside of Section 230's protections.[161]

Site Feedback, SAN JOSE MERCURY NEWS, Aug. 9, 2006, at A1, *available at* http://www.mercurynews.com/mld/mercurynews/news/local/15228670.htm.

Other countries have less categorical rules about intermediary immunity. *See, e.g.*, Eiseres/Google, Rechtbank Amsterdam [Court Amsterdam], 30 April 2007, LJN BA3941 (Neth.), *available at* http://zoeken.rechtspraak.nl/resultpage.aspx?snelzoeken=true&searchtype=ljn&ljn=BA3941&u_ljn=BA3941 (holding that Google is not obliged to prevent unlawful information from entering its search results). Discussion of this case is available at Joris van Hoboken, *The Duty of Care of Search Engines* (May 15, 2007), http://www.jorisvanhoboken.nl/?p=32.

157. *See* Fair Housing Council of San Fernando Valley v. Roommate.com, LLC, 489 F.3d 921, 926 (9th Cir. 2007) (holding a roommate search service not immune under Section 230 for information supplied by users in drop-down menus on the service's site). *See generally* Ken S. Myers, *Wikimmunity: Fitting the Communications Decency Act to Wikipedia*, 20 HARV. J.L. & TECH. 163, 174–78 (2006) (discussing Section 230 cases and the extent to which a service can encourage users to provide defamatory content without losing its immunity).

158. *See, e.g.*, Chris Sherman, *Puzzling Out Google's Blogger Acquisition*, SEARCH ENGINE WATCH, Feb. 18, 2003, http://searchenginewatch.com/showPage.html?page=2161891 (discussing Google's purchase of Pyra Labs, the creator of Blogger); Press Release, Google To Acquire YouTube for $1.65 Billion in Stock, (Oct. 9, 2006), http://www.google.com/press/pressrel/google_youtube.html (announcing that Google will acquire YouTube); Michael Arrington, *Windows Live Spaces Launches, Replaces MSN Spaces*, TECHCRUNCH (Aug. 1, 2006), http://www.techcrunch.com/2006/08/01/windows-live-spaces-launches-replaces-msn-spaces/.

159. *See* Search King, Inc. v. Google Tech., Inc., No. CIV-02-1457-M, 2003 U.S. Dist. LEXIS 27193, at *6 (W.D. Okla. May 27, 2003).

160. Yahoo's directory describes each listed site in a sentence. Yahoo! Search Directory, http://dir.yahoo.com/. The Open Directory Project does the same, with the taxonomy and descriptions being supplied by volunteers. Open Directory Project, http://www.dmoz.org/.

161. *Cf.* Benjamin Cohen & Helen Nugent, *Cole Tackles Google over Gay Link*, TIMES (LONDON), Mar. 7, 2006, at 31, *available at* http://www.timesonline.co.uk/tol/news/tech_and_web/article738224.ece (explaining that Google search on "ashley cole" returned "See results for: ashley cole gay" and describing Google spokesperson's explanation that alternative search suggestions are determined by computer algorithm).

The presence of search engines also reshapes the dynamics of the struggle between defamer and defamee. Search engines can focus attention on statements that previously would not have spread as far or as fast. Because they selectively pair a query with results, they can firmly link a name to a given piece of information. Indeed, precisely because people may wish to search on others' names, search engines regularly direct users to false claims.[162] Arguably, search engines also help diligent third parties discover unflattering information about themselves before it has spread, allowing them to move directly against the providers.[163]

In light of these dynamics, some scholars have argued that the law of search should be modified (or interpreted) to provide at least some kind of right of reply. By analogy to the Fair Credit Reporting Act, which allows individuals to correct incorrect statements about their credit history (and provides mechanisms to proceed both against original reporters and credit-record search agencies), these scholars have suggested that search engines should be required to respond to certain well-specified classes of reputational harms. For example, the search engine could allow the subject of an unflattering search result to annotate the result with an asterisked link to her reply.[164]

The broad immunity search engines and other intermediaries currently enjoy encourages reputational self-help.[165] Scholars have argued that the increasing democratization and interactivity of Internet communications technologies mean that self-help should be the response of choice; one confronted with an online falsehood should be encouraged first to propagate the truth online.[166] Given the enormous power that search engines wield in shaping which messages are heard and which are not, it is not obvious that the truth will necessarily be able to catch up with the falsehood.[167]

---

162. For a domain-specific example, consider Don't Date Him Girl, http://dontdatehimgirl.com/, which allows users to search for personal reports posted by other users about cheating men—by name. At least one alleged cheater has decided to take his chances with Section 230 in a lawsuit against the site. *See* Carl Jones, *Scorned Attorney Sues Kiss-and-Tell Web Site*, DAILY BUS. REV. (Florida), July 5, 2006, *available at* http://www.law.com/jsp/article.jsp?id=1151658319991.

163. *See* Daniel Dasey, *A Quick Self-Google Once a Day to Guard Your Reputation*, SUN-HERALD (Australia), May 23, 2004, *available at* http://www.smh.com.au/articles/2004/05/22/1085176043551.html.

164. *See* Pasquale, *supra* note 3, at 135.

165. *See generally* R. Polk Wagner, *On Software Regulation*, 78 S. CAL. L. REV. 457 (2005) (arguing that denying online actors legal recourse will encourage them to employ technical self-help).

166. *See, e.g.*, Edward A. Cavazos, Note, *Computer Bulletin Board Systems and the Right of Reply: Redefining Defamation Liability for a New Technology*, 12 REV. LITIG. 231, 243–47 (1992).

167. *See* Cho & Roy, *supra* note 69 (arguing that web ranking mechanisms favor already popular sites).

There is another name for self-help directed at search engines: SEO. Companies have been known to engage in substantial SEO to drive unflattering messages about them from search engine prominence.[168] Consider also "Googlebombing": the process of creating hyperlinks using a particular phrase and pointing to a particular page with the goal of convincing search engines to return that page on a query for that phrase. The most famous Googlebombs—George W. Bush as "miserable failure"; Andy Pressman as "talentless hack"—have been widespread bottom-up attempts to tarnish someone else's reputation. (Google has since changed its algorithms to make Googlebombs much harder to pull off.)[169] Search engines are just another media intermediary in age-old arms races to get competing messages before the public.[170]

3. Privacy

Search engines give users remarkable ability to learn about others. They can root out details that otherwise might have remained obscure and correlate information from many different sources. They democratize investigation, giving anyone the kind of power to develop a broad and deep profile of their chosen target that previously would have required a professional's help.[171]

Search engines are uncomfortable with the privacy-invading power of their own technologies. Google reacted with petulant anger to a CNet article on the privacy concerns raised by search. The article included an extensive profile of Google CEO Eric Schmidt, compiled by using Google, that disclosed Schmidt's political contributions, wife's name, and hobbies.[172] Google told CNet that it would refuse to talk to CNet reporters for a year.[173]

168. *See* Mark Glaser, *Companies Subvert Search Results to Squelch Criticism*, ONLINE JOURNALISM REV., June 1, 2005, http://www.ojr.org/ojr/stories/050601glaser/.

169. *See, e.g.*, Matt Cutts et al., *A Quick Word About Googlebombs*, OFFICIAL GOOGLE WEBMASTER CENTRAL BLOG (Jan. 25, 2007), http://googlewebmastercentral.blogspot.com/2007/01/quick-word-about-googlebombs.html (describing Google's algorithmic changes to end the effect of Googlebombs on its rankings).

170. For further discussion of Googlebombing, including these two examples, see *infra* Part II.D.1.

171. *See* Randy Cohen, *Is Googling O.K.?*, N.Y. TIMES, Dec. 15, 2002, (Magazine), at 50 (discussing the ethics of Googling potential dates). *Cf.* Kevin Poulsen, *MySpace Predator Caught by Code*, WIRED NEWS, Oct. 16, 2006, http://www.wired.com/science/discoveries/news/2006/10/71948 (discussing the use of publicly available datasets to locate registered sex offenders with Myspace profiles).

172. Elinor Mills, *Google Balances Privacy, Reach*, CNET NEWS, Aug. 3, 2005, http://news.com.com/2102-1032_3-5787483.html.

173. Saul Hansell, *Google's Chief Is Googled, to the Company's Displeasure*, N.Y. TIMES, Aug. 8, 2005, at C4. The ban was dropped *sub silentio* within a few months. *See* Elinor Mills, *Google to Yahoo: Ours Is Bigger*, CNET NEWS, Sept. 28, 2005, http://news.com.com/2102-1038_3-5883345.html (quoting from a phone interview with Schmidt).

The legal baseline when it comes to search subjects' privacy, as with defamation, is that search engines cannot be held liable for the information they pass along. But on a policy level, third-party-privacy problems are even thornier than user-privacy problems.[174] Users and search engines have a direct relationship that allows privacy-sensitive users to negotiate for better protections (at least in theory). Third parties stand in no such relationship to search engines.[175] The paradigmatic nightmare case for third-party privacy—the homicidal stalker—is a lot worse than the corresponding nightmare cases for search users—embarrassment for highly sensitive searches.

