

citation: Sperber, Lauren. "A Proprietary Protocol: How Search Defines Authority Online." M.A. thesis, New York University, 2007. [date you accessed the pdf] <<http://kenspeckle.net/papers/a-proprietary-protocol.pdf>>.

A Proprietary Protocol: How Search Defines Authority Online

Lauren Sperber

Submitted in the Department of English, Graduate School of Arts and Science, in partial fulfillment of the requirements for the degree of Master of Arts at New York University.

Approved:

Clifford Siskin

May 4, 2007

The internet presents a unique challenge in our attempt to access and determine the authoritativeness of the information it contains. The ways we access and evaluate online information depart from the standards applicable to print information not only because of the volume of material available and the low barrier to publish information oneself, but—more importantly—due to way that information is structured. In his book *Weaving the Web*, Tim Berners-Lee emphasizes the organizational structure of the web as its most revolutionary technological development, as opposed to the more commonly acknowledged democratization of the ability to publish one's ideas. Print information arrives embedded in structures we've developed throughout the history of print to easily identify its content and determine to what degree we should trust it in the form of Library of Congress cataloging information, author biographies, academic journals, and publishing house logos.

Print enforces a linear presentation of information, so our understanding of a given publication's content rests upon the way its author structures that content within the parameters of a linear progression. Chapters outline the author's argument, and any references to other publications are clarified to an extent that would allow the reader to access them only at the end. Even with the publication information provided in a works cited list or bibliography, the reader must actively pursue the works an author references, traveling to the library or a

bookstore and investing time in locating the original works. Although a single website may certainly adhere to the linear format to which we've grown accustomed in print, it might just as easily break from that setup, using hyperlinks to bring users immediately to referenced information both internally (elsewhere on the same site) or externally.

According to Berners-Lee, the thinking leading to his formulation of the web sprang from a desire to store information "without using structures like matrices or trees" (10). Drawing on the patterns of human thought, which "break[s] out" of such structures to "make intuitive leaps across [...] boundaries," Berners-Lee created a program to store information about his chaotic workplace as nodes that must be connected to each other via a specifically defined relationship. From this program he developed the idea of linking "all the information stored on computers everywhere," enabling computers to "represent associations between things that might seem unrelated" (4). This web-like structure of information, which makes the distributed network structure of the internet of which the web is a part more widely accessible, offers no single entry point, no inherent organization of its content, and no attribution of value to the indicators of relationship—links—that connect its elements to each other.

While Berners-Lee's organizational program for maintaining information about his office provided different types of links to differentiate

between, say, a relationship between an engineer and a program she runs and one between a program and a piece of hardware on which it operates, the web connects its pieces without providing any more details than the most basic structural outline: If we know that website A links to website B and C, whereas website B links only to website A and website C links to neither, we still cannot determine much about the nature of the relationship among the three sites—at least not without investigating their content as well, an undertaking that would require an impossibly enormous amount of time if we were to attempt to investigate the relationships among all the sites on the web.

Search engines, however, use mathematical algorithms based upon that link structure and the content they find on websites to determine the order in which they rank their results for users' queries. Search engines' increasing success in helping their users find information online has led us to use them as entry points to the unorganized distributed network of the internet, both "to recover that which we know exists on the Web, and to discover that which we assume must be there," in the words of John Battelle (32). This paper will investigate how the algorithms that order engines' results create and perpetuate a practical authority both for the sites that search engines link to and for the engines themselves, and how that authority shapes our access to and perception of information online.

I. What is Authority?

Before exploring the construction of authority online, we should briefly consider how authority functions offline, both in the interactions among humans and in our evaluation of information. Kant explores the complex relationship between these two primary arenas for authority (behavior and thought) in his 1784 essay "An Answer to the Question: What is Enlightenment?" Though Kant initially exhorts his readers, "Have the courage to use your own understanding!" at the essay's beginning, he also addresses the subtlety with which this command must be executed. According to Kant, "the use that anyone as a scholar makes of reason before the entire literate world" "must always be free," whereas the use of reason "which a person may make in a civic post or office that has been entrusted to him" "may [...] often be very narrowly restricted," referring to the two types of reason as public and private, respectively. Kant explains:

...in many affairs conducted in the interests of a community, [...] some of its members must conduct themselves in an entirely passive manner so that through an artificial unanimity the government may guide them toward public ends. Here one certainly must not argue, instead one must obey.

The distinction between "operational" and "noetic" authority made by Roger Shinn elucidates somewhat Kant's distinction between private and public uses of reason. Shinn describes operational authority as authority over the decisions that "society requires" (93) because they are "necessary for social functioning, even for social existence" (98)—parallel to Kant's

concept of the submission with which civil servants must exercise private reason. Noetic authority, however, corresponds to Kant's public use of scholarly reason. In Shinn's words, it "include[s] all apprehension of truth" (93) and "requires the judgment of those competent to assess the evidence and understand" (98).

