

# Through the Google Goggles: Sociopolitical Bias in Search Engine Design

by

*Alejandro M. Diaz*

Submitted to the Program in Science, Technology and Society  
in Partial Fulfillment of  
the Requirements for

Interdisciplinary Honors

at

Stanford University  
May 2005

© 2005 Alejandro Diaz  
All rights reserved

The author hereby grants to Stanford University permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.

*Signature of Author* .....

*Program in Science, Technology and Society*

*May 23, 2005*

# Through the Google Goggles: Sociopolitical Bias in Search Engine Design

by

Alejandro M. Diaz

Submitted to the Program in Science, Technology and Society  
in Partial Fulfillment of the Requirements for  
Interdisciplinary Honors

**ABSTRACT:** As much of our knowledge, news, and discourse moves online and to the Web in particular, search engines are increasingly becoming the “gatekeepers” of cyberspace. What’s more, a *single* search engine—Google—now handles the majority of Web queries. Google directs hundreds of millions of users towards some content and not others, towards some sources and not others. As with all gatekeepers (e.g., television networks), if we believe in the principles of deliberative democracy—and especially if we believe that the Web is an open, “democratic” medium—then we should expect our search engines to disseminate a broad spectrum of information on any given topic. But *unlike* most other gatekeepers, the information disseminated through modern search engines is not explicitly chosen and written by journalists, editors, and producers. It is instead largely determined by a complex system of algorithms, hardware, and software. The varied designs for search technologies *encode* certain values about what sort of content is “important,” “relevant,” or “authoritative.” In this thesis, following a hybrid approach that incorporates both media studies and STS theories, we will look at the biases of, motivations for, use of, and resistance to the Google search engine. It is hoped that through this analysis, we might start to uncover the sociopolitics of search.

Thesis Supervisor: Professor Robert McGinn  
Title: Chair, Program for Science, Technology, and Society

# Acknowledgements

I'd like to thank my advisor and mentor, Professor Robert McGinn, for supporting my work during this two-year journey, for giving me the wonderful opportunity to TA, and for teaching me all I know about STS. I am grateful to Professor Aneesh for leading the Honors College (back in 2003) and for keeping STS novel and exciting. Many thanks to Professor Winograd for his insightful comments. And to my new advisor, Professor Turner: bless you for letting me turn in that paper late.

This wouldn't be a Stanford honors thesis if I didn't thank the Undergraduate Research Program (URP) for supporting this endeavor.

Many thanks to my friends—Rebecca, Joanna, Brett, Cheryl, Tamara, Logan, Katherine, Dawn, Saleem, Leonard, Tess—for distracting me. It's your fault this took so long. To Iceland: your music is what kept me writing.

Finally, thanks to my Dad, for always being so proud of me. To my Mom: I owe it all to you. Including those five-hundred-dollars I had to pay the library because you mailed my thesis research back to Stanford and didn't tape the box properly.

# Table of Contents

1. Introduction .....	1
2. Technologies of Bias .....	5
<i>Frameworks, Foundations, and a Review of the Literature</i>	
3. Democracy, the Media, and Search Engines .....	18
<i>The Prospects for a 'Democratic' Cyberspace</i>	
4. Making Sense of Search .....	51
<i>A Technical Overview of the Web, Search Engines, and Google</i>	
5. The Politics of PageRank .....	62
<i>Do the Rich Get Richer?</i>	
6. The Transparency Dilemma .....	95
<i>How Hidden Heuristics, Commercial Optimizers, and Manual Exclusion     Silently Shape Google's Results</i>	
7. Advertising and "Mixed Motives" .....	109
<i>Exploring Search Engine (Hyper)commercialization</i>	
8. Googleopoly .....	128
<i>Consolidation, Concentration, and the Future of Search</i>	
9. Conclusion .....	158
Appendix I: The Random Surfer Model .....	A-1
Appendix II: Quantifying the Role of PageRank .....	A-4
Bibliography	

# 1 Introduction

Daniel Brandt is a bit of a conspiracy theorist. Lately, he's been busy speculating about the dark side of Google on his web site, *google-watch.org*. On a series of pages—including one entitled “Spooks on Board at Google”—Brandt links the search engine, through the inventor of the Segway scooter, with the FBI, the Bush Administration, and the NSA.<sup>1</sup> Here he readily uses phrases like “government surveillance,” “big brother,” “spies in Washington,” and “top-secret security clearance.” Brandt argues that because Google maintains detailed logs of each user's search terms, tracks IP addresses, and is a virtual monopoly on internet search, it holds private information about the online activities of almost every Internet user. This information, he argues, can be used and misused by Google, corporations, and especially post-9/11 government agencies.

But Brandt has another bone to pick with Google. In a separate document on his site, he takes issue with PageRank, the proprietary algorithm at the heart of the search engine's ranking system.<sup>2</sup> Calling PageRank Google's “original sin,” he argues that it is “uniquely tyrannical,” favoring large, corporate, and established sites over new, independent, and often more relevant pages. Because Google orders its results by PageRank, “it's frequently the case that a...perfectly relevant page [...] will get buried in the rankings because it isn't sufficiently popular.”<sup>3</sup> In addition, it allows corporations with greater budgets to optimize their sites for better placement among the results. He argues that these factors, combined with Google's overwhelming market share, effectively silence smaller sites and minority opinion on the Internet. With a pinch of alarmist flare, he calls for FTC to regulation of “advertising agencies

that parade as search engines.”<sup>4</sup> He even likens Google’s monopoly and behavior to that of Microsoft, forecasting “world domination again ... this time with cute colored letters.”<sup>5</sup>

Put lightly, Brandt’s argument seems a bit of a stretch. Google’s meteoric rise was not the result of advertising, corporate maneuvers, or illicit games. It became popular, it appears, because it simply provides what users want. And so, Brandt has become a bit of an outcast in the Web community. On the forums of the technology site *ArsTechnica*, for instance, users express little sympathy for his concerns:<sup>6</sup>

*Google rocks. Anyone who says otherwise should be shot.*

*There are two reasons that Google's status should not be concerning: (1) They've been good, so far. (2) Anyone can do it. Everyone talks like Google controls the internet, but obviously, Google's users give it that control. Anyone with bandwidth and server space (ie \$\$) could compete if they wanted. No monopolies here.*

*because in an ideal world if you want to know where to find an ATM, you should get a lecture about how all the banks are corrupt and stealing your money... Someone please tell me the guy this article is about doesn't actually believe what he says... I use google [sic] all the time, without ever worrying about it turning to the dark side. As long as it's free, I'm gonna stick with it.*

Users interviewed for a 2005 study by the Pew Internet and American Life Project expressed similar satisfaction and trust in Google, defending their exclusive reliance on the site relatively simple ways:<sup>7</sup>

*Google is clean, fast and thorough.*

*I use Google, it is fast and comprehensive.*

*Google is the search engine I use 98% of the time. I use it almost exclusively because it is fast and accurate. I go directly to vendor sites when I have that option.*

*I use Google because it gives me better searches.*

*It's fast and what I'm looking for is almost always in the top page of results.*

Some users have even gone so far as to write the company beaming “thank you letters”:

*I wanted to thank you for such a great product. If google came in a box, I'd eat it for breakfast. If Google were a waitress, I would leave 20 percent. If Google was next to me on a plane, I would let it have the window seat. How do you do it? How can you be that good and that fast? It's like magic.<sup>8</sup>*

Despite these positive sentiments, we should not write off Brandt's rant quite so swiftly. If Google does systematically disfavor new and underrepresented voices after all, then many broad social ideals—of competition, free speech, and deliberative democracy—are potentially undermined. Such a bias is problematic even if it is neither intentional nor premeditated, and even if popular use seems to condone it. As we increasingly turn to Google for information, we must ask ourselves how the search engine is influencing what we do and do not read when we go online.

Google and other search engines are, quite clearly, in part responsible for guiding us towards a particular sources of information, and away from others. Search engines therefore play a key role in mediating the interaction between Web users and content authors. These three primary stakeholders—the user, the search engine, and the sites to be searched—may have competing interests, and a negotiation among them is carried out with each search. The user, on the one hand, is interested in finding the most “relevant” information. Web pages, in contrast, are all vying to appear towards the top of the result list—even if they are not the most relevant sources for the user. Search engines are ultimately responsible for reconciling these competing interests, but they themselves may have a (commercial) stake in satisfying the largest number of searches as quickly as possible while selling as much advertising space as they can. It may often be the case, as scholars have pointed out, that the design of search engines is biased against users, content authors, or both. Even though most may not be aware of these dynamics, the central role of searching suggests that these forces do go far in shaping our online experiences.<sup>9</sup>

It is unfortunate that just as users have, for the most part, unequivocally celebrated Google's search technology, so too has “Google's information interface ... remained largely exempt from the type of social, political, and ethical criticism that other information technologies have received from scholars of technology.”<sup>10</sup> It is the goal of this thesis to turn a critical-analytical eye towards the Google search engine, and from this perspective highlight some of the sociopolitical dimensions of search engines in general. This tricky task will require

that we survey theories, research, and data from disparate fields. To understand the workings of search engines and algorithms such as PageRank, for example, we need to delve into mathematics behind this algorithm. Using some new ideas in network theory, we sketch the topology of the World Wide Web and assess how Google's technology relates to it. Most importantly, however, we turn to existing work in the fields of communications, political science, and STS in order to recognize how various social forces can be—and are—exerted through technology and the media. It is our hope that in doing so, we can determine what biases, if any, are latent in Google's search technology, and how these might further or inhibit crucial sociopolitical ideals.

## Notes

---

- <sup>1</sup> Daniel Brandt, *Spooks on Board at Google* (2004 [cited April 4 2004]), available from <http://www.google-watch.org/jobad.html>.
- <sup>2</sup> Daniel Brandt, *Why We Target Google* (2004 [cited April 2 2004]), available from <http://www.google-watch.org/bigbro.html>.
- <sup>3</sup> Ibid.
- <sup>4</sup> Ibid.
- <sup>5</sup> Ibid.
- <sup>6</sup> Hannibal, *Reading Notes: Archive Fever* (ArsTechnica OpenForum, June 27 2003 [cited May 11 2005]), available from <http://episteme.arstechnica.com/6/ubb.x?a=tpc&s=50009562&f=174096756&m=2010925275&r=2010925275>.
- <sup>7</sup> Deborah Fallows, *Search Engine Users: Internet Searchers are Confident, Satisfied, and Trusting--But They Are Also Unaware and Naive* (Pew Internet and American Life Project, January 23 2005 [cited May 22 2005]), 14, available from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf).
- <sup>8</sup> Peter Norvig, in Google Inc., *Factory Tour* (Google.com, 2005), Video, available from <http://www.google.com/intl/en/press/factorytour.html> (accessed May 20, 2005).
- <sup>9</sup> These processes are discussed in Lucas Introna and Helen Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society* 16, no. 3 (2000).
- <sup>10</sup> Michael Zimmer, *Project Description* (Graduate Student Workshop: Values in Computer and Information System Design, 2005 [cited May 8 2005]), available from <http://epl.scu.edu/~stsvvalues/students.html>.



# 2

## Technologies of Bias

Frameworks, Foundations, and a Review of the Literature

Scholars in STS have long examined the ways in which various technologies are imbued with particular social and political values, and how the consequences of these technologies are, in turn, mediated by myriad social factors. Consistent with this approach, many have looked at the motivations behind and implications of computer and information systems. Web search engines, however, have received relatively little attention from the STS community or, indeed, from the academic community in general. By conducting what we believe is the most comprehensive discussion of the sociopolitics of Google to date, we hope to update the existing literature on the design of these critical systems. In this chapter we will give an overview of our theoretical approach and briefly survey the existing work on search engine design.

### I. Foundations and Frameworks

We begin by examining the ways in which technologies, in general, can encode social values and political bias. The idea that a particular arrangement of circuits, a certain architectural design, or some lines of computer code could “have” politics of their own seems, at first glance, a bit tenuous. This is particularly so given the commonly held belief that technology is a value-neutral tool, furthering or inhibiting disparate social aims depending on how it is used. Raising a critical eye to such widespread beliefs, it makes sense to ask whether technologies themselves can presuppose, require, or instigate certain forms of social organization.

In a famous paper, Norbert Winner tackles this question, namely “Do artifacts have politics?”<sup>1</sup> Arguing against both naïve technological determinism and extreme social constructivism, he first concludes that although social forces can and do shape technological development, “technology is politically significant in its own right.”<sup>2</sup> Winner identifies two ways in which artifacts can contain political properties. One instance involves “inherently political technologies”: those that require or are strongly compatible with specific political systems. Arguments about such technologies, he admits, are “troublesome” for many of the reasons levied against technological determinism. Yet he affirms that a few, very special technics *intrinsically* require, or are at least very strongly compatible with, particular political systems. The administration and control of uranium-based nuclear power, for instance, requires a strong central authority.<sup>3</sup>

Winner’s second type of political technologies, and those most relevant for our discussion, are those in which “specific features in the design or arrangement of a device or system could provide a convenient means of establishing patterns of power and authority in a given setting.”<sup>4</sup> In these cases, the particular design of a device is variable. This flexibility allows social actors to select the features of a given implementation, possibly projecting political biases into their chosen design. His numerous examples include Robert Moses’ infamous Long Island overpasses, which were built at a low height in order to exclude public buses—used largely by low-income African Americans—from traveling to public beaches. Artifacts of this kind encode and entrench specific social structures, suggesting that technologies can be, contrary to popular wisdom, value-laden:

*If our moral and political language for evaluating technology includes only categories having to do with tools and uses, if it does not include attention to the meaning of the design and arrangements of our artifacts, then we will be blinded to much that is intellectual and practically crucial.*

Winner’s fundamental insight is that the process of technical engineering has a political dimension, as it determines what actions and actors are favored or restricted by the technology in question. This idea is echoed by noted law professor Lawrence Lessig, who argues that the “architecture” of a technology—how it is designed and built—can regulate behavior, much like

law, norms, and the market do.<sup>5</sup> In particular, Lessig describes how hardware and software code (the architectures of cyberspace) can shape what users can and cannot do, read, or say when they go online. He writes,

*Code ... constitutes a set of constraints on how you can behave. The substance of these constraints may vary, but they are experienced as conditions on your access to cyberspace. In some places you must enter a password before you gain access; in other places you can enter whether identified or not ... The code or software or architecture or protocols set these features; they are features selected by the code writers; they constrain some behavior by making other behavior possible, or impossible. The code embeds certain values or makes certain values impossible.<sup>6</sup>*

For Lessig, as for Winner, there is no question: devices can be, and often are, embedded with politics of their own. Their examples are meant to show how social actors can and do consciously and intentionally “shape” their technologies with particular aims in mind and that, in particular target contexts, these technologies in turn “shape” social relations. Thus, when the authors speak of the consequences of a particular technology, what they are really referring to is the outcome of an entire sociotechnical process of development.

While the most prominent examples of such shaping deal with intentional, premeditated bias, it is important to note that social forces can work their way into technological systems in a far more subtle, almost invisible, manner. Winner’s views on this point are instructive:

*To recognize the political dimensions in the shapes of technology does not require that we look for conscious conspiracies or malicious intentions ... Indeed, many of the most important examples of technologies that have political consequences are those that transcend the simple categories of “intended” and “unintended” altogether. These are instances in which the very process of technological development is so thoroughly biased in a particular direction that it regularly produces results heralded as wonderful breakthroughs by some social interest and crushing setbacks for others. In such cases, ... one must say that the technological deck has been stacked in advance to favor certain social interests.<sup>7</sup>*

Winner plainly acknowledges the possibility for “unintentional consequences” yet he maintains that, in the end, such political effects stem not from an “internal dynamic” but from the very social process of development. In support of his position, he describes the University of California’s development of the mechanical tomato harvester, a device that automated the

process of picking, shaking, and sorting tomatoes, and drastically reduced the cost of harvesting. The benefits, Winner points out, were not distributed equally; the initial cost could only be met by larger growers, while smaller farms could not compete with their more technologically advantaged competitors. As a result, California witnessed a dramatic concentration of its tomato harvesting industry, with the number of growers plummeting by 85% in spite of an increase in total tomato output. Moreover, tens of thousands of people—primarily farm workers—lost their jobs due to mechanization. Because the tomato harvester favored large growers at the expense of small farmers and workers, it led to “a thorough reshaping of social relationships involved in tomato production in rural California.”<sup>8</sup>

The social consequences of the automatic tomato harvester were not planned or desired by its developers. Instead, its specific design was determined by “an ongoing social process in which scientific knowledge, technological invention, and corporate profit reinforce each other in ... patterns that bear the unmistakable stamp of political and economic power.”<sup>9</sup> Consequently, it is possible—and, indeed, accurate—to say that the machine was designed in a way that favored large agribusiness interests, without directly implicating its designers for such bias. The case of the tomato harvester may seem tangential to our discussion of Google, but illuminating parallels can be made. Much like the harvester, this search engine grew out of a university research project; both technologies radically transformed their respective industries; in the end, as we will see, both ended up favoring some social groups at the expense of others. The lesson we should take from Winner’s analysis is that, when examining the social consequences latent in the design of a technology, we must also look at the broader social context of its development.

Although Winner’s argument seems convincing, useful, and insightful, it has been rather strongly criticized in the decades since its initial publication. On a purely factual note, several have called into question the facts surrounding Winner’s classic tale about Moses and his Long Island bridges.<sup>10</sup> More significant are the criticisms that he oversimplifies and distorts the complex interaction between society and its technology. At issue is Winner’s

apparent support for the theory of *soft technological determinism*, which more or less ignores the ways in which the social outcome of a technology is dependent on both its design *and* its context of adoption. As Joerges puts it, Winner adopts a “control” model (social order as a result of intentional action) rather than a “contingency” model (social order from intentional *and* reactive action).<sup>11</sup> And, as a result, he only gets the story half right. Yes, technological development stems not from an “autonomous force” or “internal dynamic” but largely from social, even political processes. But society also strongly influences how a particular technology is adopted and used, and so it too shapes the imprint left by those technologies. Although Winner makes such a concession,<sup>12</sup> it is argued that he does not take it sufficiently to heart in his analysis.

This more complete, reciprocal understanding of “technically induced social change” has gained the favor of recent scholars such as McGinn, who formalizes this idea in his IDUAR model.<sup>13</sup> McGinn identifies five interdependent variables that together determine the social effects of a given technology: the design of the *innovation* itself, its means of *diffusion*, its patterns of *use*, the degree of social *adaptation*, and society’s *resistance* to the technology. This model, which unambiguously asserts the role of society in both the design and adoption phases, runs against beliefs that “the social outcome of a...technological innovation depends solely on its inherent characteristics and momentum of introduction.”<sup>14</sup> If we are to correctly identify the sociopolitical consequences of Google’s technology, we should be careful to not make the same mistake. We must take into consideration the many powerful social forces at play.

Taken together, Winner, Lessig, and McGinn provide valuable insight, examples, and models for investigating the political effects and social design of technologies in general. Friedman and Nissenbaum, however, offer a more focused framework for looking at bias in computer systems in particular.<sup>15</sup> Information technologies, they argue, are particularly problematic since

*biases in computer systems can be difficult to identify ... Computer systems, for instance, are comparatively inexpensive to disseminate, and thus, once*

*developed, a biased system has the potential for widespread impact. If the system becomes a standard in the field, the bias becomes pervasive. If the system is complex, and most are, biases can remain hidden in the code, difficult to pinpoint or explicate, and not necessarily disclosed to users or their clients. Furthermore, unlike in our dealings with biased individuals with whom a potential victim can negotiate, biased systems offer no equivalent means for appeal.<sup>16</sup>*

Although these words were written years before the birth of Google, they perfectly summarize our own concerns surrounding the search system: its rampant dissemination, unprecedented impact and pervasiveness, enormous complexity, and utter lack of transparency.

In proposing a useful model for surmising the sociopolitical tendencies of such systems, the authors begin by clarifying what, exactly, “bias” means in this context:

*In its most general sense, the term bias ... is applied with relatively neutral content ... At other times, the term bias is applied with significant moral meaning. We use the term bias to refer to computer systems that systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others. A system discriminates unfairly if it denies an opportunity or a good or if it assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate.<sup>17</sup>*

The authors identify three general categories of “systematic and unfair” discrimination in computer systems: preexisting bias (which arises from individual or social forces shaping the design), technical bias (which stems from technical constraints or considerations), and emergent bias (which has to do with the context of use). These categories, like McGinn’s five variables, require us to take into account both technical *and* social factors when assessing the “consequences” of a particular information system.

Preexisting bias, under Friedman and Nissenbaum’s formulation, is of the same sort described by Winner, arising “when computer systems embody biases that exist independently, and usually prior to the creation of the system...through the explicit and conscious efforts of individuals or institutions, or implicitly and unconsciously.”<sup>18</sup> Technical bias, on the other hand, originates not from social forces but from the practical limitations of the technologies themselves. One salient instance of this stems from the “formalization of human constructs” through computer algorithms:

*[This is] bias that originates from attempts to make human constructs such as discourse, judgments, or institutions amenable to computers: when we quantify the qualitative, discretize the continuous, or formalize the nonformal.<sup>19</sup>*

In other words, due to the deterministic and finite nature of computers, bias may arise whenever designers seek to translate inherently ambiguous, subjective concepts into precise, unambiguous code. Doing so necessarily requires that they eschew some of the contextual, “human,” and “humane” interpretations necessary for making sound judgments about subjective matters. (Imagine, for instance, if a college used an “admissions program” to decide which applicants would be accepted. Such a program could only approximate the more subjective and nuanced opinions of admissions officers; the nature of its inevitable shortcomings—say, in taking into account difficult family circumstances—would determine whether and how the software is systematically biased against some applicants.) This point is remarkable given its direct applicability to the Google search engine—after all, the process of automatically locating and ranking “relevant” information is *precisely* a process of “quantifying the qualitative” in order to make online “discourse...amenable to computers.”<sup>20</sup>

After spelling out their framework, Friedman and Nissenbaum apply it to various contemporary computer systems, illustrating and clarifying the distinctions between their categories of bias. The most salient example they put forth is that of flight reservations systems, which return to travel agents a sorted list of flights matching their search criteria. The authors show how such systems can exhibit the various forms of bias: a preexisting profit-driven tendency of the airline companies who own the systems to favor their own flights; technical bias imposed by computer technology, which requires that search results be divided into “screens” of a few results each; and emergent bias resulting from changes in how the system was used after its initial implementation. This example indicates how we too might go about explicating bias in the Google search engine.

Our discussion of Google’s technology is informed by Friedman and Nissenbaum’s approach to computer system bias as well as by the more traditional, general purpose frameworks (e.g., McGinn’s IDUAR model). We will, however, take some liberties with respect

to terminology. Most notably, we will generally use the term “bias” in the neutral, rather than the “moralistic,” sense. We believe that it is necessary to explicitly separate the *identification* of systematic sociopolitical tendencies from *moral or political conclusions* about unfairness or inappropriateness. As Friedman and Nissenbaum concede, it is often the case that judgments of the later form are exceedingly difficult to make; quite often, what is at stake are various reasonable but incompatible ideals. This is emphatically so in the arena of search engines, where discrimination is both pervasive and necessary. The whole point of search engines, after all, is to discriminate among the billions of online documents; the designers of these systems must therefore choose which sorts of content they wish to promote, and which kinds they will bury among the results. In doing so, they will inevitably favor some valid interests over other, similarly valid interests. In this light, even “systematic and unfair bias” is unavoidable. The question, then, is not really *whether* bias exists but *what* sort of bias we should prefer. Since this normative question is beyond the scope of the STS frameworks discussed here, we leave it for the next chapter, where we propose a “deliberative” ideal along which to evaluate computer-mediated communication systems.

## II. Existing Research on Search Engine Bias

Having identified the frameworks of our discussion, we now turn to the library and information sciences community for existing work on bias latent in search systems. Salton provides a critique of the “match based” approach to information retrieval—common among search engines including Google—in which a user submits a set of keywords (a query) and is returned a set of relevant documents.<sup>21</sup> He criticizes the match paradigm on three counts: first, the search terms used may not accurately reflect the needs of users and may be skewed; second, the methods used to construct the set of relevant documents returned may be biased; and finally the size of the test collections used during development may be unrealistically small. Ellis,<sup>22</sup> Cooper,<sup>23</sup> and other writers<sup>24</sup> have similarly suggested that the measures used to



evaluate the effectiveness of information retrieval systems—recall (the number of documents returned) and precision (the relevancy of the results)—may be incorrect or insufficient. This more or less mirrors Friedman and Nissenbaum’s general conclusion that “freedom from bias should be counted among the select set of criteria—including reliability, accuracy, and efficiency—according to which the quality of systems in use should be judged.”<sup>25</sup> We can extend such arguments to search engines on the World Wide Web, suggesting how traditional metrics of effectiveness may not accurately measure a search engine’s ability to produce useful, fair, or appropriate results.

As it turns out, although there is a great deal of research in the IR community on the design and use of information retrieval systems, the work does not (in general) go very far in answering our particular research questions about the Google search engine. For one, many of the search technologies discussed actually operate on library databases, which cover a set of documents substantially different from those found on the Web. In addition, even when the research does focus on Web search engines in particular, it is usually assumed that the users are information professionals and not “typical” Web users (Gordon and Pathak, for instance, tested search engines as “used by highly effective searchers” such as librarians<sup>26</sup>). Most importantly, however, those in the IR community are largely concerned with problems of effectiveness and accuracy and not the sociopolitical biases highlighted by Friedman and Nissenbaum.

Unfortunately, once we look outside of the library sciences, we find that only a handful of articles have appeared addressing the broader issues raised by search engine bias on the World Wide Web. Mowshowitz and Kawaguchi,<sup>27</sup> in one of the few such analyses, quantitatively analyzed the statistical “bias” present in various search engines by comparing a particular search engine’s results with the “average” set of results returned by all engines. Their conception of “bias,” however, is a very different than our own; it is far more mathematical than social, determined by “measuring the deviation from the ideal of the distribution produced by a particular search engine.”<sup>28</sup> While they measure this special kind of

search engine “variance” in great detail, they do not, for instance, describe what sort of content is suppressed or supported by the various search engines. Nor do they attempt to identify the role of individual algorithms such as Google’s PageRank in determining the observed divergence. Their findings are, on the other hand, merely meant to show that search engines differ widely in the results they generate.

The sociopolitical dimensions of search engines were, however, finally discussed in 2000, when Douglas Introna and Helen Nissenbaum published their groundbreaking article on the subject.<sup>29</sup> “Search engines,” they argue, “raise not merely technical issues but also political ones,” since these systems “systematically exclude certain sites, and certain types of sites, in favor of others, systematically giv[ing] prominence to some at the expense of others.”<sup>30</sup> This article is unique among the existing scholarly literature in that it directly addresses our general concerns about technological bias in search systems and it devotes at least some attention to one aspect of Google’s unique search technology (PageRank). This article is so instructive that we consider it the primary starting point for our own research, and we will refer to the concepts, ideas, and approaches used by these authors throughout our analysis.

There are various reasons, however, why Introna and Nissenbaum’s analysis falls quite short of answering our research question. First, because it was their aim to widely survey search engine technologies in general, and because in 2000 Google had less than a 5% market share,<sup>31</sup> they do not discuss the search engine in great detail (in fact, the word ‘Google’ never appears in the article). Second, the authors focus almost entirely on the politics of search *technologies* but for the most part ignore other social forces mediating users’ responses to search engines. As we will see in the next chapter, the ideals for deliberative media promoted by Introna and Nissenbaum require that we also look, for example, at advertising practices, concentration, and ownership. This is emphatically the case given Google’s meteoric rise in market share (handling, at one point, more than three-quarters of all search traffic in the U.S.<sup>32</sup>). Third, these authors seem to rather uncritically align the Web with democratic principles. We will attempt to properly situate their concerns within recent trends in Web

content by showing that, despite early Barlovian claims of being “inherently democratic,” this space is itself developing in rather biased and anti-democratic ways. Finally, we would like to at least briefly discuss how social and regulatory pressures—in addition to commercial and technological factors—influence search results.

It should be noted that, during the course of this project, a number of articles have emerged that started to articulate these broader concerns about search engines. In 2003, Hindman and his colleagues examined the links between millions of “political” pages on the Web and surmised the implications for finding various kinds of pages via modern search engines.<sup>33</sup> In 2004, Van Couvering used a structural, political-economic approach to describe the state of advertising and consolidation in the search space.<sup>34</sup> That same year, Gerhart performed a content analysis of search results to explore whether “search engines suppress controversy.”<sup>35</sup> Although none of these articles examine these important issues from the perspective of the Google search engine, they do provide some of the necessary context, data, and theoretical motivations for our analysis.

To reiterate, although the existing research does provide informative and varied perspectives from which to embark on our discussion of Google, it does not paint a complete picture of the search engine’s biases. The literature is insufficient in several regards: it considers only a small portion of Google’s search technology; it fails to acknowledge and analyze the implications of Google’s monopoly status and recent IPO filing; it does not discuss advertising on the search engine in particular; it rests on critical but unjustified assumptions about the Web as a medium; and it focuses narrowly on a small subset of the relevant sociopolitical pressures exerted on and by search engines. We hope to address these shortcomings in the chapters to follow.

## Notes

---

<sup>1</sup> Langdon Winner, “Do Artifacts Have Politics?,” *Daedalus* 109, no. 1 (1980).

<sup>2</sup> *Ibid.*: 123.

<sup>3</sup> *Ibid.*: 128-34.

- 
- <sup>4</sup> Ibid.: 125.
- <sup>5</sup> Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York, NY: Basic Books, 1999).
- <sup>6</sup> Ibid., 89.
- <sup>7</sup> Winner, "Do Artifacts Have Politics?," 125-26.
- <sup>8</sup> Ibid.: 126.
- <sup>9</sup> Ibid.
- <sup>10</sup> Woolgar shows that, contrary to Winner's suggestion, public transit has long provided bus service from New York City to Jones Beach, though not on the parkways. Steve Woolgar, "Do Artefacts Have Ambivalence?: Moses' Bridges, Winner's Bridges, and Other Urban Legends," *Social Studies of Science* 29, no. 3 (1999).  
In another a scathing criticism of Winner's essay and his Long Island "parable," Joerges discredits Winner's assertion that Moses was a racist, classist power-monger out to embed discriminatory values in his technologies. See Bernward Joerges, "Do Politics Have Artifacts?," *Social Studies of Science* 29, no. 3 (1999)..
- <sup>11</sup> Joerges, "Do Politics Have Artifacts?," 413.
- <sup>12</sup> Winner, "Do Artifacts Have Politics?," 122.
- <sup>13</sup> Robert E. McGinn, *Science, Technology, and Society* (Englewood Cliffs, NJ: Prentice Hall, 1991), 98-101.
- <sup>14</sup> Ibid., 94.
- <sup>15</sup> Batya Friedman and Helen Nissenbaum, "Bias in Computer Systems," *ACM Transactions on Information Systems* 14, no. 3 (1996).
- <sup>16</sup> Ibid.: 331.
- <sup>17</sup> Ibid.
- <sup>18</sup> Ibid.: 332.
- <sup>19</sup> Ibid.: 334.
- <sup>20</sup> Ibid.
- <sup>21</sup> Gerard Salton, "The State of Retrieval System Evaluation," *Information Processing & Management* 28, no. 4 (1992).
- <sup>22</sup> David Ellis, "Theory and Explanation in Information Retrieval Research," *Journal of Information Science* 8 (1984).
- <sup>23</sup> W. S. Cooper, "Some Inconsistencies and Misidentified Modelling Assumptions in Probabilistic Information Retrieval," *ACM Transactions on Information Systems* 13 (1995).
- <sup>24</sup> See, for example, Renata Tagliacozzo, "Self-Citations in Scientific Literature," *Journal of Documentation* 33, no. 4 (1977). Also, R. S. Taylor, *Value-Added Processes in Information Systems* (Norwood, NJ: Ablex Publishing, 1986).
- <sup>25</sup> Friedman and Nissenbaum, "Bias in Computer Systems," 345-46.
- <sup>26</sup> Michael Gordon and Praveen Pathak, "Finding Information on the World Wide Web: The Retrieval Effectiveness of Search Engines," *Information Processing & Management* 35, no. 2 (1999).
- <sup>27</sup> Abbe Mowshowitz and Akira Kawaguchi, "Bias on the Web," *Communications of the ACM* 45, no. 9 (2002).
- <sup>28</sup> Ibid.
- <sup>29</sup> Lucas Introna and Helen Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society* 16, no. 3 (2000).
- <sup>30</sup> Ibid.: 169.
- <sup>31</sup> Danny Sullivan, *Nielsen NetRatings Search Engine Ratings, December 2000* (SearchEngineWatch.com, February 19 2001 [cited December 21 2004]), available from <http://searchenginewatch.com/mhts/9902-0012-netratings.mht>.
- <sup>32</sup> Danny Sullivan, *comScore Media Metrix Search Engine Ratings, August 2003* (SearchEngineWatch.com, August 1 2003 [cited May 1 2005]), available from <http://searchenginewatch.com/mhts/0305-mediаметrix.mht>.
- <sup>33</sup> Matthew Hindman, Kostas Tsioutsoulis, and Judy Johnson, *'Googlearchy': How a Few Heavily Linked Sites Dominate Politics on the Web* (July 28 2003 [cited April 10 2005]), available from <http://www.princeton.edu/~mhindman/googlearchy--hindman.pdf>.

- 
- <sup>34</sup> Elizabeth Van Covering, "New Media? The Political Economy of Internet Search Engines" (paper presented at the *Conference of the International Association of Media & Communications Researchers (IAMCR)*, Porto Alegre, Brazil, July 25-30 2004), available from [http://personal.lse.ac.uk/vancouve/IAMCR-CTP\\_SearchEnginePoliticalEconomy\\_EVC\\_2004-07-14.pdf](http://personal.lse.ac.uk/vancouve/IAMCR-CTP_SearchEnginePoliticalEconomy_EVC_2004-07-14.pdf) (accessed May 18, 2005).
- <sup>35</sup> Susan Gerhart, "Do Web Search Engines Suppress Controversy?," *First Monday* 9, no. 1 (2004), available from [http://firstmonday.org/issues/issue9\\_1/gerhart/index.html](http://firstmonday.org/issues/issue9_1/gerhart/index.html) (accessed May 1, 2005).

# 3

## Democracy, the Media, and Search

### The Prospects for a 'Democratic' Cyberspace

In the previous chapter, we discussed how technologies might be value laden, and we indicated that search technologies in particular should be imbued with a 'democratic' bias. But what does it mean for an item of technology to be 'democratic'? The answer to this question is surprisingly unclear when it comes to predictions and analyses of communication technologies. We frequently read about the printing press "democratizing" the West, hear pundits herald the Internet as "inherently democratic," and so on, yet we rarely get a precise explanation of what, exactly, this means. If we are to successfully make conclusions about the 'politics' of a particular communications technology, it is crucial that we first understand what the relevant ideals *are*, and why they are so important to achieve.

And that is the purpose of this chapter: to introduce the requisite political, historical, and economic foundations for our analysis of the Google search engine. Our discussion, of course, will be far from comprehensive; in some cases, it will appear naïve. Nevertheless, it is necessary that we (at least incompletely) answer several important questions: What exactly *is* democracy? What does it presuppose, require, or entail? Are the traditional media democratic? How about the Internet? How might search engines promote or undermine democracy?

#### I. What Is Democracy?

The term 'democratic'—and, indeed, the concept of democracy in general—is used, abused, and misunderstood with astounding frequency. It "is employed so widely that it has lost much of its specificity and meaning,"<sup>1</sup> and "is often used as a term of rhetoric rather than

of definite substance ... meaning virtually whatever the speaker wants it to mean.”<sup>2</sup> Our conception of what it means to be ‘democratic’ has been warped almost beyond recognition;<sup>3</sup> the word is seemingly applied to anything and everything that is deemed ‘good.’<sup>4</sup> In what can only be described as a grotesque distortion of meaning, today “a product that is consumed widely is termed a ‘democratic’ product, as opposed to a product consumed by the few.”<sup>5</sup>

In an attempt to clarify the meaning of this notoriously elastic term, we might first turn to the work of John Dewey, who writes that

*Democracy is a word of many meanings. Some of them are of ... broad and moral import ... But one of the meanings is distinctly political, for it denotes a mode of government, a specified practice in selecting officials and regulating their conduct as officials. This is not the most inspiring of different meanings of democracy; it is comparatively special in character. But it contains about all that is relevant to political democracy.*<sup>6</sup>

The most obvious strategy for understanding ‘democracy’ in this “special” political sense is to follow the etymology of the word: it literally means “rule by the people.” It is a system in which the governed govern, or, as Lincoln famously stated, “government of the people by the people for the people.” Hadenius expands on this basic conception, stating that in a democracy “public policy is to be governed by the freely expressed will of the people, whereby all individuals are treated as equals.”<sup>7</sup> It is a political system, according to Saward, in which there is “necessary correspondence between acts of governance and the equally weighted felt interests of citizens with respect to those acts.”<sup>8</sup> For McLean, democracy is simply “government by majority rule.”<sup>9</sup>

The basic idea here is that, in a democracy, the “will of the people” (in some sense) is enacted (in some sense) by their government. Such a formulation, of course, is so vague that it borders on useless. To more accurately spell out the “principles of democracy,” several have proposed sets of specific criteria for identifying democratic systems. Dahl has prominently put forth his eight minimal “institutional guarantees”;<sup>10</sup> Saward identifies 12 requirements of democratic systems;<sup>11</sup> Fishkin suggests four.<sup>12</sup> Of course, many of these “requirements” are not satisfied by governments that we nevertheless term “democratic.” This is because ‘democracy’

is what Weber called an “ideal type”: it is an abstract analytical construct for assessing the *relative level of democratization* of different political systems. While it may be the case that, according to Dahl, “no large system in the real world is fully democratized,”<sup>13</sup> these criteria allows us to examine how well “democracy” is exercised in a particular political context.

Such models will, as we will see, come in quite handy during our analysis. But for now, it makes sense to review some basic procedural aspects of democratic self-governance. As we have stated, democracy requires that the public’s expressed desires (however defined) be translated into action. How is this done? Our intuition is that it has “something to do with voting.” With a few exceptions,<sup>14</sup> voting is indeed the predominant process by which citizens direct the behavior of democratic governments. This happens either directly (in a *direct democracy*) or indirectly through elected representatives (in a *representative democracy*, or *republic*).

This distinction turns out to be rather important for several reasons. One of them is purely practical. In a true, direct democracy, every citizen is responsible for voting on all matters of government. This presents an obvious scalability problem: when there are many citizens, and the government is complex, it becomes plainly infeasible to order political affairs in this manner. By *electing* representatives to do the policy work on their behalf, citizens of representative democracies are able to have a voice (albeit an indirect one) and yet the scalability issue can be overcome. Ideally, representatives are selected based on their ability to expertly enact the wishes of their constituents, to whom they must answer to come election time.

Representative democracy, in this light, is merely an approximation of “true” direct democracy, designed to address the impracticality of implementing “rule by the people” on a large scale. It has often been suggested (or assumed) that representation sacrifices public good in favor of feasibility and is, therefore, less “democratic” than direct governance. Rousseau famously stated in his *Social Contract* that “the English people believes itself to be free; it is grievously mistaken; it is free only during the election of Members of Parliament; as soon as the



Members are elected, the people is enslaved; it is nothing.”<sup>15</sup> The only true democracy, according to Rousseau, is a direct one. “Every law which the people has not ratified in person,” he said, “is void—it is not law at all.”<sup>16</sup>

Two hundred and fifty years later, the push for direct democracy continues. Public initiatives and referendums have appeared on ballots across the country, allowing citizens to “bypass unresponsive or irresponsible political bodies and directly legislate for themselves.”<sup>17</sup> Recent advancements in communications technology have also made possible a dramatically increased degree of direct political participation. Up-to-the-minute opinion polls, for instance, permeate political coverage in the media, and by all standards have a strong influence on the actions and selection of political representatives. It has even been suggested that, for the first time in history, electronic technologies can make large-scale direct democracy feasible, allowing millions to “tell their government, every week, if not every day, what they would like it to do.”<sup>18</sup> With the “march toward direct democracy” underway, many hope that the barriers to direct democracy will crumble, allowing a truly democratic government to be realized.

But as many others have argued, the prospect of direct democracy “is no reason to celebrate.”<sup>19</sup> After all, the founding fathers chose a republic not only out of practical considerations, but also because they believed that it actually went further in promoting the public good. They feared that a direct democracy, by merely aggregating the preferences of citizens at any given moment, could not distinguish between transient impulses and reasoned, consistent beliefs. Worried about the consequences of yielding to fleeting desires, they wished to prevent temporary passions from being directly translated into law. A representative republic, according to Hamilton, would be preferable precisely because it “demands that the deliberate sense of the community should govern...but it does not require an unqualified complaisance to every sudden breeze of passion, or to every transient impulse which the people may receive.”<sup>20</sup> Representatives would be compelled to act only after an “opportunity for cool and sedate reflection” among the citizenry; laws could only be passed after adequate deliberation in the legislature.

What the Framers had in mind was what is now termed a *deliberative democracy*: one governed by the “deliberate sense” of citizens and representatives alike. They attempted to create institutions that would “refine and enlarge the public views” by encouraging discussion and compromise while mitigating the potential for unbalanced, knee-jerk policy decisions. The bicameral legislature, the Electoral College, and the system of checks and balances are all artifacts of this vision. According to Madison, the views that emerged from such a system would be “more consonant to the public good than if pronounced by the people themselves, convened for that purpose.”<sup>21</sup>

The founders’ design for our government—its “particular practice of selecting officials”—*encodes* certain aspirations about the behavior of its citizens. It is grounded on the idea that deliberation and thought are essential to the success of democracy, and it requires that citizens have time to reflect and debate the issues before changes are made. As Justice Brandeis reminds us,

*Those who won our independence believed that the final end of the State was to make men free to develop their faculties; ... that the greatest menace to freedom is an inert people; that public discussion is a political duty; and that this should be a fundamental principle of American government. They recognized the risks to which all human institutions are subject ... [and] that the path of safety lies in the opportunity to discuss freely supposed grievances and proposed remedies.*<sup>22</sup>

This “fundamental principle” of open discussion is asserted emphatically through the First Amendment, which protects in no uncertain terms the “freedom of the press.” This Amendment was designed to secure a basic freedom—the right to speak—but its aim was also to encourage public debate. If we are to fully grasp the relationship between the press, media technologies, and democracy, let us first examine the role of deliberation in democracy.

## **II. Deliberative Democracy and The Public Forum Doctrine**

Deliberative democracy, by requiring that policies be enacted only after adequate reflection and discussion, is, according to one scholar, a “necessary condition for attaining legitimacy and rationality with regard to collective decision making.”<sup>23</sup> It is indeed an

“essential part of any adequate theory of democracy,”<sup>24</sup> as many have argued forcefully.<sup>25</sup>

Benjamin Page, for example, is convincing when he writes that

*A vigorous democracy cannot settle for a passive citizenry that merely chooses leaders and then forgets entirely about politics. Such a citizenry would not know what it want its public officials to do or what they were actually doing. An ignorant public would have no way to hold its officials to account. There would be a very attenuated sort of democracy, if any sort at all. In order that the public as a whole can collectively control what its government does, the public, collectively, must be well informed. Some kind of public deliberation is required.*<sup>26</sup>

All this makes sense, but what exactly constitutes “deliberation”? Put simply, it is “reasoning and discussion about the merits of public policy.”<sup>27</sup> Such a debate need not be formal; it is not restricted to the Senate floor, nor does it require a moderator and a televised audience. Deliberation can happen anywhere: in your kitchen, when you are reading the morning newspaper; at the barbershop, while discussing current events; or over the telephone, during conversation with a relative abroad. Reflection and debate such as this—even if informal—can go far in informing and shaping our individual views. After hearing your friend’s experiences, you may change your opinion on a matter of foreign policy. Or your current beliefs may only be strengthened.<sup>28</sup> Either way, vibrant discussion helps resolve a contested issue for individuals and, indeed, for society as a whole. Ideally, after the facts have been presented and opinions shared, a consensus is reached in a community and the matter is settled. But even when the issue is not resolved (as is usually the case), deliberation can bring us closer to compromise; it can inform individual opinions such that the aggregated “majority opinion” is more reasoned; and it can raise awareness and even respect for the opinions of our fellow citizens.<sup>29</sup>

Perhaps the most important requirement of successful deliberation is the availability of diverse and antagonistic viewpoints. John Stuart Mill writes eloquently on this subject:

*He who knows only his own side of the case knows little of that. His reasons may be good ... but if he is unable to refute the reasons of the opposite side, if he does not so much as know what they are, he has no ground for preferring either opinion ... Nor is it enough that he should hear the arguments of adversaries from his own teachers ... [H]e must be able to hear them from the persons who actually believe them, who defend them in*

*earnest and do their very utmost for them. He must know them in their most plausible and persuasive form.*<sup>30</sup>

Mill stresses that citizens must be exposed to competing arguments in order to justify or correct their initial, potentially flawed (e.g., one-sided) opinions. It does not matter whether these arguments are popular or unpopular, right or wrong, offensive or pleasing. What matters is that they are heard, and heard fairly. Only by listening to the various arguments can one's position (or vote) be truly informed and valid; only if public opinion can "be set right when it is wrong"<sup>31</sup> can the public good be ensured. Constant deliberation of this sort, as Sunstein puts it, makes possible "a well functioning system of democracy [that] rests not on preferences but on reasons."<sup>32</sup>

Taken to its limit, Mill's prescription for "free discussion" approaches something like what Habermas calls the "ideal speech situation."<sup>33</sup> In this scenario, each person is allowed to speak about anything, for as long as he or she wants, and anyone else can question any assertion made by the speaker.<sup>34</sup> Free of institutional coercion, no voice is encumbered during such a discussion, and all arguments are fairly considered. According to Habermas, such a forum would allow marginalized groups—whose voices have been historically silenced—to fully participate in their democracy; it also allows others to be exposed to the potentially valid arguments made by these groups. Unconstrained by time, he argues, the "force of the better argument" eventually triumphs, and consensus is reached.<sup>35</sup>

Habermas's forum is, of course, entirely hypothetical and implausible, but it does underscore the importance of deliberation and, more specifically, of exposure to minority views. While Mill extols the virtues of access to antagonistic opinions, his core argument is that speech should not be silenced; in this respect, he is asserting a negative right (freedom to speak) rather than a positive right (access to listeners). Habermas and many others have argued that the mere "freedom to speak" does not go far enough, since it does not imply that citizens will actually be exposed to a diversity of opinions. Some ideas may be expressed but only heard by the like-minded. Or, even worse, they may not be heard at all.

Speakers wishing to access a diverse speakers have, however, long found recourse in public spaces—in our parks, streets and sidewalks. Many have stood upon proverbial soapboxes at Speaker’s Corner in London’s Hyde Park;<sup>36</sup> PETA has picketed fur retailers from nearby sidewalks;<sup>37</sup> on April 15, 1967, hundreds of thousands of people marched from Central Park to the U.N. in protest of the Vietnam War;<sup>38</sup> recently, members of the Al-Saleem mosque in San Francisco demonstrated against the granting of marriage licenses to gay couples.<sup>39</sup> For many, such experiences were unwanted, inconvenient, and, perhaps, irritating. But they present one of the rare opportunities for citizens of any class or disposition to express (and be exposed to) varied opinions. Some passersby may change their opinions as a result of such experiences; most will not.<sup>40</sup> Nevertheless, as Mill stressed, what matters is that they *were* exposed to such opinions, and that their own ideas were called into question.

Recognizing the role of “streets and parks” in promoting deliberation, the Supreme Court has long encouraged the use of these spaces to express diverse views to a cross-section of citizens. Justice Roberts declared that

*Wherever the title of streets and parks may rest, they have immemorially been held in trust for the use of the public and, time out of mind, have been used for purposes of assembly, communicating thoughts between citizens, and discussing public questions. Such use of the streets and public places has, from ancient times, been a part of the privileges, immunities, rights and liberties of citizens.<sup>41</sup>*

This idea—of effectively subsidizing speech by making public resources available for citizen use—is called the “Public Forum Doctrine.” It forms an important, but often neglected, component of First Amendment law. Sunstein explains that the Public Forum Doctrine has two principal components. “On the speakers’ side, it creates a right of general access to heterogeneous citizens. On the listener’s side, it creates an opportunity for shared exposure to diverse speakers with diverse views and complaints.”<sup>42</sup> The promotion of such views, according to Justice White, goes to the heart of The First Amendment, which “rests on the assumption that the widest possible dissemination of information from diverse and antagonistic sources is essential to the welfare of the public.”<sup>43</sup>

And yet, the Supreme Court been reluctant to extend the Public Forum Doctrine past “streets and parks,” narrowly conceived. Asserting that “not every public sidewalk is a public forum,” it has upheld restrictions on soliciting of alms and contributions outside postal offices.<sup>44</sup> It has similarly refused to extend the doctrine to include airports<sup>45</sup>—public resources where different citizens commingle—and private spaces.<sup>46</sup> But the Court has recognized that, increasingly, deliberation is not restricted to just “streets and parks.” Justice Kennedy emphasized that the “failure to recognize that new types of government property may be appropriate forums for speech will lead to a serious curtailment of our expressive activity.”<sup>47</sup> For a variety of (seemingly practical) reasons, however, the law simply has not gone very far in ensuring that we are exposed to a diversity of voices, and that seasoned deliberation actually takes place.

The limited applicability of the Public Forum Doctrine, however, should not negate the importance of deliberation for a “well-functioning democracy.” Our opinions are only as valid as they are informed. We must strive to ensure that wherever there is an opportunity for voices to be heard, a fair and diverse range of opinions is actually being expressed. If we are exposed only to arguments that have been carefully controlled and selected, if an invisible hand shapes the debate, if we see only half the story, we stop becoming an informed citizenry. Our genuine opinions become impulses “receive[d] from the arts of men, who flatter [our] prejudices and betray [our] interests,”<sup>48</sup> and we all stand to lose.

In this light, “democracy” is not just a particular “practice of selecting officials” or a curious arrangement of political power. It is “more than voting, and it should serve some purpose other than simply registering preferences.”<sup>49</sup> As activist Abbie Hoffman once said, “democracy is not something you believe in...it’s something you do. You participate.”<sup>50</sup> The term, in other words, encompasses both the design of government *and* the behavior of its citizens. It is a concept that stretches from the political to the social to the individual; indeed, its essence is in interconnecting these spheres.

In its broadest form, ‘democracy’ encapsulates a powerful sociopolitical ideal: that individuals should come together to make decisions that are in the public interest. Doing so requires that important issues be discussed in a fair, comprehensive manner—that is, it requires *deliberation*. This is precisely what media scholars suggest when they speak of “democratic” technologies. It means that voices from every corner of society can be heard, and heard fairly. It means that the discussion is not dominated or manipulated by corporations, politicians, or privileged groups. And it is a conception of ‘democracy’ that is grounded on the theory of deliberation, which in turn flows from the core principles of republican self-government.

### III. Democracy and Traditional Media

The fact of the matter is that today, deliberation is nothing like Habermas’ “ideal speaking situation.” We are a nation too large, and too distributed, to undertake any sort of singular, open, unrestricted debate. Benjamin Page calculates that

*If, for example, a nation of 250 million citizens [had] a fully equal collective discussion of some political issue, ... if each citizen insisted ... upon a rather modest two minutes of speaking time, the discussion would take ... 950 years. Extreme boredom and impatience would result. Little productive work would get done; soon ... the economy would collapse and the deliberators would run out of food to eat.*<sup>51</sup>

For obvious reasons, when we deliberate, we usually do so in small groups, with friends or family, or on the public streets. Barring any rudeness, all the participants in such discussions allowed to speak, but there are very few participants. Individual debates are often fair, but individually, they are minor in scope.