One interesting idea of the relationship between user and third-party privacy is that there should be some rough symmetry. That is, informational tools should disclose to search subjects that others are searching for information about them, possibly even including the names of the searchers. At least within closed online communities, social norms often favor some rough balance of privacy reciprocity between searchers and searchees.[176]

Third-party privacy in the age of search raises some important structural problems that set this issue apart from the intellectual-property and defamation problems discussed above. The chaotic state of privacy law means that primary liability for providers who release privacy-damaging information is sometimes available and sometimes not. Trade secrets or medical information? Well-defined causes of action will lie.[177] Your address and phone number? Probably not. One reason for this patchwork approach to primary liability is that privacy harms tend to stem from the aggregation of information, rather than from any single piece. Search engines and other information-location tools turn what would formerly have been inconsequential disclosures into the component parts of a genuine privacy violation.

174. To be analytically complete, we should consider provider privacy. In practice, however, provider privacy is not much of a problem. Providers who do not wish to be found by searching usually opt off the public Internet entirely. Accidental leaks of information found by search engines raise issues similar to those raised by third-party privacy concerns. That said, a provider that accidentally releases information harmful to itself is almost certainly the least-cost avoider as compared with a search engine.

175. Self-help is not entirely out of the question. Douglas Coupland's novel *JPod* includes a character who creates a juicy but false web site about herself (stating that she lost hundreds of pounds on the Subway diet but has cheated by sneaking junk food snacks), partly for amusement, but partly to throw searchers off the track of true information about her. DOUGLAS COUPLAND, JPOD 118–21, 166 (2006).

176. *See* Lior Strahilevitz, *Friendster and Symmetrical Privacy*, UNIV. OF CHI. L. SCH. FACULTY BLOG (Oct. 6, 2005), http://uchicagolaw.typepad.com/faculty/2005/10/friendster_and_.html.

177. *See, e.g.*, UNIFORM TRADE SECRETS ACT § 1(b)(2)(b)(ii) (defining "misappropriation" to include disclosure of information by one who acquired it "under circumstances giving rise to a duty to maintain its secrecy or limit its use"); McCormick v. England, 494 S.E.2d 431, 437 (S.C. Ct. App. 1997) (holding that "an actionable tort lies for a physician's breach of the duty to maintain the confidences of his or her patient").

It's also worth discussing the relative fault of the user and the third party. In the classic third-party intellectual-property case, the provider and user are joint tortfeasors, collaborating to rook the innocent intellectual-property holder. In the classic defamation case, the user is a deluded innocent, the third party is a besmirched innocent, and the provider is a dirty liar.[178] In a real case of privacy violation, though, the user may sometimes be the most blameworthy party: she's the one who set out to build a dossier on her subject. Interestingly, the subject may have more self-help power to prevent a privacy spill than a drive-by defamation. Anyone can make up lies about me pretty much no matter what I do, but by staying off the grid and keeping others from learning much about me, I can (imperfectly) mitigate the risks of my private information becoming widely known. That said, the choice to remain offline is less and less available, and one may have no ability to opt out of many data flows.[179]

4. User Virtue

Search censorship is on the rise around the world as more and more governments require search engines to filter results or block certain keywords from being searched.[180] The common thread uniting these various governmental demands is concern for what can best be described as their citizens' "virtue": governments fear that some content will corrupt any users who see it. That corruption can involve personal morality, as with pornography and graphic violence; it can also involve political morality, as with hate speech and dissent. Debates over whether these information flows should be suppressed are inherently political and deeply controversial.

178. Some users seek out negative information on others, it's true, but it's still hard to see those users as guilty of anything more than having an uncharitable attitude toward their fellow man. If the information is true, nothing wrong has taken place. If the information is false and the users know it's false, it may no longer be genuinely defamatory; all that remains is a kind of residual tarnishment. Additionally, if the information is false and the users think it's true, the deceit is the provider's fault. Only when a user republishes the libel—and thus becomes a provider herself—does her culpability ripen into something worthy of the name.

179. *See, e.g.*, Ryan Singel, *Sex Lube Maker's 250K Customer List Slides onto Net*, THREAT LEVEL (Apr. 24, 2007), http://blog.wired.com/27bstroke6/2007/04/sex_lube_makers.html (describing how the combination of poor security at Astroglide's web site and searchability means that Google searches on names of customers now return the name of the product they ordered).

180. Only rarely do nongovernmental actors even try to force search censorship, since they rarely have legal standing to do so. The most notable American exception is Complaint, Toback v. Google, Inc., No. 06-007246 (N.Y. Sup. Ct. May 4, 2006), *available at* http://casedocs.justia.com/new-york/nyedce/2:2006cv02692/257132/1/1.pdf, a private suit in nuisance and intentional infliction of emotional distress that accused Google of allowing traffic in child pornography. The causes of action were obviously preempted by Section 230, and the plaintiff, a New York state legislator, dropped it within two months, after Google "offered to sit down and discuss the issues."

The application of different national legal restrictions to search[181] mirrors one of the oldest and most contentious debates in Internet law: whether ISPs must, may, or must not filter out dangerous traffic flowing through their networks. Different national cultures and laws have treated different forms of Internet speech as dangerous. Child pornography touches a nerve in the United States; in France, it is hate speech; in China, it is criticism of the government. Government-mandated filtering and blocking at the network layer is not new. In the United States, a Pennsylvania law that would have required ISPs to filter content was declared unconstitutional;[182] the Chinese government engages in massive technological censorship through control over the routers that connect Chinese portions of the Internet to the rest of the world.[183]

The extension of such requirements to search engines is not exactly new, either. In 2000, French groups discovered that Yahoo!'s chat rooms and auctions included neo-Nazi material and ultimately won a court order requiring Yahoo! to block French access to such auctions, as well as "to any other site or service that may be construed as constituting an apology for Nazism or a contesting of Nazi crimes."[184] Yahoo!'s attempt to block United States enforcement of the order was ultimately inconclusive. Long before the Ninth Circuit dismissed Yahoo!'s declaratory judgment action, Yahoo! had decided that it preferred to do business in France on French terms and came into compliance with the order.[185] This has been the pattern for many United States technologies companies doing business abroad ever since—initial protestations, followed by compliance with local law.[186]

Today, China is the most aggressive search-blocking regime.[187] Search engines—like most other Internet intermediaries—are expected to block a

181. *See* CenSEARCHip, About CenSEARCHip, http://homer.informatics.indiana.edu/censearchip/about.html (describing a service that compares search results in China, France, Germany, and the United States); *see also* JACK GOLDSMITH & TIM WU, WHO CONTROLS THE INTERNET? 149 (2006) ("[W]hat we once called a global network is becoming a collection of nation-state networks . . . .").

182. *See* Cen. for Democracy and Tech. v. Pappert, 337 F. Supp. 2d 606, 663 (E.D. Pa. 2004) (holding unconstitutional a Pennsylvania anti-pornography ISP censorship law).

183. *See* Richard Clayton et al., Ignoring the Great Firewall of China 1 (2006) (unpublished manuscript), *available at* http://www.cl.cam.ac.uk/~rnc1/ignoring.pdf (describing technologies used to prevent Chinese access to disapproved Internet sites and techniques for evading this censorship).

184. *See* Yahoo! Inc. v. La Ligue Contre Le Racisme et L'Antisemitisme, 433 F.3d 1199, 1202 (9th Cir. 2006) (providing an English translation of the French order).

185. *Id.* at 1215–18. The en banc court split: some judges would have allowed the suit to go forward, some would have dismissed for lack of personal jurisdiction, and some would have dismissed for lack of ripeness. A majority of the en banc panel voted for dismissal, but neither reason for dismissal commanded a majority on its own.

186. *See* GOLDSMITH & WU, *supra* note 181, at 159–60.

187. *See* Thompson, *supra* note 63 (describing the Chinese government's Internet censorship laws and Google's operations in the country).

slightly nebulous set of forbidden content (the vagueness of the standards leads to inconsistent and unpredictable overblocking).[188] Falun Gong, Tibetan culture, and political dissidence are the principal, but by no means the only, targets.[189] These policies—particularly when adopted by "do no evil" Google—have led to a domestic political backlash, including the introduction in Congress of the Global Online Freedom Act, which would require search engines to refuse "Internet-restricting countr[ies']" governments' requests to alter results and would require search engines to notify the United States government of their blocklists.[190] The complexity of varying international standards and the concern that compliance with local law may lead to human-rights violations have led some companies to ask the United States government for help.[191]

Governments have also learned that they can ask or force intermediaries to identify Internet users who exchange forbidden content. For example, Chinese dissidents convicted and imprisoned with information that Yahoo! supplied have sued Yahoo! in the United States under the Alien Tort Claims Act.[192] Search companies have been inconsistent in their willingness to contest demands for private information from local authorities. Google strongly resisted a Department of Justice subpoena for query data in *Gonzales v. Google*,[193] but it has reached arrangements with authorities in Brazil to identify users of its Orkut social-networking service who are accused of spreading child pornography or engaging in hate speech.[194]

Government interests in censoring or finding the authors of unwanted speech are in direct tension with many user and provider interests in conducting searches. Blocking is a deliberate and heavy-handed form of bias. Government recruitment of search also directly threatens both user and provider privacy. On the user side, search queries—whether obtained from the search engine or from other sources—can be highly incriminating

188. This self-censorship system is not unique to China; search providers in Germany use a similar system to suppress hate speech. *See* Subcode of Conduct for Search Engine Providers of the Association of Voluntary Self-Regulating Multimedia Service Providers, http://www.fsm.de/en/SubCoC_Search_Engines.