A. Enforcement of Behavior

Shinn acknowledges that "the two types of authority are usually intermingled" because "effective operational decisions require expertise" (98), and Clyde Manschreck explores the effect of that intermingling on the sort of authority that enforces behavior in his introduction to his 1971 collection of essays *Erosion of Authority*, in which Shinn's essay appears. Manschreck explains Carl Friedrich's concept that any authority that allows people have to enforce the behavior of other people must necessarily be based on the authority of the enforcers to influence the beliefs of those upon whom their authority is enforced. Friedrich believes, in Manschreck's words, that "authority is not simply sheer power. A rational basis or component is crucial in authority, and when that component deteriorates, authority deteriorates" (10). But Manschreck writes from the vantage point of the U.S. in the early 1970s, which he describes as "a climate of opinion in which there are no universally acceptable standards of values" (12).

Although he agrees with Friedrich's premise that authority must have a rational basis, Manschreck fears that the then-contemporary nihilistic sense "that there is no absolute truth that *reason* can validate" (10) may be bringing upon his society "the night of meaningless violence" that Friedrich anticipated might result from a loss of "obligation with regard to the potentiality of reasoned elaboration" (10-11). "Ambiguity reigns" in Manschreck's time, and "there may not even be a valid moral order" (12). He observes that "in the course of the centuries, men have not lost only a sense of the transcendent, but also a trust in reason, for reason is finite and subjective, and cannot therefore establish ultimate values." Without this trust in reason, he concludes, "no final authority can be validated" (19).

Though concerned by the potential chaos of moral relativism brought about by the loss of "a trust in reason" (19), Manschreck ultimately finds that "man seems unable to bear the burden of absolute freedom and autonomy" (29)—a phenomenon to which he attributes "the spectacle of massive totalitarianism" (28). He cites the rise of Nazism, the despotism of Napoleon, and the restoration of the English monarchy after the Puritan Revolution as examples of people's willingness to accept even tyranny to "escape the specter of anarchy, frustration, and meaninglessness that are inherent in individual autonomy" (29). His hope for the future of authority, however, is much more optimistic: "Authority is

necessary for human community, but the obedience that undergirds authority must be made freely and responsibly" (32). This ideal aligns remarkably well with Kant's recommendation that an enlightened ruler "prescribe nothing [and] allow men complete freedom in religious matters," saying to his subjects, "Argue as much as you want and about what you want, but obey!"

B. Influence of Thought

This type of leadership, in Kant's view, would permit enlightenment, which he defines as "man's emergence from his self-imposed immaturity"—immaturity being man's "inability to use [his] understanding without guidance from another" due to a "lack of resolve and courage." Even in the scholarly public arena, where we might feel free to express our opinions without the concern of interfering with the operational authority of our leaders, Kant observes that people frequently "remain in lifelong immaturity" with "a book to serve as [their] reason." He blames this phenomenon partially upon laziness and partially upon the tendencies of the guardians of thought—such as authors, religious leaders, politicians, etc.—to perpetuate peoples' need of them by making independent thought seem "very dangerous, not to mention difficult."

The number of books—and other sources of information—at our disposal has grown exponentially since Kant's time. Since we no longer

have just one book to serve as our reason, but a seemingly infinite number of books from which to choose, merely selecting one to read, if undertaken independently, would amount to an enlightened, mature act of reason.

Kevin Kelly makes this point in *Out of Control* by quoting Richard Dawkins's assertion that, "effective searching procedures become, when the search-space is sufficiently large, indistinguishable from true creativity," along with Denis Diderot's 1755 prediction that "a time will come when it will be almost as difficult to learn anything from books as from the direct study of the whole universe," adding in his own words that "in the library of all possible books, finding a particular book is equivalent to writing it" (280). While Kant viewed the "rules and formulas" that might assist us in finding information as "mechanical aids to the [...] misuse of man's natural gifts" and thus "the shackles of a permanent authority," such formulas can lend significant insight when we apply them to understand the patterns by which information influences us, rather than to decide which information we will be influenced by.

One group of formulas by which we attempt to quantify the rather subjective and abstract workings of authority is bibliometrics, which Jon Kleinberg defines as "the study of written documents and their citation structure" (15) and uses as inspiration for his model of determining the influence of websites in his paper "Authoritative Sources in a Hyperlinked Environment." As Kleinberg explains, bibliometrics attempts to "use [...]

citations to produce quantitative estimates of the importance and 'impact' of individual scientific papers and journals." He describes Eugene Garfield's "impact factor," which is "the average number of citations received by papers published in the previous two years" of a given journal and a "more subtle citation-based measure of standing" proposed by Gabriel Pinski and Francis Narin. Pinski and Narin proposed that "not all citations are equally important" (15) and developed a measure called "influence weight" to determine the degree of influence that should be attributed to citations from a given journal (16).

II. What Defines Authority Online?

A. Protocol

Although Manschreck explained his hope for the ideal instantiation of enforced authority as one in which "the obedience that undergirds authority [is] made freely and responsibly," he might have more presciently described obedience as *undergirding* authority: The image of a grid of acceptance forming the basic structure of an authoritative system would nicely prefigure Alexander Galloway's conceptualization of the protocol power structure of a distributed network in his book *Protocol: How*

*Control Exists After Decentralization.*¹ But while Manschreck's dream for future systems of political authority relies vaguely but emphatically upon some form of centralized authority distinctly separate from the individuals who adhere to it, the structure of authority on the internet—a distributed network—collapses that distinction.