There is, however, another sort of deliberation going on in our society, one with much greater impact but open to far fewer speakers. This is *mediated deliberation*: the discussion of public policy in newspapers, television, radio, books and recently the Internet. As many have observed, mediated deliberation has emerged as the dominant form of public discourse in our society. As we drive our cars, listening to the radio, rather than walking the streets and

sidewalks, as we consume television programming in ever-increasing doses, it is the media (and *not* our face-to-face experiences) that become the primary means by which we are exposed to arguments and opinions. In this sense, Justice Kennedy observes, the media have become the “public forums” of the Information Age:

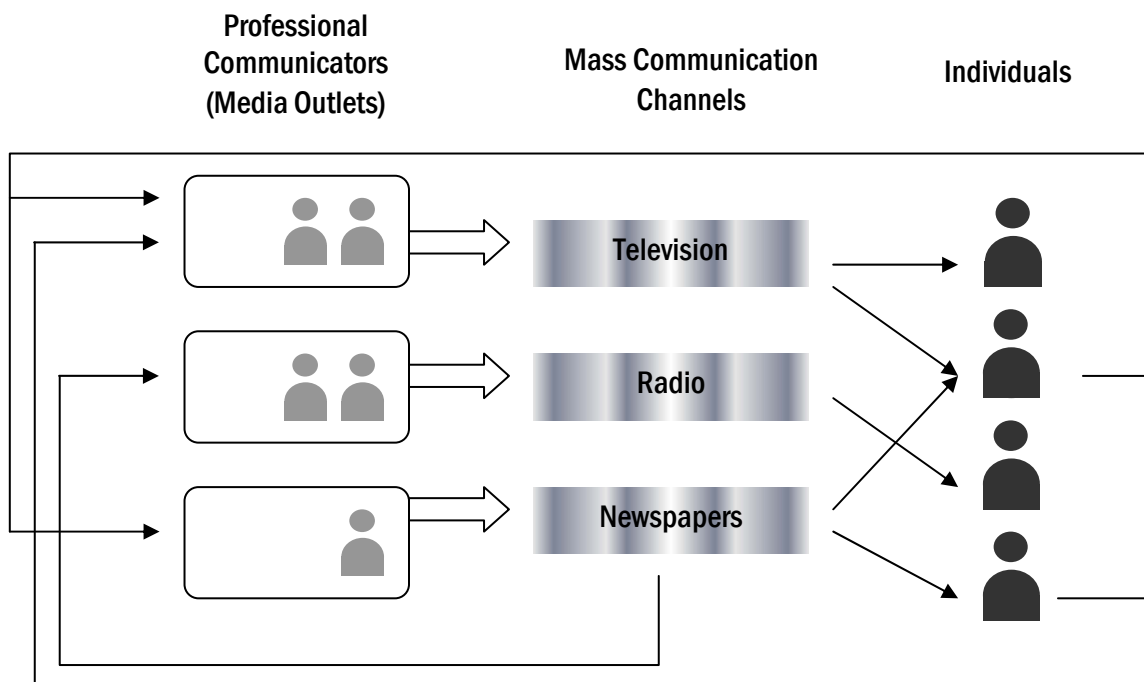
*Minds are not exchanged in streets and parks as they once were. To an increasing degree the more significant interchanges of ideas and shaping of public consciousness occur in mass and electronic media.*<sup>52</sup>

If it was once the role of “streets and parks” to ensure that speakers have access to heterogeneous citizens, if such forums served to expose citizens to minority opinions, then it follows that the mass media should be doing the same. Indeed, at its best, mediated deliberation could be a great boon for democracy. Given the enormous reach of technologies of mass communications, these could allow diverse, underrepresented, and informed opinions to reach millions of geographically and ideologically dispersed citizens like never before.

On the other hand, as Justice Kennedy adds, “the extent of public entitlement to participate in those means of communication may be changed as the technologies change.”<sup>53</sup> Even with the “celestial jukebox” and 500 cable TV channels, there is simply a fixed limit to how many voices can be aired, and how many political issues can be covered. There is also a limit to how much we, as listeners, are willing to hear. As Page explains, this “suggests the need for *professional communicators*, who not only help policy experts communicate with one another, but also assemble, explain, debate, and disseminate the best available information and ideas about public policy, in ways that are accessible to large audiences of ordinary citizens.”<sup>54</sup> Just as we entrust representatives to handle policymaking on our behalf, we rely on such “professional communicators”—journalists, reporters, writers, commentators, and television pundits—to identify the important issues, survey the various arguments, and present a reasoned, diverse set of opinions to the public. It is their responsibility, in other words, to provide a rich “forum for deliberation.”<sup>55</sup>

The decisions made by these intermediaries in selecting and covering particular issues can profoundly influence our individual opinions and, in turn, shape small-group deliberation.





**Figure 4-1.** *Hypothetical Mediated Deliberation Feedback Loop.* In this simplified diagram, “professional communicators” aggregate diverse information and opinions and present them to individuals via channels of mass communications. The opinions of these individuals, as well as those from other media outlets, are then fed back to the professional communicators, and the cycle repeats.

One study examined this “agenda-setting function of the media” and found that if a particular issue was covered, its salience was increased among audience members. In addition, the “findings indicate that...the media are functioning as a significant dimension of issue evaluation among audience members.” Put differently, “the media, by emphasizing certain attributes of an issue, tell us ‘how to think about’ the issue as well as ‘what to think about.’”<sup>56</sup> Similar studies have corroborated these findings,<sup>57</sup> emphatically with respect to election issues.

Given the power of the media in shaping democratic discourse, the onus of promoting fair deliberation and assuring the consideration of diverse views—so necessary for a “fully functioning democracy”—falls squarely on these “professional communicators” who, it should be noted, are *not* elected by the public. In determining whether these intermediaries are doing their job in presenting a “fair and balanced” democratic discourse, we must assess whether the

media is mapping the debate in a manner that benefits ordinary citizens. That is, we need to ask:

*Are the speakers and viewpoints diverse? Are voices heard from every corner of society? ... Does the public have any recourse, any way to break through the monolithic media consensus in order to inform and express itself? ... Do media organizations themselves play active parts in deliberation, pursuing their own policy objectives? Do some media outlets have sufficient influence over other media that they can affect the course of public deliberation for the country as a whole?*<sup>58</sup>

The answers to these questions are of great significance since, to recall Justice White's words, "the dissemination of the widest possible information from diverse and antagonistic sources is essential to the welfare of the public."

In a brilliant and alarming account of the development of modern mass media, McChesney answers these very questions.<sup>59</sup> Like Page, he asserts that "if democracy is genuinely committed to letting citizens have equal influence over political affairs, it is crucial that all citizens have access to a wide range of well-formulated political positions...this means that the media perform a crucial function."<sup>60</sup> He broadly but carefully analyzes the current state of mass communications, ultimately finding that

*The corporate media cement a system whereby the wealthy and powerful make the most important decisions with virtually no informed public participation. Crucial political issues are barely covered by the corporate media, or else are warped to fit the confines of elite debate, stripping ordinary citizens of the tools they need to be informed, active participants in a democracy.*<sup>61</sup>

Coverage of policy issues in the media, McChesney observes, has all but disappeared, replaced by more profitable entertainment and sensationalist content. On the rare occasions when mass media does devote time to "public service issues," the topics and arguments presented are hardly diverse. For example, "government activities that serve primarily the poor or middle class (e.g., welfare and public education) are often subject to very close scrutiny, whereas... [coverage of] intelligence, foreign policy, and military operations are conducted primarily to serve the needs of the elite."<sup>62</sup> He ultimately concludes that the "implications for democracy are...entirely negative."<sup>63</sup>

Page himself arrives at similar conclusions in his own investigation of media coverage during three early-nineties news stories.<sup>64</sup> In these cases, he finds, the arguments presented tended to be “related to the nature of particular audiences and to the ideological predispositions and/or economic interests of media owners and managers.” For certain publications—*The Nation*, *The Weekly Standard*, *People’s Weekly World*—such a finding would hardly be surprising.<sup>65</sup> But Page wasn’t looking at these kinds of publications. Rather, his analysis was focused on “mainstream” media sources: those with high popularity and, importantly, no *admitted* political slant. Together, these are the intermediaries whom we most rely upon to fairly cover a broad range issues and opinions (including arguments taken from smaller, even politically biased, outlets from whom we might not otherwise hear). Page finds that these supposedly neutral media organizations do not present a broad array of diverse, antagonistic, and minority opinions; often, they distort the issues and, to some degree, deceive the public. Page is ultimately more optimistic than McChesney about the public’s ability to “form sensible collective policy preferences” in spite of such widespread media bias,<sup>66</sup> but he is unequivocal in extolling the virtues of fair mediated deliberation, and emphatically supports the expression of minority opinions.

As Bagdikian, McChesney, Compaine, and a host of other media scholars have argued at length, the “abject failure” of the mass media and “general interest intermediaries” in promoting a balanced public discourse has been facilitated by the fact that alternative channels of mass communication have become exceedingly rare. Through a process of deregulation, mergers, and acquisitions, the media industry has essentially been consolidated into a handful of enormously powerful, vertically and horizontally integrated multinational corporations. This oligopoly has amassed control of most of the nation’s broadcast networks, television and radio stations, film studios, local and national newspapers, popular magazines, cable and satellite systems, music labels, and book publishers. Controlling both the content and the distribution channels, this tightly-knit group of about a half-dozen media corporations has made it almost impossible for alternative outlets and “antagonistic” views to reach a significant

audience.<sup>67</sup> As a result, citizens have few alternatives to inform and express their opinions. They must take what the oligopolistic media industry feeds them, or get nothing at all.

Making matters even worse, the scarcity of options has allowed media companies to pursue ever-greater profit margins with little fear of consumer retaliation.<sup>68</sup> The most obvious result of such “hypercommercialism” is the near-ubiquity of highly profitable advertising in today’s media.<sup>69</sup> We are increasingly barraged with ads each time we turn on the television, open a magazine, or even go to the movies. While this is certainly annoying, there is another, somewhat more insidious, consequence of the media’s penchant for advertising: the desires of sponsors frequently trump those of the public. The media industry would have you believe that they are “giving the people what they want,” but this is the case only when doing so actually brings in more advertising revenue.<sup>70</sup> “The media,” according to Bagdikian, “have become partners in achieving the social and economic goals of their patrons.”<sup>71</sup> Advertising invites a pervasive sameness in the media, characterized by “noncontroversial, light and nonpolitical” content carefully selected “in order to create ‘a buying mood.’”<sup>72</sup> And when there is conflict about controversial programming, when advertisers stand against the views expressed on the programs they sponsor, it is the advertisers who ultimately win—often at the cost of much-valued public discourse. A recent, troubling example of this phenomenon can be found in ABC’s recent cancellation of Bill Maher’s “Politically Incorrect.” According to one journalist, the relatively popular program provided “entertaining, politically unpredictable commentary and content of a sort that could rarely be found elsewhere in mass media.”<sup>73</sup> But a few days after the 9/11 attacks, during one particularly charged debate on the show, Maher called the actions of the American military “cowardly.” Amidst a storm of controversy, General Motors, Sears, FedEx, and Schering-Plough—some of the show’s biggest sponsors—pulled their ads.<sup>74</sup> The show was quickly cancelled.<sup>75</sup>

Such a blatant loss of public discourse is indicative of the essential problem with commercialized media: what is profitable for media companies may not be “what the public wants,” nor what’s beneficial for public discourse and democracy. Cooper makes this very

point in a convincing legal and economic analysis of traditional and “new” media.<sup>76</sup> He concludes that

*The central fact that all of these discussions share is that market forces provide neither adequate incentives to produce the high quality media product, nor adequate incentives to distribute sufficient amounts of diverse content necessary to meet consumer and citizen needs.<sup>77</sup>*

Cooper shows how free-market models—those favored by laissez-faire economists and, increasingly, the FCC—simply do not work when it comes to encouraging the development of a media industry whose aims are consonant with those of citizens and of democracy. For one, the economic advantages of appealing to a large audience encourage “middle-of-the road” mainstream content that fails to represent unpopular or diverse opinions. More broadly, the idea that the needs of citizens are best satisfied by the commercialized media is based on various incorrect assumptions about how media markets actually work. Cooper argues that the “consumers” of media are (economically speaking) not usually the readers, listeners, and viewers but the *advertisers*; that “deregulation” actually *decreases* competition by encouraging concentration and consolidation; and finally that, unlike the products of other industries, opinions are not interchangeable goods. The upshot of all this is that, while economists argue that the free “marketplace of ideas” produces the best products for consumers at the lowest cost, in media markets this means that it is the corporations and advertisers (not the citizens) reap the greatest rewards. Such a “market failure” is, clearly, a travesty for democratic deliberation.

When it comes to commercial media there does indeed seem to be a fundamental tension between capitalism and democracy. As McChesney puts it,

*The media system is linked ever more closely to the capitalist system, both through ownership and through its reliance upon advertising...Capitalism benefits from having a formally democratic system, but capitalism works best when elites make most fundamental decisions and the bulk of the population is depoliticized. ...If we value democracy, it is imperative that we restructure the media system so that it reconnects with the mass of citizens who in fact comprise “democracy.”<sup>78</sup>*

Arguments such as these, it should be noted, are hardly new. When broadcasting technologies first appeared, many worried about the consequences of allowing private corporations to have

significant control over public discourse.<sup>79</sup> During the early days of radio, in fact, Congress passed several laws that attempted to balance private interests with the need for diverse, democratic deliberation.<sup>80</sup> The Radio Act of 1927 firmly established the radio spectrum as a public resource, preventing private parties from claiming a property right to it. It also set up the Federal Radio Commission—what would later become the FCC—to regulate the broadcasting industry and limit the role of advertising. With the Telecommunications Act of 1934, Congress enacted strict ownership limits and required operators to carry a certain number of “signals of qualified noncommercial educational television stations.”

Perhaps most interestingly, in 1949 the FCC enacted a policy known as the “fairness doctrine.” Recognizing that the spectrum was a public, limited resource of great significance for democratic deliberation, this mandate attempted to “ensure that coverage of controversial issues by a broadcast station be balanced and fair.” Broadcasters were obligated to provide reasonable opportunity for contrasting points of view, and to cover various issues of relevance to local communities. Although the fairness doctrine withstood First Amendment muster, in the end it proved to be counterproductive: rather than presenting “balanced” arguments in compliance with FCC regulations, the media simply avoided controversial issues altogether. The mandate was, nevertheless, enforced until 1987, when Reagan-era deregulation put an end to the practice.<sup>81</sup>

In fact, virtually all of the original “public interest” provisions have been done away with, thanks in large part to incessant lobbying from media corporations. Public and noncommercial broadcast requirements have been practically eliminated, and ownership limits have been consistently relaxed. Even the FCC, which was originally entrusted to protect the airwaves, has now taken an entirely “free market” approach to regulating the media—meaning, of course, that it hardly regulates at all. Indeed, the entire conception of spectrum as “public resource” has seemingly been forgotten, with broadcasters continually claiming and extending their “rights” to spectrum use. The mass media, in other words, “metaphorically floss their teeth with

politicians' underpants.”<sup>82</sup> And, being in the unique position of covering their own regulation proceedings, they've been able to get away with it.

#### IV. Hope, Hype, and the Reality of the Internet

To summarize, the traditional media are in a pathetic state. Concentration, conglomeration, and hypercommercialism have squandered the potential of once-promising communications technologies to promote deliberation and democracy. The media have, in general, failed to put forth the diverse and underrepresented views necessary for individuals to make truly informed decisions. Broadcasters in particular have focused on content (usually of an “entertainment” variety) that stokes corporate media and advertising interests. Fewer than a dozen conglomerates have cornered the market on the mass dissemination of opinion, so that truly antagonistic, alternative voices have practically fallen into broadcasting oblivion. As Bagdikian puts it in his seminal book, *The Media Monopoly*, “the inappropriate fit between the country’s major media and the country’s political system has...eroded the central requirement of a democracy that those who are governed give not only their consent but their informed consent.”<sup>83</sup>

But a new medium has recently emerged, and it promises to change all this. Decentralized and distributed, the global Internet—and, in particular, the Web—allows anyone and everyone to make their views accessible, and to access anyone’s views. With a click of the mouse, you can read information and opinions that have not been “filtered” by profiteering corporations or corrupt governments. At the same time, underrepresented and unheard groups can cheaply bypass the “monolithic media empire” to have a voice. The Internet is many-to-many, all-to-all, and it promises to restore our faith in mediated deliberation. These aspirations are expressed repeatedly, and with understandable excitement:

*The Web...breaks the traditional publishing model...[It] says instead, “You have something to say? Say it. You want to respond to something that’s been said? Say it and link to it. You think something is interesting? Link to it from your home page. And you never have to ask anyone’s permission.”*<sup>84</sup>

*It's [going to be] the individual who replaces the big organization in many of our lives...You don't have to be writing for an organization to have a credible voice. The Net elevates those voices. What the large media were about was distribution capacity to communicate with hundreds of thousands of people. Now the Net does that.*<sup>85</sup>

*The masters of the media...have yet to grasp that the Internet can never be merely another profit center in their dreams of empire. Their power is based on monopoly, on controlling distribution. But the Net is built to smash monopolies. Instead of a gatekeeper, users get an open invitation to the electronic world and can choose whatever they want.*<sup>86</sup>

In this light, what Habermas once called the “ideal speech situation” now seems tantalizingly plausible. So exhilarating are the prospects that some have even jumped to label the Internet “inherently democratic.”<sup>87</sup>

We must, of course, be quite critical of such assertions. After all, a similar frenzy of delight greeted the arrival of television and radio technologies; but rather than being democratizing forces, these technologies were strongly manipulated by private interests, ultimately resulting a rather *undemocratic* state of affairs. The Internet seems to escape this fate (or so the argument goes) because it is fundamentally different than broadcast technologies. It is a highly decentralized, nonhierarchical medium of the likes never before seen. There is no scarcity of spectrum on this network. The transaction costs are negligible. The Internet, some say, is more like the printing press than it is like radio.<sup>88</sup> And, in fact, it is much more. Information on the network is not constrained by the limits of printed matter, by delivery distances, or by particular publication schedules. The Internet transcends the limits of time, space, and matter itself.<sup>89</sup>

That’s all well and good. But even after recognizing “fundamental” differences in the technology, and its limitless potential, we must still assess whether the Internet has, *in practice*, fulfilled this “democratic” promise. We should, in other words, ask the same sorts of questions we asked of traditional media: Are the opinions disseminated diverse? Even if anyone can have a voice, are these voices sufficiently loud? Is the medium dominated by commercial interests and advertising? Do a few media conglomerates control content and



access? If the network does, by design, “smash monopolies,” if it resists hierarchical control, if it protects freedom and equality, then the answers to these questions can only be encouraging.

And indeed, during the early 1990s the answers seemed downright positive. The Internet was then still in its formative years, and it was a largely noncommercial space for libraries, universities, and the government. It had grown out of a military project, and was at the time maintained by DARPA, the National Science Foundation, and various public and private universities scattered throughout the globe. Strict limits kept commercialization low, and advertising was nowhere to be seen.<sup>90</sup> This was because the university scientists who built up the network from its military roots “did so with the explicit intent to create an open and egalitarian communication environment.” They envisioned a “collaborative, noncommercial sharing community”—a public forum—that would eventually be open to all citizens.<sup>91</sup> The NSF played a large role in making this possible by extending and opening the network, promoting “universal educational access” by paying for universities’ connections to the network backbone only if the university “had a plan to spread the access around.”<sup>92</sup> Users of the network were allowed to do as they wished, provided it was in accord with the ideals of the space. When the World Wide Web was unveiled in 1991, “netizens” were able to easily publish readable content and readily get many others to look at it. The Internet of the early 90s allowed unrestricted noncommercial communication, and it was governed largely by committees representing the users themselves. It was, in short, a rather democratic system open to anyone with access.

But there’s the rub: very few *had* access. The government simply could not extend the network to all citizens. In order to expand the Internet to anyone who wanted it, many believed that commercialization and privatization were required. In 1993, Ed Krol of the Internet Engineering Task Force posted a document to the online community, announcing that

*policies which excluded or restricted commercial use...are under review and will change. As these restrictions drop, commercial use of the Internet will become progressively more common.*<sup>93</sup>

Saying that “right behind commercialization comes privatization,” he added that

*For years, the networking community has wanted the telephone companies and other for-profit ventures to provide "off the shelf" IP connections. Now...profit-oriented network purveyors complain that the government ought to get out of the network business...They've got the ear of a lot of political people, to whom it appears to be a reasonable thing.<sup>94</sup>*

“Most people in the networking community,” he continued, “think that privatization is a good idea” because it would allow wider access and more diverse uses of the technology.<sup>95</sup> And indeed, in a 1993 issue of *Wired*, Mitchell Kapor, co-founder of the Electronic Frontier Foundation, voiced his support.<sup>96</sup> But he also expressed some concern that if media conglomerates controlled access, “content will be supplied only by a carefully chosen set of providers[, ...] programming will still seek the least common denominator, and the population will be divided by income into information haves and have-nots.”<sup>97</sup> Some in the academic community also worried that by privatizing, smaller universities and secondary schools—dependent on the NSF for their free connections—would not be able to afford access.<sup>98</sup> In the end, it was suggested that the government should remain proactive in subsidizing access and ensuring that the network remained a free, open channel of communications.<sup>99</sup> Aside from this limited intervention, Kapor, Krol, and many others clearly believed that the Internet would be better off if the government stepped down, letting communications giants be free to expand the network.

And indeed, the government soon began retreating from its leadership role by handing over control of key Internet operations to a handful of corporations. The NSF began to sell its stake in the “Internet backbone,” the main lines of communication connecting various computers on the Net, to private investors. At the same time, commercial “Internet Service Providers” became increasingly common, allowing anyone (who could afford it) Internet access from their homes.<sup>100</sup> Amazon.com and other online retailers emerged as well, selling their wares over the network and in the process triggering an “e-commerce” revolution.

These developments were strongly encouraged by the Clinton Administration. Ira Magaziner, Clinton’s chief Internet advisor, predicted that “the digital economy will be an

environment or a world in which private actors lead” and that “the market-driven environment ought to be the one that governs this Internet economy.”<sup>101</sup> When Clinton proposed his plan for the “National Information Infrastructure,” the central proscription was complete deregulation—an entirely laissez-faire approach to the “marketplace of ideas.”<sup>102</sup>

The complete privatization of the Internet would be practically complete by 1995, when the NSF sold its last portion of the Internet backbone.<sup>103</sup> The following year, with the backing of the media conglomerates, Congress passed the Telecommunications Act of 1996, in which it “effectively washed their hands of the matter, apparently for all time.”<sup>104</sup> What is remarkable about these developments is that in less than three years, and with virtually no widespread public deliberation, the once government-funded, noncommercial Internet was completely and totally handed over to private corporations. As Schuler observes,

*The American taxpayer who paid for the initial Internet was never consulted on the possible directions it could take; all of the major decisions involving the development, deployment, or use of the Internet were made in a public-participation vacuum. In fact, it almost appears that many of these decisions were made with uncharacteristic speed so as to avoid public input, public input that might in fact raise uncomfortable questions about social uses or public ownership.*<sup>105</sup>

The choices, it should be emphasized, were not *either* complete public control *or* complete private control. While Kapor and Krol favored privatization, they also believed that the government needed to play a crucial role in the preservation of an open, democratic Internet. At the very least, it could prevent carriers from using incompatible protocols, or instill a content nondiscrimination mandate. It could even be the role of government to subsidize access or to build noncommercial “public forums” cordoned off from private hands.<sup>106</sup>

But Congress did none of this. You might expect Internet users to be up in arms, blasting the government for not stepping in to ensure that their beloved network would remain the “egalitarian sharing community” it once was. But they did not. You see, that very same Telecommunications Act of 1996 contained within it the Communications Decency Act (CDA),<sup>107</sup> an “absurd statute” that

*makes it unlawful, and punishable by a \$250,000 to say ‘shit’ online. Or, for that matter, to say any of the other 7 dirty words prohibited in broadcast*

*media. Or to discuss abortion openly. Or to talk about any bodily function in any but the most clinical terms.*<sup>108</sup>

On this matter, the online community *was* up in arms. Rather than criticizing the government for *not* regulating the Internet, they now wanted it gone, once and for all. A day after the bill was passed, John Perry Barlow, another EFF co-founder, published his widely circulated “Declaration of the Independence of Cyberspace,” in which he asked “Governments of the Industrial World” to “leave us alone”:

*You claim there are problems among us that you need to solve. You use this claim as an excuse to invade our precincts. Many of these problems don't exist. Where there are real conflicts, where there are wrongs, we will identify them and address them by our means. We are forming our own Social Contract...*<sup>109</sup>

This document articulated the emerging “cyberlibertarian” position that *government* is the greatest enemy of a democratic Internet, and that only *private industry* could create “a world that all may enter...[and that] anyone, anywhere may express his or her beliefs.” Helped in part by the CDA, such rhetoric soon permeated the Internet community and, as a result, there was now an agreement among Internet users, mainstream politicians, *and* the media conglomerates: the Internet was to be a private, completely deregulated space.<sup>110</sup> If the Internet was going to stay “democratic,” it would be up to private parties to keep it that way.

Given our previous observations about the traditional media, this might seem unlikely. For one, the incredible political and financial power of the media conglomerates, combined with their existing media properties, provides them unrivaled leverage over the Internet. The resources and sheer size these corporations bring to the table could allow them to entrench themselves in the communications medium before smaller companies have a fair chance to compete. Even if a formidable opponent does emerge, the media conglomerates could always buy them out.<sup>111</sup>

What’s more, the architecture of the Internet *itself* is threatened by completely privatized control. Lawrence Lessig argues forcefully in *Code and Other Laws of Cyberspace* that private actors can and do reshape Internet technologies in profoundly undemocratic ways. Market pressures can encourage key groups—ISPs, content providers, and portals—to control

traffic and content in a manner that violates the original, open, fair, “end-to-end” aspirations of the network.<sup>112</sup> Ownership, in other words, matters as much in cyberspace as it does in broadcast.

But have these concerns been warranted? A brief glance at the state of the Internet should be reason enough to worry:

- About 1 in 4 U.S. Internet users get their access from AOL-Time Warner, the world’s largest media company (and itself the product of a merger between the largest ISP and the largest media conglomerate).<sup>113</sup>
- Broadband access is usually limited to either DSL or cable modem access; that is, through the telephone or cable TV monopolies.<sup>114</sup>
- By 2001, over half of users’ online time was being spent at four sites. One third of the total time was spent at AOL-Time Warner properties.<sup>115</sup>
- The record and motion picture arms of the media oligopoly have systematically bought off or sued the creators and users of technologies that threaten their existing distribution channels (e.g., Napster, MP3.com, DeCSS, and even individual users).<sup>116</sup>
- Advertising has emerged as the primary source of revenue for online content providers, and its use has become increasingly intrusive (from small text ads, to banner ads, to pop-ups, and now to spam-a-minute email).<sup>117</sup>
- Many have reported that online content increasingly blurs the distinction between editorial and advertising content.<sup>118</sup>

All this, of course, should sound disturbingly familiar. A new, promising medium emerged, followed by privatization, deregulation, rampant consolidation, media outlet conglomeration, and hypercommercialism.

But not so fast. While all this may be true (and quite worrisome) the Internet has also had many tangible, readily observable, positive effects for diverse, democratic discussion. Many activists have been able to start and grow their own online, “grassroots” communities to pursue

particular policy objectives. A slew of “bloggers”—self-made ‘journalists’ who report their findings and solicit comments in a sort of “deliberative diary”—have gained a considerable following and a great deal of attention from even the mainstream media.<sup>119</sup> Real-world community projects have sprung up online, “evidence of an overdue renewal of interest in democracy.”<sup>120</sup>

Indeed, perhaps the most significant effect of privatization has been the explosive growth of access to the network, and the awesome ability for cheap, widespread publication of individual views. The increase in the number of Internet users has been nothing short of spectacular; the majority of Americans now have Internet connections in their homes. While there remains a significant global digital divide, recent evidence suggests that this gap is narrowing considerably.<sup>121</sup> Accompanying this growth in access, the availability of cheap (and even free) web hosting services has allowed anyone with basic computer competence to publish their own web site and instantly have it accessible by anyone around the world.

These facts suggest an interesting paradox: on the one hand, the decentralized, privatized Internet has allowed unprecedented levels of access and many-to-many communication; on the other hand, most users continue to spend their time at a handful of (largely commercial) Web sites. What gives? One rather convincing explanation is that the Internet has not done away with a fundamental *scarcity* problem of media. Recall that reason why we needed professional communicators was because, with broadcast technologies, there is simply a limited amount of available “channels.” By and large, the Internet seems to have been wildly successful at doing away with this limitation; there are literally billions of pages (“channels”) available on the Web. But there was another reason why we needed “professional communicators” to aggregate the many available voices: there is a limit to how many we, as individuals, can listen to. With television, radio, and the print media, we rely on the mass media to condense the available opinions and make them easily accessible through newspapers, the evening news, radio broadcasts, and so on.

And the same sort of mediation is required on the Internet. Indeed, there are only a few ways in which we can come across a particular piece of information amidst the sea of billions of pages: through word of mouth; via links from other pages; with referrals from traditional media sources; or through “portals” and “search engines.” While the first two factors are largely determined by individual users and content authors, influences from the latter are generally determined by private corporations. On the subject of media referrals, for instance, the advertising and media industries quickly learned that cross-promotion from TV, radio, and newspapers can drive enormous traffic to particular Web sites (witness, for instance, the MSNBC “synergy,” online voting during *American Idol*, Super Bowl dot-com advertisements, and online movie trailers).

But it is the “portals” and search engines that have emerged as the primary means by which Internet users are directed towards particular sources of information. These “entry points” of the Web are among the first and most frequently accessed pages for the vast majority of users: each one of the top 5 sites is indeed either a portal or search engine. By 2004, 84% of online Americans had used search engines, and a majority of these used them at least once a day.<sup>122</sup> According to one study, search engines were the most popular way to locate medical, governmental, and religious information online. Moreover, fully 79% of those seeking election information online began their journeys at portals and search engines,<sup>123</sup> and “Iraq” was the seventh most popular search term of 2003.<sup>124</sup> Given their capacity to direct most Web users towards particular content, these sites are clearly of enormously appealing for advertisers and media conglomerates alike. But, as the statistics above suggest, they also serve a crucial function as “intermediaries,” helping consumers *and* citizens navigate and make sense of the size and heterogeneity of the Web. As with all such intermediaries, we expect them to list the available information in a fair and diverse manner. That is, we expect search engines to be *democratic*.

## V. Democracy and Search Engines

Search engines have indeed become the primary intermediaries of cyberspace. Among the millions of online voices, these software algorithms determine which ones will be heard, and which ones will not. When Steven Levy said that “instead of a gatekeeper, users get an open invitation to the electronic world and can choose whatever they want,” he was being less than accurate. They *do* get a gatekeeper—the search engine—and they choose primarily among the sites it offers to them.

Consequently, we should scrutinize search engines just as we do the traditional media and the Internet. We need to assess whether these sites return content in a manner that is democratic, fair, and diverse. To paraphrase Page’s questions about traditional media, we now need to ask: *Is a diversity of viewpoints expressed? Can underrepresented voices be heard through the filter of search engines? Is the industry dominated by a few players, and do users seek and get a second opinion? Does the search engine manipulate its results in its own interests? What role does advertising play in the returned results? Do commercial interests trump those of the public?*

In recent years, Google has dominated the industry of Web search; with a 70% market share, it is practically a monopoly. Given our discussion in this chapter, the enormous power vested in this search engine is rather worrisome. It is, after all, a commercial venture, and is subject to market forces such as a reliance upon advertising. If the same story of the traditional media is repeated with Web search—consolidation, conglomeration, and hypercommercialism—the prospects for online democracy continue to fade. On the other hand, if Google does present a broad range of antagonistic, popular, and underrepresented voices, democratic principles will be furthered.

Google’s inventors recognized the potential problems of commercialization when the search engine was but a small, Stanford research project. In a brief appendix to their technical paper presenting their technology, they highlighted some of the “mixed motives” “inherent” in commercial search:



*Currently, the predominant business model for commercial search engines is advertising. The goals of the advertising business model do not always correspond to providing quality search to users. ... [Given] historical experience with other media ... we expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers ... [W]e believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm.*<sup>125</sup>

In this tiny appendix—inserted almost as an afterthought at the very end of a long, detailed technical paper—Brin and Page brilliantly summarize the potential politics of search engine design, and even suggest various negative implications of search engine commercialization. Google's creators, it appears, were very much aware of the issues we have discussed in this chapter, and they too worried about private interests conflicting with public discourse. So worried were they, in fact, that they called for Google to be a leader of transparent search technology, safely protected in the noncommercial, academic sphere.

But less than a year after they wrote those very words, Brin and Page were singing a very different tune. Google left its Stanford home and became its own, private corporation—and a notoriously secretive one at that. They sought venture capital, and developed a business model based on advertising. Given the founders' own admissions about the potential for negative search engine bias, we are invited to analyze its technology, its shift to commercialization, and its corporate policies. As the founders themselves make clear, the implications for online discourse can be great indeed.

## Notes

---

<sup>1</sup> Robert McChesney, *Rich Media, Poor Democracy: Communication Politics in Dubious Times* (Urbana, IL: University of Illinois Press, 1999), 4.

<sup>2</sup> Michael Sward, *The Terms of Democracy* (Cambridge, UK: Polity Press, 1998), 1-2.

<sup>3</sup> For more on the distortion of the term 'democracy,' see Gordon Graham, "The Moral Basis of Democracy," *International Journal of Moral and Social Studies* 7, no. 2 (1992): 91.

<sup>4</sup> Indeed, today, it seems that any use of the term 'democracy' has a positive connotation. See Quentin Skinner, "The Empirical Theorists Of Democracy and their Critics: A Plague on Both their Houses," *Political Theory* 1, no. 3 (1973).

<sup>5</sup> McChesney, *Rich Media*, 4.

<sup>6</sup> John Dewey, *The Public and Its Problems* (New York, NY: Henry Holt and Company, 1927), 82.

- 
- <sup>7</sup> Axel Hadenius, *Democracy and Development* (Cambridge, UK: Cambridge University Press, 1992), 9.
- <sup>8</sup> Saward, *The Terms of Democracy*, 51.
- <sup>9</sup> Iain McLean, *Democracy and New Technology* (Cambridge, UK: Polity Press, 1989), 34.
- <sup>10</sup> Robert Dahl, *Polyarchy: Participation and Opposition* (New Haven, CT: Yale University Press, 1972), 3.
- <sup>11</sup> Saward, *The Terms of Democracy*, 49-66.
- <sup>12</sup> James Fishkin, *The Voice of the People: Public Opinion and Democracy* (New Haven, CT: Yale University Press, 1995), 34.
- <sup>13</sup> Dahl, *Polyarchy: Participation and Opposition*, 8.
- <sup>14</sup> There are various (rare) cases in which “democracy” is implemented not by voting, but rather through procedures such as consensus. See, for example, Hiromi Yamashita and Christopher Williams, "A Vote for Consensus: Democracy and Difference in Japan," *Comparative Education* 38, no. 3 (2002).
- <sup>15</sup> Jean-Jacques Rousseau, *The Social Contract*, trans. Maurice Cranston (New York, NY: Penguin Books, 1968), 141.
- <sup>16</sup> *Ibid.*
- <sup>17</sup> Steven Piott, *Giving Voters a Voice: The Origins of the Initiative and Referendum in America* (Columbia, MO: University of Missouri Press, 2003), 2.
- <sup>18</sup> Cass Sunstein, *Republic.com* (Princeton, NJ: Princeton University Press, 2002), 39.
- <sup>19</sup> *Ibid.*, 36.
- <sup>20</sup> Alexander Hamilton, James Madison, and John Jay, *The Federalist Papers*, ed. Karen M. Elder (Richmond, VA: Westvaco, 1995).
- <sup>21</sup> *Ibid.*
- <sup>22</sup> *Whitney v. California*, 274 U.S. 357, 275 (1927).
- <sup>23</sup> Seyla Benhabib, "Toward a Deliberative Model of Democratic Legitimacy," in *Democracy and Difference: Contesting the Boundaries of the Political*, ed. Seyla Benhabib (Princeton, NJ: Princeton University Press, 1996), 69.
- <sup>24</sup> James Fishkin, *Democracy and Deliberation: New Directions for Democratic Reform* (New Haven, CT: Yale University Press, 1991), 36.
- <sup>25</sup> The topic of deliberative democracy has been covered extensively. For a rather balanced summary of the arguments, see Jon Elster, ed., *Deliberative Democracy* (Cambridge, UK: Cambridge University Press, 1998).
- <sup>26</sup> Benjamin Page, *Who Deliberates?* (Chicago, IL: University of Chicago Press, 1996), 5.
- <sup>27</sup> *Ibid.*, 2.
- <sup>28</sup> This description of “everyday deliberation” is borrowed from Sunstein, *Republic.com*.
- <sup>29</sup> For a critical overview of these proposed goals and effects of deliberation, see Frank Cunningham, *Theories of Democracy: A Critical Introduction* (New York, NY: Routledge, 2002), 165-66, 68-76.
- <sup>30</sup> John Stuart Mill, *On Liberty*, ed. Elizabeth Rapaport (Indianapolis, IN: Hackett Publishing Company, 1978), 35.
- <sup>31</sup> *Ibid.*, 19.
- <sup>32</sup> Cass Sunstein, "Deliberation, Democracy and Disagreement," in *Justice and Democracy: Cross-cultural Perspectives*, ed. Ronald Bontekoe and Marietta Stepaniants (Honolulu, HI: University of Hawai'i Press, 1997), 94.
- <sup>33</sup> Jürgen Habermas, *Moral Consciousness and Communicative Action* (Cambridge, UK: MIT Press, 1990), 89.
- <sup>34</sup> *Ibid.* Note that Habermas actually borrows these particular “general symmetry conditions” for mediation from Robert Alexy.
- <sup>35</sup> *Ibid.*, 90.
- <sup>36</sup> The history and philosophy of Speakers’ Corner is quite interesting in its own right. See John Roberts, "The Enigma of Free Speech: Speakers' Corner, The Geography of Governance and a Crisis of Rationality," *Social and Legal Studies* 9, no. 2 (2000): 272.
- <sup>37</sup> "Fur Special Issue: Fur and Loathing," *The Independent*, November 21 2002, 10.

- 
- <sup>38</sup> Douglas Robinson, "Throng to Parade To the U.N. Today for Antiwar Rally," *The New York Times*, April 15 1967.
- <sup>39</sup> Don Lattin, "Even in S.F., Religions Still Divided over Gay Marriage: Catholics, Muslims Steadfast Over Ban on Same-Sex Unions," *San Francisco Chronicle*, February 14 2004.
- <sup>40</sup> Again, this "streets and parks" model of deliberation is indebted to Sunstein, *Republic.com*.
- <sup>41</sup> *Hague, Mayor, et al. v. Committee for Industrial Organization et al.*, 307 U.S. 496, 515 (1939).
- <sup>42</sup> Sunstein, *Republic.com*, 30-34.
- <sup>43</sup> Justice White, quoted in Mark Cooper, *Media Ownership and Democracy in the Digital Information Age: Promoting Diversity with First Amendment Principles and Market Structure Analysis* (Stanford, CA: The Center for Internet and Society at Stanford Law School, 2004), 12, available from <http://cyberlaw.stanford.edu/blogs/cooper/archives/mediabooke.pdf> (accessed March 22, 2005).
- <sup>44</sup> *United States v. Kokinda*, 497 U.S. 720 (1990).
- <sup>45</sup> *International Society for Krishna Consciousness v. Lee*, 505 U.S. 672 (1992).
- <sup>46</sup> For more on this point (and in particular, its implications *vis-à-vis* the Internet), see Angioletta Sperti, *The Public Forum Doctrine and Its Possible Application to the Internet* (UCLA Online Institute for Cyberspace Law and Policy, 1997 [cited April 18 2004]), available from <http://www.gseis.ucla.edu/iclp/asperti.html>.
- <sup>47</sup> *Krishna v. Lee*, 563.
- <sup>48</sup> Hamilton, Madison, and Jay, *The Federalist Papers*.
- <sup>49</sup> Cunnigham, *Theories of Democracy: A Critical Introduction*, 165.
- <sup>50</sup> Abbie Hoffman, *The Best of Abbie Hoffman* (New York, NY: Four Walls Eight Windows, 1990), 386.
- <sup>51</sup> Page, *Who Deliberates?*, 4.
- <sup>52</sup> *Denver Area Educational Telecommunications Consortium, Inc., et al. v. Federal Communications Commission et al.*, 518 U.S. 727, 132 (1996).
- <sup>53</sup> *Ibid.*
- <sup>54</sup> Page, *Who Deliberates?*, 5.
- <sup>55</sup> Cooper, *Media Ownership and Democracy*, 21.
- <sup>56</sup> Discussed in *Ibid.*, 63.
- <sup>57</sup> *Ibid.*, 63-64.
- <sup>58</sup> Page, *Who Deliberates?*, 11.
- <sup>59</sup> McChesney, *Rich Media*.
- <sup>60</sup> *Ibid.*, 288.
- <sup>61</sup> *Ibid.*, 281.
- <sup>62</sup> *Ibid.*, 59.
- <sup>63</sup> *Ibid.*, 77.
- <sup>64</sup> Page, *Who Deliberates?*, 109.
- <sup>65</sup> Indeed, it might even be encouraged. Outlets with an admitted ideological slant can sometimes, according to Sunstein, "promote...the development of positions that would otherwise be invisible, silenced, or squelched in general debate." On the other hand, strongly ideologically biased outlets can lead to group polarization and extremism. This is partly because they generally do present "antagonistic" arguments "from those who actually believe them." See Sunstein, *Republic.com*, 76.
- <sup>66</sup> Page, *Who Deliberates?*, 123.
- <sup>67</sup> McChesney, *Rich Media*, 16-77.
- <sup>68</sup> *Ibid.*, 34-35.
- <sup>69</sup> See, for example, Ben Bagdikian, *The Media Monopoly*, Fourth ed. (Boston, MA: Beacon Press, 1992), 176.. Interestingly, Bagdikian argues that at the same time that concentration was allowing for increased use of advertising, advertising was allowing for increased concentration: "Ads swelled the size of the paper each day, requiring larger plants...with the result that it was no longer easy for newcomers to enter the newspaper business." Commercialization and concentration are, sadly enough, mutually reinforcing.

- 
- <sup>70</sup> McChesney, *Rich Media*, 39-42.
- <sup>71</sup> Bagdikian, *The Media Monopoly*, 151.
- <sup>72</sup> Ibid.
- <sup>73</sup> Matthew Nisbet, *Economically Incorrect: The Real Reason Bill Maher Got Canned* (The American Prospect Online, 2002 [cited April 20 2004]), available from <http://www.prospect.org/webfeatures/2002/06/nisbet-m-06-03.html>.
- <sup>74</sup> Bill Carter, "Sponsors Defect," *The New York Times*, September 26 2001.
- <sup>75</sup> Peter Marks, "An Iconoclast's Last Days on His Late-Night Soapbox," *The New York Times*, June 23 2002. It should be noted that some have argued that ABC cancelled "Politically Incorrect" not because of the 9/11 flap—which did cost the network over \$10 million—but rather because it already planned to replace his program with a less 'political' one that appealed to a more marketable demographic. Either way, it was a quest for advertising revenue, rather than for greater public discussion, that ended his show. See Nisbet, *Economically Incorrect*.
- <sup>76</sup> Cooper, *Media Ownership and Democracy*.
- <sup>77</sup> Ibid.
- <sup>78</sup> McChesney, *Rich Media*.
- <sup>79</sup> Mark Goodman faithfully recounts the Congressional debates surrounding the allocation of the radio spectrum and its implications for democratic discourse. See Mark Goodman, "The Radio Act of 1927 as a Product of Progressivism," *Media History Monographs* 2, no. 2 (1927), available from <http://www.scripps.ohiou.edu/mediahistory/mhmjour2-2.htm> (accessed August 5, 2004).
- <sup>80</sup> It is worth noting that this was hardly the first time that Congress had imposed limits on *laissez-faire* capitalism. Consider, of course, the Sherman Antitrust Act of 1890 (designed to restrict monopoly power to prevent the destruction of "free capitalism and restraint of trade").
- <sup>81</sup> Patricia Aufderheide, "After the Fairness Doctrine: Controversial Broadcast Programming and the Public Interest," *Journal of Communication* 40, no. 3 (1990).
- <sup>82</sup> McChesney, *Rich Media*, 76.
- <sup>83</sup> Bagdikian, *The Media Monopoly*, 192.
- <sup>84</sup> David Weinberger, *Small Pieces Loosely Joined: A Unified Theory of the Web* (Cambridge, MA: Perseus Publishing, 2002), viii-ix.
- <sup>85</sup> J.D. Lasica, *Interview with John Perry Barlow* (jdlasica.com, 1996 [cited April 25 2004]), available from <http://www.jdlasica.com/interviews/barlow.html>.
- <sup>86</sup> Steven Levy, "How the Propeller Heads Stole the Electronic Future," *The New York Times Magazine*, September 24 1995, 59.
- <sup>87</sup> George Gilder, quoted in Douglas Schuler, "Reports of the Close Relationship between Democracy and the Internet May Have Been Exaggerated," in *Democracy and New Media*, ed. Henry Jenkins and David Thorburn (Cambridge, MA: MIT Press, 2003).
- <sup>88</sup> Lawrence Lessig, *Code and Other Laws of Cyberspace* (Cambridge, MA: Basic Books, 2000).
- <sup>89</sup> This "transcendence" property is discussed eloquently throughout Weinberger, *Small Pieces Loosely Joined*.
- <sup>90</sup> Jesper Laursen, *The Internet: Past, Present and Future* (October 7 1997 [cited April 27 2004]), available from <http://www.vissing.dk/inthist.html>.
- <sup>91</sup> McChesney, *Rich Media*.
- <sup>92</sup> Ed Krol, *RFC 1462 - FYI on "What is the Internet"* (May 2003 1993 [cited April 27 2004]), available from <http://www.faqs.org/rfcs/rfc1462.html>.
- <sup>93</sup> Ibid.
- <sup>94</sup> Ibid.
- <sup>95</sup> Ibid.
- <sup>96</sup> Mitchell Kapor, "Where Is the Digital Highway Really Heading? The Case for a Jeffersonian Information Policy," *Wired*, Jul/Aug 1993, available from [http://www.wired.com/wired/archive/1.03/kapor.on.nii\\_pr.html](http://www.wired.com/wired/archive/1.03/kapor.on.nii_pr.html) (accessed September 19, 2003).
- <sup>97</sup> Ibid.

- 
- <sup>98</sup> Krol, *RFC 1462*.
- <sup>99</sup> *Ibid.*
- <sup>100</sup> Laursen, *The Internet: Past, Present and Future*.
- <sup>101</sup> Ira Magaziner, "Democracy and Cyberspace: First Principles," in *Democracy and New Media*, ed. Henry Jenkins and David Thorburn (Cambridge, MA: MIT Press, 2003).
- <sup>102</sup> McChesney, *Rich Media*.
- <sup>103</sup> Laursen, *The Internet: Past, Present and Future*.
- <sup>104</sup> McChesney, *Rich Media*.
- <sup>105</sup> Schuler, "Reports of the Close Relationship between Democracy and the Internet May Have Been Exaggerated," 73.
- <sup>106</sup> These policy suggestions are derived from Lessig, *Code and Other Laws of Cyberspace*.
- <sup>107</sup> *Communications Decency Act*, S.652, Second Session of the One Hundred Fourth Congress (February 1, 1996).
- <sup>108</sup> John Perry Barlow, *A Declaration of the Independence of Cyberspace* (February 9 1996 [cited April 27 2004]), available from [http://www.eff.org/Misc/Publications/John\\_Perry\\_Barlow/barlow\\_0296.declaration](http://www.eff.org/Misc/Publications/John_Perry_Barlow/barlow_0296.declaration).
- <sup>109</sup> *Ibid.*
- <sup>110</sup> Such an amazing turn of events has lead some to argue that the CDA was a "red herring": a blatantly unconstitutional component of the bill designed to divert attention from other problematic components of the law. And it seemed to have worked—the cyberporn and censorship issues siphoned off the most attention, and in 1997 the Supreme Court struck down the CDA (to no one's surprise). But by then the cyberlibertarian principles had become fully entrenched in the Internet consciousness and in the news media. See McChesney, *Rich Media*, 313-14.
- <sup>111</sup> *Ibid.*, 192-93.
- <sup>112</sup> The cable and telephone conglomerates who control most Internet connections, for instance, can design their networks to prefer their own content over those of competitors.
- <sup>113</sup> Alex Goldman, *Top 22 U.S. ISPs by Subscriber: Q4 2004* (Jupiter Communications for ISP Planet/Internet.com, March 24 2005 [cited May 22 2005]), available from [http://www.isp-planet.com/research/rankings/2004/usa\\_history\\_q42004.html](http://www.isp-planet.com/research/rankings/2004/usa_history_q42004.html).
- <sup>114</sup> Cooper, *Media Ownership and Democracy*, 169.
- <sup>115</sup> *Study: Four Sites Account for Half of Web Surfing* (CNN.com, June 5 2001 [cited April 27 2004]), available from <http://www.cnn.com/2001/TECH/internet/06/05/internet.consolidation/>.
- <sup>116</sup> This alarming trend is discussed at length in Lawrence Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* (New York, NY: Penguin Books, 2004).
- <sup>117</sup> McChesney, *Rich Media*, 172.
- <sup>118</sup> *Ibid.*, 173.
- <sup>119</sup> For a discussion of blogs from bloggers' own perspective, see John Rodzvilla, ed., *We've got Blog: How Weblogs are Changing our Culture* (Cambridge, MA: Perseus Publishing, 2002).
- <sup>120</sup> Schuler, "Reports of the Close Relationship between Democracy and the Internet May Have Been Exaggerated," 73.
- <sup>121</sup> Consider, for example, a recent World Bank study, which—although confirming that a global digital divide clearly exists—finds that that the gap is 'narrowing considerably.' See Charles Kenny, *Financing Information and Communication Infrastructure Needs in the Developing World: Public and Private Roles* (Global Information and Communication Technologies, World Bank, February 2005 [cited May 20 2005]), available from <http://lnweb18.worldbank.org/ict/resources.nsf/InfoResources/04C3CE1B933921A585256FB60051B8F5>.
- <sup>122</sup> Deborah Fallows and Lee Rainie, *The Popularity and Importance of Search Engines* (Pew Internet and American Life Project, August 2004 [cited May 10 2005]), 2, available from [http://www.pewinternet.org/pdfs/PIP\\_Data\\_Memo\\_Searchengines.pdf](http://www.pewinternet.org/pdfs/PIP_Data_Memo_Searchengines.pdf).

- 
- <sup>123</sup> Michael Cornfield and Lee Rainie, *Untuned Keyboards: Online Campaigners, Citizens, and Portals in the 2002 Elections* (Pew Internet and American Life Project, March 21 2004 [cited May 11 2005]), available from [http://www.pewinternet.org/pdfs/PIP\\_IPDI\\_Politics\\_Report.pdf](http://www.pewinternet.org/pdfs/PIP_IPDI_Politics_Report.pdf).
- <sup>124</sup> Deborah Fallows, *Search Engine Users: Internet Searchers are Confident, Satisfied, and Trusting--But They Are Also Unaware and Naive* (Pew Internet and American Life Project, January 23 2005 [cited May 22 2005]), 3, available from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf).
- <sup>125</sup> Sergei Brin and Larry Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems* (1998): 18-19, available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 10, 2005).

# 4

## Making Sense of Search

A Technical Overview of the Web, Search Engines, and Google

In order to examine the sociopolitical dimensions of the Google search engine, we must first understand how it works. As it turns out, this is no small feat; it requires a rather deep technical understanding of the Web, of search engine design, and of Google's unique search technology. In this chapter, we present a technical overview of these concepts, written with the nontechnical reader in mind. Although our discussion will be far from comprehensive—indeed, many engineering dissertations have been devoted to explaining just these matters—we wish to develop a *basic* understanding of the relevant technologies, and to encourage an appreciation for the many nontechnical challenges involved.