189. *See, e.g.*, Oxblood Ruffin, *Google, China, and Genocide*, CULT OF THE DEAD COW COMMUNICATIONS, Apr. 22, 2007, http://www.cultdeadcow.com/cDc_files/cDc-0409.php (describing Google's Chinese-search censorship as participation in "cultural genocide").

190. H.R. 4780, 109th Cong. §§ 202(1), 203 (2d Sess. 2006).

191. *See* Anne Broache, *Web Giants Ask for Feds' Help on Censorship*, CNET NEWS, Jan. 31, 2007, http://news.com.com/2102-1028_3-6154930.html.

192. *See* Complaint of Tort Damage, Wang Xiaoning v. Yahoo! Inc., No. 07-2151 (N.D. Cal. Apr. 19, 2007).

193. Gonzales v. Google, Inc., 234 F.R.D. 674, 680 (N.D. Cal. 2006) ("Google primarily argues that the information sought by subpoena is not reasonably calculated to lead to evidence admissible in the underlying litigation and that the production of information is unduly burdensome.").

194. Broache, *supra* note 48.

evidence.[195] And on the provider side, search provides another tool in the government's surveillance toolkit. Unlike with most private surveillance, the government may be interested in searching entire populations for activity it considers suspicious. Think of the National Security Agency ("NSA") surveillance program, which amounts to a gargantuan search project across the huge daily volume of phone calls, apparently seeking patterns of calls fitting specified profiles.[196]

D. SEARCH ENGINES' INTERESTS

We finish this survey with a discussion of search engines' interests. The basic search business model may be summed up quite simply: Provide high-quality results to attract users, sell ads, and rake in the bucks. This Section discusses three significant threats to this business model: SEO, click fraud, and competition. SEO manipulators and click fraudsters pose an operational threat, sucking the value out of results. Contract and commercial-fraud law provide protections from these sorts of dishonesty. Other search engines pose a competitive threat. Intellectual-property law and antitrust law both provide protection for the needs of competing search engines.

1. Preventing Search Engine Optimization

SEO can be described as a deliberate attempt by a provider to introduce results bias against its competitors. As such, it harms users, whose searches become less useful, as well as legitimate providers, whose content becomes less visible. Strong market incentives compel search engines to combat SEO because by doing so, engines can give users better results. The consequence is a technical arms race between engines and manipulators. Search engines jealously guard their ranking algorithms as a way of maintaining an edge in this race.[197] Because search results generally are public, however, manipulators can engage in extensive reverse engineering. Even if they do not know precisely how the engine computes results, they can create approximate models of the sorts of content it favors and disfavors.[198]

195. *See* United States v. Schuster, 467 F.3d 614, 617 (7th Cir. 2006) (upholding factual findings based in part on Google searches by the defendant); K.C. Jones, *Murder Suspect's Google Searches Spotlighted in Trial*, CHANNEL WEB NETWORK, Nov. 11, 2005, http://www.crn.com/it-channel/173602157 (discussing a prosecutor's claim that a murder defendant used Google to learn the "depth and topography of a lake where the body of his wife . . . was found").

196. Press Release, President George W. Bush, President Bush Discusses NSA Surveillance Program (May 11, 2006), *available at* http://www.whitehouse.gov/news/releases/2006/05/20060511-1.html.

197. *See, e.g.*, Matt Cutts, *Notifying Webmasters of Penalties*, GADGETS, GOOGLE, AND SEO (Apr. 26, 2006), http://www.mattcutts.com/blog/notifying-webmasters-of-penalties/ (discussing the tradeoffs involved in notifying providers that their sites are being penalized).

198. *See, e.g.*, Dave Tiberio, *Reverse-Engineering Search Engine Ranking Algorithms*, WEBPRONEWS.COM, http://www.webpronews.com/node/23170/print (explaining the basics of reverse engineering search engine algorithms).

Pure SEO has not generated much in the way of litigation. Because being caught engaging in SEO is the most common way to be stricken from an engine's index, the question of whether a provider is engaging in SEO has sometimes been raised as a collateral issue in placement suits.[199] Search engines use their operational need to prevent SEO as an argument against judicial oversight of their ranking decisions. Although it's not inconceivable that providers and users upset at fraud-heavy results might sue search engines, these parties are in the same boat as engines. Engines almost never make specific promises to stamp out search fraud, and there is little more that most engines could do.

It is quite possible that the next few years will see some lawsuits against providers that allege the use of SEO tactics. The trademark theories mentioned above were originally developed in direct suits by trademark holders (i.e., other content providers) against alleged manipulators.[200] Courts have recognized that some techniques of content design are deceptively manipulative and cause harm to legitimate providers, and it is possible that innovative pleading could properly state other business torts against manipulators. Similarly, luring users to one's content through SEO raises significant false-advertising concerns. In these cases, competitors, users, and consumer-protection agencies might all be proper plaintiffs.

One limit to these theories is that they see the harm as transactions facilitated or foiled by the misdirection. SEO techniques, however, are neither necessarily commercial nor necessarily centralized. Consider again Googlebombing.[201] The most famous Googlebomb of all time may be the linking of the phrase "miserable failure" to the White House biography of President George W. Bush.[202] Similarly, companies have learned how to engage in "sock puppetry," creating fake content and personae to express their point of view and suppress negative opinions in search results.[203] The misrepresentations, if any, involved in these techniques are not easily

199. *See generally, e.g.*, Search King, Inc. v. Google Tech., Inc., No. CIV-02-1457-M, 2003 U.S. Dist. LEXIS 27193 (W.D. Okla. May 27, 2003).

200. *See generally* Brookfield Commc'ns, Inc., v. W. Coast Entm't Corp., 174 F.3d 1036 (9th Cir. 1999).

201. *See* Adam Mathes, *Filler Friday: Google Bombing*, ÜBER.NU, Apr. 6, 2001, http://uber.nu/2001/04/06/.

202. *See "Miserable Failure" Links to Bush*, BBC NEWS, Dec. 7, 2003, http://news.bbc.co.uk/2/hi/americas/3298443.stm (describing the "Miserable Failure" Googlebomb). Whether the critics of George W. Bush who created this Googlebomb will feel as proud of it when the White House has its next occupant is another matter entirely.

203. *See, e.g.*, Brad Stone, *The Hand That Controls the Sock Puppet Could Get Slapped*, N.Y. TIMES, July 16, 2007, at C1, *available at* http://www.nytimes.com/2007/07/16/technology/16blog.html; *see also* Glaser, *supra* note 168; Wikipedia, Sock Puppetry, http://en.wikipedia.org/wiki/Wikipedia:Sock_puppet ("Use of sock puppets is discouraged in most cases.").

characterized. Opinions differ as to whether search engines should try to suppress Googlebombs and sock puppets.[204]

2. Preventing Click Fraud

Click fraud has led to more litigation,[205] probably because it involves both money taken directly from advertisers' pockets and the contractual relationship between advertisers and engines. The Coase theorem suggests that systematic click fraud will simply be reflected in lower prices for advertising, since click fraud increases the number of clicks charged per sale generated.[206] Competitors who target particular advertisers for fraudulent purposes present a more difficult problem. One might expect that advertisers would demand clauses relieving them of the responsibility of paying for fraudulent clicks if they are targeted.[207] Enforcing such clauses, however, requires that advertisers be able to monitor the clicks for which they are charged.[208] Search engines have been reluctant to share the

204. *See generally* Clifford Tatum, *Deconstructing Google Bombs: A Breach of Symbolic Power or Just a Goofy Prank?*, FIRST MONDAY, Oct. 2005, http://www.firstmonday.org/issues/issue10_10/tatum/index.html (suggesting that attempts at suppression will have little impact). *Compare* Saul Hansell, *Foes of Bush Enlist Google to Make Point*, N.Y. TIMES, Dec. 8, 2003, at C8 ("We just reflect the opinion on the Web . . . ."), *with* Danny Sullivan, *Googlebombing Now a "Prank" and Not Web's Opinion, Says Google*, SEARCH ENGINE WATCH (Sept. 19, 2005), http://blog.searchenginewatch.com/blog/050919-095321 (arguing that Google has changed its views about the nature of Googlebombing).

205. *See* Class Action Complaint and Jury Trial Demand, Crafts by Veronica v. Yahoo!, Inc., No. 2:06-cv-01985 (D.N.J. May 1, 2006); Class Action Complaint and Jury Trial Demand, Draucker Dev. v. Yahoo! Inc., No. CV06-2737 (C.D. Cal. May 4, 2006); Final Order and Judgment Approving Settlement, Lane's Gifts and Collectibles, Inc. v. Yahoo! Inc., No. CV-2005-52-1 (Ark. Cir. Ct. July 26, 2006), *available at* http://blog.ericgoldman.org/archives/lanegiftsac.pdf; Findings and Preliminary Order Approving Settlement of Class Action and Directing the Issuance of the Notice to the Class, Checkmate Strategic Group, Inc. v. Yahoo! Inc., No. 2:05-CV-04588-CAS-FMO (C.D. Cal. June 28, 2006); Advanced Internet Tech., Inc. v. Google, Inc., 2006 WL 889477 (N.D. Cal. stay entered Apr. 5, 2006); Complaint, CLRB Hanson Industries LLC v. Google, Inc., No. 1-05-CV-046409 (Cal. Sup. Ct. Aug. 3, 2005); *see also* David A. Vise, *Clicking to Steal*, WASH. POST, Apr. 17, 2005, at F01 (describing the lawsuit by Google against Auctions Expert alleging click fraud).