Galloway explains the organization of a distributed network by differentiating it from centralized and decentralized networks. In a centralized network, "a single central power point [...] is connected to all of the satellite nodes [in that network], which themselves are connected only to the central host" (11). Galloway later elaborates that "centralized networks may have more than one branch extending out from the center" (30). This makes the network considerably more complex, because a node may function as a mid-level hub connecting other groups of nodes hierarchically to the central power point—but, importantly, not to other hubs or nodes. Linear hierarchy makes a network "centralized," no matter how many levels may comprise it. Galloway gives the American judicial system as an example of a centralized network: "[E]ach decision of each court can always be escalated [...] to a higher level in the hierarchy" (11). The decisions of federal district courts may be appealed to federal courts of appeals, which may in turn be appealed to the Supreme Court. Most

¹ In fact, having read *Erosion of Authority* shortly after finishing *Protocol*, I initially mis-read Manschreck's phrasing as "undergrids" and had mistakenly attributed that foresight to him until my advisor, Clifford Siskin, pointed out my mistake.

corporations also function as centralized networks, with the C-level executives and the board of directors providing instructions to executives, who in turn supervise any number of levels of mid-level managers, who oversee other employees. A decentralized network functions similarly, having "*multiple* central hosts, each with its own set of satellite nodes" (11). Because these hubs communicate with and depend upon each other, there is "no single zenith point [that] exercises control over all others" (30). Galloway provides the example of airline routes, which offer service among the large number of airports by routing travelers through central hubs.

Though the structures of centralized and decentralized networks differ somewhat, both presuppose a differentiation between two types of components: hubs and nodes. This distinction limits the flow of information hierarchically by requiring that hubs purvey information and nodes receive it. Distributed networks, however, "have no central hubs and no radial nodes," making "each entity in the distributed network [...] an autonomous agent" (33). With this lack of hierarchy, "each node [...] may establish direct communication with another node, without having to appeal to a hierarchical intermediary" (11-12). Importantly, however, this structure does not require each node to maintain a continuous connection to every other node in the system—this would be highly inefficient. The interstate highway system exemplifies a distributed network because it "lacks any

centralized hubs and offers direct linkages from city to city through a variety of highway combinations": If one route from one city to another is blocked, you can take a different one (35).

Protocol, then, answers the obvious question of what keeps a distributed network of autonomous nodes from anarchy: "[I]n order to initiate communication, [any] two nodes must *speak the same language*." This language—protocol—"defines the landscape of the network," determining "who is connected to whom" (12). Although distributed networks "have no chain of command," their "autonomous agents" "operat[e] according to certain pre-agreed "scientific" rules of the system" (38). And, since protocol enables their communication, these agents must adhere to it: "Without protocol, there is no network" (11). Galloway illustrates the extent of protocol's power through a parallel with Michel Foucault's explanation of how modern governments' biopower evolved from monarch's sovereignty over their subjects' lives, quoting from *The History of Sexuality, Volume 1*: "...the ancient right to *take* life or *let* live was replaced by a power to *foster* life or *disallow* it to the point of death" (Galloway, 13). This comparison demonstrates poignantly Galloway's crucial point: "While protocol may be more *democratic* than the panopticon in that it strives to eliminate hierarchy, it is still very much structured around command and control..." (13).

Without outright contradicting Kevin Kelly's assertion a decade earlier that "[h]idden in the Net is the mystery of the Invisible Hand—control without authority" (26), Galloway reveals significant subtleties in the situation. Although the internet lacks a "chain of command" typical to centralized and decentralized networks of the most common human social structures, the protocols that simultaneously permit the internet to function and define the parameters of its workings do exist as authoritative documents, called "RFCs, or "Requests for Comments."" Galloway explains that these documents "[instruct] potential software designers or other computer scientists how to correctly implement each protocol in the real world" (38). And since RFCs must have some author(s)—both original and ongoing to keep up with changes in technology—individuals with authority over the mechanisms of the internet must indeed exist. Galloway later explains that, as we might expect, "technical protocols and standards are established by a self-selected oligarchy of scientists." Although "membership in this technocratic ruling class is open" in theory, in practice the "technical sophistication" required to contribute to the formation of protocol limits its members to "a relatively homogeneous social class: highly educated, altruistic, liberal-minded science professionals from around the globe" (122).

The internet's end-users who, in the words of long-time RFC editor Vint Cerf, "just want the system to work" (Galloway 123), implicitly accept

the authority of Cerf and other members of the Internet Society (ISOC), which, Galloway explains, "facilitates the development of Internet protocols and standards" (132). Is this the enlightened use of independent reason that Kant calls for? Maybe. As Roger Shinn points out in the final essay of Manschreck's collection, "we human beings are more dependent than ever before upon experts with skills we do not share and barely comprehend" (92-93). Most internet users indeed do not understand the technicalities of the protocols that control it, but they must accept the authority of the ISOC in order to use the technology it facilitates.