We aim, importantly, to not only describe *how* the relevant technologies work, but also to justify *why* they were designed in a particular fashion. This is critical because, by their very nature, these technologies are built on certain deeply subjective assumptions about information, semantics, and relevancy. In the end, we hope that the reader has gained not only a comprehension of the various technical solutions, but also an appreciation for the underlying—and far less technical—problems addressed by these technologies.

### I. What Is the Web?

We begin by examining the domain search engines operate on, the space they inhabit: the World Wide Web. According to its creator, Tim Berners-Lee:

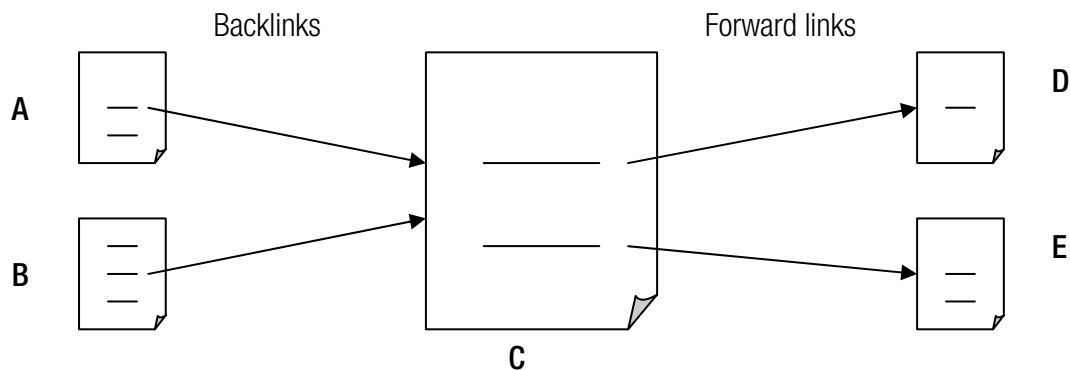
*The fundamental principle behind the Web was that once someone somewhere made available a document, database, graphic, sound, video, or screen ... it should be accessible ... by anyone, with any type of*

*computer, in any country. And it should be possible to make a reference—a link—to that thing, so others could find it.*<sup>1</sup>

In essence, the Web is a set of documents exchanged over the Internet. These documents are made portable—accessible “by anyone, with any type of computer”—because they are encoded in a particular language (HTML) and transferred via a special protocol (HTTP) invented by Berners-Lee for this purpose.

But what really makes the Web special—indeed, what gives it its name—is that last part: its interconnectedness. The Web is not just a set of any documents; it is a collection of linked *hypertext* (or, more generally, *hypermedia*) documents. Unlike books, magazines, and this thesis, documents on the Web can contain certain words that, when selected, instantly transport the user to another related document. These *hyperlinks*—or “links” for short—enable and enhance the ability to navigate the body of available information. Links may be listed like a table of contents, or they may be embedded in the text itself, allowing interested readers to “digress” to other information as they read.

Importantly, these links are unidirectional. My page, for instance, contains a link to the Stanford home page, allowing you to jump to the university’s site from my own. But you can’t “follow” this same link in the reverse direction. Indeed, unless Stanford links to my site—which, unfortunately, they do not—you simply can’t get to my page from theirs. This property of



**Figure 4-1.** *Link relationship between pages.* Page C has *forward links* to D and E; it has *backlinks* from A and B. Note that the distinction is relative: the link between C and E is a forward link relative to C, but a backlink relative to E.



hyperlinks turns out to be rather important, so we will introduce a couple of terms to distinguish between “incoming” and “outgoing” links. As illustrated in Figure 3-1, above, a *backlink* of page C is a link *to* C from some other page (i.e., an incoming link). In contrast, a *forward link* of C is a link *from* C to some other page (i.e., an outgoing link).

At this point, we have provided all the technical background about the Web necessary for the purposes of this thesis. We assume, of course, that the reader is at least a novice web user; if not, other works may provide a more adequate introduction. Note that since search engines primarily operate only with the text of Web documents, we need not concern ourselves with the semantics of “graphic, sound, and video” content on the Web. Instead, a basic, textual comprehension of the medium will suffice.

## II. An Anatomy of Search Engines

In general, a search engine such as Google allows Internet users to do keyword searches for information on the Internet. Users enter their search terms, which comprise a “query,” and the engine returns a list of links usually sorted by “relevance.” Ideally, a link leading to the desired information appears near the top of the search results. To use a popular metaphor, a search engine is like a catalog to the vast library of information available on the World Wide Web. Unlike a library catalog, however, search engines must overcome unique hurdles to be effective. For one, there is no central list of all the web pages on the Internet; a search engine must discover these pages itself. Even after doing so, it needs to somehow make sense of the text contained in these web pages. Whereas books are given summaries and assigned categories by publishers and librarians, the enormous and ever-increasing size of the Internet—combined with the lack of formal publishing mechanisms—makes such manual categorization practically impossible. Thus, search engines must automatically decipher what each page is “about,” and compare this to what it “thinks” a user is asking for. All this presents enormous economic, technological, and linguistic challenges that push the limits of what computers and software can do.

Luckily, engineers have developed a slew of techniques to meet this challenge. To build a searchable “catalog” of web pages, search engines typically use a technique known as “crawling.” Software programs called “spiders” “crawl” the “web” and, like space probes exploring unknown frontiers, gradually collect information about all the pages they come across. A spider begins its crawl on some known Web page, say *www.yahoo.com*. It extracts the text from this web page, processes it, and adds it to the search engine’s index. The spider then follows each of the links contained in that web page—for example, the link to *help.yahoo.com*—and similarly saves, processes, and adds these to the index. By recursively and systematically traversing the link structure of the Web in this fashion, spiders are able to find and index millions of web pages.

While most major search engines employ roughly the same crawling methods, there is wide variance in the way different search engines come up with results for a given query. In general, however, the process of generating search results has two phases: first, the subset of relevant pages is extracted from the search index; second, these documents are ordered by specialized ranking algorithms. The first step of reducing the index to pages that *may* be of relevance is fairly straightforward: since these pages presumably contain the user’s search terms, it is often enough to simply check the index for all pages that contain the given keywords. But the number of matches this generates, as any Internet user can testify, is quite large. Google, for instance, returns in excess of 522,000 results for a search on “Prime Minister Tony Blair.” This unwieldy number of matches may reflect the sheer volume of information on the Web, or it may stem from a design decision to err on the side of too many results rather than omitting a desired link. In any case, users are routinely inundated with so many (mostly irrelevant) results that manually weeding through them all is plainly impractical. Thus, it is essential that search engines order their results according to “relevance,” minimizing the work left to the user and maximizing the usefulness of the search tool.

As it turns out, this secondary task of ordering search results has proven to be the most difficult and important challenge in search engine design. Quite obviously, the notion of

“relevance” is profoundly subjective, influenced by myriad contextual, personal, and social factors. What one user seeks when searching for “orchids” may differ greatly from what another desires; the former may be seeking factual information about the flowers, while the latter may want to purchase them online. This difficulty is exacerbated by the inherent ambiguity of language, which is especially present in the short, underspecified queries typically entered by users. The task of ascertaining the relevance of different web pages, therefore, is a difficult problem even for humans. Programming computers to do this consistently and accurately, it would seem, is an almost insurmountable challenge.

Engineers have, however, developed a wide variety of complex and innovative rankings criteria that manage to order search results fairly well. By examining the relationships between queries and desired matches, designers have observed certain patterns and discerned particular properties common to many “relevant” matches. Once these generalized properties are discovered, they may be automatically searched for among the generated results and used to order those matches. Such “rules of thumb” are called *heuristics*, and are judged by how well they approximate the desired, or optimal, output. Keyword counting, for instance, is the most basic heuristic and is employed to some degree by many search engines. Using this simple approach, the more a document mentions the given keywords, the greater its ranking among the results. The underlying assumption here is that if a page contains many instances of the search terms, it probably deals with that topic rather extensively and is thus more useful to the user. As an example, consider our search for “Prime Minister Tony Blair.” Among the multitude of matches for this query we might find, say, the official 10 Downing Street home page, a *Washington Times* article on President Bush’s recent vacation, and an archive of political speeches. If a keyword counting technique is used to order these results, the 10 Downing Street page (which contains the words “Prime Minister” 11 times) might appear as the first result—certainly before the *Times* article which contains the phrase only once. That the prime minister’s official page is deemed more relevant agrees with most people’s

expectations; in this case, simple keyword counting is an accurate heuristic for the more subjective ideal of “relevance.”

In practice, however, keyword counting performs rather poorly. Quite often, a page that mentions certain keywords many times is, nevertheless, not the most appropriate page for those keywords. The “Terms and Conditions” page of the Southwest Airlines web site, for example, contains 65 instances of the word ‘Southwest’; the home page contains only 14. But for the vast majority of users seeking the airline’s web site—to check flight times or to buy a ticket—the home page is far more useful. To make matters worse, not only is the keyword counting heuristic often a bad indicator of relevance, it is also vulnerable to “spamming” abuses. Web designers aware of keyword counting algorithms may—and do—hide hundreds of keywords in their pages. These terms may be invisible to web users, but not to search engines. By embedding a multiplicity of popular keywords, designers can “trick” search engines into directing traffic to their otherwise irrelevant web sites. Widespread use of this “hack” diminishes the quality of search results, much to the frustration of search engine users confronted with hundreds of irrelevant and sometimes offensive sites for even the most straightforward, innocuous query.

More complex heuristics serve to increase the quality of search results and, in some cases, minimize potential abuses. One approach is to exploit *meta-text*—that is, to leverage information *about* a page’s content, rather than looking at the content itself. Page titles, for instance, often contain useful keywords that describe the entire page; hidden “description tags” can be included by designers to tell search engines what a page is about; anchor text—the words used by others when linking to the document in question—often describes a page’s content more succinctly and accurately than the page itself. Taken together, the information culled from these sources can be matched against the user’s query to more successfully decide which pages are most appropriate. Popular search engines, not surprisingly, utilize various meta-text features in their ranking algorithms. These heuristics do not replace content-based ones; rather, they complement and enhance them. Ranking systems rate each page on a range

of content and meta-text features (e.g., the number of times the keywords appear in the document, whether these terms appear in the title, etc.). These ratings are weighted and combined to produce a final, cumulative relevancy “score” for each page. The search results are then returned to the user in descending order.

While ranking systems generally work in this fashion, the implementation details of individual search engines remain largely proprietary and concealed. Heuristic models and relevancy computation algorithms are widely protected under patent law or as trade secrets. The reasons for this are two-fold. The search engine industry is a fiercely competitive one, with each company vying to increase the accuracy, usefulness, and popularity of their engine over others. These companies have learned that their traffic—and thus, their advertising revenue—is largely determined by the quality of their search results. A revolutionary ranking algorithm, if kept proprietary and obscured, can give a search site the dramatic competitive edge it needs to be profitable. In addition, by not publishing the details of their ranking algorithms, search engine companies hope to combat the detrimental effects of spamming. As noted earlier, spammers exploit detailed knowledge of a search engine’s implementation to “trick” the site into listing irrelevant and unwanted sites among its search results. Obscuring these algorithmic details theoretically reduces the risk of abuse, but reverse engineering techniques and the determination of some web publishers ensure that spamming remains a continued threat.

### **III. A Closer Look: The Google Search Engine**

In 1998, Sergey Brin and Larry Page, two computer science graduate students at Stanford University, presented their paper, “The Anatomy of a Large-Scale Hypertextual Web Search Engine” at the *International World Wide Web Conference*.<sup>2</sup> In this article, they propose a system to more effectively retrieve information from the World Wide Web. Brin and Page identify two primary design goals for their new search engine: scalability and relevancy. These, they argue, were the largest problems facing search engines of the day. By addressing

these concerns, they hoped to “improve the quality of search engines” and thus “bring order to the Web.”<sup>3</sup> Alluding to an immensely large mathematical unit, the googol, they named their prototype “Google.”

To better understand Brin and Page’s motivations in designing their search engine, we can briefly trace the state of search engines during the formative years of the Web. In the beginning—circa 1994—the World Wide Web was a very small place. Containing less than a million web pages, the Web was often navigable with little aid from search engines; users often followed links, turned to directory services such as *Yahoo!*, or simply found sites by word-of-mouth. Soon, however, the exponential growth of pages on the Web—over 100 million by the time Page and Brin published their paper—outpaced users’ abilities to locate information on the network. The problem facing a growing Internet audience shifted from whether the information they desired was “out there” to whether they could *find* it. These users, not surprisingly, increasingly turned to search engines, which promised to locate pages containing the information they desired. By the end of the decade, popular search engines were handling millions of queries a day from more users, more often.<sup>4</sup>

But just as users were struggling to keep up with the explosive growth of the Web, so too were search engines under pressure to meet demand. This challenge was two-fold: not only did the search engines have account for an ever-increasing number of documents, but they also needed to respond quickly and meaningfully to more and more queries. The search engines, to use Brin and Page’s terminology, needed to be *scalable*. This required locating and storing unprecedented volumes of information, and continuing to do so, with virtually no upper limit on the amount of data to be collected. Moreover, the contents of an engine’s vast repository needed to be accessed quickly and efficiently, even as the size of the repository was multiplied many times over. To tackle the scalability problem, Brin and Page emphasized speed, expandability, and efficiency. Their proposal leveraged techniques such as compression and optimized versions of data structures such as hit lists, lexicons, forward indices, and reverse indices to store and organize the data collected by their crawlers. They also designed

their system architecture in a distributed manner—using many, many small servers instead of a few large ones—allowing additional servers and hard disks to be easily “plugged in” as needed. This minimized cost and maximized resource utilization, while ensuring that their search engine could grow along with the size of the Web and of its audience.

Conquering the scalability problem, quite clearly, is essential to the continued success of any large-scale search engine. However, a scalable architecture is not enough; even if a search engine could index for the vast majority of pages and respond to queries quickly, it is of little use unless it provides *relevant* results. This reality was a central consideration in the design of Google, as Brin and Page articulate:

*In 1994, some people believed that a complete search index would make it possible to find anything easily. [...] However, the Web of 1997 is quite different. Anyone who has used a search engine recently, can readily testify that the completeness of the index is not the only factor in the quality of search results. “Junk results” often wash out any results that a user is interested in. [...] One of the main causes of this problem is that the number of documents in the indices has been increasing by many orders of magnitude, but the user’s ability to look at documents has not. People are still only willing to look at the first few tens of results.<sup>5</sup>*

Their means of achieving relevancy is, arguably, the most unique and interesting part of the Google project. While scalability is of great importance in the search engine space and in many other fields, it is usually a well-defined, tractable engineering challenge. In contrast, relevancy is what truly distinguishes one search engine from another in the eyes of users and is far more challenging. As Nissenbaum and Introna point out:

*Relevancy ranking is an enormously difficult task [...] Some researchers working on search technologies argue that relevancy ranking is currently the greater challenge facing search engines and that developments in technical know-how and sheer capacity to find and index sites has not nearly been matched by the technical capacity to resolve relevancy ranking.<sup>6</sup>*

It is the inherent subjectivity of relevancy that makes designing ranking algorithms trickier, requiring more than the standard, quantitative, performance-based approach to engineering.

Google’s unique ranking formula deals with the relevancy problem by heavily exploiting sources of meta-information. For instance, in a paper co-written by Larry Page, the

authors describe a heuristic called the “location metric,” which looks at the URL of a page in an attempt to guess how useful the page might be usefulness.<sup>7</sup> For example, URLs that end with “.com,” contain the string “home,” or have few slashes may be favored. Similarly, if the URL contains the user’s keywords (e.g., “Microsoft”) the page might be given a higher ranking. The basic idea here is that pages with these URLs tend to be more important and valuable to Web users, and are therefore more relevant. It is unclear how greatly the URLs influence Google’s search results, but the example does illustrate how information other than the page’s content is put to use.

Another, potentially more significant technique used by Google is its leveraging of *anchor text* information. As mentioned earlier, anchor text consists of the words used by others when linking to the page in question. The often short and underspecified link text often describes a page in much the same terms as search engine query, making it a good metric for relevancy. Consider, for instance, the “1040-EZ” link on the IRS web site or the “Miami Dolphins” link on a football site. Importantly, anchor text also has the benefit of being less vulnerable than other heuristics to “spamming” abuses. Rather than trusting the possibly false or biased keywords included by authors to describe their own pages, Google instead focuses on how *others* describe those pages to their readers. It is worth noting, however, that the anchor text heuristic is not immune to manipulation (we will return to this in Chapter 6).

Since other search engines already use location and anchor text heuristics to varying degrees, what truly sets Google apart from other search engines is its patented link analysis heuristic. This algorithmic crown jewel, simply called “PageRank,” is responsible for ascertaining the relative “popularity” or “importance” of a particular page by looking at the pages that link *to* it. This information is then used, in combination with other heuristics, to order its search results. Given the Google’s claims about the importance of PageRank in generating its search results, we will describe its workings in full in the next chapter, where we begin to surmise the sociopolitical implications of and motivations for its use.



## Notes

---

<sup>1</sup> Tim Berners-Lee, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor* (San Francisco, CA: Harper San Francisco, 1999), 37.

<sup>2</sup> Sergey Brin and Larry Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine" (paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998), available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 1, 2005).

<sup>3</sup> *Ibid.*, 3-4.

<sup>4</sup> This growth of the Web is described in Albert-László Barabási, *Linked: The New Science of Networks* (Cambridge, MA: Perseus Pub., 2002).

<sup>5</sup> Brin and Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", 2.

<sup>6</sup> Lucas Introna and Helen Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society* 16, no. 3 (2000): 178.

<sup>7</sup> Larry Page et al., "The PageRank Citation Ranking: Bringing Order to the Web," (Stanford, CA: Stanford Digital Library Technologies Project, 1999), 4, available from <http://dbpubs.stanford.edu/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=&name=1999-66.pdf>.

# 5

## The Politics of PageRank

Do the Rich get Richer?

The PageRank heuristic is, according to the Google website, the “heart of our software,” providing the “basis for all our web search tools.”<sup>1</sup> While the importance of the algorithm is stated in no uncertain terms, its workings are not articulated in any great detail. Luckily, the initial Google search engine, including the PageRank algorithm, was developed as a Stanford research project. Published technical papers, combined with the publicly available PageRank patent, provide us with the necessary information for a thorough understanding of this important algorithm.

In this chapter, we will attempt to uncover the technology and politics of Google’s PageRank algorithm. We will first delve into the mathematical principles behind the technology and get a feel for the motivations behind its particular design. From here we look at what PageRank actually ends up approximating in practice, and we will critically analyze whether it is really as ‘democratic’ as is commonly asserted. Next, we will place the technology in a broader context, examining its relationship with the structure of the Web. To explicate the connection between PageRank and Google, we will also describe our own original research quantifying PageRank’s role in shaping Google’s results. It is hoped that, in the end, we will have surmised the implications the algorithm for the dissemination of rich, diverse sources of information on the Web.

### I. The Mathematics of PageRank

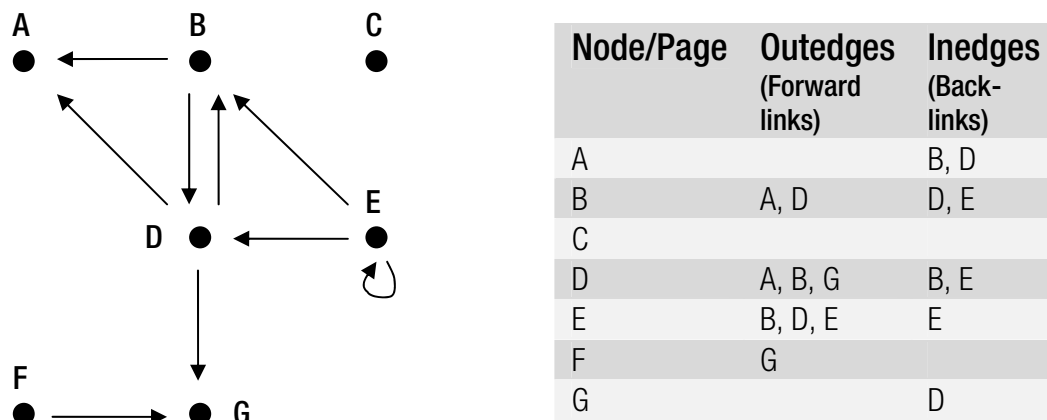
According to Brin and Page, although “the importance of a Web page is an inherently subjective matter ... there is still much that can be said objectively about the relative importance of Web pages.”<sup>2</sup> They go on to propose PageRank as a “method for rating Web

pages objectively and mechanically, effectively measuring human interest and attention devoted to them.”<sup>3</sup> This algorithm examines the link structure of the Web to determine the “importance” of each document relative to the millions of others available online. The search engine then leverages this information in an attempt more accurately ascertain “relevance” on behalf of its users.

In order to understand how PageRank works, we must first understand what is meant here by the “link structure of the Web.” Recall from the previous chapter that the Web is essentially a set of hypertext documents, each of which contains a number of unidirectional “links” to other documents. In this light, we can view the Web as a *directed graph* (Figure 5-1),<sup>4</sup> wherein

- a node represents a specific page
- an edge from node A to node B represents a link from page A to page B

By 1999, a complete graph of the Web already contained several hundred million nodes (pages) and several billion edges (links).<sup>5</sup> This graph describes how each page is interlinked and interrelated; it encodes the *link structure* of the Web. It provides a detailed account of the



**Figure 5-1.** *Hypothetical Web/hypertext graph.* Seven pages contain links as specified by the table to the right. This link structure is encoded in the directed graph to the left. A complete graph of the World-Wide Web would contain over 500 billion nodes.

many complex relationships between the billions of distributed, heterogeneous documents available online.

To uncover as much of the graph as possible, “crawlers” are used to traverse and record billions of Web pages and their links (see Chapter 4). Various ranking heuristics—which vary in complexity from simple link counting procedures to complex clustering algorithms—then exploit this information to improve the quality of search results. A straightforward example is *backlink counting*, used by the Lycos search engine, in which the search engine simply calculates the number of pages that link to a particular page  $A$  (or, to use graph theorists’ terminology, the number of inedges to node  $A$ ).<sup>6</sup> If this count is high, many pages refer to  $A$ , so we might assume that it is a “trusted” or “important” source of information. Consequently,  $A$  might be placed higher among the search results.

Google’s patented<sup>7</sup> PageRank algorithm takes the backlink heuristic one step further by recognizing that not all links are equal. With simple backlink counting, a link from my page to my professor’s site contributes as much to her ranking as a similar link from the *New York Times* website. But the more prestigious link from the *Times* site is (supposedly) a better indicator of how “important” her page is. PageRank attempts to account for this asymmetry by putting greater weight on backlinks from “important” pages. Since the *New York Times* site is deemed more “important” than my own, the former link goes much further in elevating the PageRank of my professor’s site, and thus its visibility among the results.

But if the PageRank of any given page is dependent on the “importance” of all the pages that refer to it, how do we ascertain the “importance” of all these referring pages? Isn’t

$$PR(u) = c + (1 - c) \left[ \sum_{i \in B_u} \frac{PR(i)}{|F_i|} \right]$$

**Figure 5-2. Mathematical Formula for PageRank.**  $B_u$  is the set of pages that link to  $u$  (backlinks),  $F_i$  is the set of pages  $i$  links to (forward links), and  $c$  is an arbitrary constant term. (This is actually a “slightly simplified” version of the formula, since it doesn’t handle cycles in the graph. For our purposes, however, it more than suffices.)

finding “relative importance” the very problem we set out to solve? While it may appear that we are back to square one, the solution to this question turns out to be surprisingly simple: the “importance” of each referring page is approximated by the PageRank of *that* page. The computation of PageRank is, in other words, recursive. For a particular page *u*, we take each page *A* that links to *u*, calculate the PageRank of *A*, and normalize this value by the total number of links on *A*.<sup>a</sup> Doing this for each page that refers to *u*, summing the results, and adding in some constant factors, we arrive at the PageRank of *u* (Figure 5-2).

Brin and Page indicate that their inspiration for the PageRank formula came from previous work in academic citation analysis. This provides a useful analogy for understanding the various linking heuristics: think of pages as ‘academic works,’ and links as ‘citations,’ such that ‘page A links to page B’ becomes ‘work A cites work B.’ Accordingly, backlink counting heuristics—which assume that “a page that is linked to by many [other] pages is more important than one that is seldom referenced”<sup>8</sup>—would give a higher ranking to “canonical” works, those which are heavily cited. PageRank works similarly, but it gives more weight to those citations from works which are themselves heavily referenced: “a particular paper,” they write, “is even more important if referred to by others whom are already seen as important—by other canons.”<sup>9</sup>

There is another “intuitive justification” for PageRank also worth mentioning here.

Brin and Page articulate it succinctly:

*PageRank can be thought of as a model of user behavior. We assume there is a “random surfer” who is given a web page at random and keeps clicking on links, never hitting “back” but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank.*<sup>10</sup>

Just as we have a greater chance of coming across canonical works when browsing an academic bibliography, we tend to stumble upon extensively linked pages when we go online. And a document so extensively backlinked will, by definition, have a high PageRank. It follows

---

<sup>a</sup> The reason why we “normalize” (i.e., divide) the PageRank by the number of links is simple. If a page has many links, as opposed to a few, the fact that it links to any particular page is—statistically and intuitively speaking—less “significant.” If an “important” page links to very *few* pages, those pages are probably rather important themselves, so it makes sense to give each of these links greater weight.

that we can think of PageRank as an indicator of how likely a “random surfer” is to come across to a particular page; indeed, as Brin and Page assert, the probability of him doing so is expressed precisely by the equation in Figure 5-2 (see *Appendix I* for a formal proof). PageRank assumes that documents which we are likely to visit anyhow are “important” sources of information.

These descriptions, which help us to understand how PageRank approximates “importance,” do seem to make some mathematical and intuitive sense. But what remains unclear is how this metric can be put to use by search engines. The utility of search engines lies precisely in their ability to locate documents that are *relevant to a user’s query*, yet this driving goal seems completely lost on PageRank. Nowhere does the algorithm take into account the user’s search terms,<sup>b</sup> and the ideal of “relevancy” is conspicuously absent from the mathematical and intuitive descriptions given above. Even PageRank’s metaphorical “random surfer” aimlessly jumps from link to link without any real purpose or direction. It is hard to imagine how, exactly, such meanderings “model...user behavior,” since we would hardly be productive if we ever went about finding specific information in such a haphazard fashion.

And indeed, PageRank does not *directly* ascertain relevancy at all. Nor is it meant to. Instead, the algorithm comes into play only *after* a set of relevant documents has already been identified. Given the enormous size of the Web, it is often the case that there are far too many relevant matches for the user to sift through. PageRank is used by Google to help *sort* the list of potential matches, once they have been selected using various keyword-dependent heuristics (as discussed in the last chapter). In the simplest scenario, the search engine would look up all the pages that contain the user’s search terms, and then order them by decreasing PageRank. As an example, consider a search for “Southwest Airlines.” Many, many pages—

---

<sup>b</sup> This property of PageRank is no coincidence; indeed, given the current state of technology, it is practically a necessity. Since the PageRank of each page is directly or indirectly dependent on the PageRank of millions of other pages, it can take several hours to calculate even a small number of rankings. It would, quit obviously, be impractical to do this computation each time a user enters a particular query. By designing PageRank to be completely query-independent, Brin and Page ensured that the “importance” of every page on the Web can be computed in advance, and quickly retrieved as needed.

over 900,000 in fact—contain these keywords, but very few are linked to as often, and by as many popular sites, as the airline’s home page. Sorting the matches by decreasing PageRank, we are likely to achieve the desired result, namely, that the airline’s page appears among the first few results. This example illustrates how, by using PageRank as a result ordering heuristic, we more or less rank relevant results according to their relative “importance” on the Web.<sup>11</sup>

This makes good sense. Not surprisingly, the idea of paying attention to “importance” has gained favor in the field of information retrieval. Scholars in this area have observed that when the set of documents to be searched is large and vastly heterogeneous in quality—as the Web surely is—traditional query- and content-based heuristics prove inadequate.<sup>12</sup> They usually fail to narrow the set of “relevant” results sufficiently or effectively; they are open to manipulation; and they cannot capture the relative “quality” of the various results.<sup>13</sup> These three problems do not generally arise in academic collections, wherein the number of documents on a given subject is relatively small, scholars do not usually engage in “spamming” practices,<sup>14</sup> and documents tend to be more or less comparable in quality. The unique features of the Web have lead some to propose an additional evaluative criterion of “authoritativeness,” which measures how esteemed a source is on a certain subject.<sup>15</sup> PageRank is perhaps the most famous attempt to quantify the “authoritativeness” in order to tame the wild Web of information.

## II. How PageRank Favors Popular Pages

Our description of PageRank, like that put forth by its inventors, makes heavy but unqualified use of the term “important.” This is somewhat disconcerting since importance, like relevancy, is a highly subtle, ambiguous, and subjective thing. In the broadest sense of the term, “important” means something like “worthy of consideration.”<sup>16</sup> As indicated above, we might consider a page “worthy of consideration” if it is “authoritative,” meaning that it presents a “convincing, reliable” argument “backed by evidence, and showing deep knowledge.”<sup>17</sup> What’s more, if we ascribe to ideals of democratic discourse, we should also

recognize pages that express minority and antagonistic views as important. Such voices are critical for democracy—“worthy of consideration”—because they help ensure that the people’s voice is both reasoned and informed.

PageRank, however, takes an altogether different approach. To the algorithm, being “important” simply means being “popular.” The more page authors link to your site, the higher your PageRank will be. This ranking will, in fact, be even higher if prominent pages—those which are themselves extensively linked to—reference to your site. A high PageRank, in other words, means being popular among a multitude of Web publishers, and especially among the authors of popular sites. It’s all very much like high school: being seen as “important” means being popular, and being popular means having many friends—especially popular friends.

PageRank, in this light, is really a measure of a page’s popularity among Web *authors*. But it is also (to some extent) an indicator of a page’s popularity among Web *users*. Various scholars have indicated that there is a correlation between the number backlinks a page has and its actual popularity among users.<sup>18</sup> Kavassalis, et al., describe how users “decide to visit Web sites with probabilities depending on numbers of links pointing-in to the site (in-links); conversely, sites attracting large numbers of visitors become more pointed-in than others.”<sup>19</sup> Put differently, “growth in the number of links to a site can be equated with the growth of the site’s popularity,” since, according to Adamic and Huberman, the “more a site is linked-to, the more users are aware of the site, and the more additional links it receives.”<sup>20</sup> Brin and Page’s “random surfer model”—which establishes that the PageRank of a particular document is equivalent to the probability that a “random surfer” will come across it—makes explicit the close relationship between PageRank and page popularity.

### III. Googleocracy? PageRank as a Voting Mechanism

Having suitably established how PageRank works, we can begin to examine some of the sociopolitical issues latent in its design. According to Google’s public relations literature,



PageRank is not only consonant with democratic principles, it in fact embodies the very process of democracy itself:

*PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at more than the sheer volume of votes, or links a page receives; it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important."<sup>21</sup>*

This interpretation of the algorithm as a sort of “democratic” voting mechanism is defended by celebrated Princeton computer science professor and cyberactivist Ed Felten, who writes in his blog:

*Google is a voting scheme. Google is not a mysterious Oracle of Truth but a numerical scheme for aggregating the preferences expressed by web authors. It's a form of democracy – call it Googlocracy. Web authors vote by creating hyperlinks, and Google counts the votes. If we want to understand Google we need to see democracy as Google's very nature, and not as an aberration.<sup>22</sup>*

There is, of course, a certain ring of truth to this argument. Search systems governed by PageRank determine “what people want” from what millions of Web authors themselves deem “important.” PageRank, in other words, is “democratic” precisely because it puts into action the “expressed will of the people.”

But, as convincing as the argument may seem, on close inspection it falls apart. The PageRank voting scheme, for one, does not satisfy a commonly accepted component of democracy: the “one person, one vote” principle. James Fishkin identifies this sort of “political equality” as one of the necessary conditions for ideal democracy, since it “gives equal consideration to the preferences of each citizen.”<sup>23</sup> PageRank clearly violates this requirement since “votes” cast by the authors of “important” pages weigh more heavily than those cast by the creators of unpopular or underrepresented sites.<sup>c</sup> Such a voting scheme may resemble a “democracy” in some sense—and may even, technically speaking, deserve the name—but it is

---

<sup>c</sup> To be sure, what’s an “important” page—and thus, deserving of greater voting power—is itself determined by a process of “voting.” If this second-order procedure exhibits “political equality,” the scheme may still be consonant with the “one person, one vote” ideal (consider the election of political representatives). PageRank, however, calculates importance recursively, such at each level of the recursion, “important pages weigh more heavily” in “mak[ing] other pages ‘important.’” The entire “voting scheme” is thoroughly biased towards existing, important pages.

more like a “shareholder democracy” than one in which equality and fairness are protected. In the Googleocracy, not every one is allowed to vote; only page authors, like shareholders, have any say at all. And, even among those who do have a voice, some parties are given greater influence, much as with the “one dollar, one vote” rule familiar to investors. This, quite clearly, is not the sort of democratic voting scheme we usually envision, and rightly demand, from our governments.

Even putting aside the question of whether PageRank fairly approximates “democratic” procedures, the entire concept of voting is, in any case, entirely misapplied when it comes to “governing” the dissemination of information. Sure, a political democracy generally requires that the aggregated preferences of the majority be put into practice. But this does not imply that only the majority’s views should be heard during deliberation, nor does it suggest that popular opinions should be preferred *ipso facto*. To the contrary, as discussed in chapter 3, the validity of voting—of aggregating preferences—depends precisely on the dissemination of a broad spectrum of opinions, especially those put forth by unpopular or minority groups. To recall the words of John Stuart Mill, even if “government is entirely at one with the people” and “never thinks of exerting any power ... unless in agreement with what it conceives to be their voice,” the ability to “control the expression of opinion” is outside its legitimate authority.<sup>24</sup>

What Ed Felten fails to recognize is precisely this distinction between the ideal process of “democratic” *discourse* (wherein the prevailing opinion does not dominate) and that of “democratic” *governance* (wherein the prevailing decision is the one put into action). After all, even Mill, who forcefully argued that all voices should be given equal consideration, at the same time infamously pushed for “plural voting,” wherein more “competent” citizens—the educated, or those who pass a certain literacy test—had a greater weight in the final decision-making.<sup>25</sup> On the one hand, Mill favored wholly egalitarian discussion; on the other, he supported political inequality. This asymmetry suggests that even if we ascribe to certain formal democratic voting mechanisms—including the unbalanced and rather problematic sorts

envisioned by Mill and PageRank’s proponents—we should not restrict discussion to popular opinions. A democratic “voting scheme” that decides which political actions to take should not be used to determine which opinions about those actions will actually be heard.

#### IV. Popularity, Power Laws, and PageRank

The inadequacy of the “voting scheme” argument leads us to instead examine PageRank from the perspective of *deliberative* democratic discourse. Given our discussion in Chapter 3, widespread use of the algorithm should be an immediate cause for concern. PageRank, it seems, runs against important deliberative ideals because it explicitly favors popular content and disfavors those out-of-the-mainstream voices whose importance has been widely asserted by various democratic theorists.

Making matters worse for PageRank and democracy is recent evidence that there is an enormous, preexisting bias latent in the link structure of the Web itself, the very link structure from which PageRank is computed. In a fascinating book, Notre Dame physicist Albert-László Barabási recounts how he and other scientists mapped the Web’s structure, and in so doing were able to make sense of its diversity, evolution, and sheer complexity.<sup>26</sup> At the time Barabási began his research, it was believed that the distribution of incoming links to pages on the Web followed a bell curve, meaning that some pages had many incoming links, some had very few, and the majority fell somewhere in between. If this were true, it would mean that, on average, as you surfed the web, you would be most likely to come across a “typical,” moderately popular page. Such a Web, according to Barabási, “would be the ultimate carrier of egalitarianism,” since the network would not be dominated by a few, enormously powerful players.<sup>27</sup> This, of course, would be highly encouraging from the perspective of democratic discourse. As democratic theorist Robert Goodin explains with respect to “disjointed deliberation,”

*I can see only one way in which the inputs of a plurality of groups could be blended together in a fashion that genuinely would be both directly and deliberatively democratic. Suppose each of us is a member of many different*

*“groups”...each of which approximates the deliberative ideal...Suppose furthermore that each of us overlaps any given other in only a small fraction of our group memberships. Then there might be a “web of group affiliations” which links...everyone with everyone else in a dialogue which effectively straddles the entire community.<sup>28</sup>*

Although Goodin is not speaking at all about the “Web” in the capital-W sense, this “web of group affiliations” is directly analogous to the network of hyperlinks. Note that, under his scheme, deliberative equality requires that each individual be a member of a roughly equal number of groups (i.e., have about the same amount of backlinks), and that each person (i.e., site) “overlap” (i.e., link) with any other with equal probability.

But after conducting a large-scale investigation, what Barabási instead found was that the distribution of links on the Web, rather than being “egalitarian” and “roughly equal,” actually follows a *power law* (i.e., Zipf or Pareto distribution).<sup>29</sup> This means that a small number of pages—what are now called “hubs”—collect an enormous number of backlinks, while the vast majority of documents are linked to by few or no sites at all. Recalling that a site’s traffic—its visibility and popularity—is highly correlated with its backlink count, he concludes that

*The hubs are the strongest argument against the utopian vision of an egalitarian cyberspace. Yes, we all have the right to put anything we wish on the Web. But will anybody notice? ... [Hubs] are very easy to find, no matter where you are on the Web. Compared to these hubs, the rest of the Web is invisible. For all practical purposes, pages linked by one or two other documents do not exist. It is almost impossible to find them.<sup>30</sup>*

It is worth mentioning that while a remarkably fertile body of research has largely confirmed Barabasi’s general findings,<sup>31</sup> there is still some dispute about whether *subsets* of the Web are also characterized by power law distributions. Pennock and his colleagues have found that within certain “communities” of pages—company, university, newspaper, and scientist sites—inbound links actually follow a far more egalitarian, lognormal distribution (i.e., “bell curved” on a logarithmic scale).<sup>32</sup> If this holds true for *all* subcategories of pages, it would suggest that the Web’s power law characteristic is merely an “artifact of aggregation,” and that individual sets of related pages—say, those dealing with abortion—are governed by a “considerably less

biased” distribution of links and traffic. This, of course, would dramatically mitigate concerns that a few, powerful players dominate Web discussion in any given field or on any given topic.

But, as Hindman notes, the communities studied to date “are all unusual, in that they represent groups in which there is a high degree of mutual recognition among the actors”; scientists, for example, know and habitually link to other scientists.<sup>33</sup> To get a better idea for the egalitarianism of “more typical” communities, and to hone in on direct deliberative concerns, Hindman focused on six political subcategories of pages: abortion, the death penalty, gun control, the president, the U.S. Congress, and general politics. After crawling, classifying, and tabulating almost 3 million pages, the results were “surprisingly strong and consistent”: these communities are, after all, governed by power laws. As with all such “scale free” networks, “the number of highly visible sites is small” and “comparative visibility drops off in a highly regular and extremely rapid fashion once one moves outside the core group of successful sites.”<sup>34</sup> What’s more, “almost all prominent sites are run by long-established interest groups, by government entities, by corporations, or by traditional media outlets.”<sup>35</sup> The link structure of the Web, at least when it comes to political sites, thus exhibits the same old problems of concentration and commercialization:

*[I]t suggests that it is hard for all but a few “ordinary citizens” to post their views prominently—and conversely, to read the views of other ordinary citizens, unless they are highlighted by a small number of prominent sites. Political speech posted online—particularly speech without the resources of a large organization behind it—is simply not easily accessible, because it is obscured by countless other Web pages.<sup>36</sup>*

We might hope that, as crawling and matching technology matures, search engines will be better able to do what we, as individuals, cannot: to find and catalog these millions of poorly-linked, underrepresented pages. Indeed, given the impact search engines have on directing Web traffic, they could counteract, or at least reshuffle, the enormous link inequality in the Web by drawing users’ attention to pages which are informative and relevant, but would otherwise be lost to the oblivion of cyberspace. If search engines rely heavily on PageRank, however, such aspirations will not come to fruition. Rather than mitigating the systematic

inequity present in the link structure of the Web, PageRank mirrors—and magnifies—this inequality. PageRank, is after all, a measure of the link structure’s existing bias toward a particular page, a metric for how visible (or invisible) a page is on the Web. In this light, it is disappointing, but not surprising, that Upstill, Craswell, and Hawking PageRank to be biased in favor of the sites of large, famous, technology-oriented companies—that is, in favor of popular, commercial sites.<sup>37</sup>

When PageRank is used to rank popular sites more highly, unpopular voices face a double bind: people will tend not to “randomly” stumble across their site, and their pages will not make their way into search engine listings. Search engines that rely heavily on PageRank to order their results essentially put a megaphone to the mouths of the Web’s most powerful voices, and a muzzle to those of the already underrepresented.

In addition, because PageRank is recursive, the authors of those extraordinarily popular hubs will have an even greater say in determining which *other* sites will be given a high ranking. This may allow a prominent site to dominate discourse on a given subject by linking extensively to self-reinforcing, like-minded pages. These concerns are, it should be emphasized, quite real; The Church of Scientology has, for example, recently constructed an elaborate web of self-reinforcing pages in an attempt to dominate Google’s results and silence dissenting voices (a point to which we will return in the following chapter).<sup>38</sup>

Making matters still worse, PageRank can be used—and apparently is used by Google—not only to determine which sites will get a high ranking, but also to guide their crawlers. If a previously unseen page is linked to by many popular pages it will have a much greater chance of being indexed than one with fewer backlinks.<sup>39</sup> And so, not only will users rarely find a link to these pages, but the search engines will be “biased against them, ignoring them as they crawl the Web.”<sup>40</sup> Such pages won’t get indexed. They won’t get seen. They, indeed, might as well not exist.

The use of PageRank by search engines—which ends up disfavoring unpopular, out-of-the-mainstream content on several grounds—thus seems to reproduce the same

antideliberative bias typically associated with the traditional media. Consider, for example,

Cooper's remarks:

*The general principle that ... policy should draw people into civic discourse applies with particular force to minority points of view. In the commercial model, popular, mainstream, and middle of the road ideas will almost certainly find a voice, one that is likely to be very loud. However, the unpopular, unique, and minority points of view will not.*<sup>41</sup>

Just as media outlets face enormous economic pressures to reflect “middle of the road” content, so too do “search engines wishing to achieve greatest popularity ... tend to cater to majority interests.”<sup>42</sup> PageRank was, in fact, created expressly to reflect the preferences of an “average” user. According to its creators,

*One of the design goals of PageRank was to handle the common case for queries well ... It is important to note that the goal of finding a site that contains a great deal of information ... is a very different task than finding the common case ... site. There is an interesting system that attempts to find sites that discuss a topic in detail ... this results in good results for queries like “flower”; the system will return good navigation pages from sites that deal with the topic of flowers in detail. Contrast that with the common case approach which might simply return a commonly used commercial site that had little information except how to buy flowers ... we are concentrating only on the common case approach.*<sup>43</sup>

PageRank, in other words, abandons the goals of actually reflecting a page's “importance” or “authoritativeness” on a given subject, and instead aims to mirror the “common” wishes of users. This, as the creators' own example illustrates, can have the problematic effect of promoting popular, commercial pages over more detailed, authoritative, noncommercial sources of information. Aiming only to predict what will be popular among its users, the algorithm makes no attempt to look for “reliable arguments ... backed by evidence and showing deep knowledge.” As a result, it does little to elevate *underrepresented* voices—voices that might otherwise be drowned out in the deafening din of the Web—no matter how reasoned or relevant they may be. Moreover, by taking a “one size fits all” approach, PageRank does not reflect the desires of any atypical, outside-the-mainstream users that might actually wish to see such “unpopular” content.<sup>44</sup>

But these problems are more or less typical of commercial search engines in general. Introna and Nissenbaum, in their groundbreaking overview of search engine bias, have

observed that “while markets undoubtedly would force a degree of comprehensiveness and objectivity in listings, there is unlikely to be much market incentive to list sites of interest to small groups of individuals ... or, for that matter, individuals of lesser economic power.”<sup>45</sup> PageRank, in appears, is just one embodiment of the general tendency of search engines to favor popular content of the sort preferred by the widest segment of its users.

So obvious are mainstream, commercial tendencies latent in commercial search that economists have begin to observe that “hyperlinks have already attained monetary value as incoming links to a Web site can increase [a business’] visibility on major search engines,” and some have even begun simulating “the economy of Web links.”<sup>46</sup> Although the exchange of ‘real’ for ‘link’ currency is as of yet a “black market” affair, it is becoming clear that through technologies such as PageRank, economic, organizational, and communicative influence may yet again start to work in tandem, and against the deliberative ideal that people be exposed to diverse, antagonistic sources from every corner of society. To the degree that search engines adopt popularity-biased ranking mechanisms to increase their traffic, utility, and profitability, these gatekeepers’ capacity to promote a rich online forum for public deliberation is diminished.

## V. The Rich Get Richer: New Opinions vs. the Entrenchment of the Status Quo

But if public discussion and debate—about political, economic, religious, or whatever issue—were restricted to the most popular opinions, deliberation would serve no purpose other than to entrench the prevailing views, to preserve the *status quo*. If scholars confined themselves to working with established canons, for example, it is likely that research would grind to a halt. Similarly, without dissenting voices to call into question popular opinions, those opinions would neither be modified nor supplanted. The juggernaut of “prevailing wisdom” steer society, entirely unchecked and utterly unstoppable. This would obviously be a travesty for democracy, since even if government were “entirely at one with the people,” the people’s voice may only reflect erroneous, unchallenged, and tyrannous beliefs. If history has



taught us anything, it's that "prevailing wisdom" is often wrong, and that progress emerges precisely from the subversion of the status quo.

This truism, of course, extends beyond the context political, religious, or scientific discourse. Capitalism, for instance, is grounded on an ideal of competition, which requires that a broad array of options is made available to the public, so that individuals can be free to select "the best product at the lowest cost." If there are no alternatives available, if existing monopolies leverage their enormous power to prevent completion, or if the market is biased towards existing products and against new ones, this ideal is potentially undermined. Capitalism, like democratic discourse, can only be justified when the *status quo* is fiercely and incessantly challenged.

With PageRank, unfortunately, the *status quo* is explicitly favored. The algorithm, as we have argued, entrenches a system whereby popular views are promoted, and unpopular or underrepresented views are systematically penalized. An obvious corollary of this is the fact that a site which is linked to extensively will inevitably fare better than an up-and-coming site with relatively few backlinks. If the author of a new page wishes to obtain a high PageRank, so that her views can be widely disseminated, she must first independently establish the "importance" of her page among the community of Web authors. Other writers must "certify" the importance of the views expressed on her page by linking to it, a process which Brin and Page rather vaguely liken to "academic peer review."<sup>47</sup> Once her page becomes sufficiently referenced, it is more likely to appear among the search engine's results, and from there it may, perhaps, enjoy further popularity. The onus of initially identifying and linking to new voices consequently falls not on the search engine but on a large community of Web authors. Whether this should be a cause for concern depends on whether following conditions are met: first, that a sufficient number of "important" page authors actively link to a diversity of new and underrepresented voices; and second, that such authors will be able to *find* new sources of information as yet ignored by popularity-biased search engines.

At first glance, there may seem to be little we can say regarding the first of these conditions, since it is an enormously challenging task to determine whether or not an entire community of authors systematically link to underrepresented voices.<sup>48</sup> There is, however, a rather peculiar kind of Web site that worth mentioning here, one that has often been used precisely to bring attention to interesting, “fresh” content on the Web. These so-called “weblogs” or “blogs” have achieved considerable popularity in the last few years. Their “editors” habitually “post” links to other Web sites they deem noteworthy, and usually accompany these “posts” with some sort of commentary or discussion. Rebecca Blood, herself a “blogger,” traces the history of this form, describing how these sites

*present links both to little-known corners of the web and to current news articles they feel are worthy of note...By highlighting articles that may easily be passed over by the typical web user too busy to do more than scan corporate news sites, by searching out articles from lesser-known sources, and by providing additional facts, alternative views, and thoughtful commentary, weblog editors participate in the dissemination and interpretation of the news...<sup>49</sup>*

To the extent that weblogs present a diversity of views, and are open to reflection and debate, they may be democratizing forces in their own right.<sup>50</sup> Our interest here, however, is in the narrower question of how, if at all, these sites affect PageRank’s results and encourage the dissemination of new opinions through the Google search engine.

In his own blog, John Hiler argues that “unbeknownst to most, weblogs have a significant impact on Google search results.”<sup>51</sup> PageRank, he writes, “loves links, and weblogs are all about links. Every time a blogger links to a website, its Google rank ratchets up ever so slightly. If enough bloggers pile onto that link, it can start to have a significant impact on a site’s Google rank.” According to Hiler, bloggers, by choosing which web sites they will link to, “are the *voters* in this political system,” in the Googleocracy. Their votes, moreover, are given enormous weight by PageRank, since the high degree of interlinking between the various blogs—what some have called “incestuous”—makes them appear as “important” sources to the algorithm. The end result is that “even if you never visit a blog, you’re being influence by the them. The collective votes of the weblog community are determining what sites you see on

Google.” Evidence suggests that Google is very much aware of the effect weblogs have over PageRank’s computation. In 2001, it started indexing frequently updated sites such as blogs in an attempt to cull a greater number of “fresh” links and update its billions of PageRanks accordingly.<sup>52</sup> And in 2003 Google went one step further by acquiring the Web’s most popular blog-hosting service, *blogger.com*.<sup>53</sup>

The degree to which weblogs single-handedly influence PageRank’s results may, however, be overstated; the above claims are mostly unsubstantiated and, after all, were written by a blogger himself. Some have even indicated that Google, faced with complaints that its results are dominated by blogs, has ratcheted down their ability to influence PageRank. In any case, if we wish to more adequately assess how well new sites are able to build up their ever-important backlinks, what we need is a more rigorous and comprehensive examination of the growth patterns of the Web in general.

Again, Barabási, together with his colleagues in the emerging field of “complex networks,” provides such an analysis. He reminds us that the current structure of the Web did not emerge out of thin air; it grew, page by page, from a single document (Tim Berners-Lee’s home page) to the billions we find today. One by one, each author linked to some subset of the pages available at the time of publication. To see how order ‘emerges’ through this process, suppose (for the sake of argument) that the millions of authors *randomly* selected links, with equal preference given to old and new opinions, to popular and unpopular pages. The resulting Web, he discovers, would exhibit an enormous, and ever-entrenching bias in favor of the most established sites:

*Despite the fact that we choose the links randomly and democratically, the nodes ... are not equivalent to each other. We have easily identifiable winners and losers. At each moment all nodes have an equal chance to be linked to, resulting in a clear advantage for the senior nodes. Indeed...the first nodes...will be the richest, since these nodes have the longest time to collect links. The poorest node will be the last one to join the system.*<sup>54</sup>

Put differently, even if all Web authors had the best of intentions, linking to sites in an utterly egalitarian manner, a “rich-get-richer” phenomenon would nevertheless emerge. Making matters worse, after Barabási collected and analyzed graphs of the Web at different periods, he

found that we do not link at random. Instead, “we all follow an unconscious bias, linking with a higher probability to the nodes we know, which are inevitably the more connected nodes on the Web”<sup>55</sup>:

*As a longtime reader of the New York Times, it is a no-brainer for me to choose nytimes.com. Others might prefer CNN.com or MSNBC.com. Significantly, however, the Webpages to which we prefer to link are not ordinary nodes. They are hubs. The better known they are, the easier it is to find them on the Web and so the more familiar we are with them ... The bottom line is that ... we follow preferential attachment: When choosing between two pages, one with twice as many links as the other, about twice as many people link to the more connected page.*<sup>56</sup>

The fact that authors “add links at a higher rate to those nodes that are already heavily linked”<sup>57</sup> clearly exacerbates the rich-get-richer phenomenon. Not only do senior pages have a longer time to accumulate links, but due to “preferential attachment” the most popular among them continue to “grab a disproportionately large number of links at the expense of the latecomers.”<sup>58</sup>

Barabási’s findings suggest that, when it comes to backlinks, “popularity is attractive.” The link structure of the Web develops in manner that increasingly builds up links to existing, popular sites. Thus, while bloggers may have some effect on increasing backlinks to new voices, in the grand scheme of things such influences are but “rare statistical fluctuations.” This may, at first, seem like an complete contradiction: on the one hand, many (including Google itself) seem to believe that blogs greatly affect PageRank, yet such pages seem to be insignificant relative to wider liking patterns on the Web. One way of reconciling these two perspectives is to recognize that bloggers *do* have a great effect in establishing what’s popular *in a new field*: recent news events (e.g., “Abu Gharaib prisoners”), Web “memes,” and so on. Google relies on these sites to identify what new subjects may be searched for, and which documents to return. But when it comes to established topics—e.g., “Microsoft,” “abortion,” or “flower shop”—the inertia of a rich-get-richer Web win out. In these cases, it is the existing, powerful sites that will continue to receive ever-greater PageRank scores. And, the more that search engines rely on this metric, the harder it will be for new voices to reach a sizeable audience.