206. Some have argued that search engines have no incentive to police click fraud because they can charge for illegitimate clicks. *See, e.g.*, Brian Grow & Ben Elgin, *Click Fraud: The Dark Side of Online Advertising*, BUSINESSWEEK, Oct. 2, 2006, at 46, 51 (quoting Martin Fleischmann, an Internet entrepreneur and victim of click fraud, "'I told Yahoo years ago,' he says, '"If this [fraudulent clicks] was costing you money instead of making you money, you would have stopped this."'"). Not so. The Coasean exchange returns to search engines the necessary incentive, as they will be able to increase their cost-per-click if they reduce fraud rates.

207. *But see, e.g.*, Yahoo! Search Marketing, Click Fraud FAQ, http://searchmarketing.yahoo.com/legal/clickfraud.php ("[*E*]*ven though we are not obligated to*, we voluntarily designed the Click Protection System" to identify click fraud (emphasis added)).

208. *See generally* GOOGLE, INC. CLICK QUALITY TEAM, HOW FICTITIOUS CLICKS OCCUR IN THIRD-PARTY FRAUD AUDIT REPORTS 4 (2006), http://www.google.com/adwords/ReportonThird-Party

necessary data with advertisers, fearing that it would permit reverse engineering of the search engine's fraud-detection algorithms.[209] Conversely, overly zealous enforcement of anti-fraud policies upsets advertisers, as the penalties unilaterally imposed by search engines against sources of click fraud include banishment from search advertising.[210] Independent auditing of click counts and of anti-fraud programs may be the wave of the future.[211]

Click fraud, because of its links to other ugly online practices,[212] has also embroiled some search engines in litigation involving those links.[213] Consumers and state attorneys general have begun to sue spyware makers and the advertising networks linked to them.[214] Following the chain of business relationships leads inevitably back to the search engines whose affiliate networks provide the ad delivery and billing at the heart of many of these schemes.[215] Those who distribute spyware also do so through advertisements placed on search engine sites (and on sites highly ranked due to SEO), suggesting that search engines (which often forbid these tactics in their advertiser guidelines) may bear some moral and legal responsibility for the problems created by spyware.[216] Their central

ClickFraudAuditing.pdf (arguing that third-party click-fraud-auditing firms overestimate click-fraud rates).

209. *See* ALEXANDER TUZHILIN, THE LANE'S GIFTS V. GOOGLE REPORT, http://googleblog.blogspot.com/pdf/Tuzhilin_Report.pdf (describing Google's reasons for withholding details of click fraud from advertisers). This monitoring problem creates a close connection between search engines' two biggest operational problems: SEO and click fraud.

210. *See generally* BATTELLE, *supra* note 1.

211. *See* Press Release, Interactive Adver. Bureau, The Interactive Industry Commits to the Development of Click Measurement Guidelines (Aug. 2, 2006), *available at* http://www.iab.net/news/pr_2006_08_02.asp (announcing an industry-wide group to develop a set of Click Measurement Guidelines that would standardize how invalid and fraudulent clicks are defined and counted).

212. *See* Damien Cave, *The Parasite Economy*, SALON, Aug. 2, 2001, http://archive.salon.com/tech/feature/2001/08/02/parasite_capital/index.html (coining the term); Stefan Gorling, *An Introduction to the Parasite Economy*, *in* EICAR 2004 CONFERENCE PROCEEDINGS (U.E. Gattiker ed., 2004), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=683081 (generalizing it).

213. *See, e.g.*, Crafts by Veronica v. Yahoo!, Inc., No. 2:06-cv-01985 (D.N.J. May 1, 2006).

214. *See, e.g.*, Complaint for Injunctive and Additional Relief, Washington v. Secure Computer, LLC, No. C-06-0126 (W.D. Wash. filed Jan. 24, 2006); Notice of Verified Petition, New York v. DirectRevenue, LLC (N.Y. Sup. Ct. filed Apr. 4, 2006); Settlement Agreement and Limited Release, Sotelo v. DirectRevenue, LLC, No. 05-C-2562 (N.D. Cal. Mar. 10, 2006).

215. *See* Ben Edelman, *The Spyware-Click-Fraud Connection—and Yahoo's Role Revisited* (Apr. 4, 2006), http://www.benedelman.org/news/040406-1.html (describing the role of Yahoo!'s affiliate network in placing advertisements in spyware). Edelman is co-counsel for plaintiffs in *Crafts by Veronica v. Yahoo! Inc.*, No. 2:06-cv-01985 (D.N.J. May 1, 2006).

216. *See* Ben Edelman, *Pushing Spyware Through Search* (Jan. 26, 2006), http://www.benedelman.org/news/012606-1.html (analyzing the prevalence of spyware-laden sites in Google results and advertisements for some common searches).

placement athwart many information flows may make search engines the least-cost avoiders for some of these ecological problems.[217]

3. Innovation

Search engines compete with other search engines for users.[218] They compete by offering more complete indices of the Internet, by providing more responsive results, and by integrating their searches with other features valued by users. (Their actual revenue source, advertising, is dependent on their ability to attract users, so market share by searches is a good indicator of competitive success.) All three of these techniques involve a sometimes-frenetic pace of innovation. Search engines zealously wield intellectual-property rights to protect their innovations from being appropriated.[219]

Search engines primarily use trade-secret techniques. They closely guard their ranking and indexing algorithms and routinely invoke the need to protect this secrecy in litigation that might expose operational details.[220] In practice, this secrecy is incomplete because the public disclosure of results permits limited reverse engineering. Search engines have also discovered the value of allowing nearly unlimited usage of their search facilities, including through automated application programming interfaces ("APIs");[221] they therefore are reluctant to take obfuscatory technical steps against reverse engineering. Further, the pressures of public relations encourage search engines to trumpet various advances and tweaks to their algorithms, if only in general terms.[222]

217. On search engines' centrality and connections to web spam, see YI-MIN WANG ET AL., SPAM DOUBLE-FUNNEL: CONNECTING WEB SPAMMERS WITH ADVERTISERS (2007), *available at* http://www.cs.ucdavis.edu/~hchen/paper/www07.pdf.

218. Currently, according to comScore and Nielsen NetRatings, Google is the market leader with a roughly fifty-percent share; Yahoo! has about twenty-five percent, and MSN another ten percent. AOL and Ask also have significant shares of perhaps five percent each, and the remainder is split among a great many minor search engines. *See* Press Release, comScore, comScore Releases June U.S. Search Engine Rankings (July 16, 2007), *available at* http://ir.comscore.com/releasedetail.cfm?ReleaseID=254477; Press Release, Nielsen//NetRatings, Nielsen//NetRatings Announces May U.S. Search Share Rankings (June 20, 2007), http://www.nielsen-netratings.com/pr/pr_070620.pdf.

219. Jonathan Thaw & Susan Decker, *Google to Subpoena Yahoo, Microsoft on Book Scanning*, BLOOMBERG MARKETS, Oct. 5, 2006, http://www.bloomberg.com/apps/news?pid=20601103&sid=amfuMLMq__H8.

220. *See* Gonzales v. Google, Inc., 234 F.R.D. 674, 684 (N.D. Cal. 2006) (considering Google's allegation that revealing random URLs and user queries would expose trade secrets).

221. *See* Robert D. Hof, *Mix, Match, and Mutate*, BUSINESSWEEK, July 25, 2005, at 72, *available at* http://www.businessweek.com/magazine/content/05_30/b3944108_mz063.htm (describing "mash-up" applications made by combining multiple programmatic APIs). Such mash-ups would be nearly impossible if search engines and other web companies treated their results as resources to be closely guarded.

222. *See, e.g.*, Cutts et al., *supra* note 169.

Search engines also take appropriate measures to guard the protected status of their trade secrets. They routinely require employees to sign nondisclosure and noncompetition agreements to prevent the departure of search secrets to rivals.[223] They typically require visitors to sign comprehensive nondisclosure agreements, as well.[224] Secrecy at Google, in particular, is almost a way of life.[225]

Google's patent on PageRank is best-known,[226] but all of the major search engines have patent portfolios. They aren't afraid to use them, either. In the run-up to its IPO, Google settled, for stock worth approximately $300 million, an outstanding patent-infringement lawsuit brought by Yahoo!. The suit involved a patent on the technique of displaying keyword-triggered search ads based on the bids submitted by advertisers—the technique at the heart of most search advertising today.[227] A steady drumbeat of search engine patent suits continues.[228]

Search engines also use copyright and trademark to protect their business models, although these matters are less litigated. Search engines possess valid copyrights in their software and interfaces; they have valuable

223. *See, e.g.*, Google, Inc. v. Microsoft Corp., 415 F. Supp. 2d 1018, 1020 (N.D. Cal. 2005) (discussing Microsoft's attempt to prevent Kai-Fu Lee, former vice president at Microsoft, from becoming vice president at Google).