One might argue that, given the transparency of the ISOC's processes and the open membership of the Internet Engineering Task Force, which, according to Galloway, is "the core area where most protocol initiatives begin" (132) before becoming ISOC-endorsed standards, any internet user is free to gain the necessary technological knowledge and contribute to internet protocol. Differences in socioeconomic status and access to education and technological resources make this sort of argument patently untrue—and even if all things were equal, achieving the depth of knowledge necessary to make a meaningful contribution to these standards would not be feasible for those with other roles in society, as physicians, poets, lawyers, teachers, or any other profession. In this case our reliance on others' guidance is based not in the intellectual laziness that Kant belittles but in the sheer impracticality of gaining the necessary

knowledge. As Shinn puts it, "in this age of the knowledge explosion nobody is expert on all the issues he must meet" (98).

Since protocol describes the very mechanism by which communication may occur online, the RFCs that define protocol and the technocratic oligarchy that creates and approves those RFCs wield considerable authority in shaping and maintaining the functionality of the internet. But this authority differs fundamentally from the authority structures that dominate human society. As Galloway explains, "protocol does not follow a model of command and control that places the commanding agent outside of that which is being commanded" (50). Protocol, in Galloway's words, "is endogenous" (51), meaning that it, along with the RFCs that define it and the scientists who compose the RFCs, cannot be extricated from the they control—at least not without surrendering their own authority.

But we must consider the protocol that Galloway describes as an example of operational authority—the sort that enforces users' behavior so the internet may function, although it is necessarily based upon its creators' noetic influence over internet users' beliefs. Failure to adhere to protocol doesn't result in punishment as failure to adhere to laws does in the real world—it simply isn't possible while remaining within the system. Galloway explains that "if one chooses to ignore a certain protocol, then it becomes impossible to communicate on that particular channel" (167).

B. Content Organization (DNS and Directories)

What, then, determines the noetic authority of the vast range of information available online within the strict operational authority of internet protocol (which is based in turn upon the noetic authority of its creators as technical experts)? The classificatory tree structure of a centralized or decentralized network imbues its component pieces with intrinsic information about their content and level of authority. The distributed network of the internet, however, seems to provide no inherent organization of its content or indication of its authoritativeness—all nodes of the network are equally powerful within it and universally capable of connecting to one another.

One of the most important internet protocols for facilitating ease of use, however, projects a decentralized organizational structure on top of the internet's basic distributed network: the Domain Name System (DNS). Computers identify themselves on the internet with a unique numerical Internet Protocol address (IP address), which, as Galloway explains, is "written as a group of four numbers separated by dots." Because these addresses are "very difficult for humans to remember," we translate the IP addresses for machines that host websites into more familiar web addresses, which can consist of both letters and numbers separated into groups of characters separated by dots (such as www.nyu.edu) (Galloway, 47). The first group of characters is known as the subdomain, the second

as the domain, and the last as the top-level domain. DNS, which is controlled by the Internet Corporation for Assigned Names and Numbers (ICANN), defines the translation between IP address and web address in a decentralized hierarchical database organized in a typical tree structure. At the highest level of this organization, explains Galloway, "sit the so-called root servers," which "have authority over the *top-level domains* (TLDs) such as 'com,' 'net,' 'edu,' and 'org'" (49). These TLDs are in turn represented by "name servers," which contain the information that allows a browser to locate the host for each domain name with a given TLD. Each domain's name server knows the exact IP address for all existing subdomains.

Because of this decentralized hierarchical mapping system, wherein the process of locating an IP address for a given web address "starts at the most general point, then follows the chain of delegated authority" (47), "nearly all Web traffic must submit to a hierarchical structure (DNS) to gain access to the anarchic and radically horizontal structure of the Internet" (9). In addition to the self-contradictory irony that Galloway notes in this intersection of the distributed and decentralized structures that form the internet, we can also see an attempt to insert very general categorical data into URLs of individual sites through the variety of top-level domains, which in theory would limit commercial sites to .com, educational sites to .edu, informational sites to .info, etc. In practice,

however, domain registrants have not complied with that nomenclature, and it seems to have proven impossible—or perhaps just not worthwhile—for ICANN to enforce. Many non-commercial sites reside within the ubiquitous .com TLD, and just as many sites not affiliated with any formal organization have URLs ending in .org.

The failure of TLDs to provide any meaningful organization to websites leaves their creators and users to provide the ranking and classification necessary to locate information. John Battelle recalls in *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed our Culture* that in the early days of the pre-web internet, "unless you had the exact machine address and file name, it was nearly impossible to find" the "papers, technical specs, and other kinds of documents" that "academics and technologists" stored online (39).