## VI. Quantifying the Influence of PageRank on Google's Results

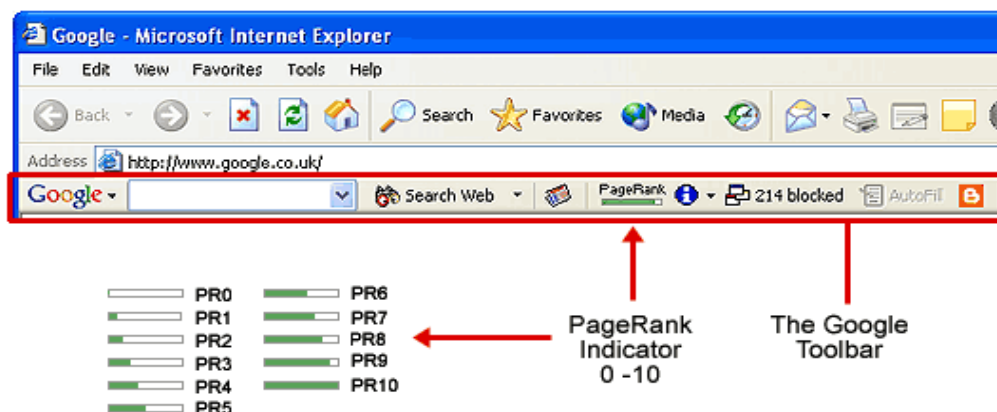
A careful reader may have noticed how, throughout our discussion, we have largely directed our attention (and criticism) at the PageRank heuristic, not at Google's ranking scheme in general. And for good reason. The PageRank algorithm, which favors prominent voices at the expense of underrepresented ones, is only worrisome *to the extent that it is used to select and order search results*. This point may seem obvious, yet existing sociopolitical analyses of PageRank<sup>59</sup> have taken it for granted that Google “tend[s] to return these most connected sites first.”<sup>60</sup> Hindman, for example, defines ‘Googleocracy’ as “the rule of the most heavily linked” without showing that such sites do, in fact, dominate the search engine's results.<sup>61</sup> Some “search engine optimizers”—those who seek, for profit or for popularity, to increase their sites' Google placement—have, in contrast, suspected that the benefit of “good PageRank” is rather slight and is decreasing as the search engine evolves. While this question is of enormous import for both SEOs and scholars alike, there has apparently been no effort at systematically answering it. We will therefore make one modest attempt at doing so.

It is first worth mentioning that despite Google's public claims that PageRank constitutes “the heart” of their search software, the algorithm is in fact only one component of an enormously multifaceted and complex ranking scheme. If PageRank were the only metric used to order search results, the quality of these results would be pitiful since, for example, the first result may be an enormously “important” site of little or no relevance to the user's specific query.<sup>62</sup> Unfortunately, while it is clear that other heuristics play a role in selecting and sorting the search results, their workings and interplay are by and large a well-guarded company secret (see Chapter 6).<sup>63</sup>

Nevertheless, anecdotal evidence does seem to support the claim that PageRank can greatly influence Google's results. Consider, for instance, a search for “Microsoft.”<sup>64</sup> There are many anti-Microsoft sites on the Web,<sup>65</sup> so we might reasonably expect at least some of these sites to show up on a search for the search engine giant. In reality, among the top 30 results, 25 are links to Microsoft owned or affiliated sites (e.g., *MSN*, *Hotmail*, *bCentral*, *Expedia*, *X-*

box), 4 are to apparently neutral developer sites, and only 1 is a link to an antagonistic site. One item that struck us as curiously out of place was the 9<sup>th</sup> result to the site of “TerraServer,” which is, apparently, a provider of satellite photographs. Given that this relatively obscure page seems to be of little relevance to the query of “Microsoft” (and certainly of less relevance than most anti-Microsoft sites), we found this perplexing—until we discovered that that the site was funded by Microsoft Research. Since TerraServer is linked to from a prominent Microsoft page, it may have been able to achieve a high PageRank, and thus, an unusually high position among the results. This case shows, as indicated earlier, how well-linked pages such as Microsoft’s not only enjoy prominence among the search results, but might also boost the visibility of favorable websites to which they link.

Such experiences are suggestive but nevertheless speculative: we do not know for certain whether PageRank, as opposed to some other variable, explains this phenomenon. To obtain a more systematic understanding of the relationship between PageRank and “Google rank,” we attempted to statistically relate these two variables. A strong, negative correlation—high PageRank predicting placement near the 1<sup>st</sup> position—would suggest that the metric is indeed important heuristic and, thus, that popular sites are preferred. Of course, the strength of the correlation may differ from query to query (for example, it may be the case that the more homogeneous the set of results, the more PageRank becomes the determining factor), so a careful analysis would need to take such variance into account.



**Figure 5-2.** *PageRank on the Google Toolbar.* This constituted our only feasible means of obtaining PageRank for our experiment.

An immediate obstacle to conducting such an analysis was the difficulty of assessing a given document's PageRank. Google does not make available its PageRank values, and due to the recursive nature of the algorithm—each value is dependent on that of millions of other Web pages—computing them on one's own requires vast computational resources.<sup>66</sup> Luckily, the publicly available “Google Toolbar” displays a very rough estimate for the PageRank of the currently displayed website as a means of letting the user see the “authoritativeness” of that page (see Figure 5-2).<sup>67</sup> Although the imprecision of this tool introduces an admittedly high margin of error, various scholars have already relied on the toolbar values in their analyses.<sup>68</sup>

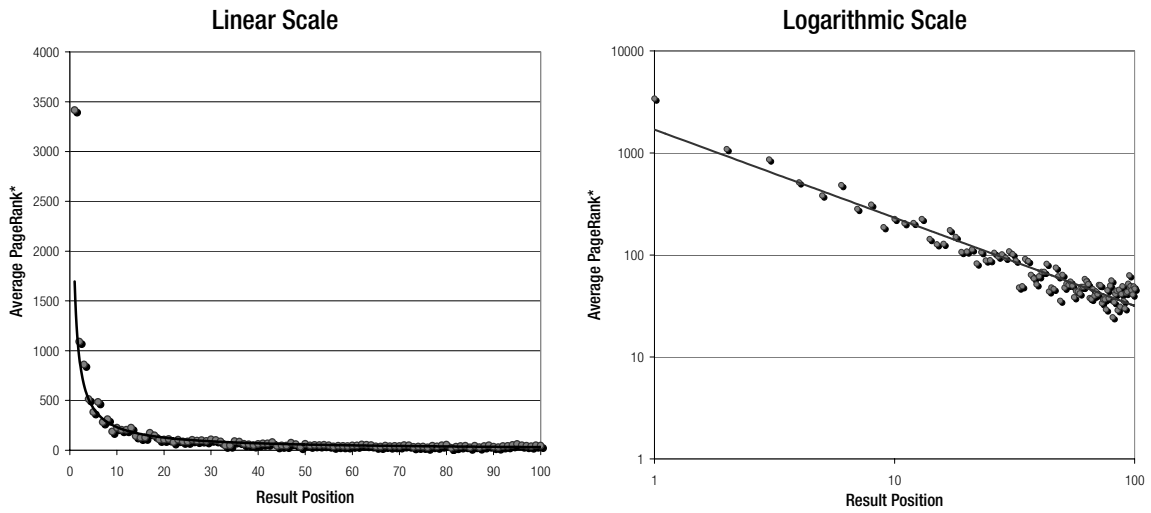
By combining this tool with various other Google toolkits, we were able to systematically obtain both of the required data. As described in *Appendix II*, we used the following methodology in our research. First, we constructed three sets of keywords: the top 100 search phrases, 100 random search phrases, and a list of 100 ‘political’ queries. For each of the 300 query words, a Google search was performed. We then recorded the position and the toolbar-reported PageRank for each of the first 100 sites returned by each query. The data—which consisted of nearly 30,000 PageRank/position pairs—were then correlated, analyzed, and graphed.

Our results were remarkably consistent. For each set of queries, there was a strong, negative correlation between Google position (e.g., “1” for the first spot) and the average PageRank of pages at that position. Put differently, documents with a high PageRank were indeed more likely to appear near the coveted first spot. Not only that, but a regression analysis revealed that the relationship followed a power law (Figure 5-2), such that the prominently placed sites tended to have a much, much greater PageRank than those less prominently placed. Although these results were highly significant in all three cases (Figure 5-4), they were more pronounced with the ‘top’ and ‘political’ queries than with the set of ‘random’ queries. This may be because random searches were often highly specific (e.g., “plastic capped jawa for sale”), producing a smaller set of relevant matches. In contrast, the frequent

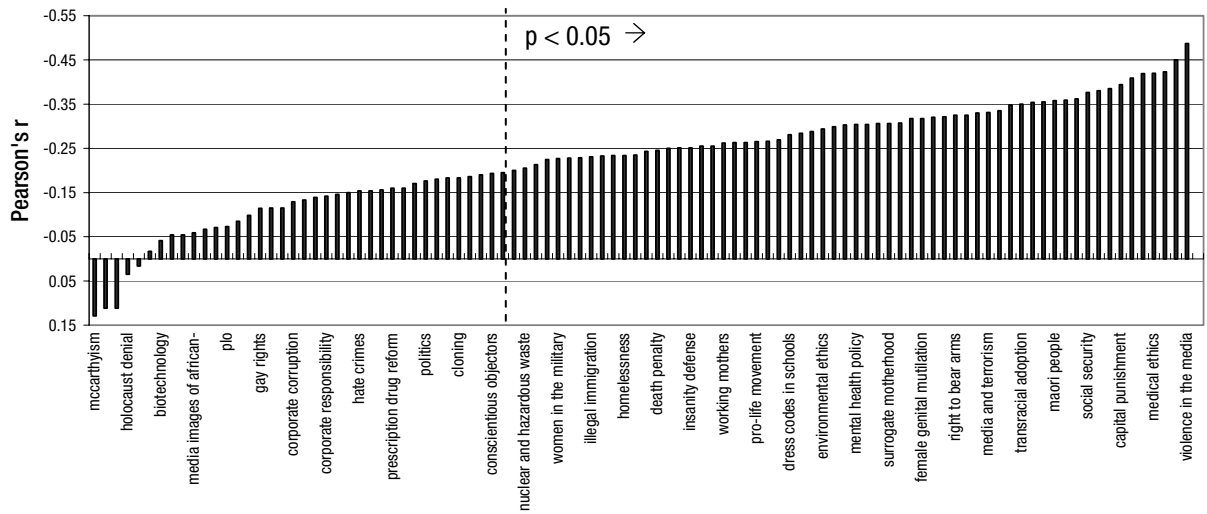
**Figure 5-2: Average PageRank\* by Position**

Combined Data Sets

Power Law Fit:  $R^2 = 0.919$ ,  $p < 0.001$



**Figure 5-3: PR-Position Correlation by Query**  
Political Topics



**Figure 5-4: General Statistics**

Query Set	Median PageRank (*)	Biv. Correlation <sup>†</sup> r	Power Law Regression <sup>†</sup> r <sup>2</sup>
Top	5 (1178.905)	-.361**	.873**
Political	4 (197.42)	-.437**	.871**
Random	3 (33.06)	-.348**	.630**
Combined	4 (197.42)	-.390**	.919**

\*\* significant at the  $p < 0.001$  level

\* the toolbar-reported PR values—which seem to indicate the *logarithm* of the “real” PageRank—have been converted via exponentiation (see Appendix II).

† the analyzed variables were the Google position and the average PageRank\* of all results observed at that position.



and political queries (e.g., “president”) usually returned an extraordinarily large set of matches, and consequently PageRank seems to have played a greater role in discriminating among the results. It is also worth mentioning that while the aggregate “average” PageRank at each position followed a predictable pattern, the PR/position correlation for individual queries was more varied and less significant, indicating that the results for individual searches were often not strictly ordered by decreasing PageRank (Figure 5-3). The overall conclusion, however, is clear: in general, enormously popular matches tend to appear most prominently in Google’s results.

In addition, the median PageRank was also quite high on all three sets of queries, suggesting the ‘typical’ result was far more popular than the ‘typical’ page on the Web. If indeed PageRank follows a power law, then “the vast majority of pages would have a toolbar PageRank of 0 or 1.”<sup>69</sup> But we found that the typical result had a PageRank of 4 on the Google Toolbar, meaning that a relatively large number of sites pointed to it (this was, again, less pronounced for the random queries, and most pronounced for the top queries). Thus, regardless of how the top 100 sites are themselves ordered, the entire set was unusually popular. In the end, it is not just that underrepresented, backlink-impooverished sites will probably not appear near the *top* of Google’s results. Rather, these sites are not likely to show up *at all*.<sup>d</sup>

---

<sup>d</sup> It is worth addressing, at least briefly, the typical “correlation vs. causation” criticism. While the end result would not change—more popular pages get greater visibility—it could be argued that PageRank does not determine ranking but *vice-versa*: that those pages that are more relevant to users and thus appear more prominently will attract greater attention, will garner more links and will, therefore, receive a higher PageRank. While we do not doubt that this does occur—indeed, we believe it leads to a positive feedback cycle—there are various reasons to suggest that Google’s use of PageRank is a causal factor. First, it is often the case that two *identical* copies of a page (stored on different servers or different locations on the same server) nevertheless have markedly different PageRanks and differ widely in their prominence (suggesting it’s not just about what the page says, but who says it). Second, it is unclear why companies would routinely pay SEOs to restructure the site’s links to increase PR unless this had a marked effect on search engine placement. Third, Google has unambiguously states that it does use PageRank to order its results. Fourth, it is unlikely that we’d get such “perfect” data unless an algorithm was explicitly ordering results in this manner. Although further experimentation would be needed to determine causality with objective certainty, we believe it is clear that PageRank does play a role in selecting and ordering query matches.

## VII. What the Politics of PageRank Mean for Google Users

In light of the concerns we raised about the PageRank algorithm, our empirical findings—which indicate that PageRank does significantly predict placement among Google’s results—are not encouraging. Taken together, our discussion suggests that when users turn to Google for information on a given topic, these users are more likely to see popular, commercial, “middle-of-the-road” content and not as likely to find what they, perhaps, “don’t want to see”: the underrepresented, less widely-recognized, antagonistic, noncommercial perspectives on a given issue.

Such a prediction, of course, rests on the assumption that popular pages primarily express popular opinion. Although this is consistent with ‘selective attention’ psychological theories—which hold that people prefer reading material that supports their already held predispositions—it may very well be that the pages deemed most “important” by the community of Web authors will be those rich, comprehensive, and even out-of-the-mainstream treatments of some controversial matter. To make any conclusions about the “politics of PageRank” we must therefore conduct a more content-specific analysis of Google’s results, one that examines whether controversial aspects of a topic do, in fact, appear among the returned pages.

In a recent article, Susan Gerhart makes a first attempt at addressing this question, namely “Do Web search engines suppress controversy?” For Gerhart, “controversies express the richness and depth of a topic” and can even “make a critical difference in life-altering decisions” (consider searches on medical treatments). As part of her experiment, she queried three major search engines—Google, Teoma, and AllTheWeb—for information on five broad topics, each of which she knew to contain some controversial subtopic that was well documented on the Web (see Table 5-1). Gerhart then recorded, in painstaking detail, whether and how such disputed perspectives were raised within the search results. She looked, for example, at whether a search for “distance learning” would return sites that shared David Noble and other academics’ concern about “the loss of control over their intellectual products,

Broad Topic Query	Controversial Subtopic
Distance Learning	" <i>Digital Diploma Mills</i> ," The trend toward commercialization of education.
Albert Einstein	Did Einstein's first wife, <i>Mivela Maric</i> , receive appropriate credit for scientific contributions to Einstein's early work?
Female Astronauts	Did the U.S. space program discriminate against " <i>Mercury 13</i> " women pilots who passed preliminary astronaut screening tests?
St. John's Wort	Does this popular herbal remedy work effectively for depression and mood improvement? Recent medical trials differ in their results.
Belize	This small Central American country has long, and ongoing, <i>border dispute with Guatemala</i> with deep historical roots in Spanish and British colonialism.

**Table 5-1. Controversial subtopics for five general queries.** Gerhart examined whether these five controversies were addressed by Google's results. (Source: Gerhart 2004)

as well as contact with students" and the tendency of these programs to act as "digital diploma mills."

Gerhart's results, though they were consistent across the three search engines, are somewhat inconclusive. For three of the topics—distance learning, Albert Einstein, and Belize—the respective disputes were to a great extent "suppressed," such that most surfers would not "be exposed to the controversies by [a general] search...alone." But for the two other topics—female astronauts and St. John's Wort—it was indeed possible for a user to "[run] across a page within search results" and "definitely recognize the existence of controversy, which this page explains in some detail."

On closer inspection, however, the findings can be at least partially reconciled. The controversies that were suppressed were, for the most part, those overrun by 'organizational clout' (e.g., official Belizean tourism sites or distance learning programs) or by pages that reflected what users "wanted to see" (e.g., Einstein quotations, 'bland' biographies for term papers, etc.). Only when a controversy was frequently discussed within a topic and widely recognized as important (e.g., the effectiveness of St. John's Wort) were disputed matters readily documented. In other words, the controversial viewpoints that perhaps matter most critical from a deliberative point of view—those antagonistic perspectives that haven't

garnered widespread attention—are precisely those that are left out of the search engine’s results. Gerhart concludes that

*Search technology tends to present the ‘sunny side’ of a topic. This bias reflects authors’ links and searchers’ choices. A few organizations often exert strong commercial (or nonprofit) influence through Web site investments and accrue high link counts through their off-Web prominence.<sup>70</sup>*

In the end, these “strategies do lead to organizationally dominated search results depriving searches of rich experience and, sometimes, of essential decision-making information.” If we really believe that through “democratic media” like the Web individuals “must have the freedom to communicate radical and unpopular ideas and opinions”—and, what’s more, that citizens should be exposed to what “they don’t want to hear”—then search engines fall short of these aspirations when they fail to disseminate those dark, uncomfortable views on a given topic. It would appear that Daniel Brandt, that search engine conspiracy theorist, might not have been so far off the mark after all.

What Brandt does not recognize, however, is that Google and other “Web search engines do not conspire to suppress controversy.” Rather, this is direct consequence of the seemingly laudable attempt to please its users. As Gerhart suggests, “On the simplest query for a topic, a searcher expects to see the most influential organizations appear, not a bundle of dirty laundry or diatribes attacking the topic’s leaders or ideas ... Searchers use a particular engine because its biases give them the results they usually want.”<sup>71</sup> By leveraging PageRank and similar technologies<sup>72</sup> to “give people what they want” (i.e., popular, organizational content), Google and other search engines make their sites useful to users in the hopes that these users will return and, perhaps, generate some ad revenue. And so, much as with the tomato harvester case described by Winner, the suppression of controversy is not the “result of a plot” by its creators but is rather a consequence of “ongoing social process in which scientific knowledge, technological invention, and corporate profit reinforce each other in deeply entrenched patterns, patterns that bear the unmistakable stamp of political and economic power.”<sup>73</sup> The tomato harvester and Google are both the result of a “technological deck that has been stacked in advance” to “reward some while punishing others.”<sup>74</sup> It is for this reason that

the same widespread communicative imbalance finds itself on Google, other search engines, and the link structure of the Web itself. All of these technological structures are the *embodiments* of a social order that gives preference to popular opinions over underrepresented voices, to the detriment of democratic discourse.

## VIII. 'The Big Picture': Conclusions on PageRank

In this chapter, we have described how the design, implementation, and use of one of Google's most popular and celebrated algorithms—PageRank—may undermine the democratic values frequently associated with the Web. The bias we have described here is, however, not the same sort of bias we usually refer to when we talk about the media. As Mowshowitz and Kawaguchi point out, the search engine's bias is not a 'content bias' since Google is not deliberately authoring skewed representations of a given topic. What we have here is, instead, an "indexical bias...exhibited in the selection of items" for distribution.<sup>75</sup>

This bias is, in turn, consistent with all three categories of 'computer bias' outlined by Friedman and Nissenbaum. PageRank embodies *preexisting bias* because it encodes the age-old tendency of the media favor mainstream popular content over antagonistic, noncommercial sources of opinion. It is likely to suffer from *emergent bias* as the link structure of the Web from which it is computed continues to grow in a rich-get-richer fashion. Finally, PageRank suffers from a number of *technical biases* that are the consequences of engineering challenges faced by its creators. Because PageRank had to be computed in advance for scalability reasons, it was designed as a query- and user-independent measure of "importance." More broadly, however, the algorithm suffers from the "formalization of human constructs" bias that occurs whenever engineers try to make "discourse...amenable to computers." Computers cannot, and perhaps never will, be able to capture the nuances inherent in selecting a diverse set of opinions on a given topic and, as a result, are not likely to ever promote a truly egalitarian selection of perspectives. Since it is impractical and maybe

even more undesirable to have humans manually deciding which results will be returned for any given query, this sort of ‘technical bias’ seems to be a necessary evil.

While the biases described here can certainly have negative political consequences,<sup>76</sup> we must keep in mind that ‘democraticness’ is not a binary property. As Dan Brinklin has pointed out, even if popular sites do get a sizeable boost for some queries, rarely do the same corporate megasites pop up across different search topics.<sup>77</sup> As a result, “small players [still] matter,” especially when we are conducting ‘typical’ searches for specialized information not easily found in the traditional media. This claim—which was partly confirmed by our own data indicating that ‘random’ queries were less thoroughly ordered by popularity—suggests that unprecedented number of “ordinary citizens” may still be reaching reach sizeable publics through the Google search engine.<sup>78</sup>

## Notes

---

- <sup>1</sup> Google Inc., *Our Search: Google Technology* (Google.com, 2004 [cited May 13 2005]), available from <http://www.google.com/technology/>.
- <sup>2</sup> Larry Page et al., "The PageRank Citation Ranking: Bringing Order to the Web," (Stanford, CA: Stanford Digital Library Technologies Project, 1999), 1, available from <http://dbpubs.stanford.edu/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=&name=1999-66.pdf>.
- <sup>3</sup> Ibid.
- <sup>4</sup> Jon Kleinberg et al., "The Web as a Graph: Measurements, Models, and Methods," in *Computing and Combinatorics: Proceedings of the 5th Annual International Conference (COCOON'99)*, ed. T. Asano, et al., *Lecture Notes in Computer Science* (Tokyo, Japan: Springer-Verlag, 1999), 1.
- <sup>5</sup> Ibid.
- <sup>6</sup> Described in Lucas Introna and Helen Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society* 16, no. 3 (2000): 172. Also in Junghoo Cho, Hector Garcia-Molina, and Lawrence Page, "Efficient Crawling Through URL Ordering," *Computer Networks and ISDN Systems* 30, no. 1-7 (1998): 3.
- <sup>7</sup> Lawrence Page, "Method of Node Ranking in a Linked Database," assigned to *The Board of Trustees of the Leland Stanford Junior University*, U.S. Patent 6285999 (September 4, 2001).
- <sup>8</sup> Cho, Garcia-Molina, and Page, "Efficient Crawling Through URL Ordering," 3.
- <sup>9</sup> Introna and Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," 173.
- <sup>10</sup> Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine" (paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998), 4, available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 1, 2005).

- 
- <sup>11</sup> While sorting results by PageRank *does* amount to ordering the results by *global* importance on the Web, it does not necessarily order these results according to their respective “importance” *for the topic in question*. This is an important distinction, since a marginally relevant page may be globally important (say, yahoo.com) but not really important within a particular field (say, “airline tickets”). PageRank ignores this distinction for the sake of efficiency.
- <sup>12</sup> These shortcomings are discussed briefly in Jon Kleinberg, “Authoritative Sources in a Hyperlinked Environment,” *Journal of the ACM* 46, no. 5 (1999).. Kleinberg is usually associated with concept of using link analysis to measure authority on the Web, and his insights apparently served as an inspiration for PageRank.
- <sup>13</sup> Page et al., “The PageRank Citation Ranking: Bringing Order to the Web,” 1-2.
- <sup>14</sup> Peer review, formatting requirements, and scholarly self-enforcement help ensure that search systems and readers are not deceived or manipulated into reading fraudulent, irrelevant, or incorrect content. A humorous counterexample can, however, be found in the Alan Sokal “Social Text Affair.” Sokal, a physicist, wrote a satirical but seemingly well-cited article which argued that reality is entirely a social construct, that there are no physical “laws” in the universe at all. He managed to get the fraudulent work published in an issue of *Social Text*. Soon after its publication, the editors realized they had been hoaxed—Sokal had simultaneously published a scathing criticism of the journal’s editorial review practices, and its inability to spot “nonsense,” in *Lingua Franca*.
- <sup>15</sup> Erzsebet Toth, “Statistical Methods in Measuring Search Engine Performance,” *Acata Mathematica Academiae Paedagogicae Nyiregyhaziensis* 20 (2004), available from [http://www.emis.ams.org/journals/AMAPN/vol20\\_1/amapn20\\_12.pdf](http://www.emis.ams.org/journals/AMAPN/vol20_1/amapn20_12.pdf) (accessed May 5, 2005).
- <sup>16</sup> “Important,” *Encarta World English Dictionary* 2005.
- <sup>17</sup> “Authoritative,” *Encarta World English Dictionary* 2005.
- <sup>18</sup> Matthew Hindman, Kostas Tsioutsoulklis, and Judy Johnson, ‘*Googlearchy: How a Few Heavily Linked Sites Dominate Politics on the Web*’ (July 28 2003 [cited April 10 2005]), 9-10, available from <http://www.princeton.edu/~mhindman/googlearchy--hindman.pdf>.
- <sup>19</sup> Petros Kavassalis et al., “What Makes a Web Site Popular?,” *Communications of the ACM* 47, no. 2 (2004): 53.
- <sup>20</sup> Lada Adamic and Bernardo Huberman, “The Web’s Hidden Order,” *Communications of the ACM* 44, no. 9 (2001): 59.
- <sup>21</sup> Google Inc., *Our Search: Google Technology*.
- <sup>22</sup> Edward Felten, *Googleocracy in Action* (Freedom to Tinker, February 3 2004 [cited May 15 2005]), available from <http://www.freedom-to-tinker.com/archives/000509.html>.
- <sup>23</sup> James Fishkin, *Democracy and Deliberation: New Directions for Democratic Reform* (New Haven, CT: Yale University Press, 1991), 31.
- <sup>24</sup> John Stuart Mill, *On Liberty*, ed. Elizabeth Rapaport (Indianapolis: Hackett Publishing Company, 1978), 16.
- <sup>25</sup> For an detailed discussion of Mill’s proposal for “plural voting,” see Joe Miller, “J.S. Mill on Plural Voting, Competence and Participation,” *History of Political Thought* 24, no. 4 (2003).
- <sup>26</sup> Albert-László Barabási, *Linked: The New Science of Networks* (Cambridge, MA: Perseus Pub., 2002).
- <sup>27</sup> *Ibid.*, 57.
- <sup>28</sup> Robert Goodin, “Democratic Deliberation Within,” in *Debating Deliberative Democracy*, ed. James Fishkin and Peter Laslett (Malden, MA: Blackwell Publishing, 2003), 57.

- 
- <sup>29</sup> Barabási explained this finding initially in Albert-László Barabási and Réka Albert, "Emergence of Scaling in Random Networks," *Science* 286, no. 5439 (1999).
- <sup>30</sup> Barabási, *Linked*, 58.
- <sup>31</sup> See: Adamic and Huberman, "The Web's Hidden Order." Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos, "On Power-Law Relationships of the Internet Topology" (paper presented at the *Special Interest Group on Data Communications (SIGCOMM)*, Cambridge, MA, August 1999). Bernardo Huberman, *The Laws of the Web: Patterns in the Ecology of Information* (Cambridge, MA: MIT Press, 2001). Broder et al 2000, Kumar et al 1999,
- <sup>32</sup> David Pennock et al., "Winners Don't Take All: Characterizing the Competition for Links on the Web," *Proceedings of the National Academy of Sciences* 99, no. 8 (2002).
- <sup>33</sup> Hindman, Tsioutsoulouklis, and Johnson, *Googlearchy*, 33.
- <sup>34</sup> *Ibid.*, 26.
- <sup>35</sup> *Ibid.*
- <sup>36</sup> *Ibid.*, 30.
- <sup>37</sup> Trystan Upstill, Nick Craswell, and David Hawking, "Predicting Fame and Fortune: PageRank or Indegree?" (paper presented at the *8th Australasian Document Computing Symposium*, Canberra, Australia, December 15 2003), available from [http://cs.anu.edu.au/~Trystan.Upstill/pubs/upstill\\_adcs03.pdf](http://cs.anu.edu.au/~Trystan.Upstill/pubs/upstill_adcs03.pdf) (accessed May 10, 2005).
- <sup>38</sup> ptsc, "*Scientology*" as a Search Phrase Analyzed by the VisIT Software (operatingthetan.com, February 12 2002 [cited April 15 2004]), available from <http://www.operatingthetan.com/google/>.
- <sup>39</sup> Google's Larry Page contributed to a paper on this subject at Stanford, which recommended using PageRank to guide crawling. See Cho, Garcia-Molina, and Page, "Efficient Crawling Through URL Ordering."
- <sup>40</sup> Barabási, *Linked*, 58.
- <sup>41</sup> Mark Cooper, *Media Ownership and Democracy in the Digital Information Age* (Stanford, CA: Center for Internet & Society at the Stanford Law School, 2003), 16, available from <http://cyberlaw.stanford.edu/blogs/cooper/archives/mediabooke.pdf> (accessed May 15, 2005).
- <sup>42</sup> Introna and Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," 176.
- <sup>43</sup> Page et al., "The PageRank Citation Ranking: Bringing Order to the Web," 10-11.
- <sup>44</sup> In the PageRank paper it is suggested that, in the future, the values could be "personalized" by starting the computation from, say, a user's bookmarks. This would mean that what Google saw as "important" for a user could more closely approximate what he or she has actually identified as important. While this may mean that a greater number of sites would be able to reach users, such personalization has a deliberative dark side. As Sunstein argues in his book, such "daily me" content can allow individuals to be exposed to more self-reinforcing and fewer 'antagonistic' opinions. In some cases—say, to make the point clear, a white supremacist with many racist bookmarks—this could lead to polarization and extremism. For more on this interesting subject, which is unfortunately beyond the scope of our analysis, see Cass Sunstein, *Republic.com* (Princeton, NJ: Princeton University Press, 2001).
- <sup>45</sup> Introna and Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," 177.
- <sup>46</sup> Boris Galitsky and Mark Levene, "On the Economy of Web Links: Simulating the Exchange Process," *First Monday* 9, no. 1 (2004), available from [http://firstmonday.org/issues/issue9\\_1/galitsky/index.html](http://firstmonday.org/issues/issue9_1/galitsky/index.html).
- <sup>47</sup> Page et al., "The PageRank Citation Ranking: Bringing Order to the Web," 15.



- 
- <sup>48</sup> This is one possible area for further empirical research. It would, however, involve enormous computational resources to crawl the millions of pages, to compare the “freshness” of these cited documents and sites, and so forth.
- <sup>49</sup> Rebecca Blood, *Weblogs: A History and Perspective* (Rebecca's Pocket, September 7 2000 [cited May 15 2005]), available from [http://www.rebeccablood.net/essays/weblog\\_history.html](http://www.rebeccablood.net/essays/weblog_history.html).
- <sup>50</sup> The academic community is only beginning to examine the connection between blogs and democracy. One initial, network-theoretic attempt can be found in Daniel Drezner and Henry Farrell, "The Power and Politics of Blogs" (paper presented at the *American Political Science Association (APSA) Annual Meeting*, Chicago, IL, 2004), available from <http://www.utoronto.ca/~farrell/blogpaperfinal.pdf> (accessed May 1, 2004). Of course, bloggers themselves have already taken up the subject. For varied perspectives from prominent bloggers, see John Rodzvilla, ed., *We've got Blog: How Weblogs are Changing our Culture* (Cambridge, MA: Perseus Publishing, 2002).
- <sup>51</sup> John Hiler, *Google (hearts) Blogs* (Microcontent News, February 26 2002 [cited May 6 2005]), available from <http://www.microcontentnews.com/articles/googleblogs.htm>.
- <sup>52</sup> Ibid.
- <sup>53</sup> David Gallagher, "Deal May Freshen Up Google's Links," *The New York Times*, February 24 2003.
- <sup>54</sup> Barabási, *Linked*, 84.
- <sup>55</sup> Ibid., 85.
- <sup>56</sup> Ibid.
- <sup>57</sup> Ibid., 86.
- <sup>58</sup> Ibid., 88.
- <sup>59</sup> E.g., Introna and Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters." Hindman, Tsioutsoulis, and Johnson, *Googlearchy*.
- <sup>60</sup> Hindman, Tsioutsoulis, and Johnson, *Googlearchy*, 27.
- <sup>61</sup> Hindman only makes the case that political sites are biased towards the most popular, and that PageRank mirrors such bias. The link between PageRank and final Google rank is left unaddressed.
- <sup>62</sup> Suppose we limit the set of results to just the pages containing the user's keywords. There will, of course, be many documents that are only of slight relevance, and there is no guarantee that the most popular Web pages are not themselves among these marginally useful pages. *Yahoo!*, for example, may happen to include the words "Southwest Airlines" on its front page, and since *Yahoo!* has a much higher PageRank than the Southwest Airline's home page, sorting merely by PageRank will return *Yahoo!* first. This, of course, does not agree with our expectations.
- <sup>63</sup> In previous years, some have proposed statistical mechanisms for "reverse engineering" a search engine's ranking scheme, but these seem of little use given the greatly increased complexity of search systems such as Google. See Glen Pringle, Lloyd Allison, and David Dowe, "What is a Tall Poppy among Web Pages?" (paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998), available from [http://www.csse.monash.edu.au/~lloyd/tilde/InterNet/Search/1998\\_WWW7.html](http://www.csse.monash.edu.au/~lloyd/tilde/InterNet/Search/1998_WWW7.html) (accessed May 12, 2005).
- <sup>64</sup> This 'experiment' was conducted on December 12, 2004.
- <sup>65</sup> To get an idea for the sort vitriol levied at the Redmond company, simply conduct a Google search for "Microsoft is evil."
- <sup>66</sup> Indeed, Google devotes tens of thousands of servers to harvesting and PageRank computation. See the discussion on "barriers of entry" in Charles H. Ferguson, "What's Next for Google?: The Search Firm Wants to Organize All Digital

---

Information. That Means War with Microsoft," *Technology Review: MIT's Magazine of Innovation*, January 2005.

- <sup>67</sup> At the time of this writing, the toolbar is available at <http://toolbar.google.com>.
- <sup>68</sup> For example, the toolbar values are used by Upstill, Craswell, and Hawking, "Predicting Fame or Fortune".
- <sup>69</sup> *Ibid.*, 2. The hypothesis that the median toolbar PR for pages on the web is between 0 and 1 is based on research such as Barabasi's, which suggests that "up to 90% of pages have few or no backlinks pointing to them." Note, however, that Upstill found a lognormal distribution of toolbar PageRank in their crawls, which would suggest that "normalization and transformation" are being done to the *toolbar* values, and that the median might actually be near 5. Their crawls, however, were small and focused on company sites—the same sort of communities Pennock found to exhibit lognormal link distributions. In addition, small crawls tend to be implicitly biased towards high-PR sites (since they follow links to find new pages, and are thus more likely to come across heavily linked pages). While Upstill et al suggest that the toolbar PR is not a true indicator of actual PR, we believe that there is not sufficient evidence to support this claim. We would need a true "random" sample of pages on the Web, and the toolbar PR for each of these, to know for certain.
- <sup>70</sup> Susan Gerhart, "Do Web Search Engines Suppress Controversy?," *First Monday* 9, no. 1 (2004): "Conclusions", available from [http://firstmonday.org/issues/issue9\\_1/gerhart/index.html](http://firstmonday.org/issues/issue9_1/gerhart/index.html) (accessed May 1, 2005).
- <sup>71</sup> *Ibid.*
- <sup>72</sup> The now-defunct DirectHit search engine used a 'collaborative filtering' technique to achieve similar results. It measured which results users were clicking and elevated the prominence of these sites under the assumption that they were more popular with users.
- <sup>73</sup> Langdon Winner, "Do Artifacts Have Politics?," *Daedalus* 109, no. 1 (1980).
- <sup>74</sup> *Ibid.*
- <sup>75</sup> Abbe Mowshowitz and Akira Kawaguchi, "Bias on the Web," *Communications of the ACM* 45, no. 9 (2002): 57.
- <sup>76</sup> Indeed, as the *New York Times* reports, political candidates and their supporters aware of PageRank are beginning to wage all-out war over securing top spots on the Google search engine. Tom McNichol, "Engineering Google Results to Make a Political Point," *The New York Times*, January 22 2004.
- <sup>77</sup> Dan Bricklin, *Why Small Players Matter* (Dan Brinklin's Web Site, 2002 [cited April 10 2005]), available from <http://www.bricklin.com/smallplayers.htm>.
- <sup>78</sup> This is, in fact, consistent with Kleinberg's work on link authority, in which he asserts that only when the user is "overloaded" by results need authoritative sources come into play. See Kleinberg, "Authoritative Sources in a Hyperlinked Environment," 2.

# 6

## The Transparency Dilemma

How Hidden Heuristics, Commercial Optimizers, and Exclusion Silently Shape Search Results

PageRank is but one piece of Google's enormous puzzle, and many other factors continually, even silently, shape the search engine's results. Some of these are technical means of approximating relevancy; others are due to the behavior of individuals outside the company; still others stem from regulatory, rather than commercial or economic, pressures. These dynamics, needless to say, have broad implications for the view of the Web promoted by the search engine.

The sheer technical complexity of Google is, however, an enormous hurdle to overcome in surmising the search engine's sociopolitical dimensions. What makes this task even more difficult is that we have at our disposal only a very limited picture of how the search engine actually works. The academic papers published by Google's founders, written almost seven years ago, describe the workings of a nascent, simple, and isolated search engine prototype. They do not outline the intricate design of today's Google, of what has in essence become the center of gravity for Web users everywhere (*see* Chapter 8). It is disappointing to find that as the significance of the search engine grew, its degree of openness decreased in tandem. Google, now fully privatized, seems to have completely abandoned its original intention of being "a competitive search engine that is transparent and in the academic realm."<sup>1</sup>

In this chapter, we will focus primarily on three factors that shape Google's search results: relevancy heuristics, rankings manipulation techniques, and the practice of manual exclusion and censorship. Taken together, these examples will shed some additional on the ways that Google's results can be, and are being, formulated. But more importantly, they will

illustrate how Google, and the public at large, faces a dilemma when it comes dealing with search engine transparency.

## I. Beyond PageRank: More Google Heuristics

As indicated in the previous chapter, PageRank is undoubtedly an enormously significant component of Google's search system, is remarkably useless on its own. To the extent that it is a *query-independent* measure of "importance," rather than of actual relevancy, it does little to approximate or model the diverse desires of millions of individual users. Google therefore relies extensively on "100-plus" additional heuristics,<sup>2</sup> including some of those introduced in Chapter 4: keyword counting, location metrics, and anchor text matching. These metrics, unlike PageRank, are *query-dependent*, meaning that they directly compare the user's keywords to the content or meta-text of a particular page.

These techniques more directly approximate "relevancy," but they nevertheless have potentially problematic biases as well. Recall, for instance, how the *location metric* looks at the URL of a page as an indicator of usefulness: "URLs ending with '.com' may be deemed more useful than URLs with other endings"; "a URL containing the string 'home' may be of more of interest than other URLs"; or a URL that contains the user's keywords may be expected to be of greater relevance.<sup>3</sup> As Introna and Nissenbaum observe, "it may therefore be of great significance 'where you are located' as to how important you are seen to be. With the URL as the basis of decision making, many things can aid you ... such as having the right domain name, being located in the root directory, and so forth."<sup>4</sup> Depending on the specific criteria used, the location metric has the potential to favor commercial and well-funded sites over personal and noncommercial pages. For instance, if a greater weight is indeed given to ".com" sites, noncommercial organizations may be disfavored, since they often reside on ".org" domains. Similarly, if a preference is given to root level pages, to domain names that contain the query

---

<sup>a</sup> A page at the "root" directory will have a URL like <http://www.stanford.edu/hello.html>, as opposed to <http://www.stanford.edu/~amd/content/articles/test/hello.html>

words, or to short URLs, those who can afford to buy and host their own domains will be preferred over sources hosted on free or shared services.<sup>b</sup> These concerns are, of course, entirely hypothetical, since we do not know what criteria Google is using for its location metric, or how strongly it affects the results. Anecdotal evidence does, however, confirm at least some of our suspicions.<sup>5</sup>

Brin and Page hint at another salient heuristic in their 1998 paper, in which they state that they “plan to support user context (like the user's location).”<sup>6</sup> Although no elaboration is given, we can assume that Google would be taking a searcher's *geographic* location into account, returning pages of particular interest to users in particular area. Interestingly, in 2000, Google licensed software to automatically geolocate its visitors.<sup>7</sup> At the present time, it appears that Google is only using this information for the purpose of displaying geographically relevant ads.<sup>8</sup> But if it were to start “localizing” its results in the manner described by Brin and Page, the implications for global democratic discourse would be ambiguous. On the one hand, increased coverage of locally salient issues would be enormously useful; on the other hand, extensive reliance on physical location could balkanize users.<sup>9</sup>

In a 2005 “Factory Tour,” a Google executive mentioned that the company was beginning to collect “real-time” information about user behavior, such as the rate at which users were clicking individual results.<sup>10</sup> By statistically identifying which matches for a given query were being visited most frequently, and raising the prominence of uncharacteristically popular pages, the ranking algorithm might be better able to reflect changing perceptions of “usefulness” or “relevancy.” The technique is likely to increase searcher satisfaction, but it is quite similar to PageRank in that it tends to elevate the visibility of relatively popular pages. This, as we have argued, may not be consistent with democratic ideals of the Web.

Beyond the heuristics already discussed—anchor text, URL metrics, PageRank, and so on—we really do not with any certainty what else might be lurking behind the scenes, nor do we fully understand how Google's various heuristics combine to produce the final rankings.

---

<sup>b</sup> The URL for a free page on Geocities might look like <http://www.geocities.com/HotSprings/9831/>.

While there has been some academic work done on reverse-engineering search engines, these methods—an “imprecise science at best”—are only successful at uncovering relatively simple systems.<sup>11</sup> The vast amount of research conducted at Google, its incessant use of testing groups, and the constant tweaks to its algorithms all seem to suggest that its technology anything but “simple” and is, in fact, becoming increasingly nuanced and abstruse. All the while, its workings are hidden from public view, “treated and steadfastly guarded as trade secrets.”<sup>12</sup>

This, of course, should be quite worrisome. Without knowledge of how these heuristics work, we really cannot systematically analyze their effects on online discourse. Our findings thus far suggest that search algorithms do encode bias, and that we should be worried about heuristics that direct users towards large, commercial sites and away from unpopular or underrepresented opinions. We were only able to articulate concerns about PageRank, however, because we happened to have access to important technical documents from Google’s academic adolescence. Without knowledge of the many, new algorithms at play, we might reasonably wonder whether they, too, are manifesting problematic biases. The need to explicate the political implications of these technologies was, after all, the reason why Brin and Page promoted Google as a “transparent” alternative to the commercial search engines.

## II. How to Manipulate Google for Fun and For Profit

But there is another side to the coin. To extent that Google’s algorithms are exposed, individuals can learn to modify their pages for the express purpose of achieving a higher ranking. On the one hand, this empowers well-meaning page authors to more effectively convey what their content is about, and allows interested readers to find these pages with greater ease. But by the same token, knowledge of how the algorithms work also allows malicious individuals to “trick” the ranking algorithm into ranking their pages higher than they deserved to be.<sup>13</sup> As Nissenbaum and Introna explain,

*Out of this strange ranking warfare has emerged an impossible situation: Search engine operators are loath to give out details of their ranking algorithms for fear that spammers will use this knowledge to trick them. Yet,*

*ethical Web page designers can legitimately defend a need to know how to design for, or indicate relevancy to, the ranking algorithm.*<sup>14</sup>

From this perspective, the issue of transparency becomes highly problematic. The threat of “spamming” is quite real, and potentially disastrous. As anyone surfing the Web in the late 1990s may recall, such practices can seriously degrade the quality of search results for everyone. By effectively dominating the results, spamming abuses seriously undermine the ability of search engines to serve as true, egalitarian spaces.

Google made great strides with its early algorithms, which were simultaneously open and (at the time) difficult to manipulate. But eventually people learned that these too can be foiled. Consider, for example, the phenomenon known as “Google Bombing.”<sup>15</sup> Using this technique, one or more authors can escalate the ranking of a particular page for a particular query—without having modified the target page itself, and without the target keywords appearing anywhere in the that page. This may seem rather odd, but given our knowledge of PageRank and anchor text heuristics, it is quite straightforward to understand. Suppose, for example, that you wanted the query “ignorant nitwit” to bring up my home page. What you could do is this: get a bunch of your friends to put a link to my page, and have them all use “ignorant nitwit” for the link text. If you get enough people to go along—especially people who host popular pages—it might just work.

The practice of Google bombing is effective because the search engine’s algorithms do not just look at the actual content of a page; they also count how many people are linking to it, and what terms they are using to do so. Armed with knowledge of Google’s inner workings, site authors have successfully set off various humorous Google bombs: “more evil than Satan itself” returns Microsoft’s page; “miserable failure” links leading to President Bush’s official biography; “weapons of mass destruction” returns a satirical “not found” page. In other cases, Web authors have deployed “Justice Bombs” against companies engaging in questionable practices. After the Critical IP corporation began to telemarket “to domain name owners by stealing their phone numbers out of an Internet database,” a Google Bombing effort was waged, and soon a page critical of the company climbed to the top spot on a query for “Critical IP.”<sup>16</sup>

This technique may seem like a fantastic means for antagonistic voices to be heard, but it can also be used to far more questionable ends. Commercial entities can, for example, place many links to their site that use common query terms. Some sites have even used the technique to bump critics out of top spots. The Church of Scientology, whose complex link manipulations we mentioned in the last chapter, detonated its own Google bomb on the term “Scientology” in the hopes of driving away visitors from two antagonistic sites.<sup>17</sup>

Google Bombing is a relatively rare phenomenon, but the broader practice of trying to manipulate results continues to be a problem for the search engine. Google has fiercely fought back against abuse, repeatedly tweaking its software in an attempt to weed out pages that compromise the integrity of its results. In response to attacks on PageRank, for instance, it routinely downgrades sites that appear on “link farms,” and every month or two it rolls out new versions of its algorithms that counter the latest attempts at ‘gaming Google.’ This “arms race” has meant that “not only is it impossible to ascertain the exact nature of its rules, but it’s also a moving target.”<sup>18</sup>

The fact that Google’s mechanisms are becoming increasingly elusive, of course, makes it even more difficult for legitimate page authors to design their sites to be “search engine friendly,” and to obtain high rankings on relevant queries. Whereas in the old days (that is, around 1996) including “META” tags was often sufficient, with Google, “optimization” requires intricate knowledge of subtle and complicated algorithms such as PageRank. Companies that wish to ensure prominence through legitimate redesigns have thus sought the services of Search Engine Optimizers (SEOs). These experts will, for a fee, redesign a site to maximize PageRank and ensure that Google picks up the right keywords from each page.<sup>19</sup> And at least one headhunting company has hired a full-time employee to work on their site’s search placement.<sup>20</sup> The upshot of all this is that well-financed organizations are able to indirectly purchase higher prominence, whereas more economically-disadvantaged entities (e.g., nonprofits) often lack both the expertise and resources needed to win the rankings game. This further increases the chances that large, commercial sites will dominate Google’s results.



Taken together, Google bombing, manipulation, and optimization illustrate how external agents can and do intentionally affect the results we all see on Google. The dynamics of site ranking consequently extend beyond the search engine itself; they also encompass the behavior of well-meaning and malicious page authors alike. This means is that, if Google wishes to promote deliberative principles, it also needs to account for—and properly attenuate—artificially inflated rankings. This is an enormously difficult problem, since no result heuristic can, it is argued, ever be immune to manipulation. Obscuring how the search engine works can mitigate the potential for externally imposed bias, and in this sense full transparency becomes a liability rather than asset.

### **III. Explicit Exclusion: Hate Speech, China, and the DCMA**

In late October 2002, Ben Edelman and Jonathan Zittrain of the Harvard Law School discovered that “Google had been filtering its own servers to block users in Germany, France, and Switzerland from accessing sites carrying material likely to be judged racist or inflammatory in each country.”<sup>21</sup> In particular, they found that the search engine was manually delisting sites that “seem to offer Neo-Nazi, white supremacy, or other content objectionable or illegal” in those countries.<sup>22</sup> After manually comparing the search engine’s listings for various known sites, Edelman and Zittrain were able to identify 113 specific cases of exclusion (though more probably exist). They speculated that Google was removing these sites “because of pressure applied or perceived by the respective governments,”<sup>23</sup> all of which have bans against hate speech. Unfortunately, these speculations could not be confirmed since, according to a 2003 article in *Wired*, “neither Brin nor anyone else at Google will talk about the preemptive self-censoring moves in Europe.”<sup>24</sup> Even today, the company has made only limited public admissions about such exclusions. Its policies on hate speech are not mentioned anywhere on its Web site, and so users are left completely ignorant about what might be missing from their search results.<sup>25</sup>

In September of 2003, Google found that it was itself the target of censorship, this time by the Chinese government. Worried about “political dissidence” appearing among the results, the Chinese had banned access to Google and AltaVista throughout the country, and were redirecting millions of its citizens to various state-sanctioned search engines instead.<sup>26</sup> Human Rights Watch and others plead with Google to “continue to resist any censorship pressure from the Chinese”:

*[S]earch engines such as Google and AltaVista play a critical role in ensuring the free flow of information to millions of users in China. Chinese users who want to read objective news, and educate themselves on such restricted topics as human rights, Tibet, religion, and the HIV/AIDS epidemic, often rely on your search engines. The Chinese government blocks access to thousands of web sites based on their content. Using Google ... remains one of the best ways to circumvent this censorship, since it permits searches that may turn up restricted information in unexpected locations that have not been blocked.<sup>27</sup>*

But only four days later, Google reached a compromise with Chinese officials, while AltaVista went public with its defiance. Access to Google was restored, but not without some modifications:

*Chinese authorities tweaked the national firewall, making the new Google China different from the site that was turned off. Today, Chinese who use Google to search on terms like “falun gong” or “human rights in china” receive a standard-looking results page. But when they click on any of the results, either their browsers are redirected to a blank or government-approved page, or their computers are blocked from accessing Google for an hour or two ... Did Google help China find or obtain the filtering technology? “We didn’t make changes to our servers,” is all [Brin will] say.<sup>28</sup>*

The degree to which Google is being censored in China—and in particular Google’s involvement in the whole matter—remains a mystery. But the company “seems to be at peace with how it all turned out,”<sup>29</sup> since Google remained accessible to Chinese users and (the cynic would add) because it was able to preserve its advertising revenues from that country. As for the censored content, Brin stated that, in any case, “political searches are not that big a fraction of searches coming out of China.”<sup>30</sup>

All these developments stem from preexisting speech restrictions in various foreign countries, so the subject of explicit censorship may seem of little relevance to users in the United States. But even here, Google has excluded a variety of pages in order to comply with

the highly controversial Digital Millennium Copyright Act (DMCA), a law that hold “service providers” liable if they merely link to unlicensed copyrighted content. A “safe harbor” provision assures that a “service provider shall not be liable” if it quickly removes any “reference or link to material or activity claimed to be infringing.”<sup>31</sup> Google, which falls into the category of a “service provider,” must therefore yank a site as soon as a complaint is made.