224. *See* Nick Denton, *Google: This NDA Never Existed*, VALLEYWAG (Jan. 22, 2007), http://valleywag.com/tech/google/this-nda-never-existed-230407.php (reprinting Google's standard nondisclosure agreement).

225. *See* Robin Sidel et al., *At Google, Mum's the Word About Almost Everything*, WALL ST. J., Apr. 27, 2007, at B1 (detailing Google's secrecy with both investors and the outside world).

226. U.S. Patent No. 7,058,628. The patent is "owned by Stanford University, but licensed exclusively to Google until 2011." BATTELLE, *supra* note 1, at 130.

227. *See* George Mannes, *Yahoo! Gets Bigger Stake in Google*, THESTREET.COM, Aug. 9, 2004, http://www.thestreet.com/pf/tech/georgemannes/10177217.html.

228. *See generally* Hyperphrase Techs., LLC v. Google Inc., No. 06-C-199-5, 2006 U.S. Dist. LEXIS 64918 (W.D. Wis. Sept. 7, 2006); Skyline Software Sys. v. Keyhole, Inc., 421 F. Supp. 2d 371 (D. Mass. 2006), *summary judgment granted to defendant Google, Inc.*, No. 06-10980-DPW, 2007 U.S. Dist. LEXIS 16053 (D. Mass. March 7, 2007); Netjumper Software, L.L.C. v. Google, Inc., No. 04-80366-CV, 2005 U.S. Dist. LEXIS 27813 (S.D.N.Y. Nov. 10, 2005); FindWhat.com v. Overture Servs., Inc., No. 02 Civ. 447 (MBM), 2003 U.S. Dist. LEXIS 2450 (S.D.N.Y. Feb. 13, 2003).

trademark and trade-dress rights in their brands.[229] Whether search results are copyrightable as such is a debatable proposition.[230]

4. Competition

Any individual search engine would love to dominate the market for search. But they share a collective interest that the market remain competitive. The law of unfair competition—unconnected with monopolization as such—provides a baseline of legitimate and illegitimate business practices in the search market.[231] Antitrust law hasn't had much to say about search, but that may change. Currently, Google's fifty-percent market share makes it the big fish among major search engines, and the obvious first target for antitrust concern. The overall web search market is dominated by the top few providers.

Are these conditions alarming? Arguments can be made both ways. On one hand, users can easily switch search engines if they are unhappy with a particular engine's practices. Search engines are trying to change that with increased personalization, but so far, little prevents users from switching. The greater competition concern may come from the rising costs of entry; the Internet continues to grow rapidly, and the SEO arms races have meant that sophisticated and computationally expensive algorithms appear to be part of the price of offering useful search. The real question may be whether one considers near technological neighbors to be good substitutes for centralized search. Reclassifying various technologies—e.g., del.icio.us's social bookmarks or eBay's product search—as "search" would greatly increase the denominator and reduce relative market shares. These near neighbors may also have lower barriers to entry than server-farm-heavy, centralized search.

229. *See* Penelope Patsuris, *The Making of a $2 Billion Brand*, FORBES.COM, Feb. 21, 2003, http://www.forbes.com/2003/02/21/cx_pp_0221google.html (describing the ubiquity of the word "google"); *see also* Frank Ahrens, *So Google Is No Brand X, but What Is 'Genericide'?*, WASH. POST, Aug. 5, 2006, at D01, *available at* http://www.washingtonpost.com/wp-dyn/content/article/2006/08/04/AR2006080401536.html (describing a baseless attempt by Google trademark attorneys to object to a newspaper's statement that "Google" was in common use as a generic verb); Beth Lipton Kriegel, *Yahoo Not Amused by Pot Parody Site*, CNET NEWS, Jan. 11, 1999, http://news.com.com/2100-1023-219986.html (describing Yahoo!'s action against yahooka.com).

230. *See generally* Dan Burk, *Method and Madness in Copyright Law*, 2007 UTAH L. REV. (forthcoming 2007), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=999433 (discussing the idea-expression dichotomy and the copyrightability of the results of automated processes).

231. *Cf.* Doug Young, *Yahoo and Former China Head in Brewing Legal Battle*, REUTERS, Aug. 17, 2006, *available at* http://services.inquirer.net/print/print.php?article_id=15801 (noting that Qihoo sued Yahoo for defamation), *with* Sumner Lemon, *Yahoo China Sues Rival Portal*, IDG NEWS SERV., Sept. 29, 2006, *available at* http://www.infoworld.com/article/06/09/29HNyahoochinasues_1.html?PORTALS (noting that Yahoo sued Qihoo for unfair competition).

Some search-ranking lawsuits filed by disappointed providers have raised antitrust theories, although they have failed to explain why the poor ranking of a provider constitutes an antitrust injury.[232] Google's proposed purchase of web-advertising powerhouse DoubleClick has raised both eyebrows and concerns.[233] Where market power in a related market—e.g., desktop operating systems or provision of broadband telecommunications—is more clearly established, search engines have appealed to the law and policy of antitrust to prevent that power from being leveraged into the search market. Thus, competing search engines have objected to the integration of a search engine into other applications—e.g., through a search box in a browser—but have not convinced antitrust regulators that the practice is worthy of concern.[234] Similarly, search engines have appealed to telecommunications regulators, asking that network operators be given little market power to discriminate among different traffic flows across their networks.[235] Search engines fear that network operators would use this power to extract rents from them.[236]

III. INTERCONNECTIONS IN SEARCH ENGINE LAW

Already, some of the connections and conflicts among these problems in search engine law should be evident. This Part examines five in more depth. In each case, a broad view of search—one that considers the interrelationships among many different doctrines—provides a clearer and more helpful analysis of what is really at stake in a given legal dispute.

First, many forms of relief against search engines are functional substitutes for one another. Some of the hardest-fought issues in search policy are all but moot in light of doctrines from other areas. In general, such doctrinal distinctions are unstable; the broad view of search forces us to recognize that the technical centrality of search engines puts strains on many different areas of law.

Second, search engines raise problems of unaccountable discretion. Two natural tools to investigate and remedy abuses of discretion are greater disclosure of the basis for ranking decisions and mandated corrections to fix

232. *See* Person v. Google, Inc., 456 F. Supp. 2d 488, 491–92 (S.D.N.Y. 2006) (alleging that Google does not allow smaller customers to purchase certain AdWords but dismissing on other grounds); Tushnet, *supra* note 75 (characterizing KinderStart's allegations of Google antitrust violations as "incomprehensible").

233. Steve Lohr, *Google Deal Said to Bring U.S. Scrutiny*, N.Y. TIMES, May 29, 2007, at C1, *available at* http://www.nytimes.com/2007/05/29/technology/29antitrust.html.

234. Ina Fried, *Vista Search Seems Fair, Regulators Say*, CNET NEWS, May 12, 2006, http://news.com.com/2102-1014_3-6071851.html.

235. *See* Eric Schmidt, A Note to Google Users on Net Neutrality, http://www.google.com/help/netneutrality_letter.html (asking concerned citizens to voice their support for net neutrality).

236. Arshad Mohammed, *Verizon Executive Calls for End to Google's "Free Lunch"*, WASH. POST, Feb. 7, 2006, at D1.

misleading rankings. But these tools run squarely up against search engines' operational interests in fighting search fraud, against their competitive interest in trade secrecy, and perhaps even against user interests in privacy. Taking a broad view of search forces us to weigh the costs and benefits of these remedies carefully.

Third, user privacy concerns must be understood in the context of search engine operations and third-party concerns. Search engines use query and clickthrough data to target advertisements, to refine search quality, and to personalize search. Prohibiting these uses outright could have significant negative effects on users, including exacerbating search engine bias. At the same time, third parties may have quite legitimate interests in learning user identities, so that query-privacy policies reflect a struggle between users and third parties for relative advantage. Given the profoundly private information users entrust to search engines, balancing these concerns will not be easy. A broader view of search makes clear the competing values at stake.

Fourth, many legal theories raised by and against search engines turn on the speech content of search recommendations. Search engines have encouraged a view that such recommendations are subjective statements of opinions about page quality and, as such, are entitled to substantial First Amendment protection. But there is a sense in which it is precisely the subjectivity of search rankings that make them problematic—the more individually tweaked and the less automatic that results are, the greater the concern that the search engine is using its privileged position to engage in unfair discrimination against particular targets. The question of search engine speech needs to be regrounded on a more stable foundation than the subjective statement-of-opinion analysis alone can provide. A broader view of search can provide a start.

Fifth, thinking about trademark disputes without placing them in context is a recipe for trouble. Search engines provide enormous value to consumers by literally reducing search costs. Trademark holders' demands for veto power over keyword sales must be understood as a tactic in the ranking wars; giving them that power would frustrate the policies of trademark law and hamper search innovation. At the same time, however, simply excluding keyword sales and search engine manipulation from trademark scrutiny altogether, as some courts have done, is also dangerous. SEO tactics by providers increase consumer confusion and are socially wasteful; it is easy to envision search engines and near relatives that flout the goals of unfair-competition law. A broader view of search can point to a healthier balance.