But internet use was fairly limited until 1993, when Berners-Lee invented a set of three new protocols that formed the basis for the collection of connected hypertext documents on the internet known as the world wide web: "universal resource identifiers (URIs), the Hypertext Transfer Protocol (HTTP), and the Hypertext Markup Language (HTML)" (Berners-Lee, 36). Galloway explains the first of these innovations as "a simple, direct way for locating any resource on the Web" and the second two as "the core suite of protocols used broadly today by servers and browsers to transmit and display Web pages" (137). Berners-Lee also

created the first "Web *client*—the program that would allow the creation, browsing, and editing of hypertext pages" (Berners-Lee, 28). This application evolved into the modern web browser, which we use to access, display, and navigate websites.

The web's more accessible interface, coupled with the decreasing cost of storing documents online and of bandwidth (the speed with which we connect to those documents), enabled it to "gr[o]w from 130 sites to more than 600,000" between 1993 and 1996," according to Battelle (40). The dramatic increase in both the usage of the internet and the volume of material it contained expanded and intensified the need for an effective means of finding documents online. By 1994, Battelle claims, "there were literally dozens of sites that organized the Web," but Yahoo!—then Jerry and David's Guide to the World Wide Web—was arguably the first to gain widespread popularity for, in Battelle's words, "organiz[ing] the web in a fashion that made sense to techies and first-time Web surfers alike" (60).

Yahoo!, which Battelle calls "one of the earliest viral success stories of the nascent Web" (58), began as a list of the websites its creators, Jerry Yang and David Filo, found interesting. This subjectively chosen list of "good" websites, by sheer force of Yahoo!'s enormous popularity, lent a certain authority to the sites included: Even though Yahoo! might not have officially pronounced the sites they linked to as authorities on any given topic, the influx of visitors gave those sites the

opportunity to influence its newly arrived readers. This means, even more importantly, that the Yahoo! directory itself had a very palpable power to influence the many web users who visited the site.

Yahoo! organized the links to the sites it listed hierarchically by subject, utilizing categories and subcategories, different in content from but similar in structure to the subject-based organization of books codified by the Library of Congress and the Dewey decimal system, which, as Battelle points out, "would be unable to scale to the enormosity of the World Wide Web" (33). Battelle cites Yahoo!'s founders' defense of their directory as offering higher-quality information than any technology-based solution to organizing online content and as being easy to understand for the majority of Web users, who were "new to the experience" (61).

Indeed, this simple continuity with the standard mechanism for organizing print information (hierarchical categories and subcategories) most likely accounts for much of Yahoo!'s phenomenal success. But Battelle's observation about the inherent lack of scalability of a subject-based system built by a single organization holds just as true for a categorized list of links like the early Yahoo! as it does for the Dewey decimal system's categorization of printed material. Furthermore, users' growing comfort with technology in general and the internet in particular led to a fascinating paradigm shift in how people navigated the web, which Battelle describes as a change "from a stance of exploration ('What's out

there?') to expectation ('I want to find something that I know is out there')." In the words of Srinija Srinivasan, a Yahoo! employee who Battelle interviews, "now we go online expecting everything we want to find will be there" (Battelle, 61). The primary application that developed along with that expectation—simultaneously satisfying it and increasing its prevalence—is search.

C. Search

Online (and offline) search technologies certainly predate the above-mentioned change in internet users' attitudes. But that shift, along with the necessarily limited breadth of a handmade list in the face of an exponentially increasing number of sites available, caused directory-based portals to decline in popularity as the point of entry to the web as the use of search engines rose. Battelle quotes engineer and entrepreneur Raymie Stata as saying that "as the amount of information available to us explodes, search has become the user's interface metaphor," concurring in his own words that "search has moved from a useful service on the edge of most internet users' experience to the de facto interface" and is now "a universally understood method of navigating our information universe" (4).

As Battelle explains in his second chapter, which covers the basics of search, internet users search "to recover that which [they] know exists on the Web, and to discover that which [they] assume must be there,"

providing more concise descriptors for these two motivators as recovery and discovery, respectively (32). Between these two very basic types of inquiries, it's easy to see why search functions as a starting point for a huge amount of online activity, especially of the information-gathering variety. Although Battelle and other business-minded search scholars such as Andrei Broder anxiously remind us that users make a good number of searches for non-informational purposes (most notably for commercial or navigational purposes—to make a purchase or find a specific site that the user already had in mind), the point remains that search engines provide an extremely logical place to begin an informational query.

When we begin an online activity with search, we enter a specific set of words, known as a query, and the search engine "produce[s] a list of URLs (and sometimes summaries of content) it believes are most relevant to [our] query" (Battelle, 19). In other words, search engines do not return a simple list of all web documents containing the keywords we enter, with all qualifying documents weighted equally. To do so would not only be extremely unhelpful in helping us achieve the intention with which we came to the search engine—it would also be physically impossible. Since the results must be ordered somehow, it's only logical that each search engine would list its results in order of decreasing applicability to the query at hand.