And therein lies the problem: results are removed not after infringement has been *proven*, but merely after it has been *alleged*. As many have rightly pointed out, this can have a chilling effect on online speech; it gives copyright holders leverage against search engines, which they can then use to silence critical, possibly legal content.<sup>32</sup> The concerns are, unfortunately, not imaginary. In 2002, after receiving a complaint from the Church of Scientology, Google removed a prominent anti-Scientology site, *Xenu.net*, from its results; the site had, previously, been the first result when users entered “Scientology.” In a letter to *Xenu.net*, Google stated, “Had we not removed these URLs, we would be subject to a claim for copyright infringement, regardless of its merits.” And what, exactly, was the infringing material? Among other things, copyrighted “photographs of Scientology founder L. Ron Hubbard and others juxtaposed with Adolf Hitler.” The real motive—to silence on of its strongest critics—is obvious. By “getting Google to delete them from its mammoth database, the Church hope[d] to remove one of the most obvious ways that Internet users can stumble across the site.”<sup>33</sup>

At first, Google just deleted potential DMCA violations from its index without informing users that some content was removed. But after pressures from several watchdog groups, the search engine began to redirect potentially illegal links to the “Chilling Effects Clearinghouse,” a site operated by various cyberlaw clinics at prominent law schools around the country. According to Larry Page, “It’s a nice compromise. In general, though, few things get removed in this way. It’s not a practical problem.”<sup>34</sup>

All these cases—hate speech in Europe, political speech in China, and copyrighted content in the US—have several things in common. First, they are again examples of how

search engine policies can have a strong impact on the dissemination of antagonistic or political views. Second, they establish that Google does, under certain circumstances, deliberately exclude particular pages from its index. Finally, these cases illustrate how governmental and regulatory pressures, in addition to commercial and economic forces, can restrict the composition of its search results.

#### IV. The Veil of Secrecy

But what is perhaps most fascinating—and certainly most puzzling—about these tales of explicit exclusion is Google’s (at least initial) silence on the subject. While it may be that Google didn’t want to risk jeopardizing its relationship with Chinese officials, it is not immediately clear what Google has to gain by trying to keep so quiet about the exclusion of hate speech or by not immediately divulging that content was being removed in compliance with the DMCA. It is not, after all, as if Google is losing a competitive advantage, or inviting spamming abuses, by “giving away” the fact that it happens to comply with German or U.S. law.

It is likely that Google’s behavior here is not motivated by the specifics of these cases but by the company’s more general presumption in favor of secrecy. At the Google headquarters, visitors and interviewees are generally required to sign a non-disclosure agreement (NDA) upon entering the premises.<sup>35</sup> Google recently fired an employee who wrote about the company’s benefits package on his blog.<sup>36</sup> Even scholars and journalists have found it quite difficult to deal with the company. As Elizabeth Van Covering told us, echoing the difficulties we faced in contacting Google for information<sup>37</sup>:

*they are quite difficult to get hold of, especially if you don't want to speak to the press people. They are extremely concerned about confidentiality, and I think people are prevented by a policy from talking to anyone. I would say they are much more difficult than the other search engines to talk to.*<sup>38</sup>

Given the success of Google’s technology and the millions of dollars being spent by Microsoft and other competitors bent on overthrowing the search giant, no doubt the company believes its extreme confidentiality is necessary to maintain its crucial technological advantage.

These policies, however, may also serve a purpose beyond ‘mere’ competitive maneuvering. It appears that, despite reports of intense and persistent *internal* politics and debate at Google,<sup>39</sup> the company strives quite hard to foster a public image of Google that is entirely stripped of its political dimension. It may be the case that, in order to protect the public perception of the company as an unbiased, objective source of information, the company refuses to *publicly* engage with—or even acknowledge—controversial issues surrounding its policies or technologies. This would, of course, explain Google’s silence on the European exclusions: since any disclosure about the hate speech exclusions in Europe would likely attract significant mainstream media attention and perhaps call into question the apolitical aura of the company, Google keeps quiet. The drive to protect the Google image may also explain why Brin refuses to divulge his political predispositions; he may be worried that, if he did so, “people who don’t care about Google users might start gaming him the way they try to game his search engine.”<sup>40</sup>

At the moment, this strategy seems to have worked. A 2005 report by Pew confirms that Web users continue to place enormous trust in the search engine, and no doubt the public’s confidence in the search engine contributes to its continued success.<sup>41</sup> It is unfortunate, however, that this trust appears not to be derived from transparency and honesty but from a carefully constructed, incomplete image of the search engine. And, as a result, Google’s lack of transparency may again serve to drive users to Google while, at the same time, keeping these same users in the dark about the politics of its search technology.

## V. The Overarching Problem of Transparency

In a *New York Times* column last year, Thomas Friedman wrote, “Google is a little bit like God.”<sup>42</sup> And indeed, many of us have a great deal of faith in the search engine. It has such an amazing ability to divine what we want that it does seem sort of infallible. Even though we really know so little about how results are generated, we may be content to simply believe that

'Google works in mysterious ways.' The search engine may bless some and strike down others, we say, but it always does so for valid reasons. As *Salon* observes,

*For good reason, Google doesn't talk about its ranking mechanisms; if folks knew what Google was doing, the search engine would be easy to trick. But in the absence of information from the company, rumors, theories, and groundless speculation run free. On the Web, Google has taken on the aura of a god – enigmatic, arbitrary, worthy of our fear and our love. Everyone's watching it for signs of anger and of embrace; we know that whatever it does will affect us profoundly, and so people watch it, and they worry.*<sup>43</sup>

Google is, emphatically, not God. Sure, its advances in search technology may earn our admiration and respect, but this does not mean that it should be beyond reproach. Google's search technology is, after all, shaped by a real corporation, governed by real laws, and ultimately developed by real people with particular ends in mind. If we believe in protecting fairness, diversity, and egalitarianism—if we really care about the founding ideals of the World Wide Web and, indeed, of our nation—we must demand that Google's architects do not undermine this vision.

It may, of course, very well be true that Google shapes its technologies with the explicit intent to secure a democratic cyberspace. But it may also be the case that economic and political forces drive it to sacrifice certain principles for practical, political, or economic reasons (as suggested in the previous chapter). On the one hand, the search engine may be generating the results that satisfy our interests, but on the other it may be keeping us from seeing critical antagonistic, diverse, and even unwanted views on a given subject. The fact of the matter is that we simply do not fully know how this complex sociotechnical system is shaping the Web, and is doing so right under our noses. Nor can we, as long as its inner workings remain hidden from view and users are not informed of what kind of content is deliberately omitted from the search results.

But the answer, as we have seen, is not so simple. When it comes to assessing sociopolitical bias in the various heuristics, for example, a lack of disclosure seems to be a necessary evil for protecting against egregious attacks on fairness and the integrity of search results. Similarly, Google seems to believe that ensuring users' confidence necessitates a public

image devoid of much detail or controversy—this despite the fact that commercial search systems are, according to the founders themselves, “inherently biased against the needs of users.”<sup>44</sup> The dilemmas surrounding search transparency suggest that Brin and Page may have abandoned their original vision for a “competitive search engine that was transparent” not because they “sold out,” but because they have come to see this as a contradiction.

## Notes

---

- <sup>1</sup> Sergey Brin and Larry Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine" (paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998), 18, available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 1, 2005).
- <sup>2</sup> Google Inc., *Factory Tour* (Google.com, 2005), Video, available from <http://www.google.com/intl/en/press/factorytour.html> (accessed May 20, 2005).
- <sup>3</sup> Junghoo Cho, Hector Garcia-Molina, and Lawrence Page, "Efficient Crawling Through URL Ordering," *Computer Networks and ISDN Systems* 30, no. 1-7 (1998): 4.
- <sup>4</sup> Lucas Introna and Helen Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society* 16, no. 3 (2000): 17.
- <sup>5</sup> Search engine advertisers in particular have noted that URLs do, in fact, matter. See, for example, *Slashes, URL Length and Google Ranking* (Webmaster World Forums, 2003 [cited October 23 2003]), available from <http://www.webmasterworld.com/forum3/4881.htm>.
- <sup>6</sup> Brin and Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", 14.
- <sup>7</sup> Gary Rivlin, "Google to Offer Companies Ads that Focus on a Region," *The New York Times*, April 15 2004.
- <sup>8</sup> *Ibid.*
- <sup>9</sup> See, more broadly, Cass Sunstein, *Republic.com* (Princeton, NJ: Princeton University Press, 2002), 9,194.
- <sup>10</sup> Marissa Mayer in Google Inc., *Factory Tour*.
- <sup>11</sup> Glen Pringle, Lloyd Allison, and David Dowe, "What is a Tall Poppy among Web Pages?" (paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998), available from [http://www.csse.monash.edu.au/~lloyd/tilde/InterNet/Search/1998\\_WWW7.html](http://www.csse.monash.edu.au/~lloyd/tilde/InterNet/Search/1998_WWW7.html) (accessed May 12, 2005).
- <sup>12</sup> Introna and Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," 173.
- <sup>13</sup> *Ibid.*: 174.
- <sup>14</sup> *Ibid.*
- <sup>15</sup> The discussion of Google Bombing is derived from John Hiler, *Google (hearts) Blogs* (Microcontent News, February 26 2002 [cited May 6 2005]), available from <http://www.microcontentnews.com/articles/googleblogs.htm>, Tom McNichol, "Engineering Google Results to Make a Political Point," *The New York Times*, January 22 2004..
- <sup>16</sup> Hiler, *Google (hearts) Blogs*.
- <sup>17</sup> John Hiler, *Church v. Google: How the Church of Scientology is Focing Google to Censor its Critics* (Microcontent News, March 21 2002 [cited May 6 2005]), available from <http://www.microcontentnews.com/articles/googlechurch.htm>.

- 
- <sup>18</sup> Justin Sanger, quoted in Farhad Manjoo, "The Google Backlash," *Salon.com*, June 25 2003, available from <http://www.salon.com/tech/feature/2003/06/25/google/> (accessed May 21).
- <sup>19</sup> For more on the role of SEOs, see Elizabeth Van Covering, "New Media? The Political Economy of Internet Search Engines" (paper presented at the *Conference of the International Association of Media & Communications Researchers (IAMCR)*, Porto Alegre, Brazil, July 25-30 2004), available from [http://personal.lse.ac.uk/vancouve/IAMCR-CTP\\_SearchEnginePoliticalEconomy\\_EVC\\_2004-07-14.pdf](http://personal.lse.ac.uk/vancouve/IAMCR-CTP_SearchEnginePoliticalEconomy_EVC_2004-07-14.pdf) (accessed May 18, 2005).
- <sup>20</sup> Manjoo, "The Google Backlash."
- <sup>21</sup> Josh McHugh, "Google Sells Its Soul," *Wired*, January 2003, 133.
- <sup>22</sup> Jonathan Zittrain and Benjamin Edelman, *Localized Google Search Result Exclusions* (Berkman Center for Internet & Society at Harvard Law School, October 26 2002 [cited November 2 2004]), available from <http://cyber.law.harvard.edu/filtering/google/>.
- <sup>23</sup> *Ibid.*
- <sup>24</sup> McHugh, "Google Sells Its Soul," 135.
- <sup>25</sup> Zittrain and Edelman, *Localized Google Search Result Exclusions*.
- <sup>26</sup> McHugh, "Google Sells Its Soul," 132.
- <sup>27</sup> Kenneth Roth, *Google, Alta Vista: Resist Chinese Censorship* (Human Rights Watch, September 7 2002 [cited January 10 2005]), available from <http://www.hrw.org/press/2002/09/china0907.htm>.
- <sup>28</sup> McHugh, "Google Sells Its Soul," 133.
- <sup>29</sup> *Ibid.*, 134.
- <sup>30</sup> *Ibid.*
- <sup>31</sup> *Digital Millenium Copyright Act*, 17 US 5 § 512, (October 1998).
- <sup>32</sup> The chilling effect of the DMCA is argued forcefully in Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York, NY: Basic Books, 1999).
- <sup>33</sup> McHugh, "Google Sells Its Soul," 134.
- <sup>34</sup> David Sheff, "Playboy Interview: Google Guys," *Playboy*, September 2004.
- <sup>35</sup> The NDA policy is mentioned in McHugh, "Google Sells Its Soul." In addition, various Google job candidates have related this policy to me as well, noting that you "practically need to sign it at the gate."
- <sup>36</sup> "Bloggers, Beware if Your Boss is Reading," *The Seattle Times*, February 14 2005.
- <sup>37</sup> At various points during our research, we emailed the company for various data, API tips, and general opinions about certain developments. These requested were either refused or ignored.
- <sup>38</sup> Elizabeth Van Covering, Personal Communication (e-mail), May 20 2005.
- <sup>39</sup> At the Gartner symposium, for example, CEO Eric Schmidt recounted one debate at the company. After a business executive proposed some change, "One of the engineers says, 'That's evil.' It was like setting off a bomb in the middle of the table ... They concluded it was, and this poor person was thrown out of the room." This is quoted in Stephen Shankland and Dawn Kawamoto, *Google CEO Defends Privacy Policies* (C|Net News, May 19 2005 [cited May 20 2005]), available from [http://news.com.com/2100-1032\\_3-5713639.html](http://news.com.com/2100-1032_3-5713639.html).
- <sup>40</sup> McHugh, "Google Sells Its Soul," 133.
- <sup>41</sup> Deborah Fallows, *Search Engine Users: Internet Searchers are Confident, Satisfied, and Trusting--But They Are Also Unaware and Naive* (Pew Internet and American Life Project, January 23 2005 [cited May 22 2005]), 12, available from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf).
- <sup>42</sup> Thomas Friedman, "Is Google God?," *The New York Times*, June 29 2003.
- <sup>43</sup> Manjoo, "The Google Backlash."
- <sup>44</sup> Brin and Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", 17.



# 7

## Advertising and “Mixed Motives”

### Exploring Search Engine (Hyper)commercialization

Advertising is, by and large, how the commercialized media make money. Newspapers, magazines, radio, and television outlets provide free or inexpensive content to their readers, listeners, and viewers. In exchange, these outlets sell advertising space to those who will pay a premium for access to their audiences. Advertising is thus a mixed blessing. On the one hand, it makes it viable to disseminate information to a broad audience at a low cost; on the other hand, there is the persistent threat that the wishes of sponsors will subtly work their way into the content itself, narrowing the range of opinions that can be profitably and widely expressed (see Chapter 3). Advertising allows for widespread deliberation, yet it can also undermine objectivity, distort public discourse, and threaten crucial democratic ideals.

These competing forces come strongly into play in the arena of search engines. Because search engines are the primary gatekeepers of the Web, they are important to advertisers for the very same reason they are central to online discourse. To the extent that search engines direct users towards some voices and not others, they can also direct consumers towards particular products and services. Search advertising also has unique qualities that may make it more desirable than “traditional” forms of advertising. First, unlike television, radio, and newspaper advertisements, search engine users can *instantly* go from seeing an ad to actually buying the product. What’s more, because the ads that are displayed are those that are deemed relevant to a user’s query, the target audiences are already highly segmented—a very attractive property to advertisers—and consist largely of users that have explicitly expressed an interest in specific topic or product. Given these advantages, it is not surprising to find one industry report conclude that “worldwide search revenue estimates of \$7B by 2007 are conservative.”<sup>1</sup>

*Advertising Age* has even gone so far as to predict that the 2005 ad revenues of Yahoo! and Google *alone* could “could rival the combined prime-time ad revenues of ABC, CBS, and NBC...a stunning achievement.”<sup>2</sup>

Like most media companies, Google relies heavily on advertising to stay in business. In a recent filing with the SEC,<sup>3</sup> the company reported that advertising is indeed its “principle source of revenue,” accounting for over \$900 million—or over 95 percent—of its gross income in 2003. This enormous dependency on advertising leads us to question what (if any) effect this has on Google, on the integrity of its results, and on the behavior of its users. We can easily conjure up many troubling scenarios that indicate how sponsors’ interests might subtly and inconspicuously work their way into the search engine’s results. Google could, for example, raise the prominence of its advertisers’ pages among the matches it provides to users. Or it could bias its crawlers to have a greater affinity for the sites of its sponsors. If Google does engage in such practices, if those with money pay their way to prominence, if what users see is a vision of the Web skewed by commercial interests, the implications for egalitarian, democratic discourse would be, needless to say, quite negative.

## I. A Survey of Search Engine Advertising Schemes

It is important to first recognize that although the characteristics of search engine use suggest new and unprecedented possibilities for the dissemination of targeted advertising, search companies also face various unique challenges in monetizing their services. These challenges have, in turn, lead to the development of unique, but sometimes problematic, solutions. It is useful to first put our discussion into context by providing a brief overview and analysis of the search engine advertising industry. What we find will allow us to more accurately assess Google’s approach to this thorny issue.

We begin with what is perhaps the most recognizable form of advertising historically used by search engines: the ubiquitous “banner” ad. This is a large, usually animated, clickable image placed above the main content of a particular page. Banner ads can be found on many

pages around the Web, including those of the most popular news and entertainment destinations. Yet there are various reasons why, despite their size and graphical richness, these ads tend to not work well, especially in the context of search engines. First, only a few banners can reasonably be placed on each page, so there is a fixed limit to how many ads of this form a site can sell.<sup>4</sup> Second, due to the consistency with which banner images are sized and placed, users have developed an uncanny ability to unconsciously spot and ignore them.<sup>5</sup> Finally, a banner ad can take quite a while to load (especially for those on modem connections), which decreases the likelihood that the searcher will actually see it. This last problem has particularly negative consequences for search engines. Whereas users of content-oriented sites tend to stick around for a while to actually read the page (and thus have a chance to see its ads), searchers will typically find a result, click on it, and leave the page in but a few seconds—often before the advertisement has even finished loading.<sup>6</sup>

In light of these shortcomings, search engines have increasingly turned to text-based advertisements, which seem to escape many of the limitations of their graphical brethren: they load quickly, many more can be placed on a given page, users aren't as "blind" to them, and people seem to "trust" text information more than graphics.<sup>7</sup> But to the extent that users still identify them as ads, these too may not be very effective at capturing surfers' attention. A 2001 *New York Times* article describes the situation:

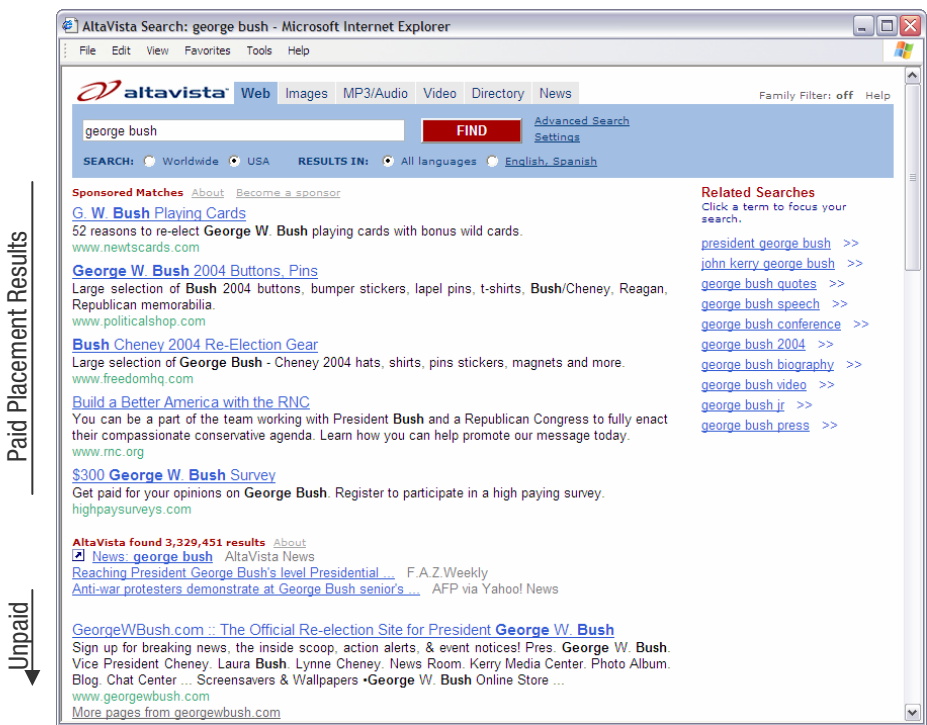
*For a while, advertisers were eager to place their messages above and to the right of these objective search results because they could key their messages to the search terms. But the effectiveness, and thus the prices, for such ads have plummeted as users have trained their eyes to ignore them.*<sup>8</sup>

The inescapable problem here is that searchers usually focus—with "laser beam accuracy"—on what they perceive to be the actual search results, and generally disregard any ads or content at the periphery. Their behavior suggests that, if sponsors wish to be noticed, their solicitations must look like, and appear amongst, the actual results. As the CEO of one search engine company puts it, "The money is in the search results themselves, not the billboards on the side of that road. The question is how do you profit from the search results, when they have been given away for free."<sup>9</sup>

The way many of these sites have gone about “profiting from their results” is by offering various kinds of “paid listings.” In essence, these programs allow advertisers to purchase greater exposure among the results themselves. The most common scheme is called *paid placement* (Figure 7-1):

*With paid placement, advertisers are guaranteed a high placement or top position in search results, usually in relation to specific keywords or a broad range of words. Positioning for paid listings on the page can vary, but they will usually run in three areas: (1) at the top of the page's search results, (2) on the side (usually the right-hand side) of the page, or (3) at the bottom of the page. Sometimes they are clearly marked as “sponsored listings” and segregated from editorial results, and other times they may be hard for the average user to distinguish.<sup>10</sup>*

A second form of paid listings is the *paid inclusion* service, which allows one to pay so that “Web sites or URLs are included in a search engine’s index” that “might otherwise not have been included.”<sup>11</sup> Search engines offering paid inclusion claim that these services will not elevate a site’s rankings; they will merely ensure that the site will get crawled, or crawled more often. Taken together, paid inclusion and paid placement programs give advertisers greater influence during both the indexing and ranking stages of Web search.



**Figure 7-1.** *Paid placement results on the Altavista Search Engine. The top 5 results for the query “George Bush” are sponsored matches. The normal, “editorial” results actually begin towards the bottom of the screen.*

It is not surprising to find that paid listings—and in particular paid placements—have proven to be vastly more effective than previous methods at drawing users’ attention towards sponsors’ sites. To the degree that these “matches” walk, talk, and act like relevant results, users click them. As *Business Week* puts it, paid placements have become “the Holy Grail of Internet advertising, and no wonder.”<sup>12</sup> These ads have caught on, in some form or another, among virtually all of Web’s most popular search engines (Altavista, AOL, AskJeeves, Hotbot, Google, Lycos, MSN, and Yahoo!, to name a few). Overture, the enormously profitable company that “pioneered” paid listings, actually indexes nothing *but* sponsored links. In addition to displaying these “results” on its own site, it licenses them out to a multitude of other search engines; these sites, in turn, display the listings in exchange for a cut of the profits. The demand for paid listings has indeed become so great that, according to *The Economist*, they “are leading the recovery in advertising expenditure on the Internet.”<sup>13</sup>

## II. The Problem with Paid Listings

While paid listings may be a bonanza for search companies, investors, and advertisers alike, their implications for online, egalitarian discourse are depressingly obvious:

*[The] concept that Web sites should be able to buy their way to the top of search listings is being copied in one way or another by every major search and portal site. As they do, the search engines, which are still the most popular gateways to the Web, are transforming themselves from infinite electronic encyclopedias to the more prosaic, if profitable, role of universal commercial directories.*<sup>14</sup>

To the extent that search engines are widely used, and to the extent that the commercial interests of the rich dominate the results of even noncommercial queries, the practice of selling prominence can seriously distort what the Web consists of for millions of users. The Web, as seen through the filter of search engines, becomes less and less like a global, egalitarian public forum described by its proponents.

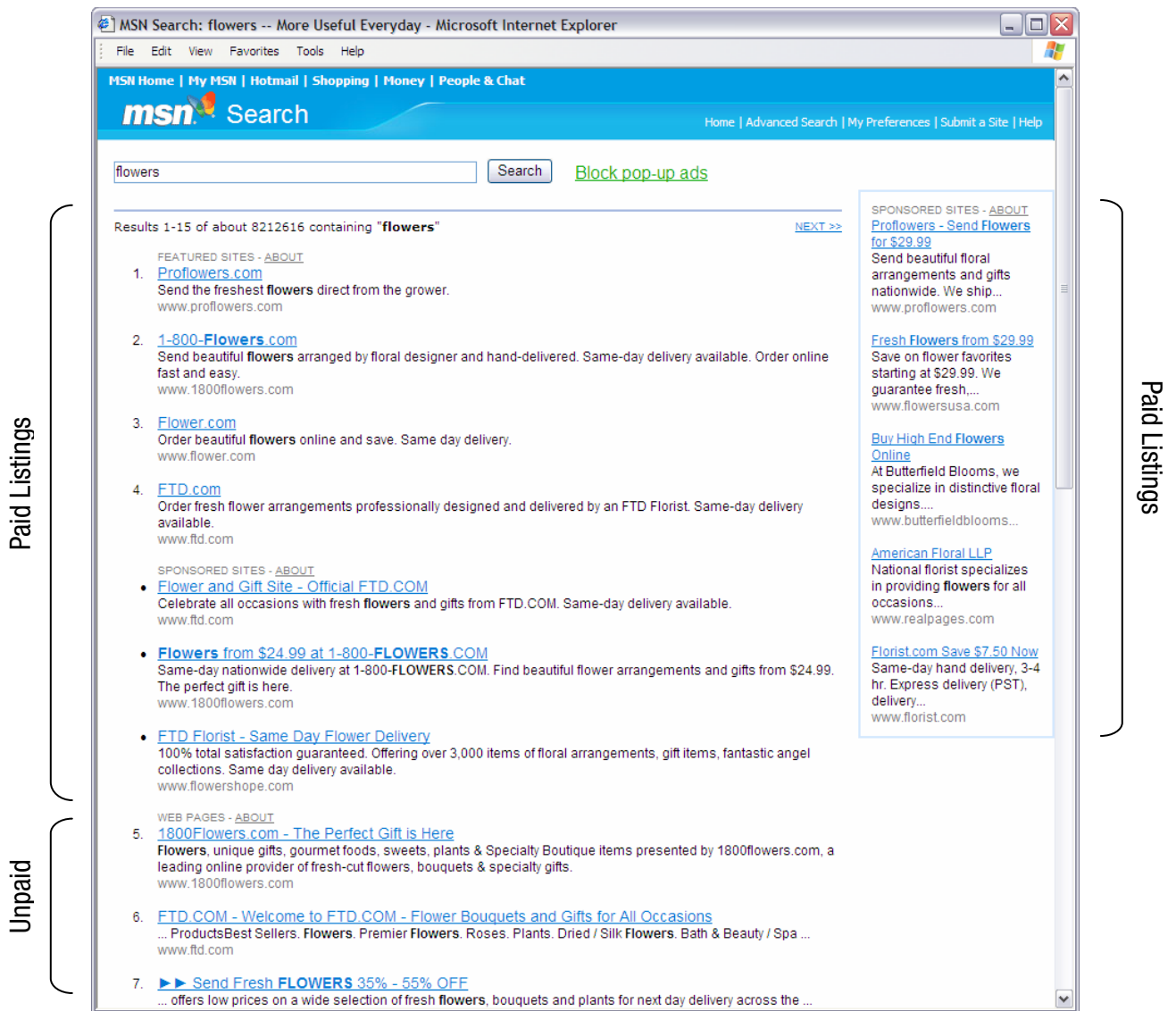
To be sure, just as market forces drive search engines to sell prominence, so too do market forces push back. If, as commercial listings become more numerous, the relevancy of a

search engine's results decline, dissatisfied users may switch to a competitor. If enough users do this, the search engine would soon suffer an overall decline in advertising revenues. From this angle, the amount of paid listings to include is a straightforward optimization problem. Economists Bhargava and Feng respond to it by proposing "a mathematical model for optimal design of a paid placement strategy"<sup>15</sup> that would "give a search engine the best balance between revenues from content providers and revenues based on user base."<sup>16</sup> Such economic models are, however, not very comforting. We might reasonably wonder whether users will actually see what's missing from their search results, or whether the entire search industry will become so utterly swamped with paid listings that there are no real alternatives available. Most disturbingly, the effects of paid listings on discourse in general are chalked up as 'externalities' and wholly ignored.

Bhargava and Feng also assume "that search engines cannot hide the fact that they perform paid placement."<sup>17</sup> In reality, it appears that the majority of Internet users remain completely unaware of such practices. In 2002, a study commissioned by Consumers Union found that *fewer than one in four* Internet users had ever heard of search engines "taking fees to list some sites more prominently than others."<sup>18</sup> After being told that, in fact, most search engines do exactly this, "a solid majority (80%) say it is important for search engines to tell users about their fee details, including 44 percent who say it is *very* important."<sup>19</sup> At the time, these findings suggested that, even though sponsored links were technically being labeled in some form or another, the fact that these were *paid* listings was not being conveyed to users very effectively. Several search engine companies were using—and continue to use—remarkably vague and misleading terminology to demarcate their paid listings (e.g., "Featured Sites," "Products and Services"). It is not surprising that search engines would do this; it is, after all, in their own best interests to at least *appear* noncommercial, objective, and relevant in order to direct as much traffic as possible to sponsors' sites.

Recognizing that paid listings were not being disclosed adequately, in 2001 the watchdog group Commercial Alert filed a complaint with the FTC alleging that seven search

companies were engaging in “deceptive advertising” practices.<sup>20</sup> They argued that “without clear and conspicuous disclosure that ads are ads, concealment may mislead search engine users to believe that search results are based on relevancy alone, not marketing ploys.”<sup>21</sup> In June 2002, the FTC responded with an open letter to various search engines.<sup>22</sup> While it did not call for immediate action against the search engines named in the complaint, it did agree that



**Figure 7-2.** Egregious use of paid listings by Microsoft’s MSN search engine. Here, sponsored results dominate all the prominent spots; unpaid results begin so far down the page that we had to lengthen the browser window just to capture them. If not for our brackets, you might have had trouble spotting where, exactly, the unpaid results actually begin. It appears that MSN intentionally blurs the line between sponsored and “normal” results by demarcating the various kinds of listings using relatively small print and vague terminology (e.g., “Featured Sites”). This image does not even capture the full extent to which MSN is “hypercommercialized.” On this particular query, a large “popup” ad also appeared, and the sponsored links persisted on subsequent pages of the results.

there was a “need for clear and conspicuous disclosures of paid placement, and in some instances paid inclusion.”<sup>23</sup> The “purpose of such a demarcation is to advise consumers as to when they are being solicited, as opposed to being impartially informed.”<sup>24</sup> Although many search engines have since ‘cleaned up their act’ by indicating which results are “sponsored,” some continue to intentionally blur the line between “editorial” and “paid” results<sup>25</sup> (Figure 7-2). It remains to be seen whether the FTC’s warnings about paid placement will, in the long run, have any teeth.

On the issue of paid inclusion, wherein search engines direct their crawlers towards particular sites for a fee (but do not give those sites a rankings boost), the FTC was far less stringent. It did not require that those listings be marked or separated, arguing that

*[t]o the extent that paid inclusion does not distort the ranking of a Web site or URL, many of these programs provide benefits to consumers, by incorporating more Web sites...into an individual search engine...giv[ing] consumers a greater number of choices in search results lists.<sup>26</sup>*

A similar argument is echoed by Yahoo!’s Tim Cadogan: “We find paid inclusion helps drive quality of search results up” since “sites that have deep content, not normally accessed by crawlers, can have their information presented to the users.”<sup>27</sup> But while it may very well be the case that inclusion programs allow search engines to index more of the Web, *charging* for inclusion is another matter entirely, and the practice effectively biases crawlers towards the sites of the economically advantaged. While the spiders are busy going longer and deeper into paid sites, the pages of underrepresented individuals and underfunded organizations—already penalized by crawlers that ignore “unpopular” content—are left out. Crawlers guided do paid inclusion paint a distorted picture of what the Web, a picture that is then passed on to users with no disclosure whatsoever.

But disclosure alone does not solve the problems of paid listings. If we really wish to promote ideals of democratic discourse, then any search policy that allows those with money to buy greater influence with their money should be seen as highly problematic. The fact that paid listings, disclosed or not, make it possible for the economically advantaged to have a greater influence on what Web users see is self-evident; if advertisers did not enjoy a greater



share of search-directed traffic, it is hard to imagine why they would be spending so much money on paid listings in the first place.

Nor is it enough to say, as Yahoo!'s Diana Lee does, that "as long as consumers are getting what they want, that's all that matters."<sup>28</sup> This, too, goes against the very entailments of deliberative discourse. Unexpected, even unwanted, exposure to competing points of view on a subject, especially coming from those who have been traditionally underserved, are of enormous importance for democracy. A search for "Microsoft," or "John Kerry," or "George Bush" should ideally bring up relevant pages from the powerful as well as from the underrepresented. This sort of interaction is exactly what a true, public forum is meant to accomplish; it is why the Supreme Court has defended speakers' access to parks and sidewalks; it is what Justice White means when he states that the First Amendment is meant to secure "the widest possible dissemination of information from diverse and antagonistic sources."<sup>29</sup> If commercial interests can push their way to the top of the results, and if by the same token they penalize those wishing to express antagonistic or unpopular views, ideals of deliberation are not served.

Our concern, it should be emphasized, is not with advertising in general. It is with a particular type of advertising that masquerades as actual, relevant content; it is with advertising that supplants, rather than complements, what individuals might otherwise see. Paid listings, even if disclosed, are not "just like" advertising in the traditional media. Industry reporter Danny Sullivan, however, disagrees:

*Think newspapers. Newspapers have both "editorial" copy, which is not supposed to be influenced by advertising, as well as ads themselves. You may read the paper primarily for the articles, but there are certainly times when you may find the advertisements useful, as well ... In "old" media ... most people can readily identify ads because they look or act so very different from "content."<sup>30</sup>*

But therein lies the problem. In the new media of search engines, paid listings (as opposed to banner ads) *don't* "look or act so very different" from normal results. Search engines with paid listings are hardly like a newspaper with lots of informative, unbiased content and obvious, product-oriented ads sprinkled here and there. They are much more like a newspaper in which

the articles on the front page are all written and paid for by commercial groups. Or like television if all the primetime spots were allotted to infomercials. No wonder, despite Sullivan's claim that users will eventually "learn" to distinguish paid and unpaid content, a 2005 study continued to find that

*While most consumers could easily identify the difference between TV's regular programming and its infomercials, or newspapers' or magazines' reported stories and their advertorials, only a little more than a third of search engine users are aware of the analogous sets of content commonly presented by search engines, the paid or sponsored results and the unpaid or "organic" results. Overall, only about 1 in 6 searchers say they can consistently distinguish between paid and unpaid results.*<sup>31</sup>

This finding, the report adds, "is particularly ironic, since nearly half of all users say they would stop using search engines if they thought engines were not being clear about how they present their paid results."<sup>32</sup>

In some cases, paid listings, like the Yellow Pages, do provide a service for those seeking commercial content. One industry analyst found that that over 30% of search queries were, in fact, seeking commercial content.<sup>33</sup> But here, too, the availability of noncommercial opinion is critical to making informed decisions (consider, for example, the not-for-profit Consumer Reports). But more than that, it needs to be recognized that searchers are not just *consumers*, looking for products, but also *citizens*; that search engines should not be just a *market* for wares but should also serve as unbiased *public forums*. It is obviously difficult to distinguish when a search engine should don the hat of "commercial directory" (serving consumers), and when it should don the hat of "public forum" (serving citizens). But this is not an insurmountable challenge. At least one search engine has already confirmed that it has "spent a lot of R&D" to develop algorithms attempt to distinguish between commercial and noncommercial searches, to ensure that its paid results "only show up under paid queries."<sup>34</sup> Such innovative solutions are what we need to pursue if search engines are to serve the needs of both citizens and of consumers. It is what we need if search engines are to serve democracy, while remaining economically viable.

### III. Advertising on Google

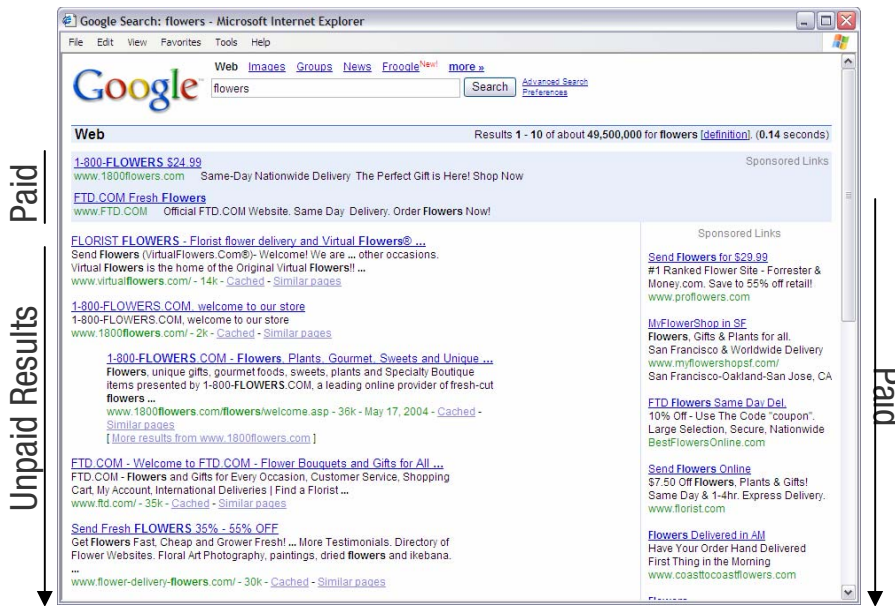
Google's inventors were well aware of these problems since the inception of the search engine at Stanford. To recall their appendix on "advertising and mixed motives" (a smaller except of which, you may recall, was included in Chapter 3):

*Currently, the predominant business model for commercial search engines is advertising. The goals of the advertising business model do not always correspond to providing quality search to users. For example, in our prototype search engine one of the top results for cellular phone is "The Effect of Cellular Phone Use Upon Driver Attention", a study which explains in great detail the distractions and risk associated with conversing on a cell phone while driving ... It is clear that a search engine which was taking money for showing cellular phone ads would have difficulty justifying the page that our system returned to its paying advertisers. For this type of reason and historical experience with other search engines we ... expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers.*

*Since it is very difficult even for experts to evaluate search engines, search engine bias is particularly insidious. A good example was OpenText, which was reported to be selling companies the right to be listed at the top of the search results for particular queries ... This type of bias is much more insidious than advertising, because it is not clear who "deserves" to be there, and who is willing to pay money to be listed. This business model resulted in an uproar, and OpenText has ceased to be a viable search engine. But less blatant bias are likely to be tolerated by the market ... we believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm.<sup>35</sup>*

As noted in Chapter 3, less than one year after they wrote those very words, Brin and Page incorporated the search engine on an influx of venture capital. When CEO Eric Schmidt came on board shortly thereafter, the company began to rely primarily on advertising and, in particular, on its AdWords paid listings program. So what happened to all that talk about "inherent bias"? Do advertisers now dominate its results? Did Google, as on Wired article puts it, "sell its soul"?

First the good news. Google has been emphatic about making objectivity and users' needs its primary concern. It has emphasized its commitment to "keeping user trust and not accepting payment for search results."<sup>36</sup> It includes dramatically ads than Yahoo!, MSN, Altavista, or any of the other major search engines. Most of the sponsored listings it does display (through a program paid placement program called AdWords) are relegated to a box



**Figure 7-3.** Paid placement results on the Google Search Engine. Even on this highly commercial query, only two sponsored listings appear above the editorial results (for most queries, none appear at all). The sponsored links are usually confined to the right.

on the right-hand side of the search results, where they are not as easily confused with the unpaid results. In the rare cases where sponsored links appear directly above the actual search results—and, indeed, appear quite similar to the ‘actual’ results—they are unambiguously labeled as “sponsored,” set apart using a darker background color, and limited to two per page (see Figure 7-3).<sup>37</sup> Interestingly, Google does not directly ‘sell’ this premium advertising space; rather, when one of the ‘normal’ ads to the right of the editorial results receives a great deal of clicks (suggesting it is useful to many users), its algorithms promote the ad, “for free,” to the more prominent position.<sup>38</sup> So, to the extent that Google uses paid placement schemes, it does so with relatively little intrusion.

Moreover, Google has voiced intense opposition to paid inclusion—the practice of accepting payment from a site for more comprehensive or frequent crawling of its content—and has maintained from day one that its “search technology can index the Web without outside help, and that such a program would undermine confidence in its search results.”<sup>39</sup> Indeed, it was for some time the only major search company that steadfastly refused to guide its crawlers towards the sites of its sponsors.<sup>40</sup> MSN and AskJeeves have recently discontinued their paid inclusion programs as well (they say they made this switch to clarify the distinction between its commercial and noncommercial results, but pressure from consumer interest

groups may also have played a part). Yahoo! is the last hold-out, continuing to champion its SiteMatch paid inclusion program and contending that “a direct relationship” with commercial sites allows the company “to ensure that [its] information ... is fresh and relevant.”<sup>41</sup>

Most observers agree that Google is setting the example for responsible search advertising practices by both maintaining the integrity of its ‘editorial’ listings and by keeping the advertising it does carry separate, clearly labeled, and to a minimum.<sup>42</sup> The search engine may have the luxury of doing this, in part, because keyword-search advertising comprises less than half of its total revenue (most comes from its AdSense program, under which authors agree to let Google place ads on their sites for a cut of the profits)<sup>43</sup>. It may also be that Google’s enormous popularity allows it to make more money off of fewer ads. But in any case, in interview upon interview, Google has made it abundantly clear that it is well aware of the problems of hypercommercialism—characterized by McChesney as a continued increase in the level of advertising and conflation of the interests of editors and sponsors—and does its best to avoid them.

Even though the relationship between editorial and paid listings on Google is relatively unproblematic, much controversy has surrounded Google’s ‘bias’ in the selection of advertisements themselves. As Brin admits, “We don’t try to put our sense of ethics into the search results, but we do when it comes to advertising.”<sup>44</sup> The resulting scheme is a patchwork of proscriptions: the search engine doesn’t accept ads for beer, but it does for wine;<sup>45</sup> ads for pornography are fine, but ads for guns are not;<sup>46</sup> you can promote T-shirts depicting the cannabis leaf and drug paraphernalia, but you may not advertise water pipes.<sup>47</sup> Most worrisome, perhaps, is that ads have been rejected because the sponsoring site—or even a page it links to—advocates *against* an individual or group. For example, one entrepreneur discovered that Google rejected his ad since, among other things, he sold a T-shirt that read “Kerry Sucks (Too).”<sup>48</sup> Ads for the nonprofit environmental advocacy group Oceana were similarly rejected because the organization’s site was critical of Royal Caribbean Cruise Lines.<sup>49</sup> The full extent of Google’s seemingly arbitrary advertising standards finally came to

light when, in August 2004, the *San Francisco Chronicle* obtained internal documents detailing the company’s advertising policies. Among other things, these the policies prohibited ads for sites that bashed politicians, gave special scrutiny to ads by the Church of Scientology, and allowed sites to advertise on the keyword ‘abortion’ only if they made no reference to religion.<sup>50</sup> Google’s more recent disclosure about its advertising standards, though less specific, dramatically illustrates its bias away from controversial content and products (Table 7-1).

Content Restrictions																			
<b><i>Anti and Violence</i></b>	<p>Advertisements and associated websites may not promote violence or advocate against a protected group. A protected group is distinguished by their race or ethnic origin, sex, color, age, national origin, veteran status, religion, sexual orientation, or disability.</p> <p>Ad text advocating against any organization or person (public, private, or protected) is not permitted. Stating disagreement with or campaigning against a candidate for public office, a political party or public administration is generally permissible.</p> <p>This standard applies to everyone who wants to advertise on Google, whether we agree with their viewpoint or not.</p>																		
<b><i>Hacking and Cracking Sites</i></b>	Advertising is not permitted for the promotion of hacking or cracking. For example, sites must not provide instructions or equipment to illegally access or tamper with software, servers, or websites.																		
<b><i>Solicitation of Funds</i></b>	Only government-registered charities may solicit funds. Political fundraising is generally permitted.																		
Products and Services																			
<b><i>Banned</i></b>	<table border="0"> <tr> <td>Aids to Pass Drug Tests Marketing Tools</td> <td>Gambling</td> </tr> <tr> <td>Cable Descramblers</td> <td>Hard Alcohol</td> </tr> <tr> <td>Beer</td> <td>Miracle Cures</td> </tr> <tr> <td>Black Boxes</td> <td>Mod Chips</td> </tr> <tr> <td>Counterfeit Designer Goods</td> <td>Prostitution</td> </tr> <tr> <td>Dialers</td> <td>Child and other Non-Consensual Pornography</td> </tr> <tr> <td>Drugs and Drug Paraphernalia</td> <td>Tobacco and Cigarettes</td> </tr> <tr> <td>Fake Documents</td> <td>Traffic Devices</td> </tr> <tr> <td>Fireworks/Pyrotechnic Devices</td> <td>Weapons</td> </tr> </table>	Aids to Pass Drug Tests Marketing Tools	Gambling	Cable Descramblers	Hard Alcohol	Beer	Miracle Cures	Black Boxes	Mod Chips	Counterfeit Designer Goods	Prostitution	Dialers	Child and other Non-Consensual Pornography	Drugs and Drug Paraphernalia	Tobacco and Cigarettes	Fake Documents	Traffic Devices	Fireworks/Pyrotechnic Devices	Weapons
Aids to Pass Drug Tests Marketing Tools	Gambling																		
Cable Descramblers	Hard Alcohol																		
Beer	Miracle Cures																		
Black Boxes	Mod Chips																		
Counterfeit Designer Goods	Prostitution																		
Dialers	Child and other Non-Consensual Pornography																		
Drugs and Drug Paraphernalia	Tobacco and Cigarettes																		
Fake Documents	Traffic Devices																		
Fireworks/Pyrotechnic Devices	Weapons																		
<b><i>Restricted</i></b>	Prescription Drugs and Related Content (pharmacies must be SquareTrade-certified).																		

**Table 7-1.** *Restrictions on Google Advertising.* The search engine’s “sense of ethics” is reflected in the restrictions it places on the content and the commerce of its sponsors. (Google Inc., “Content Policy”).

Yahoo! and MSN, by many accounts, impose far fewer restrictions on the content of the ads they run.<sup>51</sup>

It is not immediately clear what this flagrant bias—based in part on what Brin calls “personal preference”<sup>52</sup>—means from the perspective of democratic discourse. Media scholars like McChesney, Bagdikian, and Mazzocco are, after all, not so much worried about biased advertising standards but by the dissolution of the boundary between editorial content (which is supposed to be ‘objective’) and advertising (which, almost by definition, is not). Google has seemingly adopted a similar position, steadfastly reminding the press, its users, and advertisers that its advertising biases in “no way affect the results [they] deliver”—as if that puts an end to the matter.<sup>53</sup> This is largely consistent with the policies of other media organizations like newspapers which, in Google’s defense, are even more ambiguous and far less transparent about the sort of ads they reject.

Nevertheless, it is not too difficult to see how advertising selectivity may have political and deliberative implications, as Lawrence Lessig suggests in his latest book:

*Say you want to run a series of ads that try to demonstrate the extraordinary collateral harm that comes from the drug war. Can you do it?*

*Well, obviously, these ads cost lots of money. Assume you raise the money. Assume a group of concerned citizens donates all the money in the world to help you get your message out. Can you be sure your message will be heard then?*

*No. You cannot. Television stations have a general policy of avoiding “controversial” ads. Ads sponsored by the government are deemed uncontroversial; ads disagreeing with the government are controversial ... Thus, the major channels of commercial media will refuse one side of a crucial debate the opportunity to present its case.<sup>54</sup>*

By recognizing that advertising may be used as a tool not only for promoting products, services, political candidates but also as medium to voice antagonistic opinions about these subjects, Lessig and other scholars have argued that advertising too may serve as a kind of deliberative forum (though, of course, one largely confined to the well-heeled). It should be further pointed out that, advertising always competes with ‘editorial’ content for the viewer’s attention and, in this sense, the two can never be fully separated. In newspapers or television, radio or the

Internet, whenever editorial content is interspersed with paid content that deals with similar topics, the spectrum of views put forth on that subject encompasses both types of material, for better or for worse.

Taken together, these observations indicate that it is not enough to worry about the diversity of “editorial” content. Rather, effective deliberation requires that we keep as level a playing field as possible in advertising as well. This is partly the motivation behind Senator McCain’s proposal to give political candidates free airtime,<sup>55</sup> and it is precisely why Google itself donates millions of dollars of free ad space to nonprofit groups.<sup>56</sup> It is ironic that the search company recognizes the importance of leveling the *economic* playing field in advertising—making money is the *point* of advertising!—but still refuses to allow many controversial, antagonistic, and critical voices an opportunity to advertise so that they may level the playing field of *opinion* on a given topic. The motivations behind this policy are thus quite unclear. It may be the case, for instance, that Google does not want to be placed in the awkward position of putting its sponsors’ ads (e.g., those for Royal Caribbean) in close proximity to ads for groups (e.g., Oceana) that criticize these same sponsors. Or it may be that the company is worried about trademark, ‘brand disparagement,’ or libel lawsuits if it does accept “anti” ads.<sup>57</sup> Maybe Google, as indicated in the preceding chapter, simply wants to avoid being associated with controversy in general.

In any case, the conclusion remains the same: the more advertisements Google includes under the constraints of this policy—especially if these ads are hard to distinguish from the editorial listings—the more likely it is that users will find mainstream, commercial sites promoting a particular position, product, or service, and the less likely it is that they will hit noncommercial, antagonistic, ‘controversial’ voices. These voices, so critical for deliberative discourse, consequently face a triple bind: they are less likely to appear in the ‘editorial’ listings that tend to suppress controversy (see Chapter 5), they are less likely to have the financial means to buy a prominent advertising spot, and, even if they had the money, their message may not conform to Google’s content standards. The good news, we should reiterate, is that



Google includes a relatively small number of ads, so the more liberal policies dictating its selection of editorial content do tend to dominate.

#### IV. Conclusion

In this chapter, we have observed how, despite the fact that search specific advertising bears comparatively little resemblance to advertising in other mass media,<sup>58</sup> the typical concerns about hypercommercialism continue to apply. We have seen how, with the advent of paid listing and paid inclusion programs, paid content on search engines is becoming more prevalent while the line between advertising and with editorial content is being increasingly blurred. And, as a result, it is possible that well-backed interests are unduly influencing what we see when we go online.

Google, like all search engines, found it necessary to advertise in order to stay afloat. But unlike the other portals, it has hardly invited hypercommercialism in with open arms. Indeed, the primary concern expressed here is not that advertisers are influencing its contents, but rather that it is perhaps excessively dictating what its advertisers themselves can and cannot say.

#### Notes

---

<sup>1</sup> Safa Raschtchy and Jason Avilio, "Industry Note: Search Symposium Shows Bigger Role for Search in Advertising," (U.S. Bancorp Piper Jaffray, 2003).

<sup>2</sup> Kris Oser, "New Ad Kings: Yahoo, Google," *Advertising Age*, April 25 2005.

<sup>3</sup> Google Inc., *United States Securities and Exchange Commission Form S-1* (United States Securities and Exchange Commission, April 29 2004 [cited May 7 2005]), 32, available from <http://www.sec.gov/Archives/edgar/data/1288776/000119312504139655/ds1a.htm>.

<sup>4</sup> This point is made by Danny Sullivan, *Buying Your Way In: Search Engine Advertising Chart* [Web] (SearchEngineWatch, May 30 2003 [cited May 20 2004]), available from <http://searchenginewatch.com/webmasters/article.php/2167941>.