A. CLAIMS AGAINST SEARCH ENGINES AS FUNCTIONAL SUBSTITUTES

Multiple legal lines of communication exist between search engines and other parties. Those concerned with one particular form of harm are not

limited to legal theories directly addressing that harm. If they can gain relief against a search engine[237] on another theory, it may be just as good. Wherever in law this multiplicity appears, it raises a concern that parties not be allowed to subvert one doctrine by appealing to another.

Consider first providers' desire to prevent searching. Suits based on unwanted access can often be interchanged with unfair-competition suits. Many lawsuits by providers run together both forms of harm in their complaints—the search engine both imposes technical burdens and interferes with providers' intellectual property.[238] On the one hand, competitive concerns motivated some of the most famous unauthorized-access suits.[239] On the other, providers who cannot use the law or self-help to prevent access often turn to intellectual-property arguments. Thus, for example, the Authors Guild has no ability to prevent Google from obtaining physical access to books.[240] Its lawsuit against Google is therefore exclusively based in copyright infringement.[241]

A closer look at theories of unauthorized access shows this instability in action. Pro-access advocates won a hard-fought victory when the California Supreme Court in *Intel v. Hamidi* held that trespass to chattels would not lie without proof of actual harm.[242] Their victory is largely symbolic, given that courts have been giving broad scope to anti-intrusion statutes and have been willing to uphold browsewrap contracts.[243] *Hamidi* notwithstanding, the

237. Or, in the engine's case, gain a blanket immunity from suit.

238. *See, e.g.*, Kelly v. Arriba Soft Corp., 336 F.3d 811, 817 (9th Cir. 2003) (raising copyright theories predicated both on initial copying and on framing); Ticketmaster Corp. v. Tickets.com, Inc., No. CV 99-7654 HLH(BQRX), 2000 WL 525390, at *2, *4 (C.D. Cal. Mar. 27, 2000) (raising copyright and trespass-to-chattels theories).

239. Register.com, Inc. v. Verio, Inc., 356 F.3d 393, 397 (2d Cir. 2004); eBay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1063 (N.D. Cal. 2000). The most promising unfair-competition theories are not available to plaintiffs who provide factual information not protectable by copyright. *See* Sw. Airlines Co. v. FareChase, Inc., 318 F. Supp. 2d 435, 440 (N.D. Tex. 2004) (stating that "[f]are, route and scheduling information are all facts and thus not copyrightable"); *Ticketmaster Corp.*, 2000 WL 525390, at *2 (denying a copyright claim).

240. *See* 17 U.S.C. § 109(a) (2000) (authorizing the owner of a lawful copy the right to re-sell that copy without the permission of the copyright holder). *But see* Sharon Billington, *Relief from Online Used Book Sales During New Book Launches*, 29 COLUM. J.L. & ARTS 497, 497–98 (2006) (arguing against empirical evidence and without considering the effects on libraries and critics that publishers should be allowed to prevent resale of used books within six months of first publication).

241. *See* Complaint at 2, Authors' Guild v. Google, Inc., No. 05-CV-8136 (S.D.N.Y. Sept. 20, 2005), *available at* http://pub.bna.com/eclr/05cv8136comp.pdf. *Cf.* Ty, Inc., v. Publ'ns Int'l Ltd., 292 F.3d 512, 515 (7th Cir. 2002) (considering a copyright suit against a collectors' guide to Beanie Babies containing allegedly infringing photographs thereof); Kevin Emerson Collins, *Cybertrespass and Trespass to Documents*, 54 CLEV. St. L. REV. 41, 62–68 (2006) (discussing cases involving the subsequent use of information obtained through improper access to physical documents).

242. Intel Corp. v. Hamidi, 71 P.3d 296, 305–06 (Cal. 2003).

243. *See supra* Part II.B.1.

major search engines have accepted that a content provider who wants search engines out can keep them out.[244]

Private third parties have also exploited overlapping doctrines; copyright claims provide stronger relief against intermediaries like search engines than do other private-law claims.[245] Two ISP cases with search engine angles show this phenomenon at work. In the pre-DMCA case of *Religious Technology Center v. Netcom On-Line Communications Services, Inc.*, the Church of Scientology attempted to prevent the distribution of unpublished L. Ron Hubbard manuscripts by asserting copyright claims against an ISP.[246] The Church has been a regular user of DMCA notices under Section 512 against search engines.[247] In *Diebold v. Online Policy Group*, an electronic-voting-machine company attempted to use the Section 512 subpoena process to suppress distribution of the source code to its voting machines.[248] The company relied on Section 512(d), the "information location tools" provision applicable to search engines.[249] Both plaintiffs were acting principally to avoid the release of embarrassing secrets but were unlikely to succeed on secrecy-based claims. Instead, copyright theories promised more legal leverage.

Lawyers in search engine suits will not respect boundaries between legal fields when framing their cases. Those who make law and policy for search engines must be alert to these overlaps and end-runs. Considering the various strands of search engine law together will help make such possibilities clear.

B. THE PROS AND CONS OF DISCLOSURE AND MANDATED RESULTS

If we were certain we could identify instances in which search engines returned incorrect results, the natural response would be to demand that they return instead the results we knew to be correct. Even if we are not certain we know which results are correct, we might know that some particular results are incorrect and demand their deletion. Frank Pasquale claims that subjects of a search deserve not to have search engines return misleading information about them and identifies two cases in which the correct search results can be specified precisely enough that he claims legal intervention is warranted: searches on proper names and searches on

244. *See supra* notes 93–94, 108 and accompanying text.

245. *See, e.g.*, 47 U.S.C. § 230(e)(2) (2000) ("Nothing in [the CDA intermediary immunity] shall be construed to limit or expand any law pertaining to intellectual property.").

246. Religious Tech. Ctr. v. Netcom On-Line Commc'ns Servs., Inc., 907 F. Supp. 1361, 1365 (N.D. Cal. 1995).

247. *See* Chilling Effects, Keyword:Scientologists, http://www.chillingeffects.org/dmca512/keyword.cgi?KeywordID=10 (documenting a number of instances in which the Church used such notices against search engines).

248. Diebold v. Online Policy Group, 337 F. Supp. 2d 1195, 1200 (N.D. Cal. 2004).

249. *Id.* at 1201.

trademarked terms. Although he stops short of claiming that such searches should instead return content provided by the person named or the holder of the trademark, he does propose that they be allowed to annotate misleading results with an asterisk.[250]

We can also make a related claim that doesn't depend on having an objective measure of "correct" and "incorrect" results. Start instead from the observation that users depend on search engines to find information for them. To know whether they should trust a search engine, they need to answer the same question they asked the search engine to solve: *What content is available?* Sometimes asking another search engine or having another user ask the same search engine will suffice, but sometimes—as in personalized search or when only one search engine indexes a particular kind of content—those options aren't available. In general, users need information about a search engine's inputs and its reasoning to make informed choices. This claim leads, therefore, to an argument that search engines should disclose detailed information about their algorithms.

Providers and search subjects also have legitimate reasons to want greater disclosure. Google has been accused of manipulating search results during litigation to make the judge at a crucial hearing unable to replicate the behavior of which Google's adversary complained.[251] Providers who feel they have been unfairly ranked for reasons unrelated to the quality of their content may need access to operational details to evaluate whether they really have been targeted. Third parties may want explanations as to why unflattering content appears prominently. And many disputes about click fraud cannot be evaluated without examining details of search engine billing. Again, disclosure seems the natural remedy.

Natural, perhaps, but not necessarily a good idea. Excessive mandating and disclosure can have dangerous consequences for the entire search ecosystem. Search engines' innovative interests mean that too much disclosure can suppress their incentives to create new algorithms, reduce the diversity of options, and limit users' genuine choice among engines. Mandated results are even more restrictive for competition and diversity, in that they enforce uniform policies about results across engines. Moreover, too much transparency in relationship to personalized search could lead to the disclosure of user queries, raising privacy concerns.

250. Pasquale, *supra* note 3, at 117. *But see* Goldman, *supra* note 3, at 195–98. Pasquale's precise specification of these cases responds to Goldman's critique that regulators cannot identify correct results as well as search engines can by restricting intervention to clearly identified mistakes.

251. *See* BATTELLE, *supra* note 1, at 184–86 (describing American Blinds' allegations that Google altered its search results to influence the judge's ruling on Google's motion to dismiss). Search engine watchers know that there are many innocent potential explanations for such behavior. Major search engines have multiple data centers, which may be out of sync with each other. An update might have taken place between one query and another the next day. And so on.

Most serious of all are the consequences of mandated results and disclosure for the SEO arms race. Search engine manipulators make their living by reverse engineering search algorithms. Search engines are able to preserve a layer of genuine, useful results through a combination of keeping precise algorithmic details secret and changing their algorithms to foil detected SEO techniques. Mandated disclosure undermines the former; mandated results undermine the latter. Legal interventions here threaten to hand search engine algorithms to manipulators on a platter. Even Pasquale's limited proposals are partially vulnerable to manipulation. What proof would a search engine require of one's real name before awarding an asterisk? And what would stop manipulators from registering trademarks on popular search terms on unlikely categories of goods? Consider a registration of REAL ESTATE as used to sell lip gloss (a product category for which it is fanciful, and thus registrable)—perhaps a pointless trademark but excellent for search engine placement.