The term "relevance," which most search engines use to describe their default ordering of results, contains enough ambiguity to implicitly place the burden of good judgment in the use of those results upon the user. All that a search engine means when it deems a result "relevant"—at least according to the *O.E.D.*'s primary definition thereof—is that the page in question is "bearing upon, connected with, pertinent to, the matter at hand." The choice of relevance of a description of its results clearly communicates the search engine's lack of responsibility for the content of those pages: They bear *some relation* to your query, but you have no assurance of the validity of their content. The engines' creators, like Kant, hope you will use your reason to independently evaluate the authoritativeness of the sites they provide. And yet, the pages that an engine deems to be most relevant for a given term will indeed acquire a very real authority over the engine's users for the same reasons that the early Yahoo! directory could be seen as imbuing the sites it deemed worthwhile with authoritativeness: The visitors' attention offers these sites a chance to influence them, and the prevalence of search makes that authority even stronger than that offered by Yahoo!'s listing. And this very influence gives the engines themselves an extraordinary authority in answering users' implicit question: Which websites have the best—or most authoritative—information about my query?

i. Google

The level of power that search engines wield online leads naturally to fierce competition among different engines. If search engines are purveyors of authority, their creators and users (not to mention the companies that spend billions advertising on them) all want to know: Which authority on authoritativeness has the most authority? The current answer is almost unquestionably Google. For this reason Battelle spends the majority of his book *The Search*, which is supposed to be a broad investigation of search's implications for both culture and business, discussing the story and workings of Google. To defend this bias, he cites market share data from a study done by Piper Jaffray & Co. in the first three months of 2005, which indicates that Google is used for 51% of searches worldwide, with its closest competitor—the search engine that Yahoo! developed in the face of Google's sudden and momentous popularity—claiming only 24% of the market share, followed by MSN's search at 13%, and AOL's and Ask Jeeves's at 5% each (Battelle, 30).

Cultural perceptions of search and anecdotal wisdom offer an even stronger sense of Google's dominant authority among search engines. As Battelle observes, "it seems that the words 'Google' and 'search' are now nearly one and the same" (37)—and although he reminds his readers that this equivalence is most likely a transitory phenomenon, Google's current popularity as a search engine and reputation as a "good" and "trustworthy"

company currently lend it much more authority than its competitors. Even the *O.E.D.*, a monolithic edifice of print-culture authority, acknowledges Google's singular authoritativeness in the sphere of online search: In June 2006, the *O.E.D.* added the search-engine sense of the verb form of Google (the verb google already existed with a lower-case g as a descriptor for certain moves in a cricket game). According to the *O.E.D.*, Google means "to use the Google search engine to find information on the Internet" as an intransitive verb and "To search for information about (a person or thing) using the Google search engine" as a transitive verb.

As SearchEngineWatch.com contributor Barry Schwartz noted at the time, Google was "already [...] a verb in some other dictionaries," but he considered the *O.E.D.*'s update momentous because it "is considered the most authoritative dictionary of the English language." We interpret the *O.E.D.*'s acknowledgement as an indication that a word has 'officially' entered the English language—which immediately suggests a widespread and widely accepted absorption of whatever the word represents into English-speaking culture. The online communities of search technology and linguistic bloggers alike posted numerous commentaries on this update precisely because a long-standing authoritative source of information about our culture had made a very official proclamation of Google's relatively new authority therein.

Notably, none of the other major search engines have *O.E.D.* entries, as verbs or otherwise. Even the definitions for "yahoo" have no updates indicating the online authority of the website named when its founders, according to Battelle, "pulled out a dictionary and started at 'Y'" to look for a "more memorable name" than Jerry and David's Guide to the World Wide Web in 1995 (58-59). Of course, we could argue that Google made it to the dictionary simply because it has a catchier name (and one that's more easily made into a verb) than its competitors and that its addition to the dictionary is really an indication that Google is becoming a generic term for online search meaning nothing more about its authoritativeness than Kleenex's *O.E.D.* entry means about its profit margin. Indeed, when *Washington Post* reporter Frank Ahrens wrote that Google was on its way to joining the "proper nouns that have moved beyond a particular product to become descriptors of an entire sector," he actually received a complaint letter from Google's trademark lawyer explaining in great detail the inappropriateness of using Google generically.

Although the verb "to Google" may indeed become generic over time, the recency of its addition to the dictionary does indicate the Google engine's current prevalence in the cultural consciousness and its status as the benchmark against which we evaluate other search engines. When a character on the popular television show *The O.C* awkwardly claimed to have "A9.com'd" someone, online cultural commentators such as Andrew

Baio immediately interpreted the incident as poorly executed product placement by Amazon.com in an attempt to raise its A9 search engine to Google status. Amazon denies paying for the mention, but the reaction speaks to the extent of Google's dominance. Furthermore, its close competitor Yahoo! touts what's arguably an equally catchy and verb-ready name and has since before Google came into existence. The phrase "I Yahooped it" sounds about as natural as the phrase "I Googled it," but Yahoo! has no dictionary entry.

Perhaps the most powerful anecdotal evidence of Google's authority comes from a recent wave of criticism toward the engine. When a large number of blogs about human sexuality and adult entertainment inexplicably disappeared from Google results for basic queries that normally listed such sites, online coverage of the incident gave an almost unanimous explanation for the extent of users' outrage: Google's indispensability in finding things online. Tech culture journalist Xenia Jardin wrote that "all of us rely on one single service [Google] to access so much of the information we need each day," and blog network mogul Nick Denton concurred: "You think there are other search engines, so that's okay? There are no other search engines."