<sup>5</sup> Magnus Pagendarm and Heike Schaumburg, "Why Are Users Banner-Blind? The Impact of Navigation Style on the Perception of Web Banners," *Journal of Digital Information 2*, no. 1 (2001), available from <http://jodi.ecs.soton.ac.uk/Articles/v02/i01/Pagendarm/>.

<sup>6</sup> Ibid.

<sup>7</sup> "A Matter of Trust: What Users Want From Web Sites. Results of Internet Users for Consumer WebWatch," (Princeton Survey Research Associates, 2002), available from <http://www.consumerwebwatch.org/news/report1.pdf>. The point about less 'blindness' is

- 
- made by usability expert Jakob Nielsen in Alan Stewart, "Searching for a Better Spot on the Web," *The Financial Times*, January 15 2003.
- <sup>8</sup> Saul Hansell, "Clicks for Sale: Paid Placement Is Catching On in Web Searches," *The New York Times*, June 4 2001.
- <sup>9</sup> Evan Thornley, chief executive of LookSmart, quoted in *Ibid.*
- <sup>10</sup> Grant Crowell, *The "Secret System" of Search Engine Advertising* (December 18 2003 [cited 2004 April 28]), available from [http://searchenginewatch.com/\\_subscribers/articles/article.php/3289361](http://searchenginewatch.com/_subscribers/articles/article.php/3289361).
- <sup>11</sup> Heather Hipsley, *Re: Complaint Requesting Investigation of Various Internet Search Engine Companies for Paid Placement and Paid Inclusion Programs* [Web] (U.S. Federal Trade Commission, June 27 2002 [cited May 20 2004]), "Paid Placement and Paid Inclusion", available from <http://www.ftc.gov/os/closings/staff/commercialalertletter.htm>.
- <sup>12</sup> Andy Reinhardt, "And You Thought the Web Ad Market Was Dead: Sales of So-called "Paid Placement" Listings Are Soaring.," *BusinessWeek International*, May 5 2003.
- <sup>13</sup> "Spiders in the Web," *The Economist*, May 15 2004.
- <sup>14</sup> Hansell, "Clicks for Sale: Paid Placement Is Catching On in Web Searches."
- <sup>15</sup> Hemant Bhargava and Juan Feng, "Paid Placement Strategies for Internet Search Engines" (paper presented at the *Eleventh International World Wide Web Conference (WWW2002)*, Honolulu, Hawaii, 2002), 122.
- <sup>16</sup> *Ibid.*, 118.
- <sup>17</sup> *Ibid.*
- <sup>18</sup> "A Matter of Trust: What Users Want From Web Sites. Results of Internet Users for Consumer WebWatch," 17.
- <sup>19</sup> *Ibid.*
- <sup>20</sup> Leslie Miller, "How to Net Results in Search-Site Seas," *USA Today*, August 13 2001.
- <sup>21</sup> Commercial Alert, quoted in the FTC's response to their complaint. See Hipsley, *Re: Complaint*.
- <sup>22</sup> David Gallagher, "U.S. Warns Web Sites to Label Sponsorships," *The New York Times*, July 2 2002.
- <sup>23</sup> Hipsley, *Re: Complaint*.
- <sup>24</sup> *Ibid.*
- <sup>25</sup> According to Danny Sullivan of *searchenginewatch.com*, many search sites (including Netscape, Teoma, and Yahoo!) continue to violate the FTC's guidelines on paid listings. See the table at the end of Sullivan, *Buying Your Way In: Search Engine Advertising Chart*.
- <sup>26</sup> Hipsley, *Re: Complaint*.
- <sup>27</sup> Grant Crowell, *How Search Engines Make Money* (SearchEngineWatch, December 16 2003 [cited March 29 2005]), available from [http://searchenginewatch.com/\\_subscribers/articles/article.php/3289341](http://searchenginewatch.com/_subscribers/articles/article.php/3289341).
- <sup>28</sup> Stefanie Olsen, *Are Search Engines Confusing Users?* (C|Net News, October 13 2003 [cited May 20 2005]), available from [http://news.com.com/2100-1024\\_3-5090701.html](http://news.com.com/2100-1024_3-5090701.html).
- <sup>29</sup> Justice White, quoted in Mark Cooper, *Media Ownership and Democracy in the Digital Information Age: Promoting Diversity with First Amendment Principles and Market Structure Analysis* (Stanford, CA: The Center for Internet and Society at Stanford Law School, 2004), available from <http://cyberlaw.stanford.edu/blogs/cooper/archives/mediabooke.pdf> (accessed March 22, 2005).
- <sup>30</sup> Sullivan, *Buying Your Way In: Search Engine Advertising Chart*.
- <sup>31</sup> Deborah Fallows, *Search Engine Users: Internet Searchers are Confident, Satisfied, and Trusting--But They Are Also Unaware and Naive* (Pew Internet and American Life Project, January 23 2005 [cited May 22 2005]), 3, available from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf).
- <sup>32</sup> *Ibid.*
- <sup>33</sup> Raschtchy and Avilio, "Industry Note: Search Symposium Shows Bigger Role for Search in Advertising," 3.
- <sup>34</sup> *Ibid.*, 4.

- 
- <sup>35</sup> Sergey Brin and Larry Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine" (paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998), 17-18, available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 1, 2005).
- <sup>36</sup> Google Inc., *United States Securities and Exchange Commission Form S-1*, 32.
- <sup>37</sup> Danny Sullivan, *FTC Recommends Disclosure To Search Engines* [Web] (SearchEngineWatch.com, 2002 [cited May 20 2004]), available from <http://searchenginewatch.com/sereport/article.php/2164891>.
- <sup>38</sup> AdWordsRep, *Top Position? Impossible?* (Search Engine Watch Forums, November 24 2004 [cited 2005 May 8]), available from <http://forums.searchenginewatch.com/showthread.php?t=2952>.
- <sup>39</sup> Chris Ulbrich, *Paid Inclusion Losing Charm?* (Wired News, July 5 2004 [cited February 10 2005]), available from <http://www.wired.com/news/business/0,1367,64092,00.html>.
- <sup>40</sup> Safa Raschtchy and Jason Avilio, "Silk Road Weekly: Google, Yahoo, and MSN: A Search Update," (US Bancorp Piper Jaffray, 2003).
- <sup>41</sup> Ulbrich, *Paid Inclusion Losing Charm?*
- <sup>42</sup> See, for example, Danny Sullivan, *Paid Content Disclosure Ratings: June 2002* (SearchEngineWatch, July 2 2002 [cited July 22 2003]), available from <http://searchenginewatch.com/sereport/article.php/2164881>.
- <sup>43</sup> Google Inc., *United States Securities and Exchange Commission Form S-1*, 54.
- <sup>44</sup> David Sheff, "Playboy Interview: Google Guys," *Playboy*, September 2004.
- <sup>45</sup> *Ibid.*
- <sup>46</sup> Jeff Johnson, *Google Accepts Porn Ads but Refuses Those for Guns* (CNSNews.com, October 7 2003 [cited July 23 2003]), available from <http://www.cnsnews.com/Culture/archive/200310/CUL20031007c.html>.
- <sup>47</sup> Verne Kopytoff, "Google Ad Rules Complex, Controversial: Documents Reveal Details about what the Popular Search Engine Accepts, Rejects," *The San Francisco Chronicle*, August 9 2004.
- <sup>48</sup> *Ibid.*
- <sup>49</sup> Michael Liedtke, *Google Bans Environmental Group's Ads* (USA Today Online, February 12 2004 [cited June 10 2004]), available from [http://www.usatoday.com/tech/news/2004-02-12-google-bans-ad\\_x.htm](http://www.usatoday.com/tech/news/2004-02-12-google-bans-ad_x.htm).
- <sup>50</sup> Kopytoff, "Google Ad Rules Complex, Controversial."
- <sup>51</sup> *Ibid.*
- <sup>52</sup> Sheff, "Playboy Interview: Google Guys."
- <sup>53</sup> Google Inc., *Google Adwords: Content Policy* (Google, 2005 [cited March 21 2005]), available from <https://adwords.google.com/select/contentpolicy.html>.
- <sup>54</sup> Lawrence Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* (New York, NY: Penguin Books, 2004), 167.
- <sup>55</sup> John Samples and Adam Thierer, *Why Subsidize the Soapbox? The McCain Free Airtime Proposal and the Future of Broadcasting* (CATO Intstitute, Analysis 480, August 6 2003 [cited May 8 2005]), available from <http://www.cato.org/pubs/pas/pa-480es.html>.
- <sup>56</sup> Sheff, "Playboy Interview: Google Guys."
- <sup>57</sup> This latter possibility is documented in Naomi Klein, *No Logo* (New York, NY: Picador, 2002), 176.
- <sup>58</sup> The similarities and differences between advertising in the mass media and search engines are described in detail in Elizabeth Van Covering, "New Media? The Political Economy of Internet Search Engines" (paper presented at the *Conference of the International Association of Media & Communications Researchers (IAMCR)*, Porto Alegre, Brazil, July 25-30 2004), 13, available from [http://personal.lse.ac.uk/vancouve/IAMCR-CTP\\_SearchEnginePoliticalEconomy\\_EVC\\_2004-07-14.pdf](http://personal.lse.ac.uk/vancouve/IAMCR-CTP_SearchEnginePoliticalEconomy_EVC_2004-07-14.pdf) (accessed May 18, 2005).

# 8

## Googleopoly

### Concentration, Conglomeration, and the Future of Search

Thus far, we discussed three key influences on Google’s search results: PageRank, the secrecy shrouding the search engine’s technology, and the practice of advertising. But we have largely left out of our analysis a consideration of how Google’s ownership and market dominance mediates the sociopolitical effects associated with these technologies and policies. It is necessary that we investigate these matters here, since we cannot fully understand the consequences of Google—or, indeed, of any technological development—until we have examined the forces, especially economic forces, shaping how it is designed and adopted.

We will, in this chapter, examine these processes of creation and diffusion—what Robert McGinn calls the “innovation life cycle”<sup>1</sup>—by providing some additional economic and social context. To this end, we will first examine the state of competition and concentration of the search engine industry in general. Tracing the mechanism, character, and speed of Google’s rising dominance, we will then discuss some of the sociopolitical implications of the search engine’s overwhelming rate of adoption. Next, we focus on how control and ownership of the company—an important concern highlighted by various media scholars—may shape its corporate goals and strategic direction, and what effects this may have for the Internet community. Finally, we will take a peek at the future, making some predictions about where Google, and the rest of the industry, is headed.

#### I. Towards ‘Coke and Pepsi’? Consolidation in the Search Sector

Over the past few decades, concerns over media consolidation have reached a fever pitch. McChesney, for example, writes that

*Concentrated corporate control of the media has produced a broadcast journalism that is great at generating profit, pleasing advertisers, and protecting powerful institutions from scrutiny, but lousy at what it's supposed to do: informing the citizenry and confronting abusers of power.<sup>2</sup>*

Consolidation makes perfect sense for media companies because it allows them to leverage economies of scale (via horizontal integration) and to develop mutually-reinforcing, cross-promotional “synergies” (through vertical conglomeration). These routinely increase profit margins, but they are also deployed to raise barriers of entry for newcomers who may not be able to effectively engage in wage price wars or gain access to cross-promotional outlets. Thus, while capitalism’s success often depends on rich competition, “all entrants seek to become a monopoly.”<sup>3</sup> To the extent that a few firms succeed in amassing control of the media, media scholars have argued, the dissemination of diverse and antagonistic views is potentially undermined (see Chapter 3).

Applying these concerns to the field of search engines, we might suppose, as Kawaguchi and Mowshowitz do in their study of variance among the search engines, that “too few intermediaries spells trouble”:

*The only real way to counter the ill effects of search engine bias on the ever-expanding Web is to make sure a number of alternative search engines are available. Elimination of competition in the search engine business is just as problematic for a democratic society as consolidation in the news media. Both search engine companies and news media firms act as intermediaries between information sources and information seekers.<sup>4</sup>*

Unfortunately, the issue of concentration in the search engine industry has received relatively little attention.<sup>5</sup> This is, perhaps, because the nascent, “transitioning” state of the industry—scarcely ten years have passed since the introduction of the first Web search engine—makes it difficult to distinguish long-term patterns from the normal wax and wane of competitors in new markets.

But as the dot-com dust settles, and as consistent, comparative market data become available, a fairly clear pattern of consolidation starts to emerge. Users are, first of all, increasingly converging on a smaller set of search engines. In 1998, each of the top 8 search engines was used by at least 10% of the online audience and, on average, reached about 23% of

all Web users.<sup>6</sup> But today, the top three sites—Yahoo!, Microsoft, and Google—handle over four-fifths of all search traffic,<sup>7</sup> and almost half of Web users frequent a single search site.<sup>8</sup> Thus, whereas users were once distributed across many portals and individually relied on several different search engines, today they stick to a few, overwhelmingly popular sites. (see Figure 8-1<sup>9</sup>).

Not only has traffic begun to center around fewer sites, but over the last ten years, ownership of the various search sites has been consolidated into the hands of a decidedly smaller number of companies. These developments were predicted as early as 1996, when Jupiter Communications, an industry research firm, forecast an imminent “shake-out” in the sector. “There are simply too many players,” they warned investors, “offering similar

### Search Engine Reach (Adjusted\*)

Source: Nielsen//NetRatings for Search Engine Watch

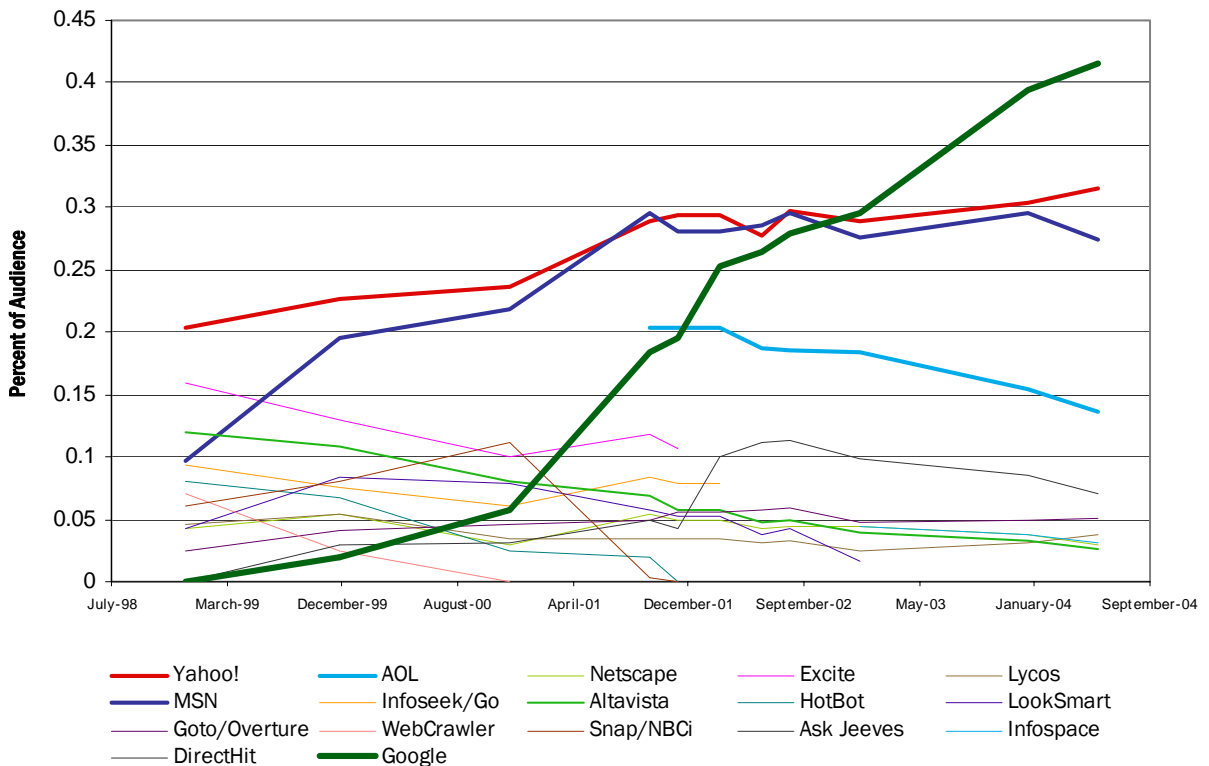
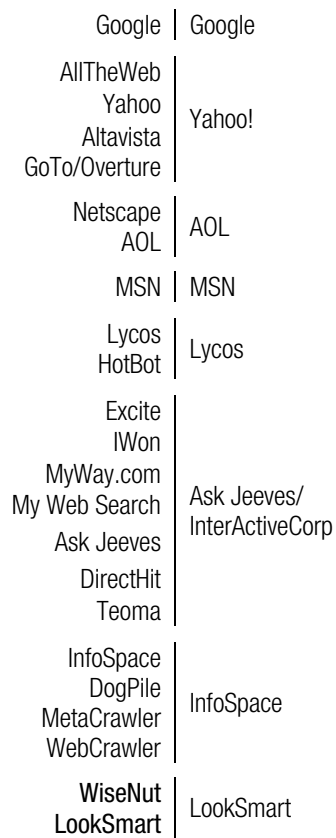


Figure 8-1. Search Engine Audience Reach, 1999-2004. Over the past five years, a smaller number of search engines have risen to receive the majority of Web search traffic.

\*Due to a change in Nielsen’s methodology, we adjusted data after 03/2001 for continuity.

functionality and features, competing for a limited number of advertising dollars and users.”<sup>10</sup> Even Excite’s CEO, George Bell, was pessimistic about the chances for survival: “There are a lot of ‘two’ examples out there ... There’s Pepsi and Coke, Time and Newsweek ... the third always tends to struggle, the fourth tends to get bought. I think [Yahoo and Excite] will make it.”<sup>11</sup> Excite, of course, did not make it. After a steady decline in profitability and traffic, it was ultimately acquired by Ask Jeeves,<sup>12</sup> which also gobbled up DirectHit, Teoma, iWon, MyWay, and MyWebSearch. Yahoo! was more lucky, maintaining a sizeable market share despite the dot-com crash, and became the acquirer as opposed to the acquired. It recently bought technology firm Inktomi and search advertising pioneer Overture—which, in turn, had only months prior bought the AllTheWeb and AltaVista search sites.<sup>13</sup> In the end, while Robin Kellet of MSN UK believes that the “period of consolidation is probably almost over,”<sup>14</sup> it is clear that the search market has been radically transformed (see Figure 8-2). It is no longer composed of a large number of startups but is instead dominated by a few, relatively large corporations.

In light of these developments, it is not surprising to find many referring to the emerging ‘search oligopoly.’<sup>15</sup> And, economically speaking, that characterization seems apt. Under Kaysen and Turner’s determination of oligopoly, for example, “type I” oligopoly is achieved when “the eight largest firms have 50% of receipts and the 20 largest at least 75%”<sup>16</sup>—a threshold easily exceeded regardless of whether we look at ad revenues or traffic share.<sup>17</sup> Applying the more complex Herfiendhal-Hirshman Index (HHI), which “reflects ... the number and size distribution of firms in a market, as well as concentration of the output,”<sup>18</sup> we see that search achieves a



**Figure 8-2. Search engine consolidation.** Today, the search arena comprises far fewer players (right), many of which have acquired competitors (left).

high level of concentration exceeding those in the much-ballyhooed traditional media sectors.<sup>19</sup> For Sheu and Carley, these indicators suggest that “the industry looks close to being plagued by anticompetitive practices.”<sup>20</sup>

Their conclusion is consistent with many economists’ concerns about consolidation in general. Oligopolies are anticompetitive because they provide both the financial incentive and the means to collude, to “charge prices and make profits above competitive levels,” and to raise barriers of entry. “Oligopolists,” writes Gomery, “are mutually interdependent ... [and] work together to fashion positive governmental policies ... Simply put, [they] tend to seek and agree on an informal set of rules for competition, restricting the game of profit maximizing to themselves.”<sup>21</sup> But unlike the oil and music oligopolies (controlled by OPEC and RIAA, respectively), there is not much evidence today suggesting the existence of ‘search engine cartel.’ It appears, for one, that the major players, are not conspiring but are instead embroiled in a winner-takes-all fight to the finish.<sup>22</sup> And as each one tries to outdo the other, they must all watch out for the various newcomers like Acoona.com.<sup>23</sup> There may therefore be more competitiveness in the industry than meets the eye. The situation may be in this respect like the book publishing industry, in which “one can properly lament some concentration...but this market structure was nowhere near as tightly controlled as movies and music.”<sup>24</sup> This is what Gomery rather optimistically calls a “loose and open oligopoly,” wherein a few firms dominate but many participate.<sup>25</sup>

One distinguishing feature of open oligopolies is the presence of relatively low barriers to entry.<sup>26</sup> Given our accustomed, romantic visions of tech innovation—billion-dollar companies sprouting from Silicon Valley garages—this condition may appear satisfied in the context of search engines. Google, after all, began its ascent to the top, and eventually overtook multi-billion-dollar giant Yahoo!, on a ‘mere’ \$25 million dollar investment.<sup>27</sup> And as noted above, competitors continue to appear (although very few been able to obtain a significant market share).<sup>28</sup> It also seems that search engines, unlike operating systems, are highly



'substitutable'<sup>29</sup>: dissatisfied users, we may suppose, need only point their browser to a new site. Thus, we may conclude that the market is, as Compaine argues, "oligopoly proof."<sup>30</sup>

Nevertheless, many are beginning to call into question newcomers' abilities to engage successfully in the search space. First, search companies are finding it hard to keep up with the exponential growth of the Web, which demands highly complex technical systems and enormous expertise to manage:

*Today, the wholesale search market has significant barriers to entry. Economies of scale have asserted themselves, secondary competitors have folded, and the creation of new search engines by startups is becoming prohibitively expensive. Consider: to crawl, index, and search more than eight billion pages—still only a fraction of the Web—Google now operates a global infrastructure of more than 250,000 Linux-based servers of its own design, according to one Google executive I spoke with, and it is becoming a major consumer of electrical power, computer hardware, and telecommunications bandwidth.<sup>31</sup>*

These economic hurdles, according to an executive at Ask Jeeves, are "likely to lead to more consolidation rather than competition from new entrants."<sup>32</sup>

In addition, dominant firms are seeking to maintain market share by differentiating their portals and attempting to reduce the 'substitutability' of their sites. A typical strategy here is the 'bundling' of features such as personalization, email, stock quotes, and driving direction tools (Yahoo!, Google, and MSN are all feverishly implementing such features).<sup>33</sup> In some cases, this has the intended effect of "locking in" users to a particular site; if a user already has, for example, an email account at Yahoo!, then MSN's site will not, in fact, "do just as well." The strategy presents a two-fold challenge for newcomers: they must devote additional development and infrastructure costs to supporting these advanced features, and they must somehow get users to 'defect' from their preferred site (particularly difficult if, for example, this requires switching email addresses).<sup>34</sup>

So while search is still a maturing industry, several trends are clear. First, there is an increase in consolidation: more users are "seeing the Web" through the eyes of a smaller number of search engines. These search engines are, in turn, controlled by a handful of companies, suggesting the possible emergence of a 'search oligopoly.' Finally, there are doubts

about the “openness” of this oligopoly, since “[t]he entry barrier of running a search engine is increasing over time due to the increasing size of the web, and increased switching costs resulted from products (e.g., email) that function to lock in users.”<sup>35</sup> Although the prospects aren’t very promising—to reiterate Mowshowitz and Kawaguchi’s warning, “too few intermediaries spells trouble”—it remains to be seen whether, after the dust has settled, the search industry will be a monopoly, an oligopoly, or a richly competitive industry.

## II. Googleopoly?

Google is, at least for now, perched firm and tall at the top of the search oligopoly. While it was certainly a latecomer in the industry—only 5% of Web users had accessed the site as of December 2000<sup>36</sup>—it now handles almost half of all U.S. Web searches and its users are far and away the most loyal (56% of them use nothing else).<sup>37</sup> The company is also highly profitable, cornering nearly a quarter of the *entire* \$8.7b-a-year online advertising market in 2004.<sup>38</sup> And, in what is perhaps the “ultimate measure of impact” on the public consciousness, its name has become a verb: potential mates “google” each other before a date, recruiters “google” job applicants, and schoolchildren “google” for everything from encyclopedia articles to games for their graphing calculators.

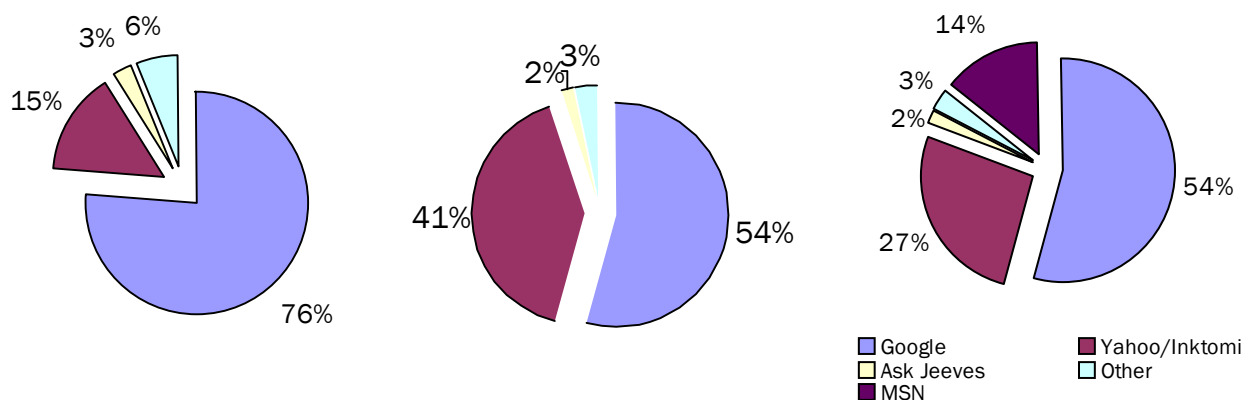
Amazingly, Google rose to the top without the aid of mergers, acquisitions, or even a large advertising budget, and it did so amidst a stock market crash that was decimating the dot-coms. Its success was simply attributed to the quality of its results, a product of its unique and groundbreaking technologies like as PageRank. Google’s algorithmic superiority clearly caught the eye of users, but it also garnered the attention of “portal” operators such as Yahoo! and AOL, who had previously ignored the importance of “good” search and were now clamoring to license Google’s technology to power its own sites.<sup>39</sup> And so, by May 2003, after Yahoo! and AOL outsourced their search technology to Google, the Web’s top three search destinations were all powered by the little company from Mountain View. Taken together, this

meant that its servers were fielding a whopping 76% of all Web search queries performed in the United States.<sup>40</sup>

Google's technology thus had an enormous influence on virtually all online discourse and communication, since its results were shaping what the Web would consist of for the vast majority of Internet users. Its dominance is of great political, social, and economic import, whether one is searching for information about the invasion of Iraq or looking to purchase a digital camera. As Jonathan Zittrain of the Harvard Law School explains, Google had quickly become the "the traffic cop at the main intersection of the information society."<sup>41</sup>

Given our original observations about the media oligopoly, this should be a cause for concern. A dominant intermediary like Google could choose to unfairly wield its power to further a self-interested sociopolitical agenda, much like Robert Moses arguably did with his Long Island bridges.<sup>42</sup> But even without malicious intent, the effects of any negative, antidemocratic bias is intensified in proportion to how widely Google's search technology is diffused.<sup>43</sup> For instance, if users find websites primarily through search engines (they do), if Google handles the vast majority of these search queries (it does), and if the use of PageRank does result in popular, mainstream opinions dominating the search results (*see* Chapter 5), then Google's monopoly could make it considerably difficult for 'ordinary' sites to be seen by a significant population of Web users.

But concern over Google's dominance need not hinge on whether or not the company has illicit motives, or on whether its results are 'democratically' selected. All intermediaries, even the fairest ones, must have biases; they must all somehow choose to elevate some issues, opinions, and voices and to ignore others. When many intermediaries can reach a sizable chunk of the public—each encoding its own opinions about what is interesting, relevant, or valid—these biases can counteract each other and, taken as a whole, a broad array of opinions can be disseminated (in part, through a second step of interpersonal communication<sup>44</sup>). In contrast, when only one or a few outlets have any significant reach, there is enormous inequality in what is transmitted: some views garner lots attention, and those left out receive



**Figure 8-3. Algorithmic Search Share.** Google, whose algorithms fielded the vast majority of search queries by 2003 (left), lost market share when Yahoo! switched to its own search technology in 2004 (center). Yahoo!'s algorithmic share was itself cut down when rival Microsoft switched to its own search technology in 2005 (right). *Source:* Danny Sullivan, "comScore Media Metrix Search Engine Ratings," 2003, 2004, and 2005.

are not heard at all. Consequently, for Bagdikian, it is consolidation—irrespective of commercialization problems—that is the real enemy:

*The threat does not lie in the commercial operation of the mass media. It is the best method there is and, with all its faults, it is not inherently bad. But narrow control, whether by government or corporations, is inherently bad. In the end, no small group, certainly no group with as much uniformity of outlook and as concentrated in power as the current media corporations, can be sufficiently open and flexible to reflect the full richness and variety of society's values and needs. The answer is not elimination of private enterprise in the media, but the opposite. It is the restoration of genuine competition and diversity.<sup>45</sup>*

More search intermediaries means a more equitable distribution of attention, since different search engines do have different biases.<sup>46</sup> It is troubling to find one writer claiming that "so powerful has Google become that many ... view it as the Web itself: if you're not listed on its indexes, they say, you might as well not exist."<sup>47</sup> While the opposite extreme—a highly balkanized audience with little "common ground"—is also problematic,<sup>48</sup> a World Wide Web that consists only of what appears at the top of Google's results is, frankly, a very attenuated sort of deliberative public forum.

Assuaging these concerns slightly, in 2004, Google's power was brought down a notch when Yahoo! stopped using the Mountain View search engine to power its results, opting instead to use the technology of its recently-acquired search developer Inktomi, whose

algorithms were also used at Microsoft's MSN portal. Although this dramatically increased the reach of Inktomi's technology, Google continued to power the majority of Web searches, including those performed by millions of AOL users.<sup>49</sup> When Microsoft debuted its own technology in 2005, algorithmic control was further diversified, though Google still handles the lion's share of search queries<sup>50</sup> (see Figure 8-3). None of these developments, of course, negate the proposition that this is a highly consolidated industry, one that should continue to draw the attention of media observers and concerned citizens alike.

### III. Google Ownership, from Academic to Private to Public

When it comes to the effects of media consolidation, ownership—who funds and controls a company—is as important as market share. “Media ownership equals power,” writes Mazzucco, “and always has.”<sup>51</sup> The owners of the media can, and often do, brandish their companies as tools for achieving particular financial or political ends. In some cases, they may push sensationalist coverage in order to increase profits; in other cases, conglomerates may use their agenda-setting power to influence policy makers. It is therefore of enormous political, economic, and social consequence how ownership is structured not only in the sector as a whole, but within a specific organization as well. Nonprofit, private, or public organizations can differ dramatically in their objectives and the compromises they are willing to make for financial gain. Google, having evolved from an academic project to a private company to a multi-billion dollar publicly-traded institution, is positioned the very heart of these issues.

In its early days, Google was just a collection of makeshift computers messily stored in the basement of the computer science building at Stanford University.<sup>52</sup> But the founders expressed a strong, almost ideological, conviction to keep it that way. Aware that “the issue of advertising causes enough mixer incentives,” they felt it crucial that Google be “a competitive search engine that is transparent and in the academic realm.”<sup>53</sup> Cynics may see their almost immediate turn to the private sector—in the midst of a dot-com boom, no less—as indication

that its founders abandoned these beliefs in favor of financial reward. Perhaps. But it must be emphasized that, by 2000, running a “competitive” search engine was already an extremely expensive endeavor. In order to keep up with demand, the project would require millions of dollars worth of computer equipment, and somehow have to afford ongoing personnel, bandwidth, and maintenance costs. These sorts of expenses are difficult for any academic institution or research endowment to justify, since it is hardly the most cost-effective way to develop new, publishable research.

Indeed, while at Stanford, Brin and Page had to come up with creative ways of staying operational. Without sufficient funds to purchase equipment, they found themselves engaging in guerilla-style computing: they apparently got in the habit of taking computers off the loading dock—ostensibly to help ‘set up’ the machines for their recipients—and used them to power Google until a new shipment came in (at which point the ‘old’ machines would finally be delivered to their intended users). Google was, in the words of one employee, “floating” on the resources of the Stanford CS department.<sup>54</sup>

As a private company infused with venture capital, Brin and Page now had at their disposal the financial resources necessary to pursue their technological objectives—to, in their words, “make the world a better place.”<sup>55</sup> But there were strings attached. Unlike the academic project from which it sprung, Google, Inc. had to worry about profitability. It introduced text ads, which despite their “unobtrusiveness” still constituted a clear and unambiguous breaking of the founders’ noncommercial promise (see previous chapter). By 2001, the extraordinary revenues generated by search advertising made the three-year-old company one of the few tech startups still in business and turning a profit<sup>56</sup> (in contrast, it took Amazon.com seven years to only barely climb out of the red<sup>57</sup>). This had the laudable side-effect of giving the founders considerable freedom to innovate, and even to eschew profit potential, for the public good. Google deployed many new, useful, and usually free services—even before the company could figure out how it could make money off of them.<sup>58</sup> The company was applauded for providing “excellent employee benefits” which included healthcare for domestic partners and

generous maternity and paternity leaves (it even began delivering meals home after the baby's arrival).<sup>59</sup> And, while other companies were sneaking paid listings into its results, Google refused to do so.<sup>60</sup>

But everyone knew that Google's private days were numbered. The venture capitalists and "angel investors" who retained a large stake in the company—among them, Arnold Schwarzenegger and Tiger Woods<sup>61</sup>—were antsy to cash out, and the public was eager to buy a piece of the "the next Microsoft."<sup>62</sup> And so, in 2004, Google filed for an Initial Public Offering (IPO).<sup>63</sup> This shift from private to public ownership is could mean a dramatic shift in the way a company does business, for better or for worse since

*private and public corporations may operate differently. Public companies generally have access to greater financial resources, and can therefore be more flexible choosing to expand, contract or stay pat. On the other hand, public companies have to report on their performance every three months ... encourag[ing] a greater emphasis on short-term profit maximization. Private companies ... can ignore outsider advice and look to the long run ... [and] may be able to pursue editorial objectives without having to answer to a stockholder-voted board.*<sup>64</sup>

During its period of private ownership, Google seemed to enjoy "the best of both worlds": it had a relatively large degree of economic *and* "editorial" autonomy. For the most part, Brin, Page, and finally Schmidt had to answer to very few, and shied away from informing the public about financial, technological, or strategic developments. Such secretiveness obviously has a dark side (see Chapter 6), but it allowed the company to make some of the laudable decisions mentioned above, without having to defend itself at stockholder meetings or in the *Wall Street Journal*.<sup>65</sup>

The move to public ownership could, as Gomery suggests, require that profit be put front-and-center. Google would, by law, have a fiduciary responsibility to act in the economic interest of its stockholders, and would be expected to continually post sizeable profits. When it comes to media companies, these demands are often achieved in familiar ways: through concentration, conglomeration, and hypercommercialization. And it is precisely these practices, McChesney reminds us, that bring the economic aims of public media companies in

conflict with the interests of citizens, the needs of consumers, and the aspirations of a democratic society.

As a public company, Google could feel increased pressure to roll out more insidious, obtrusive, and “editorialized” advertising of the sort unapologetically promoted by Yahoo! and Microsoft, both public companies (see Chapter 7). But this change in ownership could also threaten through conglomeration. Media giants would, needless to say, find a property like Google to be of enormous value, helping these firms extend their vertical and horizontal empire into the new frontiers of cyberspace. Such an acquisition would be entirely consonant with their broader and highly profitable drives for “synergy”:

*[T]he pressure to become a conglomerate is ... stimulated by the desire to increase market power by cross-promoting and cross-selling media properties or “brands” across numerous, different sectors of the media that are not linked in the manner suggested by vertical integration. ... “When you make a movie for an average cost of \$10 million and then cross promote and sell it off of magazines, books, products, television shows out of your own company,” Viacom’s Redstone said, “the profit potential is enormous.”<sup>66</sup>*

“Synergy” was the motivation behind the 2005 acquisition of publicly-traded Ask Jeeves by InterActiveCorp, which sought to leverage the search engine in order to “cross-promote IAC services such as Expedia, Hotels.com, Evite and Lending Tree.”<sup>67</sup> The goal, in the words of an Ask Jeeves executive, was to create “a traffic echo system, sending traffic back and forth between IAC sites, from Ask Jeeves to our brands and vice versa.”<sup>68</sup>

Given the economic promise of cross-promotion, it is not surprising to find the traditional media getting in on the action as well. NBC (with Snap and Xoom), Disney (with Infoseek and Go), and Time Warner (with, of course, AOL) have over the last few years made forays into the portal market, with varying degrees of success.<sup>69</sup> This affinity between portals and the traditional media is, however, often mutual. “[P]ortal sites,” in the words of one author, “more and more taken over by major media companies ... see this as a key strategy” for success.<sup>70</sup> As early as 1997, one industry magazine was even wondering whether survival was possible without the help of a media giant: “Can Excite make it without a deep-pockets old-media partner like Bertelsmann?”<sup>71</sup> Another observer concludes that “[t]here is every reason to



predict that there will be a continuation of mergers and alliances of the older and the newer media companies.”<sup>72</sup> As a private company safely in the black, Google’s leaders were able to safely avoid a buyout (it was apparently received offers Microsoft, among others<sup>73</sup>). But as a public company, stakeholders would have trouble resisting the stability, efficiency, and inflated stock prices associated with takeover bids. This is cause for concern since, if history is any guide, the media giants would be more than happy to leverage Google for commercial gain by biasing its ranking algorithms and hypercommercializing its content. This would, as we have stressed, threaten to undermine those romantic ideals of the Web as a “democratic space.”

But in Google’s IPO filing, Brin, Page, and Schmidt exhibited a remarkable concern about these issues. In a letter to its investors, they write:

*[T]he time has come for the company to move to public ownership. This change will bring important benefits for our employees, for our present and future shareholders, for our customers, and most of all for Google users. But the standard structure of public ownership may jeopardize the independence and focused objectivity that have been most important in Google’s past success and that we consider most fundamental for its future. Therefore, we have implemented a corporate structure that is designed to protect Google’s ability to innovate and retain its most distinctive characteristics.*<sup>74</sup>

Restating their dual goals to remain profitable and to “make the world a better place,” the founders describe at length how the company has no intention of sacrificing users or employees in order to post short-term profits. Recognizing that “the media and technology industries...have experience considerable consolidation and attempted hostile takeovers,” they “set up a corporate structure that will make it harder for outside parties to take over or influence Google.”<sup>75</sup> This structure is a two-class voting system, also used by the *Washington Post* and *The New York Times*, wherein “common” stock gets one vote per share and “Class B” stock—that held by Google employees, including Brin, Page, and Schmidt—get ten votes per share. This ensures that even if founders sell a sizeable portion of their 30-percent stake, they will retain majority control over the company, making hostile takeovers and outsider control of the company virtually impossible.

It is difficult to overstate the impressiveness of Google’s approach to going public. Whereas many startups could not wait to “go public” and “cash in,” Google’s founders have been almost apologetic about the need to transition to private, and then public, ownership. They have long been acutely aware of the problems of the commercial media (even citing Bagdikian in their original Stanford paper!). While Yahoo! and Ask Jeeves have been teeming with excitement about the potential for synergies, commercialization, and advertising, Google’s founders have struggled to strike a compromise between profitability and the public good—even as they turned, overnight, into multibillionaires.

#### IV. “Don’t Be Evil”: Enlightened Despotism in the Digital Age

In the 18th century, Frederick II of Prussia led his country through three wars, passed the first German legal code, expanded compulsory education, and deployed a more efficient bureaucratic system. He was an eminent *philosophe*: not only did he support the arts and sciences, but he himself contributed several works. Frederick was a sympathizer of the American Revolution, and is remembered for instituting religious tolerance, granting freedom of the press, and abolishing torture throughout the kingdom.<sup>76</sup> Although recent scholars have complicated this romantic image of the king,<sup>77</sup> to this day he continues to be admired as a military genius, and is credited with triggering the modernization of Germany.<sup>78</sup>

Frederick was an absolute ruler, of course, but he was convinced that his role was to act as “a servant to the state.” In his *Essay on Forms of Government*, he wrote:

*The sovereign is attached by indissoluble ties to the body of the state; hence it follows that he, by repercussion, is sensible of all the ills which afflict his subjects ... There is but one general good, which is that of the state, [and] the sovereign represents the state ... it is his duty to see, think, and act for the whole community, that he may procure it every advantage of which it is capable ... [H]e ought ... to be the last refuge of the unfortunate; to be the parent of the orphan, and the husband of the widow.*<sup>79</sup>

Frederick the Great is, perhaps, the personification of one strand of Enlightenment thinking (favored by Kant and Voltaire) which held that an “enlightened despot,” rather than a

democratic polity, would be most fit to secure individual liberty and promote the public good. This idea, of course, has its origins at least as far back as Plato, who famously preferred the philosopher-king to rule by the masses.<sup>80</sup>

Two hundred and fifty years later, reading over Google's IPO filing, the analogy is almost irresistible. In their "Letter from the Founders," Brin and Page state that they "aspire to make the world a better place" by "improv[ing] the lives of as many people as possible." Because "searching and organizing all the world's information [is] an unusually important task," it "should be carried out by a company that is trustworthy and interested in the public good ... Google therefore has an enormous responsibility." Their dual-class structure—which gives majority control to the Brin-Page-Schmidt triumvirate—"helps ensure that this responsibility is met."<sup>81</sup> The founders, in other words, believe that they can best serve the public if they have broad powers to pursue their own, presumably laudable, objectives. It is no surprise, therefore, to find many within and outside the company refer to Google's corporate governance as a "benevolent dictatorship."<sup>82</sup>

While critics have characterized these statements as "arrogant,"<sup>83</sup> the founders can hardly be attacked for at least espousing ideals of fairness and objectivity, for understanding their function as information intermediaries, and for expressing a strong desire to act in the interest of the public good. After years of tanking dot-coms, corporate scandals, and laser-focused pursuit of profit by media moguls, it is a great comfort to many to find a prominent company controlled by such apparently "enlightened" figures. And yet, we should not be so quick to beatify them. Even amidst public outcries over Google's biases, its advertising policies, and its threats to privacy, Google has remained notoriously secretive, asking users to trust their judgment, often on blind faith, or go elsewhere. This fits well with Kant's famous characterization of 'enlightened' despots: "Argue as much as you will, and about what you will, but obey!"<sup>84</sup>

With such enormous power and relatively limited accountability, we might be quick to question whether, regardless of their initial aspirations, the company's 'integrity' will not

compromised. What if profitability begins to decline? What if investors complain about the dual-class structure? What if the company faces increased competition from Yahoo!, Microsoft, firms less troubled by Google's ethical dilemmas? What if the founders' multi-billion-dollar fortunes are put into jeopardy? Will Brin and Page have to make more uncomfortable compromises to stay alive? And if the founders leave the company—a risk highlighted in the IPO itself—what then?

For Rousseau, these questions go to the heart of the problem of “enlightened absolutism.” “When someone is brought up to command others,” he wrote, “everything conspires to rob him of justice and reason.”<sup>85</sup> More recently, and in the context of media consolidation, Bagdikian echoes these concerns:

*Today the integrity of news and other public ideas depends on corporate self-control, on the hope that the large corporations that now control most of the media will never use that power as an instrument to shape society to their liking. The history of those who hold great power inhibited solely by self control is not reassuring. It was the morbid record of absolute power left to its own devices that lead to the formation of democracies in general and in the United States in particular.*<sup>86</sup>

Indeed, in light of the failure of enlightened despotism, the views of Hobbes, Plato, Kant, and Voltaire eventually gave way to those of Rousseau, Paine, Locke and Mill. The latter thinkers defended democracy on the grounds that it was the only realistic and sustainable means of governing for the “public good.”<sup>87</sup>

The problem with companies, however, is that there is no equivalent of a “democracy,” anyway. “Public” companies—ones ruled by “shareholder democracy”—do not act in the interests of the public but rather in the interests of their stockholders. Through our elected representatives we can, of course, impose regulations on companies in order to secure the public good (indeed, the FTC has already mandated that search engines label paid listings as such). Regulation, however, can only go so far, especially in this post-Reagan, lobbyist-influenced era. And so, a benevolent dictator model seems to be the only one in which the interests of the public might stand a chance against the interests of profit. But there is still something very troubling about leaving such a critical Web institution to three, enormously

wealthy individuals. How can we, as the public, be sure that Google is acting in our best interests? What do Brin and Page think 'the public good' is about? What will they promise to do and not do? How, in short, can they be held accountable?

Companies seeking to mitigate such concerns often spell out their objectives in the form of a corporate code of conduct. "General Electric," for example, "devotes 15 pages on its Web site to its integrity policy" while "Nortel's site has 34 pages on guidelines."<sup>88</sup> In contrast, Google's ethical imperative is commonly encapsulated by the aphorism "don't be evil."<sup>89</sup> When *Wired* asked Schmidt what, exactly, this means he responded, in the simplest way possible: "Evil is whatever Sergey says is evil."<sup>90</sup> And what, we may ask, does Sergey say is evil? Well, like a journalist who refuses to acknowledge his personal political orientation, Brin keeps quiet:

*He tells me he listens to NPR on his morning drive to work. I think Democrat and ask about his voter affiliation. He says he votes across party lines. Independent? He smiles and tells me there's no easy shortcut toward figuring out how he comes to his decisions about good and evil. And even if there was, he wouldn't let me in on it. If I succeed in figuring out what exactly he considers good and evil, people who don't care about Google users might start gaming him the way they try to game his search engine.<sup>91</sup>*

The "whatever Sergey says is evil" line quickly became associated with the opaqueness and vagueness of Google's policies. So, in a 2004 interview with *Playboy*, Brin retorted by saying that the quote was "not one of [Schmidt's best]."<sup>92</sup> And, when pressed to provide an alternative definition, one that characterized "evil" in more precise terms, he explained how there was "no hard and fast rule." The founders, he said, "collect input," but ultimately they "have to make a decision." Although this often comes down to "personal preference," Brin believes they "do a good job of deciding."<sup>93</sup> Benevolent dictatorship, indeed.

To be fair, Google has made some attempts at more precisely spelling out their sense of corporate ethics. As early as 2001, Google had posted a page on its web site clearly stating that it would not accept money for prominent placement among its search results (e.g., paid placement).<sup>94</sup> In their IPO, they suggested that not being evil meant doing "good things for the world"—like keeping its results "unbiased and objective"—"even if we forgo some short term gains."<sup>95</sup> And more recently, in July 2004 (right around the time Google filed for its IPO), a 14-

page “Code of Conduct” finally appeared on its website. Available to the public but directed towards its employees. it goes:

*The Google Code of Conduct ... isn't merely a set of rules for specific circumstances but an intentionally expansive statement of principles meant to inform all our actions; we expect all our employees ... to study these principles and do their best to apply them to any and all circumstances which may arise.*

*The core message is simple: Being Googlers means striving toward the highest possible standard of ethical business conduct. This is a matter as much practical as ethical; ... our most important asset by far is our reputation as a company that warrants our users' faith and trust. That trust is the foundation upon which our success and prosperity rests, and it must be re-earned every day, in every way, by every one of us.<sup>96</sup>*

Unfortunately, the code says very little by way clarifying how the core values of the Web—espoused by Brin and Page in their 1998 paper—will be upheld. It details, among other things, Google’s nondiscrimination policy, its accounting guidelines, its satisfaction of export controls, and even its stance on allowing dogs in the workplace (it’s encouraged). While topics such as “Avoiding Conflicts of Interest,” “Preserving Confidentiality,” “Protecting Google’s Assets,” and “Obeying the Law” dominate the Code and are covered in great detail, only a half-page is devoted to the subject of “Serving Our Users.” This section merely states the obvious: Googlers should make their products “useful,” should be “honest” and “responsive,” and should not be afraid to “take action.”<sup>97</sup>

Of course, even if we knew *what* principles really guide Google’s practices, we have almost no way of confirming whether these principles are being followed. How can we, for example, assure ourselves that new voices can “rise to the top” of Google’s search results? How can we be sure that paid placement does not occur? As Naomi Klein argues in her book *No Logo*, companies often fail to follow even their most explicit codes of codes of conduct—especially when there is no independent oversight.<sup>98</sup> Nike, for instance, famously kept the locations of its overseas factories under wraps in order to prevent activists from exposing working conditions at its “sweatshops.” Given our discussion of Google’s “transparency dilemma” (chapter 6), it does not seem to unreasonable to question whether, when the going gets tough, Google’s corporate responsibility will be held under proper scrutiny.

And so, the implications of a Google’s “benevolent dictatorship” are ambiguous at best. On the one hand, it can be argued that the triumvirate, rather than a body of profit-seeking shareholders, is best able to govern the company in an “enlightened” manner. This notion is supported by the leaders’ outstanding recognition of the problems of media commercialization, their rejection of short-term profit-seeking, and their insistence on “editorial integrity.” At the same time, it is quite troubling that the company’s stance on so many crucial issues—regarding advertising, ‘democratic’ values, censorship—are decided by a few powerful figures who insist on keeping their policies secret. Google claims, quite convincingly, to take core ethical principles to heart, but we neither know what these principles are, nor can we confirm whether they are being followed. This may serve to shelter Google from criticism and “political” controversy, but it is important to keep in mind these *are* political issues, ones that have enormous consequences for the Web, and ones that Netizens should be addressing. Anything else would be, for lack of a better word, undemocratic.

## V. What Next? The Future of Search

The concerns we have expressed thus far are, of course, entirely hypothetical; they merely suggest how and why Google *might* “turn its back” on the democratic ideals of the Web, the same ideals its founders espoused less than a decade ago. There is thus an obvious retort to everything we’ve been saying: if Google does betray the values of the Web and the needs of its users, the quality of its product will decline, and its users will just switch to another search engine. After all, it’s so easy: users can just type in a different URL or, with a few clicks of the mouse, can change their browser’s start page. We should, in other words, should “just let the market decide.”

But there are a few problems with this argument. First, the complexity and opacity of search engine technology makes it almost impossible for users to notice what is ‘missing’ from their search results. And so, in general, users are likely to continue to use whatever

search engine they have already used frequently.<sup>99</sup> In addition, as noted above, bundled services such as email tend to “lock” users to a search engine, and Google is presumably no exception. One unique way Google is building brand loyalty is through tight integration into the major browsers. Mozilla’s Firefox and Apple’s Safari—the second and third most popular browsers, respectively—come with a nifty Google search bar installed by default, and, with the release of the Google Toolbar, a large number of Internet Explorer users now enjoy similar functionality.<sup>100</sup> Users accustomed to these tools, we might suppose, may be less inclined to switch to another site.

But more troubling, perhaps, is that ‘better’ alternatives are becoming increasingly rare. As discussed earlier, the enormous costs associated with running a search engine—not to mention the industry-wide tightening of venture capital—are leading to greater industry consolidation and increased pressures for commercialization and conglomeration. With over 90% of queries handled by the top three search engines, the only significant competition Google faces comes from Yahoo! and Microsoft. These two companies, which appear to be more fully embracing the business practices of their counterparts in the traditional media, are investing millions of dollars in research and advertising in the hopes of chiseling away at Google’s market share.<sup>101</sup>

Currently, many analysts predict that Microsoft will be the most formidable opponent, since it is likely to leverage its existing software monopolies in order to drive users to its site.<sup>102</sup> We need only be reminded how, only a few years ago, the Netscape browser was far and away the most popular one, but by bundling Internet Explorer with its operating system, Microsoft gained another, almost complete monopoly.<sup>103</sup> Similarly, while Real Player was once the de facto standard for streaming audio and video, Microsoft overtook that market as well, by including the software free with Windows.<sup>104</sup> “Microsoft,” writes one observer, “has been good at letting others pioneer a technology before taking over, exploiting its dominance in desktop operating systems.”<sup>105</sup> Sure Google may have its little search boxes, but Microsoft controls the operating system and productivity suite used by billions of users worldwide.