This is not the place to evaluate when disclosure and results mandates are appropriate policy. For present purposes, it should suffice to note that these remedies raise concerns that cut across many areas of search engine law. There are reasons why they may be useful interventions and reasons why they are dangerous. Considering one without considering the other would be reckless.

C. USER PRIVACY CONCERNS IMPLICATE OTHERS' INTERESTS

The privacy problems posed by private stockpiles of user data are well-understood and do not require extensive rehashing here.[252] What does require attention is how the solutions to these problems may have different inflections in the search engine context than in other domains. Users' privacy interests must be understood in relation to other interests in query data.

First, users' own computers disclose to providers the users' use of a search engine and the query terms used. The "referer" information that a browser by default gives to any site from which it requests a web page includes the URL of the web page that referred the user to the provider's page. Most search engines include in the URL of the page displaying the results the query terms that a user entered. This automatic leak of query information—which can be blocked by technical measures either at the search engine or at the user's computer—means that search engines are not the only institutions that can easily accumulate query data.[253] Put another

252. *See generally* SIMSON GARFINKEL, DATABASE NATION (2000) (cataloging the threats to privacy posed by database technology); DANIEL SOLOVE, THE DIGITAL PERSON (2004) (examining the dangers of "digital dossiers" and proposing legal solutions).

253. Simply cutting and pasting a search result URL into one's address bar will hide from the provider the search query that led one to their site. One can also install software to much the same end. *See, e.g.*, Mozilla, RefControl, https://addons.mozilla.org/firefox/953/

way, the interaction of engine-to-user result flows with provider-to-user content flows creates an additional user-to-provider query-data flow.

Second, third parties harmed by search may have legitimate interests in learning some private information about users. This tension has been most clearly noted in the case of flows of copyrighted information in the ISP context; the Recording Industry Association of America ("RIAA") has been particularly active in attempting to learn users' identities.[254] Those whose privacy is breached by a search also have an interest in learning about it: they may need to take precautions against stalkers, they may need to take action against the provider releasing this information, and they may need to discourage searchers from searching.[255] There is something uncomfortable about a rule that assigns different weights to the privacy interests of search users and search subjects.

Third, the uses to which search engines put their query-data warehouses are relevant to users' other interests in search. Massive stockpiles of queries are useful for improving search. (Indeed, the AOL data release was neither the product of poor security nor a concession to corporate pressure for valuable data—it was an ill-advised attempt to further academic research into better search technologies.)[256] Extensive collection of query data is also a prerequisite for personalizing search.[257] Personalization of information reception and its concomitant promotion of diversity, in turn, can be an important technique for countering media bias.[258] Thus, a privacy-mediated

(download page for RefControl, a program to hide referer [sic] information). Although the blocking of such query leaks by search users is possible, it also seems to be quite rare.

254. *See* Sonia K. Katyal, *The New Surveillance*, 54 CASE W. RES. L. REV. 297, 297–98 (2003) (discussing the RIAA's attempts to locate infringing users through cell-phone records).

255. David Brin has written about this tension and has argued that the solution is greater transparency in general. We cannot keep people from surveilling others but can at least let those being surveilled know about it. *See generally* DAVID BRIN, THE TRANSPARENT SOCIETY (1998).

256. *See* Katie Hafner, *Tempting Data, Privacy Concerns; Researchers Yearn to Use AOL Logs, but They Hesitate*, N.Y. TIMES, Aug. 23, 2006, at C1 (discussing the academic ethics of research using publicly released AOL query data).

257. Consider this in light of Amazon's practice of recommending books by linking to books that users searched for or bought in the same browsing session. On such "collaborative filtering" systems, see generally JOHN RIEDL ET AL., WORD OF MOUSE: THE MARKETING POWER OF COLLABORATIVE FILTERING (2002). Of course, ads can be personalized just as easily as organic search results can. *See, e.g.*, William Marra, *Yahoo's SmartAd Raises Privacy Concerns*, ABC NEWS, July 4, 2007, *available at* http://abcnews.go.com/Technology/Story?id=3342775 (describing Yahoo's search-advertising program, which allows advertisers to tailor ad placement based on demographic information of users).

258. *Cf.* Yochai Benkler, *Siren Songs and Amish Children: Autonomy, Information, and Law*, 76 N.Y.U. L. REV. 23, 73–74 (2001) (arguing that personal autonomy and diversity are aided when individuals can use a medium to pursue their own informational choices).

concern with preventing individual manipulation by search engines[259] is in tension with a concern with preventing manipulation by monolithic one-size-fits-all information sources.

To repeat, problems of online privacy protection are subtle and tangled. Considering the various threads of search engine law all at once reveals just how many of them are connected to privacy in one way or another. Once again, any rational attempt to make sensible policy (here, privacy policy) in the search context demands careful engagement with these many interests and pressures.

D. SEARCH ENGINE RESULTS AS SPEECH

Many legal questions involving search require a theory of search engine speech. The First Amendment rights of search engines, users, and providers may provide defenses to third parties' attempts to impose liability for harmful content flows. Search engines' speech-facilitating roles may give them a thumb on the scales in debates over access. And the conflicting speech claims of search engines and providers will have a significant effect on how we think about search engine rankings. Ultimately, a theory of search engine speech will need to integrate all of these concerns.[260]

Such a theory is beyond the scope of this Article. Instead, this Section considers the cross-cutting problems raised by one attempt at framing the question. Google has asserted a theory of search rankings as subjective statements of opinion. Under a claim of tortious interference with contract, such as that raised in the *Search King* suit, it is a complete defense if the allegedly harmful act consisted of protected speech. Statements of opinion on matters of public concern are protected, unless provably false. Thus, for example, a negative bond rating is not a statement that could be proven "true" or "false" and thus cannot support defamation liability.[261] Since a search engine's rankings are merely a claim about the engine's subjective assessment of pages' relevance to particular users' queries, goes the reasoning, the search engine is not making a claim that could be shown false—and is therefore protected.[262]

259. *See* Tal Zarsky, *Online Privacy, Tailoring, and Persuasion*, *in* PRIVACY AND TECHNOLOGIES OF IDENTITY: A CROSS-DISCIPLINARY CONVERSATION (K. Strandburg & D. Stan Raicu eds., 2006), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=946498.

260. Substantial guidance may come from telecommunications law, which has long struggled with intermediaries' dual roles as speakers and as conduits. The greater interactivity of search engines—whose rankings are driven by providers' attempts to seek placement, by users' queries, and by search engine algorithms themselves—may require even more subtle analyses.

261. Jefferson County Sch. Dist. No. R-1 v. Moody's Investor's Servs., Inc., 175 F.3d 848, 855 (10th Cir. 1999).

262. Search King, Inc. v. Google Tech., Inc., No. CIV-02-1457-M, 2003 U.S. Dist. LEXIS 27193, at *11–12 (W.D. Okla. May 27, 2003).

The relationship of subjective opinion to objective fact, however, is not simple. Thus, for example, *Milkovich v. Lorain Journal Co.*,[263] while stating the rule that the Constitution shields opinions, leaves in place two significant exceptions. A statement of opinion may imply an underlying fact (the Court's example: "In my opinion John Jones is a liar."[264]), and even a statement of opinion may be false if not honestly held (the Court's example: "I think Jones lied," where the speaker thought nothing of the sort[265]).

In this light, the *Search King* rule suggests several counterarguments. First, as Search King alleged, the purported "opinion" is in fact the output of a computer algorithm, and the computer is perfectly predictable and objective. The court rightly dismissed this argument, distinguishing process from result. The subjectivity entered the algorithm when it was programmed, and although the intervening process is mechanical, what emerges at the end are the subjective judgments made by Google programmers about web page relevance and quality.[266]

Second, as Search King and KinderStart have alleged, search engines themselves emphasize the objective quality of their results and should be held to those statements.[267] This argument has slightly more bite, given that search engines have not been careful in their public statements.[268] By alleging claims of objectivity, provider plaintiffs also nudge their pleadings closer to consumer-fraud causes of action in which the initial claim of evenhanded objectivity is false, rather than the later ranking decision. Still, this argument also fails. Search engines haven't explicitly claimed that their results are objectively correct and can craft their public-relations materials to say that they don't imply it, either.[269]

A more troubling counterargument, however, combines the first two. It argues that the problem is hand manipulation of results.[270] This theory

263. Milkovich v. Lorain Journal Co., 497 U.S. 1 (1990).

264. *Id.* at 18.

265. *Id.* at 20.

266. This view of the subjectivity of search results is consistent with an argument that they involve sufficient selection and arrangement to satisfy copyright's originality requirement.

267. *See also* Maughan v. Google Tech., Inc., 49 Cal. Rptr. 3d 861, 874 (Cal. Ct. App. 2006) (noting Google's emphasis that its search processes are completely automated).

268. Rebecca Tushnet has noted this tension. Tushnet, *supra* note 75. Raymond Nimmer observes that *Search King* did not involve a claim by a user "who detrimentally relied on the rankings themselves" and contrasts cases in which safety-ratings services "voluntarily assumed [a] role [that] invited reliance by the public." RAYMOND T. NIMMER, INFORMATION LAW § 10:77 (2006).