III. How Search Works

Search's status as an entry point to the internet gives engines the ability to purvey authority to other sites in the form of the opportunity to influence visitors interested in a given subject. This authority comes in different sizes, based on the applicability of a given site to a user's search as determined by the search engine. As mentioned above, when we make a query, the engine returns a list of links in order of relevance from greatest to least. Some users might weed carefully through the titles and summaries before selecting a site to visit, but most will not, and sites listed first will naturally get more opportunity to influence than those listed later in the results, making the rules that determine this ordering powerful in determining online authority. In fact, Brian Pinkerton found in an analysis of user behavior on his WebCrawler search engine in October 1995 that "only 17%" of users viewed the second page of search results (Pinkerton 52).

Whereas we might excuse the algorithms of bibliometrics described above from Kant's scathing review of "rules and formulas" as "the shackles of a permanent immaturity" because we use those algorithms to understand attributions of authority that are created by non-algorithmic means, a search algorithm is precisely the sort of "mechanical aid" to our own thought process that Kant decries reliance upon, especially if we so instinctually provide the highly ranked sites with our attention. The

unfeasibility of sifting through the information available online using only our own reason means that the enlightened internet user (in the Kantian sense) ought to endeavor to understand the algorithms behind the search engine rankings that she relies upon for online navigation. Furthermore, in an investigation specifically into the mechanisms of authority online, we must certainly explore the workings of the algorithms that provide so much authoritativeness to the sites they rank highly. Only by learning how search technology works may we understand the factors that determine which sites an engine will attribute with more authority—as well as the factors that determine which engine will wield the most authority in conferring that authoritativeness.

A. The Basic Workings of Search

Most search engines function by crawling the internet, indexing online documents located by the web crawling application, and using the index it creates to provide users with results to queries. For an application to "crawl" the internet simply means, as Battelle explains, that it "send[s] out vast numbers of requests to pages on the internet" (20), "tak[ing] note of any links it has found on [a given] page, and queue[ing] those links in its request file" (21). A search engine crawler then passes the information it receives from those requests to the search engine's database, known as its index. As Michael Maldin explains in his essay on the Lycos search

engine he created, "all Web spiders use essentially the same algorithm to locate documents on the Web," which he describes as:

1. Create a queue of pages to be explored, with at least one Web page in the queue.
2. Choose a page from the queue to explore.
3. Fetch the page chosen in Step 2 and extract all the links to other pages. Add any unexplored page links to the exploration queue.
4. Process the page fetched in Step 3 to extract information such as title, headers, key words, or other information. Store this information in a database. (Mauldin)

According to Mauldin, "various spiders differ mainly in how they choose which page to explore next (Step 2) and what information they keep about each page (Step 4)." The high degree of similarity in the crawling process of search engines means that the breadth of their indices end up being fairly similar. The differences among search engines lie primarily in the type of information stored in the database, as Mauldin notes, in the ways they interpret that information, and in the interface through which users make queries and access results.

A search engine's index generally lists "all the pages on [a given] site, as well as all pertinent information about those pages: the words on the page, the links, the anchor text (text around and within a link), and so on" (Battelle 21). The index, then, is a database of URLs and the words, phrases, and other websites and documents associated with those URLs. When you query a search engine, the results it provides are URLs it has associated with your query terms in its index.

i. The Beginning of Internet Search

Archie, which Battelle and Mauldin both cite as "the first internet search engine" (Battelle 39), did not crawl the web like most contemporary search engines, but "buil[t] a database of retrievable files by performing recursive directory listings" once a month at "approximately 1,100 UNIX anonymous FTP archives world-wide" (Schwartz, et al. 3), keeping only the titles (not the full text) of the files it found. Although Battelle describes Archie as being "quite popular" among the academic and technical crowd, its interface was not user-friendly—Archie could initially only be queried by "connecting directly to an Archie server via a command-line interface" (40). Its output was equally limited. Since Archie only indexed the titles of files it found online, users "had to know—or infer—the title of the document [they] were looking for" (40). Furthermore, it only provided users with the location of the machine on which the file in question could be found, rather than the exact location of the file itself—a defect corrected by a subsequent internet search engine called Veronica, which was based on the Gopher file-sharing protocol instead of FTP.

A plethora web-based search engines improved upon Archie and Veronica with user-friendly interfaces, indices that included the full text of documents as well as ancillary information like links, and crawling programs that queried websites and other internet documents automatically as described at the beginning of this section instead of

relying upon monthly updates. As the amount of material online grew, these engines also analyzed their indices to increase the relevance of their results to users' queries.

B. Analysis, Relevance, and Ranking

According to a 1999 *Smart Computing* article by John Lalande that provides instructions for searching FTP systems, Archie results could only be sorted "alphabetically by the name of the FTP server" or by the date the file was updated, with the most recent files appearing first. A user's best shot at determining the relevance of Archies results was sorting by date—with only the title in Archie's index, there was no other information upon which the engine (or the user) could base judgments of relevance.