Taking advantage of this fact, Microsoft is expected to integrate its search interface into future versions of Windows and Office, and also to allow developers to build applications on top of its search technology.<sup>106</sup> That Microsoft—a company whose business practices have been roundly criticized, whose synergies with traditional media are extensive (e.g., MSNBC), and whose relentless pursuit of profit is legendary<sup>107</sup>—might become the de-facto gatekeeper for the Web seems, to put it mildly, a cause for concern. Charles Ferguson of MIT's *Technology Review* writes,

*If the firm dominating the search industry turned out to be Microsoft, the implications might be more disturbing still. The company that supplies a substantial fraction of the world's software would then become the same company that sorts and filters most of the world's news and information, including the news about software, antitrust policy, and intellectual property.*<sup>108</sup>

And so, it is important to keep in mind that although we have extensively focused on Google as a market leader, far more troubled waters may lie ahead.

At the same time, the Mountain View company will not sit idly by as Yahoo! and Microsoft attempt to take over. The firm's strategy seems to be 'information integration,' namely, to "search and organize all the world's information."<sup>109</sup> It has rolled out, and integrated into its main search tool, a news portal, a comparison shopping site and, a video locator. It has even begun to make printed books searchable from the Web and, in 2004, teamed up with Stanford University Libraries to scan and make available millions of out-of-copyright books.<sup>110</sup> In 2003, Google acquired Blogger, the world's most popular blog-hosting service.<sup>111</sup> And it is increasingly deploying technologies to collect and present personal data, and not just public information from Web sites and newsgroups. To this end, it recently announced its free e-mail service, Gmail<sup>112</sup>; it acquired photo software organization firm Picasa<sup>113</sup>; and it made available a picture sharing application called Hello.<sup>114</sup> Some say that Google ultimately seeks to become the new "platform" for computing: a place where the postmodern, networked, on-the-go information user will seek and send all forms of text and new media from any computer.<sup>115</sup>

But with Google starting to crawl, index, correlate, and present “all the world’s information,” it is not surprising to find it attracting the attention of many privacy activists. Indeed, the implications of a single firm knowing so much about you—about your searches, your communications, your files—are so great that it is precisely this subject which has begun to tarnish the company’s otherwise spotless public image. When it was announced that gMail would mine your emails to present “relevant” advertisements, for example, many Internet users expressed concerns that their extremely private communications were being inappropriately ‘read’ by Google and third party advertisers.<sup>116</sup> Similar concerns have been raised about such services as “My Search History” (which records and reminds users of their past searches), “Web Accelerator” (a plug-in that speeds up Web browsing by routing all your traffic through Google’s servers), and “Google Desktop Search” (a tool that mines your computer’s files in order supplement your Google results with relevant documents found on your computer).<sup>117</sup> And with its acquisition of “Keyhole”<sup>118</sup>—a firm whose software can provide a high-resolution satellite image of almost any spot in the US—the comparisons to “Big Brother” do, indeed, become all too easy to make.<sup>119</sup>

Google’s responses to these criticisms were, at first, characteristically shy. For instance, after the Electronic Frontier Foundation and other groups suggested that, given the post-9/11 climate, government officials might be accessing Google’s detailed, user-identifiable search logs for surveillance purposes,<sup>120</sup> Brin refused to comment on whether the company had been subpoenaed for such information. “I’ve taken the policy,” Brin said in 2002, “of not to responding publicly about government requests.”<sup>121</sup>

But with pressures from the public intensifying, the company eventually began to articulate its privacy policies in greater detail. In 2004, the search engine launched the “Google Privacy Center,”<sup>122</sup> a site that meticulously specifies, for each of Google’s major products and services, “the types of personal information gather[ed] ... as well as some of the steps ... take[n] to safeguard it.”<sup>123</sup> The company promises, for example, not to record personally-identifiable search histories, not to sell one’s information to third parties, and to secure personal data. Yet

it reserves the right to collect and aggregate usage data by, for example, “storing user preferences in cookies and by tracking user trends and patterns of how people search.”<sup>124</sup> Moreover, there is still no word on whether the government is, in fact, using Google to snoop on suspects (though other technology companies, including rival Yahoo!, have publicly confessed to being ‘deluged’ with requests from law enforcement agencies<sup>125</sup>). Although an adequate treatment of this debate is beyond the scope of this chapter, suffice it to say that as Google continues to dominate search and other related information services, it—and the entire Internet community—will need to come to terms with some very serious matters of personal privacy.

## VI. Conclusions

Similar to our conclusions about advertising in the last chapter, we again find that search engines are susceptible the same problematic forces originally observed in the context of traditional media. The search space is characterized by very high levels of consolidation, and search engines are prime candidates for absorption into the empires of media conglomerates. Given the enormous costs of running a successful search engine, and the relatively high barriers to entry, it is expected that this trend will continue. Google is, for now, far and away the market leader, fielding the majority of search queries performed in the United States.

As McChesney, Bagdikian, Gomery and Champaine have argued, we should be worried whenever a single firm wields so much influence over an entire media industry. The relative dearth of alternative search engines provides both the opportunity and the incentives to hypercommercialize content and to bias results in a self-interested manner. What makes Google remarkable is that it at least appears to recognize these problems, and has actually warned investors that it reserves the right to sacrifice profit in order to further the public good. And to prevent being bought out or overruled by stockholders, Google left majority control in the hands of Brin, Page, and Schmidt . Although this ownership structure grants the “enlightened” triumvirate sufficient freedom to steer the company as it sees fit, valid concerns

may be raised about whether the public is entitled to sufficient transparency and protection under this scheme. In the end, however, most observers agree that it is Microsoft and Yahoo! that pose the greatest threat to the dissemination of noncommercial, antagonistic voices on the Web.

## Notes

---

- <sup>1</sup> Robert E. McGinn, *Science, Technology, and Society* (Englewood Cliffs, NJ: Prentice Hall, 1991), 76.
- <sup>2</sup> Robert McChesney, "Oligopoly: The Big Media Has Fewer and Fewer Players," *The Progressive*, November 1999, 23.
- <sup>3</sup> Benjamin Compaine and Douglas Gomery, *Who Owns the Media?: Competition and Concentration in the Mass Media Industry*, 3rd ed. (Mahwah, NJ: Lawrence Erlbaum Associates, 2000), 521.
- <sup>4</sup> Abbe Mowshowitz and Akira Kawaguchi, "Bias on the Web," *Communications of the ACM* 45, no. 9 (2002): 60.
- <sup>5</sup> For one exception, see Tair-Rong Sheu and Kathleen Carley, "Monopoly Power on the Web: A Preliminary Investigation of Search Engines" (paper presented at the 29th Telecommunications Policy Research Conference, Alexandria, VA, 2001), available from <http://arxiv.org/abs/cs.CY/0109054> (accessed May 7, 2005). Note that this work is a bit dated. It is not, for instance, aware of the meteoric rise of Google.
- <sup>6</sup> Danny Sullivan, *NetRatings Search Engine Ratings, June 1998* (SearchEngineWatch.com, December 1998 [cited December 20 2004]), available from <http://searchenginewatch.com/mhts/9806-10-netratings.mht>.
- <sup>7</sup> Danny Sullivan, *comScore Media Metrix Search Engine Ratings, December 2004* (SearchEngineWatch.com, February 11 2005 [cited May 8 2005]), available from <http://searchenginewatch.com/reports/article.php/2156431>.
- <sup>8</sup> Deborah Fallows, *Search Engine Users: Internet Searchers are Confident, Satisfied, and Trusting* (Pew Internet & American Life Project, April 23 2005 [cited May 8 2005]), available from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf).
- <sup>9</sup> We obtained from the *SearchEngineWatch* archives the Nielsen//Netratings search engine figures between January 1999 and September 2004. Unfortunately, in 2001 Nielsen changed its methodology: rather than measuring what percentage of users visited a particular search engine (sum > 100%), it began to directly measure the percentage of queries performed by each search engine (sum==100). To account for this change—for illustrative purposes only—we assumed that the figures did not change between December 2000 (before the change) and March 2001 (after the change) for each search engine. Then, we looked at, for each search engine, its ratio of Nielsen percents before and after the change. We then multiplied the post-2001 numbers by this ratio and plotted them as seen in this Figure 8-1.
- <sup>10</sup> Jupiter Communications, quoted in Danny Sullivan, *The End for Search Engines?* (ClickZ Experts, February 21 2001 [cited December 23 2004]), available from <http://www.clickz.com/experts/search/opt/article.php/837281>.
- <sup>11</sup> George Bell, quoted in *Ibid.*
- <sup>12</sup> Richard Waters and Amy Lee, "Ask Jeeves to Join Excite Internet," *The Financial Times*, March 5 2003.
- <sup>13</sup> Mary Anne Ostrom, "Pasadena, Calif., Commercial Search Firm to Buy Web Search Properties," *San Jose Mercury News*, February 26 2003.
- <sup>14</sup> Caspar van Vark, "Search Engines: Search Still Sets the Pace," *Revolution*, April 21 2004.

- 
- <sup>15</sup> See, for example, Stephen Arnold, "In Search of Clicks that Make Cash: Three Search Companies are All Chasing the Same Pool of Advertising Spend," *World Information Review*, April 1 2003.
- <sup>16</sup> Compaine and Gomery, *Who Owns the Media?*, 555-56.
- <sup>17</sup> This threshold, in fact, is exceeded by Yahoo! and Google alone.  
In terms of traffic, the two companies fielded over 67% of all Web search queries, according to Sullivan, *comScore Media Metrix Search Engine Ratings, December 2004*.  
In addition, both companies announced, in their annual reports for 2004, overall advertising revenues of roughly \$3b a piece. According to our calculations, about \$2 (Google) to \$2.7 (Yahoo!) billion of this was in domestic sales. Since U.S. Internet ad expenditures that year totaled about \$8.7b, these companies control over half of the online advertising market in the United States. Sources: Yahoo Inc., *United States Securities and Exchange Commission Form 10-K* (United States Securities and Exchange Commission, 2005 [cited May 7 2005]), 32,38, available from <http://yhoo.client.shareholder.com/edgar.cfm?DocType=Annual>.; Google Inc., *United States Security and Exchange Commission Form 10-K* (United State Securities and Exchange Commission, March 30 2005 [cited May 8 2005]), 25,27, available from <http://www.edgar-online.com/bin/cobrand/?doc=A-1288776-0001193125-05-065298>.; John Markoff and Nat Ives, "Web Search Sites See Clicks Ad Up to Big Ad Dollars," *The New York Times*, February 4 2005.
- <sup>18</sup> S. A. Rhodes, quoted in Compaine and Gomery, *Who Owns the Media?*, 558-59.
- <sup>19</sup> According to Sheu and Carley, search engines have an HHI of 1163 (based on audience traffic, presumed to be correlated with advertising revenue). See Sheu and Carley, "Monopoly Power on The Web", 14. A comparison with traditional media sectors is provided by Mark Cooper and Steven Cooper, *Hope and Hype v. Reality: The Role of the Commercial Internet in Democratic Discourse and Prospects for Institutional Change* (Stanford Law School Center for Internet and Society, 2003 [cited May 7 2005]), available from <http://cyberlaw.stanford.edu/blogs/cooper/archives/HOPEALL.pdf>.
- <sup>20</sup> Sheu and Carley, "Monopoly Power on The Web", 22.
- <sup>21</sup> Compaine and Gomery, *Who Owns the Media?*, 514-15.
- <sup>22</sup> At least, this is how the situation is portrayed in the popular and business press. See, for example, Jim Hu and Mike Ricciuti, *Search and Destroy: Microsoft, Google May Go Head-to-Head* (C|Net News.com, June 25 2003 [cited July 9 2003]), available from [http://news.com.com/2102-1032\\_3-1020641.html](http://news.com.com/2102-1032_3-1020641.html).
- <sup>23</sup> Saul Hansell, "Search Sites Play a Game of Constant Catch-Up," *The New York Times*, January 31 2005. Lev Grossman, "Search and Destroy: A Gang of Web Search Companies is Gunning for Google," *Time*, December 22 2003.
- <sup>24</sup> Compaine and Gomery, *Who Owns the Media?*, 518.
- <sup>25</sup> *Ibid.*, 517.
- <sup>26</sup> The following discussion of 'barriers to entry' is guided by Picard's excellent overview of media economics. See Robert Picard, *The Economics of Financing Media Companies* (New York, NY: Fordham University Press, 2002), 72-80.
- <sup>27</sup> Matt Marshall, "Google Founders' Brashness Sparks Debate," *San Jose Mercury News*, August 18 2005.
- <sup>28</sup> Hansell, "Search Sites Play a Game of Constant Catch-Up."
- <sup>29</sup> For a discussion of substitutability in general economic terms, see Chad Syverson, "Product Substitutability and Productivity Dispersion," *The Review of Economics and Statistics* 86, no. 2 (2004).
- <sup>30</sup> Compaine and Gomery, *Who Owns the Media?*, 476.
- <sup>31</sup> Charles H. Ferguson, "What's Next for Google?: The Search Firm Wants to Organize All Digital Information. That Means War with Microsoft," *Technology Review: MIT's Magazine of Innovation*, January 2005.

- 
- <sup>32</sup> Adrian Cox, quoted in Tony Glover, "Search Engines Power New Dotcom Boom," *Sunday Business*, March 27 2005.
- <sup>33</sup> Compaine and Gomery, *Who Owns the Media?*, 460.
- <sup>34</sup> Sheu and Carley, "Monopoly Power on The Web", 17-18.
- <sup>35</sup> *Ibid.*, 17.
- <sup>36</sup> Danny Sullivan, *Nielsen NetRatings Search Engine Ratings, December 2000* (SearchEngineWatch.com, February 19 2001 [cited December 21 2004]), available from <http://searchenginewatch.com/mhts/9902-0012-netratings.mht>.
- <sup>37</sup> Deborah Fallows and Lee Rainie, *Data Memo: The Popularity and Importance of Search Engines* (Pew Internet & American Life Project, August 2004 [cited March 5 2005]), available from [http://www.pewinternet.org/pdfs/PIP\\_Data\\_Memo\\_Searchengines.pdf](http://www.pewinternet.org/pdfs/PIP_Data_Memo_Searchengines.pdf).
- <sup>38</sup> According to Google's most recent annual report, the company's domestic advertising revenues in 2004 were \$2.1 billion (it collected \$3.14 billion total, but a third of this came from abroad). By our calculations, this represents about a quarter of the \$8.7b spent on Internet ads in the U.S. that year (the latter figure reported in the *New York Times*). See Google Inc., *United States Security and Exchange Commission Form 10-K*, 25,27. Also, Markoff and Ives, "Web Search Sites See Clicks Ad Up to Big Ad Dollars."
- <sup>39</sup> By 1998, Scott Rosenberg was already pointing his readers to Google—"a search engine that actually works"—and proclaiming it was "one more reason to distrust the conventional view that the portals have the future of the Web sewn up." Scott Rosenberg, "Yes There Is a Better Search Engine: While the Portal Sites Fiddle, Google Catches Fire," *Salon.com*, December 21 1998, available from <http://archive.salon.com/21st/rose/1998/12/21straight.html>. Not surprisingly, in light of Google's strong brand and growing popularity, the major portals abandoned their view that "search was not that important" and began to rely on Google to provide its algorithmic results. See Michelle Prather, "Ga-Ga for Google," *Entrepreneur Magazine*, April 2002, available from <http://www.entrepreneur.com/article/0,4621,297807,00.html>.
- <sup>40</sup> Danny Sullivan, *comScore Media Metrix Search Engine Ratings, August 2003* (SearchEngineWatch.com, August 1 2003 [cited May 1 2005]), available from <http://searchenginewatch.com/mhts/0305-mediаметrix.mht>.
- <sup>41</sup> Jonathan Zittrain, quoted in John Markoff and G. Pascal Zachary, "In Searching the Web, Google Finds Riches," *The New York Times*, April 13 2003.
- <sup>42</sup> In his article on the social shaping of technology, Langdon Winner describes how Robert Moses, master builder of New York City, built bridges over the Southern State Parkway (in Long Island) at a low height in order to prevent public transportation from accessing Jones Beach. This had the effect of instituting *de facto* class segregation over that publicly owned destination. See Langdon Winner, "Do Artifacts Have Politics?," *Daedalus* 109, no. 1 (1980): 125.
- <sup>43</sup> This is consistent with Robert McGinn's IDUAR model for "technological change ensuing social change (TCSEC)." See McGinn, *Science, Technology, and Society*, 99.
- <sup>44</sup> Elihu Katz, "The Two-Step Flow of Communication: An Up-To-Date Report of an Hypothesis," in *Marketing Classics: A Selection of Influential Articles*, ed. Ben Enis and Keith Cox (Boston, MA: Allyn and Bacon, 1973).
- <sup>45</sup> Ben H. Bagdikian, *The Media Monopoly*, 4th ed. (Boston: Beacon Press, 1992), 223-24.
- <sup>46</sup> Mowshowitz and Kawaguchi, "Bias on the Web."
- <sup>47</sup> Stefanie Olsen, *The Google Gods: Does Search Engine's Power Threaten Web's Independence?* (C|Net News.com, October 31 2002 [cited May 6 2005]), available from [http://digitalcity.com.com/2009-1023\\_3-963618.html](http://digitalcity.com.com/2009-1023_3-963618.html).
- <sup>48</sup> Cass Sunstein, *Republic.com* (Princeton, NJ: Princeton University Press, 2001), 91-99.
- <sup>49</sup> "The Weakness of Google," *The Economist*, May 1 2004, 63. Danny Sullivan, *comScore Media Metrix Search Engine Ratings, February 2004* (SearchEngineWatch.com, April 28 2004 [cited March 20 2005]), available from <http://searchenginewatch.com/mhts/0402-mediаметrix.mht>.

- 
- <sup>50</sup> Sullivan, *comScore Media Metrix Search Engine Ratings, December 2004*.
- <sup>51</sup> Dennis W. Mazzocco, *Networks of Power: Corporate TV's Threat to Democracy* (Boston: South End Press, 1994), 166.
- <sup>52</sup> Jeff Dean, *Google: A Behind the Scenes Look* (Seattle, WA: University of Washington Television, 2004), Video, available from <http://www.uwv.org/programs/displayevent.asp?rid=2459>.
- <sup>53</sup> Sergey Brin and Larry Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems* 30, no. 1-7 (1998): 18-19, available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 1, 2005).
- <sup>54</sup> Dean, *Google: A Behind the Scenes Look*.
- <sup>55</sup> Google Inc., *United States Securities and Exchange Commission Form S-1* (United States Securities and Exchange Commission, April 29 2004 [cited May 7 2005]), 32, available from <http://www.sec.gov/Archives/edgar/data/1288776/000119312504139655/ds1a.htm>.
- <sup>56</sup> *Ibid.*, 50.
- <sup>57</sup> Ariana Eunjung Cha, "After Years in the Red, Amazon Posts Profit," *The Washington Post*, January 23 2002.
- <sup>58</sup> Sergey Brin discusses this product-before-monetization pattern in Grossman, "Search and Destroy: A Gang of Web Search Companies is Gunning for Google." So does Google's CEO in Eric Schmidt, *Keynote Address at the 2004 Conference on Entrepreneurship* (Stanford, CA: Stanford University Graduate School of Business, 2004), available from <http://wesley.stanford.edu/multimedia/events/entrepconference/google.ram> (accessed May 10, 2005).
- <sup>59</sup> Indeed, so impressed was *Workforce* that gave the company its 2003 General Excellence award. Todd Raphael, "At Google, the Proof is in the People," *Workforce*, March 2003.
- <sup>60</sup> Alex Salkever, *The Search Engines' Little secret* (BusinessWeek Online, August 6 2001 [cited July 7 2003]), available from [http://www.businessweek.com/print/bwdaily/dnflash/aug2001/nf2001086\\_115.htm?tc](http://www.businessweek.com/print/bwdaily/dnflash/aug2001/nf2001086_115.htm?tc).
- <sup>61</sup> Gary Rivlin, "Google Goes Public? Search for 'Rich Get Richer'," *The New York Times*, April 25 2004.
- <sup>62</sup> "Google: Up For Grabs," *The Economist*, May 8 2004.
- <sup>63</sup> Google Inc., *United States Securities and Exchange Commission Form S-1*.
- <sup>64</sup> Compaine and Gomery, *Who Owns the Media?*, 527.
- <sup>65</sup> Indeed, the founders appear to be nostalgic for the 'old days' before Google was on everyone's computers and so many investors' minds. As a recent issue of *Fortune* describes Brin's state: "I preferred things during the bubble'—when Google was not so famous—'and we could go on doing our own thing.' At the end of his presentation he repeated the thought, saying that coping with Google's torrid growth had become the worst part of his job: 'It's a distraction from pure technology, which is what I love.'" See Fred Vogelstein, "Can Google Grow Up?," *Fortune*, November 24 2003.
- <sup>66</sup> Robert McChesney, *Rich Media, Poor Democracy: Communication Politics in Dubious Times* (Urbana: University of Illinois Press, 1999), 22.
- <sup>67</sup> Glover, "Search Engines Power New Dotcom Boom."
- <sup>68</sup> Adrian Cox, quoted in *Ibid.*
- <sup>69</sup> These moves are discussed in Compaine and Gomery, *Who Owns the Media?*, 471,522,84-85. and McChesney, *Rich Media, Poor Democracy*, 180-81.
- <sup>70</sup> Compaine and Gomery, *Who Owns the Media?*, 521.
- <sup>71</sup> Unspecified *Red Herring* article, quoted in Sullivan, *The End for Search Engines?*
- <sup>72</sup> Compaine and Gomery, *Who Owns the Media?*, 421.
- <sup>73</sup> John Markoff and Andrew Sorkin, "Microsoft And Google: Partners Or Rivals?," *The New York Times*, October 31 2003.
- <sup>74</sup> Google Inc., *United States Securities and Exchange Commission Form S-1*, 27.
- <sup>75</sup> *Ibid.*, 29.

- 
- <sup>76</sup> See R.R. Palmer and Joel Colton, *A History of the Modern World* (New York, NY: Alfred A. Knopf, Inc., 1984), 307,22-23., and John Gagliardo, *Enlightened Despotism* (New York, NY: Thomas Y. Crowell Company, 1967), 22,42-53,78..
- <sup>77</sup> Leonard Krieger, *An Essay on the Theory of Enlightened Despotism* (Chicago, IL: University of Chicago Press, 1975), 19.
- <sup>78</sup> For more on the history and legacy of Frederick II, see Robert B. Asprey, *Frederick the Great: The Magnificent Enigma* (New York, NY: Ticknor & Fields, 1986).
- <sup>79</sup> Frederick the Great, "Benevolent Despotism," in *The Portable Enlightenment Reader*, ed. Isaac Kramnick (New York, NY: Penguin Books, 1995), 454,58.
- <sup>80</sup> Palmer and Colton, *A History of the Modern World*, 314-15.; Krieger, *An Essay on the Theory of Enlightened Despotism*.; Gagliardo, *Enlightened Despotism*.
- <sup>81</sup> Google Inc., *United States Securities and Exchange Commission Form S-1*, 29-30.
- <sup>82</sup> The phrase 'benevolent dictator' is usually associated with Linus Torvalds, founder of the Linux operating system. I have, however, heard it used by Google employees as well. And *Developer Shed* refers to Brin as a 'benevolent big brother' in Norbert Cartagena, "Google: A God or a Devil?," *Plugin Magazine*, November 2004, 22, available from [http://www.developershed.com/plugin/PIM\\_200411.pdf](http://www.developershed.com/plugin/PIM_200411.pdf) (accessed May 9, 2005).
- <sup>83</sup> Investors in particular saw Google's approach to the IPO as "arrogant." See, for example, Henry Blodget, "Gambling on Google--The Thrilling Conclusion," *Slate*, August 19 2004, available from <http://slate.msn.com/id/2105418/>. and Vogelstein, "Can Google Grow Up?."
- <sup>84</sup> Immanuel Kant, "What Is Enlightenment?," in *The Portable Enlightenment Reader*, ed. Isaac Kramnick (New York, NY: Penguin Books, 1995), 3.
- <sup>85</sup> Jean-Jaques Rousseau, *The Social Contract*, trans. Maurice Cranston (New York: Penguin Books, 1968), 120.
- <sup>86</sup> Bagdikian, *The Media Monopoly*, 223.
- <sup>87</sup> Robert Dahl, *Democracy and Its Critics* (New Haven, CT: Yale University Press, 1989), 1-33.
- <sup>88</sup> Josh McHugh, "Google Sells Its Soul," *Wired*, January 2003, 132.
- <sup>89</sup> Google Inc., *Google Code of Conduct* (Google.com: Investor Relations, August 18 2004 [cited May 7 2005]), available from <http://investor.google.com/conduct.html>.
- <sup>90</sup> McHugh, "Google Sells Its Soul," 132.
- <sup>91</sup> *Ibid.*, 133.
- <sup>92</sup> David Sheff, "Playboy Interview: Google Guys," *Playboy*, September 2004.
- <sup>93</sup> *Ibid.*
- <sup>94</sup> Google Inc., *Why We Sell Advertising, Not Results* (Google.com, 2004 [cited November 20 2004]), available from <http://www.google.com/honestresults.html>.
- <sup>95</sup> Google Inc., *United States Securities and Exchange Commission Form S-1*, vi.
- <sup>96</sup> Google Inc., *Google Code of Conduct*.
- <sup>97</sup> *Ibid.*, "Serving Our Users".
- <sup>98</sup> Naomi Klein, *No Logo: Taking Aim at the Brand Bullies* (New York: Picador, 2002).
- <sup>99</sup> Rahul Telang, Tridas Mukhopadhyay, and Ronald Wilcox, "An Empirical Analysis of the Antecedants of Internet Search Engines Choice" (paper presented at the Workshop on Information Systems and Economics, Charlotte, NC, 1999).
- <sup>100</sup> Michael Miller, "Internet Explorer Still Dominates the Web, but Firefox is Growing Fast and Igniting Innovation," *PC Magazine*, March 22 2005. At the time of this writing, the Google Toolbar for Internet Explorer is available at <http://toolbar.google.com>, Firefox is available from <http://www.mozilla.org>, and Safari is included with Apple's Mac OS X operating system.
- <sup>101</sup> Hansell, "Search Sites Play a Game of Constant Catch-Up."
- <sup>102</sup> See, for example, Hu and Ricciuti, *Search and Destroy: Microsoft, Google May Go Head-to-Head*. Also, Simon Sharwood, *Microsoft Brains Take on Google Brawn* (C|Net News.com, July 9 2003 [cited July 10 2003]), available from <http://news.com.com/2100-1032-1024038.html>.



- 
- <sup>103</sup> Michael Cusumano and David Yoffie, *Competing on Internet Time: Lessons from Netscape and Its Battle With Microsoft* (New York, NY: Free Press, 2000).
- <sup>104</sup> The issue regarding Real Player was the subject of an antitrust investigation by the EU, which in March 2004 found Microsoft liable for illegally 'bundling' its own player with the Windows operating system. See Ian Ayres and Barry Nalebuff, "Going Soft on Microsoft? The EU's Antitrust Case and Remedy," *The Economists' Voice* 2, no. 2 (2005), available from <http://www.bepress.com/ev/vol2/iss2/art4>.
- <sup>105</sup> "How Good is Google?," *The Economist*, October 23 2003, available from [http://www.economist.com/displaystory.cfm?story\\_id=2173573](http://www.economist.com/displaystory.cfm?story_id=2173573).
- <sup>106</sup> Hu and Ricciuti, *Search and Destroy: Microsoft, Google May Go Head-to-Head*.
- <sup>107</sup> McChesney, *Rich Media, Poor Democracy*, 162-63.
- <sup>108</sup> Ferguson, "What's Next for Google?," 46.
- <sup>109</sup> Google Inc., *United States Securities and Exchange Commission Form S-1*, 30.
- <sup>110</sup> Adair Lara, "Googleizing' Libraries Won't Replace Books," *The San Francisco Chronicle*, December 18 2004.
- <sup>111</sup> David Gallagher, "Deal May Freshen Up Google's Links," *The New York Times*, February 24 2003.
- <sup>112</sup> Michael Bazeley, "Google to Offer E-Mail for Free," *San Jose Mercury News*, April 1 2004.
- <sup>113</sup> Cade Metz, "Picasa is Better, and Now Free," *PC Magazine*, March 8 2005.
- <sup>114</sup> *Ibid.*
- <sup>115</sup> For one account of this theory from the mainstream press, see Michael Bazeley, "Google Keeps Ambitions Quiet," *San Jose Mercury News*, May 1 2004.
- <sup>116</sup> See, for example, Hiawatha Bray, "Google Cofounder Draws Wrath with Personal E-Mail Service," *The Boston Globe*, April 26 2004. Or, Michael Bazeley, "Jumbo Inboxes for Gmail Cause Privacy Fears," *San Jose Mercury News*, April 3 2004.
- <sup>117</sup> For coverage of Google's privacy battles, especially regarding Web Accelerator, see Stephanie Olsen, *Google Speed Bump Draws Scorn* (C|Net News.com, May 6 2005 [cited May 8 2005]), available from [http://news.com.com/2100-1032\\_3-5698477.html](http://news.com.com/2100-1032_3-5698477.html). On the desktop search controversy, see Danny Sullivan, *A Closer Look at Privacy & Desktop Search* (SearchEngineWatch.com, October 14 2004 [cited May 6 2005]), available from <http://searchenginewatch.com/sereport/article.php/3421621>.
- <sup>118</sup> Michael Bazeley, "Google Buys Map Maker," *San Jose Mercury News*, October 28 2004.
- <sup>119</sup> Google was, in fact, nominated for a Big Brother Award, though it was eliminated from 'competition' in the first round. See Danny Sullivan, *Google and the Big Brother Nomination* (SearchEngineWatch.com, April 2 2003 [cited May 6 2005]), available from <http://searchenginewatch.com/sereport/article.php/2175251>.
- <sup>120</sup> This Orwellian connection between corporate and governmental surveillance, brought to light in the form of the U.S. Government's recent Total Information Awareness project, is described in Robert O'Harrow, *No Place to Hide: Behind the Scenes of Our Emerging Surveillance Society* (New York, NY: Free Press, 2005).
- <sup>121</sup> Sullivan, *Google and the Big Brother Nomination*.
- <sup>122</sup> Google Inc., *Google Privacy Center* (Google.com, July 7 2004 [cited May 7 2005]), available from <http://www.google.com/privacy.html>.
- <sup>123</sup> *Ibid.*, "Google Privacy Policy".
- <sup>124</sup> *Ibid.*, "Data Collection".
- <sup>125</sup> Miles Benson, "Tech Firms Feel Heat as U.S. Snoops on Citizens," *The Times-Picayune*, May 6 2002.

# 9 | Conclusions

In 1994, Jacques Derrida gave a lecture in London, entitled “Memory: The Question of Archives.”<sup>1</sup> In it, he imagined how the recent developments of electronic and digital communications might have impacted Freud’s psychoanalytic research. Derrida argued that the state of the technology would not only have facilitated communication among researchers, but would have shaped the very events *themselves*, at least as seen retroactively through differing technologies of “archivation.” These technologies, he said,

*would have transformed this history from top to bottom and in the most initial inside of its production, in its very events. This is another way of saying that the archiving, printing, writing, prosthesis, or hypomnesic technique in general is not only the place for stocking and for conserving an archivable content of the past which would exist in any case, such as, without the archive, one still believes it was or will have been. No, the technical structure of the archiving archive also determines the structure of the archivable content even in its very coming into existence in its relationship to the future. The archivation produces as much as it records the event. This is also our political experience of the so-called news media.*<sup>2</sup>

Ten years later, Google CEO Eric Schmidt spoke at the Stanford Graduate School of Business.<sup>3</sup> He mentioned in his keynote address how a student once approached him to ask whether the company’s news service had any bias. Schmidt responded:

*This is not really a newspaper. These are computers. You know, we have a tagline: “no humans were harmed in the production of this newspaper.” You know, it’s computers that are reading it, and they’re assembling it using algorithms.*

But the student insisted. “No,” he said, “you don’t understand. All news has a bias, yours must, too.” Schmidt would not back down:

*I can assure you. It has no bias. These are computers, they’re boring. I’m sorry you just don’t get it.*

Throughout this thesis, we have tried to call into question this general perception that information systems are objective, “boring” purveyors of information.<sup>4</sup> We have tried to show how, to use Derrida’s terminology, the “process of archivation” not only shapes our “political experience” with the news media, but with Google as well.<sup>5</sup> To the extent that the “technical structure” of the Google archive selectively filters some content and not others, it conveys a particular version of online events. To the extent that Google acts as a “gatekeeper” for the Web, it profoundly influences what consists of for us, and what we are exposed to when we switch on our modems. Google, in other words, shapes the Web as much as it records it.

By integrating reports from the press, public message boards, SEC filings, industry consultants, interviews, and extensive interdisciplinary research spanning economics to complex network theory, we have hopefully shown that:

- *Search engines do have biases.* With Google, these biases include an affinity towards extremely popular, often commercial sources (Chapter 5). While this may be a case of ‘indexical’ bias (rather than the ‘content bias’ Schmidt envisions), we have argued that it is also important and worthy of rich, critical analysis.
- *Search engine design is profoundly shaped by technological, commercial, and regulatory forces.* Google’s design was motivated by engineering challenges (Chapter 4), but the enormous costs of running a competitive search engine also necessitated the introduction of advertising (Chapter 7) and the practice of catering to majority interests (Chapter 5). The company has also explicitly manipulated its results in response to, for instance, European prohibitions on hate speech (Chapter 6).
- *Search engine bias has implications for democratic discourse on the Web.* As discussed in Chapter 3, a normative vision of Web grounded on ‘deliberation’ requires that intermediaries like Google disseminate a wide spectrum of opinions on a given issue. This prescription stresses, in particular, that antagonistic, outside-the-mainstream voices are critical for the establishment of a well-informed citizenry. To

the extent that commercial, popular, or technologically-savvy entities dominate Google's search results (via advertising or algorithmic bias), these ideals are undermined.

- *These consequences are contingent on social variables* such as patterns of use, “gatekeepership,” and industry consolidation. The social and political implications of bias in the Google search engine is dependent, quite obviously, on how the search engine is used and the extent to which it serves as a primary gatekeeper. Because Google currently handles the majority of Web queries, and because the industry as a whole is facing rapid consolidation, its influence on Web users may indeed be substantial (Chapter 8). This is particularly troubling since the search engine is so notoriously secretive (Chapter 6), perhaps limiting the degree to which users can voice resistance.

Echoing earlier conclusions about the politics of search engines, we have found that the design, economic context, and use of these technologies presents a fundamental challenge to utopian visions of the Web. It is important that we recognize these tensions as we start to move much of the world's information, knowledge, and discourse online.

## **I. Back to the Basics: Google under Disparate Theoretical Approaches**

Our analysis, as these findings indicate, was informed by prominent theories of STS (e.g., social shaping), political science (e.g., deliberative democracy), and studies of communication (e.g., mass media structures). Despite the fact that all these approaches predate the widespread adoption of the Internet, our results are very much consistent with existing research within these traditions. Our conclusions about the PageRank algorithm, for example, mirror Winner's discussion of the automated tomato harvester. Likewise, search engine advertising is similar in function, if not form, to practices of hypercommercialism in the

traditional media. And, of course, prevailing views about the Web itself bear striking resemblance to the centuries-old ideal of deliberative democracy.

We may be able to better organize these findings by explicitly reformulating them in terms of the frameworks outlined at the beginning of our discussion. This already proved useful, you might recall, in Chapter 5, when we discussed how PageRank reflects all three categories of bias outlined by Friedman and Nissenbaum in their discussion of ‘computer bias.’ Similarly, McGinn’s IDUAR model<sup>6</sup> allows us to focus in on the many factors that may be contributing to the apparent tendency of Google to elevate commercial, mainstream voices over antagonistic, minority sources of information. This model suggests that the social change is not just a result of Google’s technology *per se*; it also depends on the motivations behind its design (nature of the innovation), the ability for free, substitutable adoption (diffusion), the fact that millions of users turn to Google when looking for various kinds of information (use), the relative ease with which the search engine allowed even novice users to find what they were looking for (adaptation), and various instances of governmental and watchdog opposition to some of the search engine’s features (resistance). These factors, which we have already discussed in the preceding chapters, and are detailed in Table 9-1.

## II. Is Google Evil?

Given the critical-analytic lens with which we have approached the sociopolitics of search, it may seem that we are promoting the view that Google is, in fact, “evil.” But this is certainly not the case. Sure, Google’s policies on secrecy and advertising content are worrisome and, from such a supposedly ‘enlightened’ company, downright puzzling. But it is our view that the observed tensions between the search engine and democratic aspirations are, for the most part, not the product of malicious or even profiteering intent. We instead believe that they stem from the inherent limitations of commercialized search. It is hard to imagine a search company staying afloat, after all, if it does not present what its users want; it is difficult to

<b>Innovation</b>	<p>Salient technological characteristics of the Google search engine include PageRank, its crawlers, the 100+ ranking metrics, the implementation of its AdWords advertising generator, the Google Toolbar and so on. The design of these material technics was shaped by, among other things:</p> <ul style="list-style-type: none"> <li>• the preexisting technical characteristics of the Web (e.g., decentralization of content, use of HTML, etc.).</li> <li>• difficulties finding content online, due in part to the spamming abuses plaguing prior search technologies.</li> <li>• theeconomic costs of operating a large-scale search engine, which necessitated a commercial (rather than academic) ownership model dependent upon advertising</li> <li>• design values that place “usefulness” (i.e., giving users what they want) before deliberativeness or comprehensiveness. This may explain PageRank’s bias towards popular, highly-linked content.</li> <li>• ethical principles that reflect the founders’ worries about the problems of the traditional media.</li> </ul>
<b>Diffusion Process</b>	<p>Google was always ‘free’ to users, but the <i>mechanism</i> by which the service was funded has varied. With the widespread popularity and effectiveness of the system, it eventually outgrew its academic roots and was reborn as corporate entity. Today, the search engine is operated by a publicly-owned company and paid for by advertising, as opposed to research grants. Brin and Page, however, remain at the helm.</p> <p>The <i>character</i> of the diffusion is remarkable in that it consisted of no marketing whatsoever (other than word-of-mouth). Google’s diffusion was, it appears, motivated by the quality of its “revolutionary” technology.</p> <p>As documented in the previous chapter, Google was widely diffused with extraordinary <i>speed</i>. Less than five years after its introduction, it was handling three-fourths of all search queries in the United States (no doubt due to low switching and use costs).</p>
<b>Use Pattern</b>	<p>According to a recent study by Pew, about half of Americans with access to the Internet frequent search engines once or twice a week, and an additional 35% do so daily. Search engines are, moreover, a primary means by which users find information about political, medical, commercial and religious matters. Google is far and away the most popular search engine, and it is clear that millions of users rely on this search engine—often exclusively—for both important and trivial information.</p>
<b>Adaptation</b>	<p>Like most forms of ‘new’ media, use of the Google search engine presupposes basic computer and Internet skills. But most accounts, Web searching before Google was a far more arduous affair, one that involved using myriad search engines, scanning many results, and constructing Boolean queries. It could be argued that Google was partly a success because it <i>reduced</i> the social adaptation necessary to navigate through cyberspace.</p>
<b>Resistance</b>	<p>Institutional and social agents have objected to various of characteristics of the search engine. The German and Chinese governments called on Google to remove listings which were in violation with its policies. Privacy groups, on the other hand, have objected to the companies’ aggregation of personal data. These relatively rare criticisms appear to have had little effect curbing diffusion and use, but they have sometimes motivated a change in the innovation itself (e.g., manual removal of hate speech from <i>google.de</i>).</p>

**Table 9-1.** *Application of McGinn’s IDUAR Model.* This table summarizes the five socioculturally-dependent “intermediary variables” that partly explain the potential changes brought about by the Google search engine.

make money if it does not display advertising; it cannot remain competitive and prevent its results from being ‘hijacked,’ many argue, if its algorithms operate in full view.

The “take away” point here is that in many ways Google does an exemplary job at promoting democratic ideals *given the limitations of the commercial model*. Although we have not attempted to make a comprehensive, comparative analysis of Google relative to other search engines, our research suggests similar problems (e.g., the popularity bias) pervade successful search engines in general. These search engines also seem to be more likely to embrace hypercommercialism, to synergize with traditional media, to see searchers and consumers rather than as citizens, and to constantly pursue a fatter bottom line at the expense of democratic ideals.

It would be quite difficult to suggest that we are better off without Google or, for that matter, without any of the other search engines. Awash in a sea of bits, we may be tempted to look at ‘democraticness’ and ‘bias’ as binaries, as things you either have or you don’t. It makes more sense to take a step back, and to think of Google as *one more way* in which people can get information. Only the most hardened cynic would think that the success of Google has resulted in a *net loss* of sources to which we are exposed.<sup>a</sup> And so, as Compaine reminds us,

*the questions to ask yourself are: Are there more or fewer voices available today than 15, 25 or more years ago? And, is it easier or harder, are the regulatory barriers higher or lower, is it more expensive or less expensive, to gain access, in whatever format, to a large audience ... than in 1900? in 1950? in 1990?'*

We believe that the answers to all these questions are emphatically “yes,” and that Google—certainly more than the traditional broadcast media—is making it possible for more people to hear and contribute to a broader spectrum of opinions. The deliberative standard is quite clearly an extremely difficult (some might say impossible) one to meet. And so, it’s not that haven’t moved forward. It’s just that we aren’t quite “there” yet.

---

<sup>a</sup> Indeed, I have relied on this very search engine to find a great deal of the recent research and news in this area. My analysis would have been far less complete had I been confined only to ‘books’ (you know, those things printed on dead trees and cloistered in old buildings at old universities).

### III. Moving Forward I: Alternatives and Regulation

In the spirit of “moving forward,” we might ask whether alternative search engines might emerge, ones that do indeed promote a more balanced cross-section of Web sources. Unfortunately, it seems safe to conclude that—to reiterate a point made in the last chapter—if left to the market, such sites are unlikely to appear. The reasons for this are manifold. First, traditional economic models tend to assume that consumers—who are supposed to express their preferences by selecting among alternatives—are ‘rational’ and well-informed. But as noted earlier, users are largely unaware of how search engines work,<sup>8</sup> and with the increased complexity and secretiveness surrounding these systems, it is likely that this will continue to be the case. Second, while the market mechanism is intended to most fully satisfy the preexisting preferences of consumers, the deliberative ideal requires that individuals also be exposed to material that is *contrary* to these predispositions. What’s good for consumers is, in other words, not always what’s good for citizens. Third, the barriers to entry are becoming increasingly high, limiting the number of alternatives among which consumers can choose in the first place. All this leads Introna and Nissenbaum to conclude that “thorough and wide-ranging access to the Web lies within the category of goods ... that should not be left entirely (if at all) to the marketplace.”<sup>9</sup>

Articulating a similar concern about the undermining of the ideals of the Web, Larry Lessig finds great comfort in open source software. He writes, in *Code and Other Laws of Cyberspace*,

*Open code ... functions as a Kind of Freedom of Information Act for network regulation. As with ordinary law, open code requires that lawmaking be public, and thus that lawmaking be transparent ... open code is a foundation to an open society.*<sup>10</sup>

We might be encouraged, then, to find Nutch, “an open source search engine founded specifically in response to concerns about bias in commercial search engines.”<sup>11</sup> Nutch developers, aware of the typical criticism about the spamming vulnerabilities associated with open code, respond as follows:



*With an open-source search engine, this will still happen, just out in the open. This is analogous to encryption and virus protection software. In the long term, making such algorithms open source makes them stronger, as more people can examine the source code to find flaws and suggest improvements. Thus we believe that an open source search engine has the potential to better resist manipulation of its rankings.<sup>12</sup>*

Even so, it should be clarified that Nutch is not an operational search engine but a development project (i.e., writing code, not running it). Since the project cannot do away with the enormous bandwidth and infrastructure costs necessary to run a large search engine of any stripe, its developers expect that more well-funded organizations will adopt Nutch to power large-scale search engines. It cannot ensure, however, that these companies will not modify the code, say, to reinstitute the biases typical to commercial search engines (the ability to modify code is, after all, part of the philosophy of open source).

Thus, the preexisting concerns about the “openness” of search systems continue to hold. In response to this problem, Introna and Nissenbaum call for regulation that

*would demand full and truthful disclosure of the underlying rules (or algorithms) governing indexing, searching, and prioritizing, stated in a way that is meaningful to the majority of Web users. Obviously, this might help spammers. However, we would argue that the impact of these unethical practices would be severely dampened if both seekers and those wishing to be found are aware of the particular biases inherent in any given search engine. ... Those who favor a market mechanism would perhaps be pleased to note that disclosure would move us closer to fulfilling the criteria of an ideal competitive market in search engines.<sup>13</sup>*

But as the authors admit, “disclosure, by itself, may not sustain and enhance Web offerings in the way we would like—that is, by [promoting] less popular sites.” This recognition suggests that if we want to regulate existing search engines to counteract the deliberative concerns raised here, we would need to somehow prescribe how these algorithms should work.

Although these authors focused only on algorithmic bias (ignoring, for example, ownership and advertising biases), we could easily imagine similar policies that dictate not only how results should be ‘democratically’ selected, but that also prevents Google and other search engines from enforcing, for example, biased advertising standards.

This kind of tight regulation and restriction of commercial search engines, needless to say, “is likely to be neither practically appealing nor wise, and might smack of cultural elitism

or paternalism.”<sup>14</sup> Even if the policies were worthwhile and effective—which is doubtful<sup>15</sup>—they are not, in any case, likely to pass constitutional muster (at least in the United States). Lessig, for example, describes how “the courts will defend the rights of the stations to be this biased” about their advertising standards.<sup>16</sup> More broadly, van Covering has pointed out how “[w]ithin the US, search engine results...are protected by the constitutional guarantee of freedom of speech [and] are entitled to publish whatever content they like.”<sup>17</sup>

And so, failing a complete overhaul of media law, exiting regulatory proposals seem quite unrealistic. One remaining option is to echo Google’s call for a search engine “that is transparent and in the academic realm”<sup>18</sup>—one funded not by advertising but by government or research grants, one that is designed not with commercial interests, but with citizen interests, in mind. Such a search engine might explore new research in natural language processing, perhaps relying on algorithms that look more fully at the informational richness of a page (as opposed to its popularity or commercial orientation).<sup>19</sup> Due to the financial costs of operating such a system, however, the prospects are increasingly grim. As one search engine manager points out, making such a system available would “cost you a ton of money”:

*This is why ever since 2000, 2001, most of the search research done at the universities is what I call Metacrawler-esque, which is people not building a search engine but doing something on top of a search engine, because they just can’t afford to build their own. Which is a shame, because you’re not getting these big engines coming out of academia any more.*<sup>20</sup>

The virtually insurmountable regulatory and economic challenges plaguing existing proposals for more “egalitarian” search engines suggest that the market mechanism—despite all its problems—may be, at least for now, the only practical means of getting a competitive search engine off the ground.<sup>21</sup> It could be that our best hope is for a search engine that, although commercial, does its best to avoid the pitfalls of commercialism, sacrificing democratic principles only when doing so is reasonably necessary. This may sound like wishful thinking, but we believe that Google is, for the most part, one such company. This is why Google is not “evil”—at least for now.

## IV. Moving Forward II: Future Research Directions

When we began this project two years ago, there was a troubling lack of attention given to the social and political problems raised by the design of search engines in general and Google in particular. Despite the unprecedented growth in Web access and content, search engines largely failed to capture the interest of both media scholars (who, perhaps, saw them as just another kind of Web site) and scholars of STS (who, perhaps, saw these sites as media, not as value-laden systems).

It is encouraging to find that we are not alone in trying to broaden the handful of initial investigations to include content analyses of search results,<sup>22</sup> structural economic analyses of the search industry, and discussions of advertising practices.<sup>23</sup> And the trend appears to be accelerating, with two conference panels devoted exclusively to the sociopolitics of search engines set to take place later this year.<sup>24</sup> We hope that research in this area will continue, and that technologists, political scientists, and media observers alike will continue to come together to examine how search engines might further the democratic potential of the Web.

## Notes

---

<sup>1</sup> Jacques Derrida, *Archive Fever: A Freudian Impression*, trans. Eric Prenowitz (Chicago, IL: University of Chicago Press, 1996).

<sup>2</sup> *Ibid.*, 16-17.

<sup>3</sup> Eric Schmidt, *Keynote Address at the 2004 Conference on Entrepreneurship* (Stanford, CA: Stanford University Graduate School of Business, 2004), available from <http://wesley.stanford.edu/multimedia/events/entreconference/google.ram> (accessed May 10, 2005).

<sup>4</sup> To be fair, Schmidt describes how he later spoke with the developer of Google News, who confirmed that there was a bias—in favor of cricket (a sport he enjoyed) and international stories. Nevertheless, Schmidt's tale does not recognize the more subtle ways in which search engine bias always operates.

<sup>5</sup> The connection between Google and Derrida's musings on the archive are borrowed from Hannibal, *Reading Notes: Archive Fever* (ArsTechnica, June 27 2003 [cited July 22 2003]), available from <http://arstechnica.com/news.ars/post/20030627-130.html>.

<sup>6</sup> Robert E. McGinn, *Science, Technology, and Society* (Englewood Cliffs, NJ: Prentice Hall, 1991), 98-101.

<sup>7</sup> Benjamin Compaine and Douglas Gomery, *Who Owns the Media?: Competition and Concentration in the Mass Media Industry*, 3rd ed. (Mahwah, NJ: Lawrence Erlbaum Associates, 2000), 576.

- 
- <sup>8</sup> For example, most do not know that some links are paid for, even though they are frequently labeled as “sponsored sites.” Deborah Fallows, *Search Engine Users: Internet Searchers are Confident, Satisfied, and Trusting--But They Are Also Unaware and Naive* (Pew Internet and American Life Project, January 23 2005 [cited May 22 2005]), available from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf).
- <sup>9</sup> Lucas Inrona and Helen Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," *The Information Society* 16, no. 3 (2000): 181.
- <sup>10</sup> Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York, NY: Basic Books, 1999), 108.
- <sup>11</sup> Elizabeth Van Covering, "New Media? The Political Economy of Internet Search Engines" (paper presented at the *Conference of the International Association of Media & Communications Researchers (IAMCR)*, Porto Alegre, Brazil, July 25-30 2004), 20, available from [http://personal.lse.ac.uk/vancouver/IAMCR-CTP\\_SearchEnginePoliticalEconomy\\_EVC\\_2004-07-14.pdf](http://personal.lse.ac.uk/vancouver/IAMCR-CTP_SearchEnginePoliticalEconomy_EVC_2004-07-14.pdf) (accessed May 18, 2005).
- <sup>12</sup> Nutch Project, *Nutch FAQ* (Apache Incubator, March 22 2005 [cited May 2 2005]), available from <http://incubator.apache.org/nutch/faq.html>.
- <sup>13</sup> Inrona and Nissenbaum, "Shaping the Web: Why the Politics of Search Engines Matters," 183.
- <sup>14</sup> *Ibid.*
- <sup>15</sup> As Terry Winograd pointed out to me, a prime motivation behind Google’s approach to search was to remove ‘subjective’ human judgment from the process as much as possible. While we contend that ‘subjective human judgment’ does work its way even into Google’s system (social shaping, values in design), it would seem that such regulation would need to involve humans more directly if it is to be feasible. It is unclear how we might get the sort of deliberative balance necessary without specifying such ‘subjective’ criteria (cf., the Fairness Doctrine). After all, the previous Web navigation paradigm—directories—have also been roundly criticized as antideliberative, for the wishes and biases of editors can more directly work themselves in to the results. See V. Phua, "Towards a Set of Ethical Rules for Search Engines" (MSc Dissertation, London School of Economics, 1998)..
- <sup>16</sup> Lawrence Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* (New York, NY: Penguin Books, 2004), 152.
- <sup>17</sup> Van Covering, "New Media? The Political Economy of Internet Search Engines", 23.
- <sup>18</sup> Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine" (paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998), 18, available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 1, 2005).
- <sup>19</sup> See, for instance, B. Galitsky, *Natural Language Question Answering System: Technique of Semantic Headers* (Adelaide: Advanced Knowledge International, 2003).
- <sup>20</sup> Van Covering, "New Media? The Political Economy of Internet Search Engines", 10.
- <sup>21</sup> Note that in countries with an independent but government-sponsored media, there is the possibility for a noncommercial, egalitarian search engine. Apparently the BBC, amidst controversy over ‘tainted’ paid placement, tried this in 2002. See note 3, Heather Hipsley, *Re: Complaint Requesting Investigation of Various Internet Search Engine Companies for Paid Placement and Paid Inclusion Programs* [Web] (U.S. Federal Trade Commission, June 27 2002 [cited May 20 2004]), available from <http://www.ftc.gov/os/closings/staff/commercialalertletter.htm>.
- <sup>22</sup> Susan Gerhart, "Do Web Search Engines Suppress Controversy?," *First Monday* 9, no. 1 (2004), available from [http://firstmonday.org/issues/issue9\\_1/gerhart/index.html](http://firstmonday.org/issues/issue9_1/gerhart/index.html) (accessed May 1, 2005).
- <sup>23</sup> Van Covering, "New Media? The Political Economy of Internet Search Engines".
- <sup>24</sup> The Association of Internet Researchers will be holding a panel entitled “Search Engines: Their Logics, Their Politics” in Chicago as part of their *Internet Research 6.0* conference in October 2005 (<http://conferences.aoir.org/viewabstract.php?id=38&cf=3>). A similar discussion will take place at the *Graduate Student Workshop on Values in Information System Design* (Santa Clara University, August 2005, <http://epl.scu.edu/~stsvales/>).