269. Moreover, this counterargument misses a basic reality of search engine business. Search engines are trying hard to maximize the subjective satisfaction of users with their search results, and on a query-by-query basis, it is extremely hard to find an objective "fact" in the degree of user satisfaction.

270. *See* James Grimmelmann, *Google Replies to Search King Lawsuit*, LAWMEME (Jan. 9, 2004), http://research.yale.edu/lawmeme/modules.php?name= News&file=article&sid=807, *available*

abandons any claim to object to broad algorithmic decisions but argues that specific deviations from algorithmic choices—or specific algorithmic tweaks to hurt particular providers—should be actionable. The same programmers both write the algorithms and tweak the results,[271] so while there may not be a clear line between the two, hand tweaks feel much slimier.

Why? Perhaps because the generic ranking algorithm provides a baseline against which claims of later manipulation can be objectively measured. In terms of the *Milkovich* analysis, the search engine is lying not about the poorly ranked page's quality, but about its own belief that the page is of low quality. *The ranking algorithm believes that the page is important, so returning a worse rank is a lie about what the ranking algorithm believes.* This way of phrasing the argument captures its strong intuitive appeal. Even while defending their rankings as subjective assessments, search engines have been highly reluctant to make hand adjustments.[272] There is something unsettling about hand tweaks, whether or not that "something" rises to a level that would permit a suit in tort.

Even this brief survey of one issue's implications has stinted on other significant connections. A concern with hand tweaks suggests that greater transparency is a necessary tonic. The concern is also intimately related to the ongoing struggles against SEO, since the most frequent algorithmic changes are counters deliberately targeted at new SEO techniques. It connects to fears of government censorship, to the individual deletions of DMCA takedowns, and to broader questions of individual (and personal) actions versus massive parallelism that arise in privacy and access to content. A proper theory of search engine speech should consider these issues together, rather than in isolation.

E. TRADEMARKS AND SEARCH ENGINES IN CONTEXT

The problem of searches on trademarked terms is, as noted above, one of the most litigated issues in search. It also provides a case study in the

at http://web.archive.org/web/20040612081746/research.yale.edu/lawmeme/modules.php?name= News&file=article&sid=807; Pasquale & Bracha, *supra* note 3.

271. *See, e.g.*, William Slawsky, *20 Ways Search Engines May Rerank Search Results*, SEO BY THE SEA, Oct. 14, 2006, http://www.seobythesea.com/?p=334 (describing the systematic changes that search engines might make to results after the basic relevancy algorithm but before showing results to users).

272. *See* Google, An Explanation of Our Search Results, http://www.google.com/explanation.html (stating that search results are generated completely objectively and are independent of the beliefs and preferences of those who work at Google). Google drafted this text to respond to public complaints that the first-listed result for the search "Jew" was an anti-Semitic web site. This incident is significant in two ways. First, it revealed Google's extreme aversion to hand-tweaking its results, even though it would have been very easy to remove the anti-Semitic web site from results or move it further down the list. Second, this annotation required substantial public hand-wringing. (And note that Google now faces similar questions about what standards it will use every time it must decide whether to add this annotation to a results page.)

respective perils of too much and too little deference to search engines' decisions.

Courts dealing with the trademark implications of provider and search engine behavior have been much influenced by the *Brookfield* analogy of a highway billboard. A user who is misdirected by an information-location tool during a search for a trademarked term, the analogy asserts, is like a driver who has been misdirected by a billboard to take the wrong exit. She may not ultimately be confused about the source of the goods she acquires, but the additional search costs of going back to find the true source outweigh her desire for the "real McCoy."[273]

This analogy misses two important features of search engines. First, it misapprehends some of the technical realities of how various advertising techniques actually translate into user visits. In particular, if keywords in hidden metatags are billboards, they are invisible ones; most search engines now ignore them.[274] Similarly, contextual ads that are clearly disclosed as such and do not use the trademarks in their text never appear to users as billboards would; one might analogize them instead to billboards placed near the plaintiff's store but not using the plaintiff's trademarks. Second, the search costs involved in going back to find the originally desired source are much lower online than offline—a few seconds of clicking rather than a few minutes of driving—so far fewer users, even if diverted, will actually be locked into the wrong source. Indeed, these advertisements reduce search costs, first by offering users information about alternatives possibly responsive to their queries and second by funding search itself. Contextual ads are actually a substantial improvement, from a consumer-confusion point of view, over earlier search business models such as direct results-placement purchases and generic noncontextual banner ads.

Commentators and courts, however, have articulated a slightly questionable basis for finding no liability. Instead of finding no consumer confusion, they have found no use of the trademark, cutting off the trademark inquiry at the threshold. The *interactivity* of the search engine's dealings with the user has created analytical confusion: any possible "use" of the trademark is in the user's search query, rather than in the results. It has

273. Brookfield Commc'ns, Inc., v. W. Coast Entm't Corp., 174 F.3d 1036, 1064 (9th Cir. 1999).

274. *Compare* Pop Warner Little Scholars, Inc. v. N.H. Youth Football & Spirit Conf., No. 06-cv-98-SM, 2006 WL 2591480, at *3 (D.N.H. Sept. 11, 2006) ("Google and other search tools continue to associate defendants' web site with plaintiffs' marks [due to metatags] . . . ."), *with* Eric Goldman, *Outdated Metatags Don't Infringe*—Pop Warner v. N.H. Youth Football & Spirit Conference, TECH. & MKTG. L. BLOG (Sept. 25, 2006), http://blog.ericgoldman.org/archives/2006/09/outdated_metata.htm (criticizing *Pop Warner* and stating, "[S]earch engines aren't that inefficient or inaccurate given that THEY ARE SMART ENOUGH NOT TO RECOGNIZE KEYWORD METATAGS IN THE FIRST PLACE").

therefore seemed plausible to say that triggering ads or results based on a trademarked query is not a "use in commerce" by the search engine.[275]

To see why this blanket rule may be inappropriate, consider a line of interactive offline cases: trademark suits against restaurants that serve one cola when a customer has requested another.[276] The customer who receives a Pepsi after ordering a Coke is a victim of passing-off; whether the deception falls within the Lanham Act should not depend on whether the restaurant has used the COCA-COLA trademark on its menu or whether the server repeated "Coke" to confirm the customer's order.

The search engine's proper defense is that it is not misleading users, not that it is not using the trademark. It is easy to imagine search engines that deliberately cause serious confusion. Think of what would happen if Froogle or Amazon—search engines specifically oriented toward finding particular goods for purchase—were to adopt a policy of steering all searches for COCA-COLA to purchase pages for substitute brands. The law should not wholly ignore this possibility. Similarly, blatant SEO tactics cause enormous consumer confusion—particularly when they push genuine results entirely out of view—and a rule that such tactics are categorically immune from trademark scrutiny because search engine spamming is not trademark use seems perverse.

Finally, as in so many other contexts, the degree of concern increases with the opacity of the search engine's processes and the paucity of useful alternative search engines available to users. Decisions that affect these other matters affect the trademark inquiry. Thus, at the risk of sounding like a broken record, I reiterate the theme of this Part: Looking at various strands of search engine law together makes important connections clear. Trademark law itself tries to incorporate many of these concerns, so awareness of their practical effects improves the clarity of the doctrinal trademark inquiry itself.

V. CONCLUSION

This Article has argued that search engine law is important and that it is complicated. It is important because it exists at the point of convergence of many strands of Internet law. It is complicated for the same reason. The bulk of this Article has been an examination of the many doctrines from which we must assemble a coherent law of search engines and of their many interrelationships. Search engine law is a system with many moving parts but

275. *See generally* Stacey L. Dogan & Mark A. Lemley, *Trademarks and Consumer Search Costs on the Internet*, 41 HOUS. L. REV. 777 (2004) (arguing that the "use in commerce" requirement includes a requirement that a trademark be used *as* a trademark). *But see* Graeme B. Dinwoodie & Mark D. Janis, *Confusion Over Use: Contextualism in Trademark Law*, 92 IOWA L. REV. 1597, 1609, 1629–36 (2007) (arguing that "use in commerce" incorporates no such requirement and that such a requirement would be unwise, particularly for search engines).

276. *See, e.g.*, Coca-Cola Co. v. Overland, Inc., 692 F.2d 1250, 1252 (9th Cir. 1982).

few degrees of freedom, and the challenge will be to satisfy as many competing policy demands at once as possible.

Legal scholars have much work to do in the search space. Isolated patches—the trademark law of keyword-advertising sales, access to computer systems, intermediary liability for the flow of copyrighted materials, and a few others—have received careful and sustained scholarly attention. But these efforts must be connected and supplemented with equally thoughtful analyses of the many other specific conflicts created by search. And they must be connected with more overarching studies of larger themes in search engine law.

This Article has repeatedly referred to some of these themes: the tension between transparency and secrecy in search engine operations; the relationship of competition among providers and among search engines; the power of search engines to promote and infringe upon the privacy of users, providers, and third parties; the role of search engines in enhancing and inhibiting free speech; and the political economy of innovative freedom and others' claims upon search engines. A fuller discussion of these themes must await other days and other articles. The need for such further study should by now be apparent.

As of this writing, Google lists 15,800 results for "search engine law." That number will only increase.