# Appendix

## Mathematical Proof of PageRank's "Random Surfer" Model

Recall that the PageRank of a particular page  $u$  is defined by Brin and Page as:

$$PR(u) = c + (1 - c) \left[ \sum_{i \in B_u} \frac{PR(i)}{|F_i|} \right]$$

where  $F_x$  is the set of pages  $x$  links to (forward links),  $B_u$  is the set of pages that link to  $u$  (backlinks/referring pages), and  $c$  is an arbitrary constant term.<sup>1</sup> Brin and Page give an "intuitive justification" of this formula in one of their papers:

*PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank.<sup>2</sup>*

The authors do not, however, show that this description is equivalent to the mathematical form given above. Using some basic concepts from probability, however, we can do so ourselves. For clarity, let's again use my professor's home page as an example. How likely is it that a user, randomly surfing the web, will end up on her site?

In answering this question, notice that a "random surfer" can land on her page in exactly one of two ways: either by following a link, or by getting "bored" and jumping directly to it. Consequently, we can express the probability of arriving at her page ( $p$ ) as follows:

$$\begin{aligned} \Pr\{\text{lands on } p\} &= \Pr\{\text{jumps to } p \text{ or doesn't jump, and instead follows link to } p\} \\ &= \Pr\{\text{jumps to } p\} + \Pr\{\text{doesn't jump to } p\} \times \Pr\{\text{follows link to } p\} \end{aligned}$$

For simplicity, we let  $c$  denote the likelihood that he will get bored and jump right to the professor's page. Our equation becomes:

$$\Pr\{\text{lands on } p\} = c + (1 - c) \times \Pr\{\text{follows link to } p\}$$

Now all that's left for us to figure out is the probability that our user, aimlessly following links from page to page, will arrive at my professor's cyber abode.

As it turns out, this is not too difficult. Recall from Chapter 4 that there are only two links *to* the professor's page: one from that *New York Times* article and one from my home page. Since our user must have followed one of these links, we need only consider two cases:

- *He follows link from my page to the professor's.* In this case, the user must "randomly" end up on my page, and then "randomly" click on the link to my professor's site.
- *He follows link from the Times page to the professor's.* Here, he must "randomly" end up at the *New York Times* article, and then "randomly" click on the link to my professor's site.

The probability of arriving at the professor's page via a link, clearly, is just the likelihood that one of the above events occurs.

So let's look at the first case. Here, the probability of arriving at the professor's page via my page is the likelihood of the surfer randomly landing on my page *and* randomly clicking on the link to my professor's page. By definition, the probability of a user randomly landing on my page is its PageRank. Once the user has arrived at my page, moreover, the chances of him "randomly" clicking the link to my professor's site depends on how many links I have on my page; the more links I have, the greater the chances he will follow some other link and miss the professor's page. This suggests that the likelihood of him clicking on that link is inversely proportional to the number of links on my page. Combining these facts, we see that

$$\begin{aligned}\Pr\{\text{follows link from } p \text{ to } m\} &= \Pr\{\text{lands on } m \text{ and clicks link to } p\} \\ &= \Pr\{\text{lands on } m\} \times \Pr\{\text{clicks on link to } p\} \\ &= \text{PageRank}(m) \times \frac{1}{\# \text{ links}(m)} \\ &= \frac{\text{PageRank}(m)}{\# \text{ links}(m)}\end{aligned}$$

We can follow the same steps when looking at the *New York Times* link, giving us:

$$\Pr\{\text{follows link to } p \text{ from } Times\} = \frac{PageRank(Times)}{\# \text{ links}(Times)}$$

Finally, we combine these two cases, giving us the likelihood that a user follows any link to the professor's page:

$$\begin{aligned} \Pr\{\text{follows link to } p\} &= \Pr\{\text{follows link to } p \text{ from } m \text{ or follows link to } p \text{ from } Times\} \\ &= \frac{PageRank(m)}{\# \text{ links}(m)} + \frac{PageRank(Times)}{\# \text{ links}(Times)} \\ &= \sum_{i \in B_p} \frac{PageRank(i)}{\# \text{ links}(i)} \end{aligned}$$

Replacing this result in our previous formula, we see that

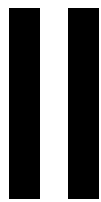
$$\Pr\{\text{lands on } p\} = c + (1 - c) \left[ \sum_{i \in B_p} \frac{PageRank(i)}{\# \text{ links}(i)} \right]$$

This, of course, is identical to our previous formula for PageRank, confirming that the “random user” interpretation is equivalent to the mathematical form. This is beneficial for at least two reasons. First, it improves our understanding of PageRank, clarifying both descriptions of the algorithm. Moreover, by reconciling the two forms, we are now free to rely on either one during our analysis.

## Notes

<sup>1</sup> Actually, this is a “slightly simplified” version of the PageRank equation, but for our purposes the formula is sufficient. For a complete version modified to take into account cycles in the link structure, see Lawrence Page et al., “The PageRank Citation Ranking: Bringing Order to the Web,” (Stanford, CA: *Stanford Digital Library Technologies Project*, 1999), 1, available from <http://dbpubs.stanford.edu/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=&name=1999-66.pdf>.

<sup>2</sup> Sergey Brin and Lawrence Page, “The Anatomy of a Large-Scale Hypertextual Web Search Engine” (paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998), 4, available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 1, 2005).



# Appendix

## Quantifying PageRank's Influence on Google Ranking

In order to properly situate our discussion of PageRank within the context of the Google search engine, we need quantify the role this algorithm plays in shaping the search results. Unfortunately, although there is wide speculation on this matter, the relationship between PageRank and result position has not been systematically analyzed in the existing literature. We therefore made an initial exploration of this question.

### I. Data Collection Method

Our approach was to obtain, for a large number of queries actually performed by users, the PageRank of each of the first 100 results returned by Google. We did this in three steps: first, we collected a reasonably large and representative number of queries (300 in all); second, we performed a Google search for each of these queries and recorded the results; finally, we acquired the PageRank of each results (30,000 in all).

To make our analysis fairly comprehensive, and to partially capture differences among types of queries (e.g., political vs. 'typical' searches), we used three large query sets as input:

- *TOP*. This set consisted of 100 extremely popular search terms. We obtained from marketing firm WordTracker the 1000 queries performed most frequently between February 12 and April 8, 2004 (this data was, in turn, compiled from *Metacrawler* logs and cleansed to remove pornographic search terms).<sup>1</sup> After removing invalid (e.g., URLs) and duplicate entries, we selected the first 100 queries from this list.



- *POLITICAL*. This set of queries focused on controversial matters that have been subjected to extensive debate—the very subjects for which a broad spectrum of opinions and rich deliberation is arguably most important (e.g., ‘affirmative action’). We compiled this list from two sources. First, we included the six subjects chosen by Hindman, et al., in their analysis of the link structure of the political Web. We also included 94 “controversial topics” listed on the Web site of online library Questia.<sup>2</sup> These, according to the site, are issues “about which people have very strong yet divergent feelings and opinions, which lead to much debate and argument. These feelings and opinions may be due to religious, social, or political beliefs. Controversial topics tend to polarize people, who often make strong arguments for and against the subjects.”<sup>3</sup>
- *RANDOM*. The last set was a more random sample of queries performed by real users, and it was intended to capture the ‘typical’ searches performed by real Web surfers. We obtained from the Stanford’s Linguistic Data Resource Center a log of about 2.4 million queries performed on the Excite search engine (released by Jack Xu on December 20, 1999).<sup>4</sup> Again we removed extraneous entries (e.g., URLs inadvertently entered into the search field), and with a random number generator selected 100 entries.

In addition, we performed some analyses on the union of all these search terms; this 300-query set is referred to hereafter as the *ALL* set. The three query sets are fully enumerated at the end of this appendix.

Once we had these sets of queries, our next step was to obtain the top 100 results returned by Google. For this, we wrote a Java program that utilized the publicly-available Google API<sup>5</sup> to automatically extract the results. The URL, title, and position of each result were stored in a separate spreadsheet for each query. There were 300 files with slightly under 30,000 results in all (some queries returned fewer than 100 matches).

Next we had to obtain the PageRank for each page returned by Google in the step above. As mentioned in Chapter 5, the immediate obstacle is that does not make available its PageRank values, and due to the recursive nature of the algorithm—each value is dependent on that of millions of other Web pages—computing them on one’s own requires vast computational resources which we did not have available. Luckily, the publicly available “Google Toolbar” displays a very rough estimate for the PageRank of the currently displayed website as a means of letting the user see the “authoritativeness” of that page.<sup>6</sup> According to reports,<sup>7</sup> moreover, the values it displays are actually the logarithm of the “real” PageRank (such that a page with PR 6 is some number *X times* more popular than than one with PR 5).

Given the large size of our data set, however, we could not manually visit and record the toolbar value for each of the 30,000 pages. We instead used the SEMonitor search engine optimization program (SEO),<sup>8</sup> which includes a tool to automate this process (it essentially emulates the toolbar’s interaction with the Google server). We ran each query file through SEMonitor and added the PageRank values to its spreadsheet file. Finally, we exponentiated the toolbar values to get the “real” PageRank values and added these as well. The latter “scaled up” PageRanks are the ones used in the analysis below.<sup>9</sup>

## II. Results

With the necessary data stored away, the next step was to find out whether, in fact, Google PageRank was somehow associated with position among the results. For each set of queries—TOP, POLITICAL, RANDOM—we merged the query spreadsheets into one file; this file had two critical columns: PageRank and position. Each line on this file—which represents one of the matches for some query in that set—constituted an observation.

Since queries vary widely in their median PageRank, however, no overall trends were able to be deduced by plotting these observations on a Cartesian graph. We suspect that this was due to the overall variance in median PageRank for the various queries and not due to there being no relationship between PR and position within a given query. With this in mind,

we computed the average PageRank at each position (from 1 to 100), using the SPSS statistical software we obtained the correlation values for each of these query sets. As can be gleaned from Table I-1, the results were strong and consistent: higher PageRank is correlated with low position number (i.e., appearance at the top of the list).

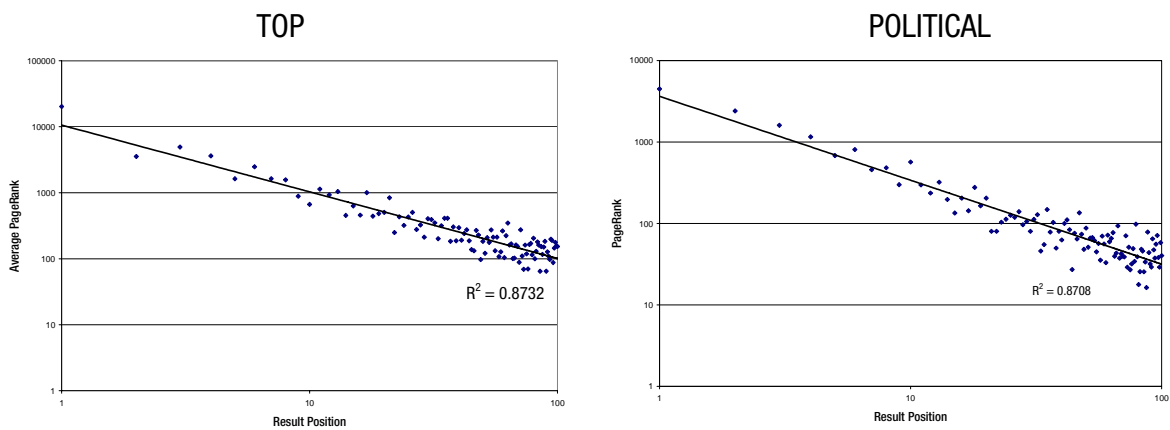
Query Set	Median Average PR	Biv. Correlation r	Power Law Regression $r^2$
Top	5 (1178.905)	-.361**	.873**
Political	4 (197.42)	-.437**	.871**
Random	3 (33.06)	-.348**	.630**
Combined	4 (197.42)	-.390**	.919**

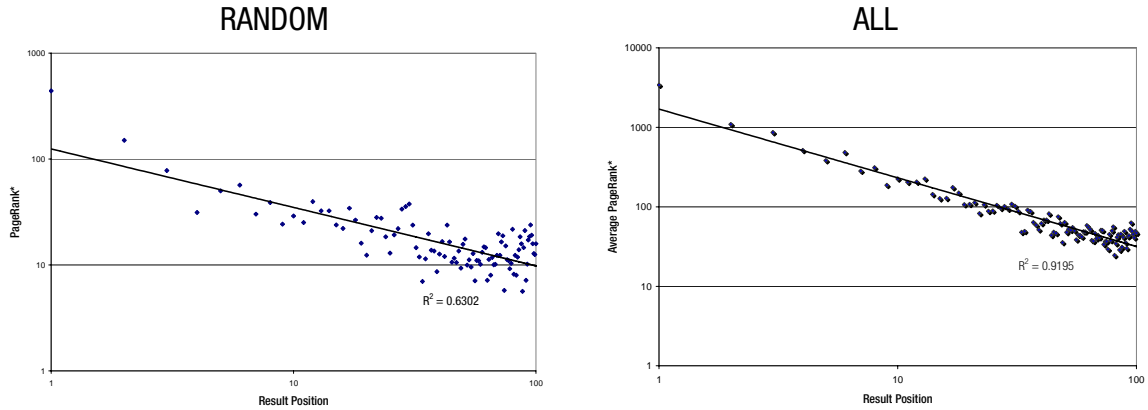
\*\* significant at the  $p < 0.001$  level

**Figure 5-4. Result Statistics.** In all cases there was a strong relationship between average PageRank and position.

We also plotted this data (average PageRank vs. position) and, as the graphs below indicate, there was indeed a very strong relationship. The graphs fit a power law regression almost perfectly:

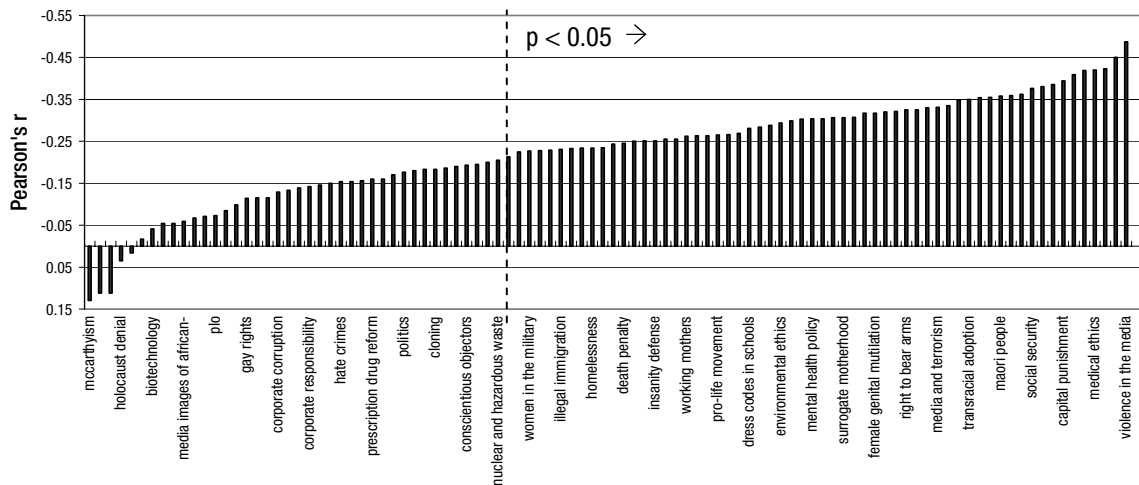
### Average PageRank by Position





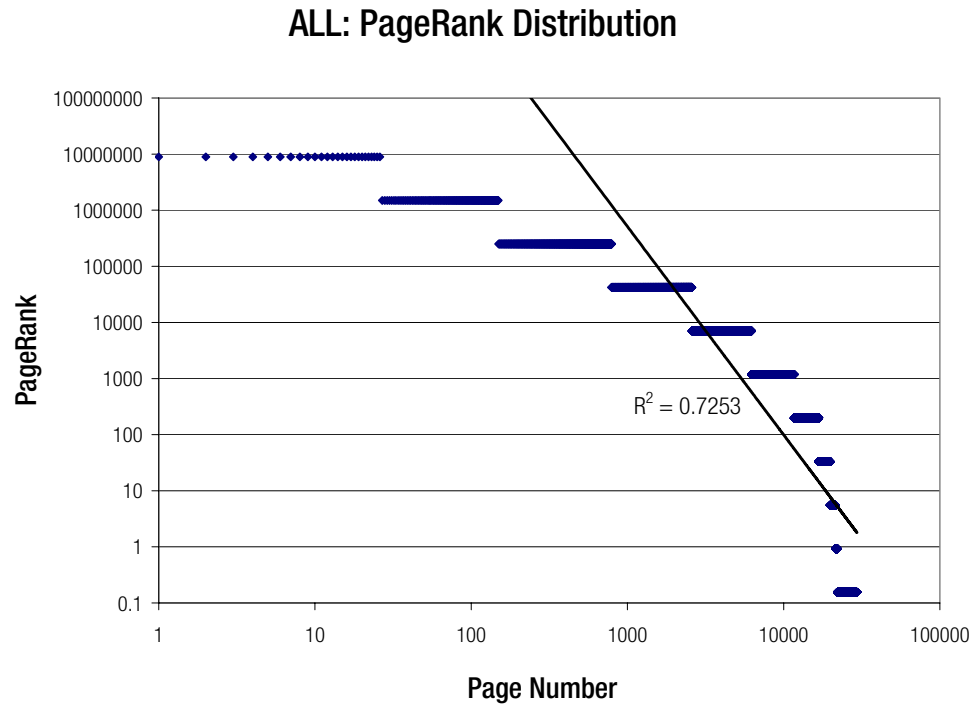
For further confirmation that this was not an “artifact of aggregation” we correlated PR and PageRank *for each query*. A very negative pearson’s *r* for all queries would suggest that even individual queries are strictly ordered by decreasing PR (as opposed to “on average”). We found that the majority of queries had a significant *r* value (on average,  $r=0.26$ ,  $p < 0.05$ ), though a good number of queries were not significantly ordered by PageRank. The graph below illustrates the correlation values for queries in the political set.

### POLITICAL: PR/Position Correlation by Query



Finally, we note that the PageRanks for all the URLs in general of were somewhat power-law *distributed*. This is consistent with prior research which suggests taking a randomly

selected set of pages on the Web, order them by decreasing PageRank, and graphing this will produce a power law curve. And indeed it does (high  $r^2$  of 0.72):



### III. Summary of Findings

For each set of queries, there was a strong, negative correlation between Google position (e.g., “1” for the first spot) and the average PageRank of pages at that position. Because the PageRank tends to decrease according to a power law, pages near the “front” of the results tend to have an exponentially greater PageRank than those less prominently placed. Although these results were highly significant in all three cases (Figure 5-4), they were more pronounced with the ‘top’ and ‘political’ queries than with the set of ‘random’ queries. This may be because random searches were often highly specific (e.g., “plastic capped jawa for sale”), producing a smaller set of relevant matches. We also found that while the aggregate “average” PageRank at each position followed a predictable pattern, the PR/position correlation for

individual queries was more varied and less significant, indicating that the results for individual searches were often not strictly ordered by decreasing PageRank.

We also note that the median PageRank was quite high on all three sets of queries, suggesting the 'typical' result was far more popular than the 'typical' page on the Web. We know according to research in complex network theory that the vast majority of pages would have a toolbar PageRank of 0 or 1.<sup>10</sup> But we found that the typical result had a PageRank of 4 on the Google Toolbar, meaning that a relatively large number of sites pointed to it (this was, again, less pronounced for the random queries, and most pronounced for the top queries). Thus, regardless of how the top 100 sites are themselves ordered, the 'typical' result had, in any case, a relatively high PageRank value.

#### IV. Query Lists

TOP Queries (In order)			
1. paris hilton	27. thongs	53. msn	79. pregnancy
2. google	28. sublime directory	54. hair styles	80. mardi gras
3. ebay	29. tattoos	55. women	81. freeones
4. yahoo	30. parent	56. best buy	82. johnny depp
5. jokes	31. jobs	57. clip art	83. irs
6. health	32. music	58. lord of the rings	84. blink 182
7. milf	33. cars	59. flowers	85. custom mini choppers
8. games	34. wallpaper	60. topless	86. map quest
9. mapquest	35. recipes	61. driving directions	87. 50 cent
10. dictionary	36. free games	62. kids	88. hairstyles
11. search engines	37. song lyrics	63. kelly blue book	89. lowes
12. thong	38. home depot	64. linkin park	90. online games
13. britney spears	39. spanking	65. used cars	91. furniture
14. hotmail	40. spybot	66. zip codes	92. free music downloads
15. kazaa lite	41. white pages	67. real estate	93. dragons
16. april fools	42. orlando bloom	68. southwest airlines	94. christina aguilera
17. lyrics	43. kazaa	69. sears	95. american idol
18. anime	44. movies	70. funny	96. spring break
19. paris hilton video	45. baby names	71. jessica simpson	97. tattoo
20. weather	46. eminem	72. free clip art	98. camel toe
21. maps	47. walmart	73. priceless	99. amazon
22. ask jeeves	48. author	74. poems	100. circuit city
23. dogs	49. janet jackson	75. south beach diet	
24. msn emotions	50. airline tickets	76. funny pictures	
25. carmen electra	51. quotes	77. thumbzilla	
26. prom dresses	52. birthdays	78. yellow pages	

POLITICAL/Controversial Keywords (Alphabetical)			
1. abortion	2. death penalty	3. gun control	4. president

- |  |                                       |                                 |                                    |
|--|---------------------------------------|---------------------------------|------------------------------------|
| 5. u.s. congress                       | 29. environmental ethics              | 56. media images of women       | 79. school violence                |
| 6. politics                            | 30. euthanasia                        | 57. media in wartime            | 80. school vouchers                |
| 7. advertising ethics                  | 31. evolution                         | 58. medical ethics              | 81. search and seizure             |
| 8. affirmative action                  | 32. false memory                      | 59. medical marijuana           | 82. separation of church and state |
| 9. animal rights                       | 33. female genital mutilation         | 60. medical malpractice         | 83. sex education                  |
| 10. anti-semitism                      | 34. foreign aid                       | 61. mental health policy        | 84. smoking laws                   |
| 11. assisted suicide                   | 35. free speech                       | 62. minimum wage                | 85. social security                |
| 12. australian aborigines              | 36. freedom of the press              | 63. miranda rule                | 86. standardized testing           |
| 13. bilingual education                | 37. gay marriage                      | 64. multiculturalism            | 87. stem cell research             |
| 14. bioethics                          | 38. gay parents                       | 65. nafta                       | 88. steroids                       |
| 15. biotechnology                      | 39. gay rights                        | 66. nuclear and hazardous waste | 89. surrogate motherhood           |
| 16. business ethics                    | 40. gays in the military              | 67. nuclear disarmament         | 90. tax reform                     |
| 17. capital punishment                 | 41. globalization                     | 68. pharmaceutical industry     | 91. teenage pregnancy              |
| 18. censorship                         | 42. hate crimes                       | 69. plo                         | 92. tibet                          |
| 19. civil liberties and anti-terrorism | 43. hate speech                       | 70. police interrogation        | 93. tobacco industry               |
| 20. class action lawsuits              | 44. health care reform                | 71. political ethics            | 94. transracial adoption           |
| 21. cloning                            | 45. holocaust denial                  | 72. pornography debate          | 95. violence in the media          |
| 22. conscientious objectors            | 46. holocaust reparations             | 73. prescription drug reform    | 96. whistleblowing                 |
| 23. corporate corruption               | 47. home schooling                    | 74. privacy                     | 97. women in the military          |
| 24. corporate responsibility           | 48. homelessness                      | 75. pro-choice movement         | 98. working mothers                |
| 25. creationism                        | 49. illegal immigration               | 76. pro-life movement           | 99. workplace diversity            |
| 26. dress codes in schools             | 50. insanity defense                  | 77. right to bear arms          | 100. year round school             |
| 27. drug legalization                  | 51. legal drinking age                | 78. second amendment            |                                    |
| 28. ebonics and black english          | 52. maori people                      |                                 |                                    |
|  | 53. mccarthyism                       |                                 |                                    |
|  | 54. media and terrorism               |                                 |                                    |
|  | 55. media images of african-americans |                                 |                                    |

### RANDOM Sample of Excite Queries

- |   |  |                                      |  |
|---|--|--------------------------------------|--|
| 1. Blue Stinger game walk-through   | 11. calendar year year 2000  | 21. puzzlers                         | 35. program is not registered properly |
| 2. vasectomy  | 12. Grand National   | 22. "ara mina"                       | 36. spleen art                         |
| 3. Nannie Little Rose   | 13. erotic massage pictures  | 23. toronto raptors                  | 37. Appliance Repair Book              |
| 4. Virginia department of education   | 14. grammar and writing help on spelling rules, suggestions, quizzes | 24. merrill lynch                    | 38. Latex-free diapers                 |
| 5. William Gates III  | 15. heisman  | 25. back that ass up                 | 39. michellin                          |
| 6. stock market company   | 16. thumbnail post   | 26. Travis Coe                       | 40. made to order memories             |
| 7. can I put a 3 inch shotgun shell in a 2 3/4 chamber on my winchester 1200 shotgun? | 17. prada luggage  | 27. MOVIE REVIEWS                    | 41. El Plebeyo cancionero              |
| 8. cnbc   | 18. where can i find a world map                                     | 28. mujeres nuas brasileiras         | 42. amway                              |
| 9. nullmodem and msdos  | 19. where can i find casual sports coats on line                     | 29. and/or                           | 43. riverdance                         |
| 10. golden globes   | 20. is chris gaines real?  | 30. recipes for shrimp gumbo         | 44. plastic caped jawa for sale        |
|   |  | 31. Does wine taste better with age? | 45. where can i buy a pocketreader     |
|   |  | 32. astrology                        | 46. phillips magnavox remote controls  |
|   |  | 33. WHAT LINUS GAMES ARE GOOD        |  |
|   |  | 34. what is usb                      |  |

- |  |  |   |  |
|--|--|---|--|
| 47. frogs                              | 60. What is an application service provider? | 71. WORDS TO TLC SONGS                        | 87. public exhibitionism                 |
| 48. sex AND story                      | 61. "Council of the Arts" +Norwich           | 72. grafico                                   | flashers +free                           |
| 49. costume dresses from 1940's movies | 62. latin babes                              | 73. grenadier regiment                        | 88. museum                               |
| 50. sport mart                         | 63. how do you do a NSLOOKUP in windows 95   | 74. "mermaid"                                 | 89. population                           |
| 51. Sex Dvd                            | 64. da vinci                                 | 75. yellow pages                              | 90. Scrooge by Albert Finney             |
| 52. CHEERLEADERS NUDE CHEERS           | 65. lockwood vineyards                       | 76. Urban Nissan                              | 91. white shadow                         |
| 53. indian motorcycles                 | 66. THUMBNAI LS GAY MEN FUCKING UP THE ASS   | 77. Fcc ID lookup                             | 92. rocky mountain spotted fever in dogs |
| 54. vanagon gas heater                 | 67. lincoln vehicles                         | 78. benefit dynamics pension                  | 93. led zeppelin font                    |
| 55. vcs computers                      | 68. Where is Lake Wawasee?                   | 79. big charts                                | 94. standing rib roast                   |
| 56. Sparkle Power Supply website       | 69. art teacher resource site                | 80. shingles, treatment                       | 95. wespentaille                         |
| 57. richards realm                     | 70. hermaphrodite                            | 81. kelly brook                               | 96. hun                                  |
| 58. Russian computer languages         |  | 82. painting fiberglass                       | 97. build a bridge truss                 |
| 59. grand admiral doenitz              |  | 83. what else happened besides the Y2K rumor? | 98. airfares to hawaii                   |
|  |  | 84. pine needles                              | 99. conxion                              |
|  |  | 85. Carribean sailing                         | 100. adaptive clothing wheelchair skirts |
|  |  | 86. screensavers                              |  |

## Notes

- <sup>1</sup> The keywords were obtained via a paid subscription to WordTracker's service, which, in addition to various metrics useful to search marketers, provides all its members a detailed list of the Web's most popular keywords. See <http://www.wordtracker.com>.
- <sup>2</sup> *Questia.com*, <http://www.questia.com/library/sociology-and-anthropology/social-issues/controversial-topics.jsp>
- <sup>3</sup> *Ibid.*
- <sup>4</sup> Originally posted on <ftp.excite.com/pub/jack>. Available via Stanford AFS as <afs/ir/data/linguistic-data/IR/Excite>.
- <sup>5</sup> For information on the Google API, see <http://www.google.com/apis/>.
- <sup>6</sup> The Google toolbar is available at <http://toolbar.google.com>.
- <sup>7</sup> See Bob Wakfer, "Google PageRank & How to Get it: Part 1." *InfoPool Online News and Views* (June 2004 [cited May 10, 2005]). Available at <http://www.compar.com/infopool/articles/PR-calculation.html>.
- <sup>8</sup> Semonitor is available from FlamingoSoft at <http://www.semonitor.com>.
- <sup>9</sup> Wakfer, "Google PageRank."
- <sup>10</sup> Note, however, that Upstill found a lognormal distribution of toolbar PageRank in his crawls of corporate, which would suggest that "normalization and transformation" are being done to the *toolbar* values, and that the median might actually be near 5. These crawls, however, were small and focused on company sites—the same sort of communities Pennock found to exhibit lognormal link distributions. In addition, small crawls tend to be implicitly biased towards high-PR sites (since they follow links to find new pages, and are thus more likely to come across heavily linked pages). While Upstill et al suggest that the toolbar PR is not a true indicator of actual PR, we believe that there is not sufficient evidence to support this claim. We would need a true "random" sample of pages on the Web, and the toolbar PR for each of these, to know for certain. We attempted contacting various crawlers but they are all either outdated (OCLC) or biased by PageRank (Stanford CS WebVac).



# Bibliography

- Adamic, Lada, and Bernardo Huberman. "The Web's Hidden Order." *Communications of the ACM* 44, no. 9 (2001): 55-60.
- Albert, Réka, Hawoong Jeong, and Albert-Laszló Barabási. "Diameter of the World-Wide Web." *Nature* 401, no. 9 (1990): 130.
- Asprey, Robert B. *Frederick the Great: The Magnificent Enigma*. New York, NY: Ticknor & Fields, 1986.
- Aufderheide, Patricia. "After the Fairness Doctrine: Controversial Broadcast Programming and the Public Interest." *Journal of Communication* 40, no. 3 (1990): 47-72.
- Ayres, Ian, and Barry Nalebuff. "Going Soft on Microsoft? The EU's Antitrust Case and Remedy." *The Economists' Voice* 2, no. 2 (2005). Available from <http://www.bepress.com/ev/vol2/iss2/art4>
- Bagdikian, Ben. *The Media Monopoly*. Fourth ed. Boston, MA: Beacon Press, 1992.
- Barabási, Albert-Laszló. *Linked: The New Science of Networks*. Cambridge, MA: Perseus Pub., 2002.
- Barabási, Albert-Laszló, and Réka Albert. "Emergence of Scaling in Random Networks." *Science* 286, no. 5439 (1999): 509-12.
- Benhabib, Seyla. "Toward a Deliberative Model of Democratic Legitimacy." In *Democracy and Difference: Contesting the Boundaries of the Political*, edited by Seyla Benhabib. Princeton, NJ: Princeton University Press, 1996.
- Berners-Lee, Tim. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. San Francisco, CA: Harper San Francisco, 1999.
- Bhargava, Hemant, and Juan Feng. "Paid Placement Strategies for Internet Search Engines." Paper presented at the *Eleventh International World Wide Web Conference (WWW2002)*, Honolulu, Hawaii 2002.
- Blood, Rebecca. *Weblogs: A History and Perspective*. Rebecca's Pocket, September 7, 2000 [cited May 15 2005]. Available from [http://www.rebeccablood.net/essays/weblog\\_history.html](http://www.rebeccablood.net/essays/weblog_history.html).
- Brin, Sergey, and Larry Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." Paper presented at the *Seventh International World Wide Web Conference*, Brisbane, Australia, April 1998. Available from <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed May 1, 2005).
- Compaine, Benjamin, and Douglas Gomery. *Who Owns the Media?: Competition and Concentration in the Mass Media Industry*. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- Cooper, Mark. *Media Ownership and Democracy in the Digital Information Age: Promoting Diversity with First Amendment Principles and Market Structure Analysis*. Stanford, CA: The Center for Internet and Society at Stanford Law School, 2004. Available from <http://cyberlaw.stanford.edu/blogs/cooper/archives/mediabooke.pdf> (accessed March 22, 2005).
- Cooper, Mark, and Steven Cooper. *Hope and Hype v. Reality: The Role of the Commercial Internet in Democratic Discourse and Prospects for Institutional Change*. Stanford

- Law School Center for Internet and Society, 2003 [cited May 7 2005]. Available from <http://cyberlaw.stanford.edu/blogs/cooper/archives/HOPEALL.pdf>.
- Cooper, W. S. "Some Inconsistencies and Misidentified Modelling Assumptions in Probabilistic Information Retrieval." *ACM Transactions on Information Systems* 13 (1995): 100-11.
- Cornfield, Michael, and Lee Rainie. *Untuned Keyboards: Online Campaigners, Citizens, and Portals in the 2002 Elections*. Pew Internet and American Life Project, March 21, 2004 [cited May 11 2005]. Available from [http://www.pewinternet.org/pdfs/PIP\\_IPDI\\_Politics\\_Report.pdf](http://www.pewinternet.org/pdfs/PIP_IPDI_Politics_Report.pdf).
- Cunnigham, Frank. *Theories of Democracy: A Critical Introduction*. New York, NY: Routledge, 2002.
- Dahl, Robert. *Democracy and Its Critics*. New Haven, CT: Yale University Press, 1989.
- Dewey, John. *The Public and Its Problems*. New York, NY: Henry Holt and Company, 1927.
- Drezner, Daniel, and Henry Farrell. "The Power and Politics of Blogs." Paper presented at the *American Political Science Association (APSA) Annual Meeting*, Chicago, IL 2004. Available from <http://www.uts.utoronto.ca/~farrell/blogpaperfinal.pdf> (accessed May 1, 2004).
- Ellis, David. "Theory and Explanation in Information Retrieval Research." *Journal of Information Science* 8 (1984): 25-38.
- Elster, Jon, ed. *Deliberative Democracy*. Cambridge, UK: Cambridge University Press, 1998.
- Fallows, Deborah. *Search Engine Users: Internet Searchers are Confident, Satisfied, and Trusting--But They Are Also Unaware and Naive*. Pew Internet and American Life Project, January 23, 2005 [cited May 22 2005]. Available from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf).
- Fallows, Deborah, and Lee Rainie. *The Popularity and Importance of Search Engines*. Pew Internet and American Life Project, August, 2004 [cited May 10 2005]. Available from [http://www.pewinternet.org/pdfs/PIP\\_Data\\_Memo\\_Searchengines.pdf](http://www.pewinternet.org/pdfs/PIP_Data_Memo_Searchengines.pdf).
- Faloutsos, Michalis, Petros Faloutsos, and Christos Faloutsos. "On Power-Law Relationships of the Internet Topology." Paper presented at the *Special Interest Group on Data Communications (SIGCOMM)*, Cambridge, MA, August 1999.
- Ferguson, Charles H. "What's Next for Google?: The Search Firm Wants to Organize All Digital Information. That Means War with Microsoft." *Technology Review: MIT's Magazine of Innovation*, January 2005, 38-46.
- Fishkin, James. *Democracy and Deliberation: New Directions for Democratic Reform*. New Haven, CT: Yale University Press, 1991.
- Frederick the Great. "Benevolent Despotism." In *The Portable Enlightenment Reader*, edited by Isaac Kramnick, 452-59. New York, NY: Penguin Books, 1995.
- Friedman, Batya, and Helen Nissenbaum. "Bias in Computer Systems." *ACM Transactions on Information Systems* 14, no. 3 (1996): 330-47.
- Gagliardo, John. *Enlightened Despotism*. New York, NY: Thomas Y. Crowell Company, 1967.
- Galitsky, Boris, and Mark Levene. "On the Economy of Web Links: Simulating the Exchange Process." *First Monday* 9, no. 1 (2004). Available from [http://firstmonday.org/issues/issue9\\_1/galitsky/index.html](http://firstmonday.org/issues/issue9_1/galitsky/index.html)
- Gerhart, Susan. "Do Web Search Engines Suppress Controversy?" *First Monday* 9, no. 1 (2004). Available from [http://firstmonday.org/issues/issue9\\_1/gerhart/index.html](http://firstmonday.org/issues/issue9_1/gerhart/index.html) (accessed May 1, 2005).
- Goodin, Robert. "Democratic Deliberation Within." In *Debating Deliberative Democracy*, edited by James Fishkin and Peter Laslett, 54-78. Malden, MA: Blackwell Publishing, 2003.
- Google Inc. *United States Securities and Exchange Commission Form S-1*. United States Securities and Exchange Commission, April 29, 2004 [cited May 7 2005]. Available via EDGAR database from <http://www.sec.gov/Archives/edgar/data/1288776/000119312504139655/ds1a.htm>.
- . *Factory Tour: Google.com*, 2005. Video. Available from

- <http://www.google.com/intl/en/press/factorytour.html> (accessed May 20, 2005).
- . *United States Security and Exchange Commission Form 10-K*. United State Securities and Exchange Commission, March 30, 2005 [cited May 8 2005]. Available from <http://www.edgar-online.com/bin/cobrand/?doc=A-1288776-0001193125-05-065298>.
- Gordon, Michael, and Praveen Pathak. "Finding Information on the World Wide Web: The Retrieval Effectiveness of Search Engines." *Information Processing & Management* 35, no. 2 (1999): 141-80.
- Graham, Gordon. "The Moral Basis of Democracy." *International Journal of Moral and Social Studies* 7, no. 2 (1992).
- Gutmann, Amy, and Dennis Thompson. *Democracy and Disagreement*. Cambridge, UK: Belknap Press, 1998.
- Habermas, Jürgen. *Moral Consciousness and Communicative Action*. Cambridge, UK: MIT Press, 1990.
- Hadenius, Axel. *Democracy and Development*. Cambridge, UK: Cambridge University Press, 1992.
- Hamilton, Alexander, James Madison, and John Jay. *The Federalist Papers*. Edited by Karen M. Elder. Richmond, VA: Westvaco, 1995.
- Hannibal. *Reading Notes: Archive Fever*. ArsTechnica OpenForum, June 27, 2003 [cited May 11 2005]. Available from <http://episteme.arstechnica.com/6/ubb.x?a=tpc&s=50009562&f=174096756&m=2010925275&r=2010925275>.
- Hiler, John. *Google (hearts) Blogs*. Microcontent News, February 26, 2002 [cited May 6 2005]. Available from <http://www.microcontentnews.com/articles/googleblogs.htm>.
- Hindman, Matthew, Kostas Tsioutsoulouklis, and Judy Johnson. *'Googlearchy': How a Few Heavily Linked Sites Dominate Politics on the Web*. July 28, 2003 [cited April 10 2005]. Available from <http://www.princeton.edu/~mhindman/googlearchy--hindman.pdf>.
- Hipplesley, Heather. *Re: Complaint Requesting Investigation of Various Internet Search Engine Companies for Paid Placement and Paid Inclusion Programs* [Web]. U.S. Federal Trade Commission, June 27, 2002 [cited May 20 2004]. Available from <http://www.ftc.gov/os/closings/staff/commercialalertletter.htm>.
- Huberman, Bernardo. *The Laws of the Web: Patterns in the Ecology of Information*. Cambridge, MA: MIT Press, 2001.
- Introna, Lucas, and Helen Nissenbaum. "Shaping the Web: Why the Politics of Search Engines Matters." *The Information Society* 16, no. 3 (2000): 169-85.
- Jenkins, Henry, and David Thorburn, eds. *Democracy and New Media*. Cambridge, MA: MIT Press, 2003.
- Jeorges, Bernward. "Do Politics Have Artifacts?" *Social Studies of Science* 29, no. 3 (1999): 411-31.
- Kant, Immanuel. "What Is Enlightenment?" In *The Portable Enlightenment Reader*, edited by Isaac Kramnick, 1-7. New York, NY: Penguin Books, 1995.
- Kapor, Mitchell. "Where Is the Digital Highway Really Heading? The Case for a Jeffersonian Information Policy." *Wired*, Jul/Aug 1993. Available from [http://www.wired.com/wired/archive/1.03/kapor.on.nii\\_pr.html](http://www.wired.com/wired/archive/1.03/kapor.on.nii_pr.html) (accessed September 19, 2003).
- Katz, Elihu. "The Two-Step Flow of Communication: An Up-To-Date Deport of an Hypothesis." In *Marketing Classics: A Selection of Influential Articles*, edited by Ben Enis and Keith Cox, 175-93. Boston, MA: Allyn and Bacon, 1973.
- Kavassalis, Petros, Stelios Lelis, Muhamoud Rafea, and Seif Haridi. "What Makes a Web Site Popular?" *Communications of the ACM* 47, no. 2 (2004): 50-55.
- Kenny, Charles. *Financing Information and Communication Infrastructure Needs in the Developing World: Public and Private Roles*. Global Information and Communication Technologies, World Bank, February, 2005 [cited May 20 2005]. Available from <http://lnweb18.worldbank.org/ict/resources.nsf/InfoResources/04C3CE1B933921A5>

85256FB60051B8F5.

- Klein, Naomi. *No Logo*. New York, NY: Picador, 2002.
- Kleinberg, Jon. "Authoritative Sources in a Hyperlinked Environment." *Journal of the ACM* 46, no. 5 (1999): 604-32.
- Kleinberg, Jon, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins. "The Web as a Graph: Measurements, Models, and Methods." In *Computing and Combinatorics: Proceedings of the 5th Annual International Conference (COCOON'99)*, edited by T. Asano, H. Imai, D.T. Lee, S. Nakano and T. Tokuyama, 1-17. Tokyo, Japan: Springer-Verlag, 1999.
- Krieger, Leonard. *An Essay on the Theory of Enlightened Despotism*. Chicago, IL: University of Chicago Press, 1975.
- Lessig, Lawrence. *Code and Other Laws of Cyberspace*. New York, NY: Basic Books, 1999.
- . *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. New York, NY: Penguin Books, 2004.
- Magaziner, Ira. "Democracy and Cyberspace: First Principles." In *Democracy and New Media*, edited by Henry Jenkins and David Thorburn. Cambridge, MA: MIT Press, 2003.
- Mazzocco, Dennis W. *Networks of Power: Corporate TV's Threat to Democracy*. Boston: South End Press, 1994.
- McChesney, Robert. *Rich Media, Poor Democracy: Communication Politics in Dubious Times*. Urbana, IL: University of Illinois Press, 1999.
- McGinn, Robert E. *Science, Technology, and Society*. Englewood Cliffs, NJ: Prentice Hall, 1991.
- McHugh, Josh. "Google Sells Its Soul." *Wired*, January 2003, 129-35.
- McLean, Iain. *Democracy and New Technology*. Cambridge, UK: Polity Press, 1989.
- Mill, John Stuart. *On Liberty*. Edited by Elizabeth Rapaport. Indianapolis, IN: Hackett Publishing Company, 1978.
- Miller, Joe. "J.S. Mill on Plural Voting, Competence and Participation." *History of Political Thought* 24, no. 4 (2003): 647-67.
- Mowshowitz, Abbe, and Akira Kawaguchi. "Bias on the Web." *Communications of the ACM* 45, no. 9 (2002): 56-60.
- O'Harrow, Robert. *No Place to Hide: Behind the Scenes of Our Emerging Surveillance Society*. New York, NY: Free Press, 2005.
- Page, Benjamin. *Who Deliberates?* Chicago, IL: University of Chicago Press, 1996.
- Page, Larry, Sergey Brin, Rajeev Motwani, and Terry Winograd. "The PageRank Citation Ranking: Bringing Order to the Web." Stanford, CA: Stanford Digital Library Technologies Project, 1999. Available from <http://dbpubs.stanford.edu/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=&name=1999-66.pdf>
- Pagendarm, Magnus, and Heike Schaumburg. "Why Are Users Banner-Blind? The Impact of Navigation Style on the Perception of Web Banners." *Journal of Digital Information* 2, no. 1 (2001). Available from <http://jodi.ecs.soton.ac.uk/Articles/v02/i01/Pagendarm/>
- Palmer, R.R., and Joel Colton. *A History of the Modern World*. New York, NY: Alfred A. Knopf, Inc., 1984.
- Pennock, David, Gary Flake, Steve Lawrence, Eric Glover, and C. Lee Giles. "Winners Don't Take All: Characterizing the Competition for Links on the Web." *Proceedings of the National Academy of Sciences* 99, no. 8 (2002): 5207-11.
- Phua, V. "Towards a Set of Ethical Rules for Search Engines." MSc Dissertation, London School of Economics, 1998.
- Picard, Robert. *The Economics of Financing Media Companies*. New York, NY: Fordham University Press, 2002.
- Piott, Steven. *Giving Voters a Voice: The Origins of the Initiative and Referendum in America*. Columbia, MO: University of Missouri Press, 2003.
- Pringle, Glen, Lloyd Allison, and David Dowe. "What is a Tall Poppy among Web Pages?" Paper presented at the *Seventh International World Wide Web Conference*, Brisbane,

- Australia, April 1998. Available from [http://www.csse.monash.edu.au/~lloyd/tilde/InterNet/Search/1998\\_WWW7.html](http://www.csse.monash.edu.au/~lloyd/tilde/InterNet/Search/1998_WWW7.html) (accessed May 12, 2005).
- Raphael, Todd. "At Google, the Proof is in the People." *Workforce*, March 2003, 50-51.
- Roberts, John. "The Enigma of Free Speech: Speakers' Corner, The Geography of Governance and a Crisis of Rationality." *Social and Legal Studies* 9, no. 2 (2000): 271-92.
- Rodzvilla, John, ed. *We've got Blog: How Weblogs are Changing our Culture*. Cambridge, MA: Perseus Publishing, 2002.
- Rousseau, Jean-Jacques. *The Social Contract*. Translated by Maurice Cranston. New York, NY: Penguin Books, 1968.
- Salton, Gerard. "The State of Retrieval System Evaluation." *Information Processing & Management* 28, no. 4 (1992): 441-49.
- Samples, John, and Adam Thierer. *Why Subsidize the Soapbox? The McCain Free Airtime Proposal and the Future of Broadcasting*. CATO Institute, Analysis 480, August 6, 2003 [cited May 8 2005]. Available from <http://www.cato.org/pubs/pas/pa-480es.html>.
- Saward, Michael. *The Terms of Democracy*. Cambridge, UK: Polity Press, 1998.
- Schmidt, Eric. *Keynote Address at the 2004 Conference on Entrepreneurship*. Stanford, CA: Stanford University Graduate School of Business, 2004. Available from <http://wesley.stanford.edu/multimedia/events/entrepconference/google.ram> (accessed May 10, 2005).
- Schuler, Douglas. "Reports of the Close Relationship between Democracy and the Internet May Have Been Exaggerated." In *Democracy and New Media*, edited by Henry Jenkins and David Thorburn. Cambridge, MA: MIT Press, 2003.
- Sheff, David. "Playboy Interview: Google Guys." *Playboy*, September 2004.
- Sheu, Tair-Rong, and Kathleen Carley. "Monopoly Power on the Web: A Preliminary Investigation of Search Engines." Paper presented at the *29th Telecommunications Policy Research Conference*, Alexandria, VA 2001. Available from <http://arxiv.org/abs/cs.CY/0109054> (accessed May 7, 2005).
- Skinner, Quentin. "The Empirical Theorists Of Democracy and their Critics: A Plague on Both their Houses." *Political Theory* 1, no. 3 (1973): 287-306.
- Sperti, Angioletta. *The Public Forum Doctrine and Its Possible Application to the Internet*. UCLA Online Institute for Cyberspace Law and Policy, 1997 [cited April 18 2004]. Available from <http://www.gseis.ucla.edu/iclp/asperti.html>.
- Sunstein, Cass. "Deliberation, Democracy and Disagreement." In *Justice and Democracy: Cross-cultural Perspectives*, edited by Ronald Bontekoe and Marietta Stepaniants, 93-117. Honolulu, HI: University of Hawai'i Press, 1997.
- . *Republic.com*. Princeton, NJ: Princeton University Press, 2002.
- Syverson, Chad. "Product Substitutability and Productivity Dispersion." *The Review of Economics and Statistics* 86, no. 2 (2004): 534-50.
- Tagliacozzo, Renata. "Self-Citations in Scientific Literature." *Journal of Documentation* 33, no. 4 (1977): 251-65.
- Telang, Rahul, Tridas Mukhopadhyay, and Ronald Wilcox. "An Empirical Analysis of the Antecedants of Internet Search Engines Choice." Paper presented at the *Workshop on Information Systems and Economics*, Charlotte, NC 1999.
- Van Covering, Elizabeth. "New Media? The Political Economy of Internet Search Engines." Paper presented at the *Conference of the International Association of Media & Communications Researchers (IAMCR)*, Porto Alegre, Brazil, July 25-30 2004. Available from [http://personal.lse.ac.uk/vancouve/IAMCR-CTP\\_SearchEnginePoliticalEconomy\\_EVC\\_2004-07-14.pdf](http://personal.lse.ac.uk/vancouve/IAMCR-CTP_SearchEnginePoliticalEconomy_EVC_2004-07-14.pdf) (accessed May 18, 2005).
- Weinberger, David. *Small Pieces Loosely Joined: A Unified Theory of the Web*. Cambridge, MA: Perseus Publishing, 2002.
- Winner, Langdon. "Do Artifacts Have Politics?" *Daedalus* 109, no. 1 (1980): 121-36.
- Woolgar, Steve. "Do Artefacts Have Ambivalence?: Moses' Bridges, Winner's Bridges, and

Other Urban Legends." *Social Studies of Science* 29, no. 3 (1999): 433-49.

Yamashita, Hiromi, and Christopher Williams. "A Vote for Consensus: Democracy and Difference in Japan." *Comparative Education* 38, no. 3 (2002): 277-89.

Zittrain, Jonathan, and Benjamin Edelman. *Localized Google Search Result Exclusions*. Berkman Center for Internet & Society at Harvard Law School, October 26, 2002 [cited November 2 2004]. Available from <http://cyber.law.harvard.edu/filtering/google/>.