As crawlers began to index the full text of internet documents and non-textual meta-information, search engines creators developed ways to analyze that information to provide meaningful sorting to their search engines' results. Various mechanisms of analysis developed along with a large number of search engines in the mid- to late-90s. Two main types of analysis can be distinguished: ad-hoc analysis that an engine performs at the moment a user enters search terms to determine the *relevance* of a document for the particular query at hand, and more stable index analysis, which an engine performs as it creates its database to apply an inherent *rank* value to all URLs it contains independent of any specific search terms.

i. WebCrawler

The contemporaneous development of a good number of web search engines makes an attempt to name the first search engine to sort its results difficult if not futile. Battelle points to Brian Pinkerton's WebCrawler, created in 1994, as "the first to index the full text of the web documents it found" (42)—a point that Pinkerton clarifies in his 2000 dissertation "the first full-text search engine *to broadly catalog the web*" (5, emphasis added), indicating that other full-text engines had existed but were limited to smaller sets of documents. Regardless, full-text indexing would be a prerequisite for determining relevance, since it provided search engines' creators with the information upon which such judgments might be made.²

Pinkerton explains that his WebCrawler first determined the relevance of its search results with an information retrieval system "based on [Gerard] Salton's vector-space retrieval model" (45), a classic (pre-internet) information retrieval algorithm that Salton developed for his digital search engine SMART (Salton's Magical Automatic Retriever of Text), and which, according to Mike Thelwall's 2004 book, "all commercial search

² In his 1997 essay describing his Lycos search engine, Michael Mauldin names WebCrawler and the Repository-Based Software Engineering (RBSE) spider, which launched a few months before WebCrawler in 1993, as "the first Internet search engines to implement ranked-relevance retrieval." Mauldin describes RBSE spider as "supporting relevance feedback using WAIS (Wide-Area Indexing Service)," but I have not found detailed information on RBSE spider's ranking system in general or WAIS in particular.

engines seem to use [...] in conjunction with other approaches" (121).

Thelwall describes the vector-based model (as it applies to websites) in layman's terms:

The model converts all web pages into "bags of words" in no particular order. Each page is represented primarily by a list of how often each of its words occur. These word counts are later converted to weights using a mathematical formula that gives higher weights to words that occur often in a document compared to other words in the same document. It also gives higher weights to rare words, ones that do not occur very often in other web pages. (121)

Pinkerton's WebCrawler introduced other features now common in search engines: "a stop list to filter common words from the queries and index" (such as 'and' or 'or') and reduction of "plurals to their root form" (Pinkerton 47). The quality of WebCrawler's sorting of its results decreased, however, "as the size of the index and the search volume grew." Pinkerton's relatively simple algorithm was challenged by an increase in both the "number of documents in the index" and the "breadth of subjects covered" therein, and by serving a user base with decreasing pre-existing familiarity with search technology.

Pinkerton aimed to improve his search results by licensing information retrieval software called Callable Personal Librarian (CPL) from Personal Library Software (53). Pinkerton explains that "CPL supported basic boolean operators (AND, OR, NOT), as well as the critical phrase operator (ADJ, for adjacent) that was missing in the first index" without

requiring users to enter "advanced query operators" (53). CPL also "weight[ed] documents in which the query terms occurred as phrases more highly than documents where they did not" (54) and gave higher relevance to results in which query terms were present "the word towards the beginning of the document" (58).

Although these changes improved the quality of WebCrawler's results, Pinkerton worked on "other ways in which the search results could be improved," namely "influencing ranking by external weighting and shortcuts." These mechanisms were "present in several different search engines in many different ways" by the time Pinkerton completed his dissertation in 2000, but he claims that "at the time, [they] were new to search engines and [...] the field of information retrieval" (58). Pinkerton explains that "with so many pages on the Web," a "measure of a document's quality" becomes necessary, especially when ordering the results for "common" or "imprecise" queries (59).

Admitting that "what constitutes a document of high quality is in the eye of the beholder," (58), Pinkerton explains that he based his ranking system on "the number of distinct sites that linked to a particular document," purposefully choosing "to use the number of sites rather than the number of pages because the number of pages could be heavily influenced by a lot of links emanating from, say, a friend's site" (60). To avoid disadvantaging new websites in his ranking system, Pinkerton also

specified that "pages that appeared more recently would have their link scores boosted to allow them to compete more favorably with pages that had been around for a long time" (61). Although Pinkerton did not release a commercial version of WebCrawler incorporating the index ranking system he designed, he emphasizes the importance of combining full-text based relevance sorting with this sort of ranking, both for the quality of the sorting of common searches and the efficiency of performing that sorting. Pinkerton's link analysis process "only needed to be run once a week because the link count information was slow to change," thus providing better results than relevance analysis without requiring much time or processor power for their delivery. The importance of these effects became readily apparent with the successful web-wide implementation of similar ranking mechanisms to those with which Pinkerton had experimented.

ii. Lycos and Excite

Battelle refers to AltaVista as "the first truly good search engine" (42), but it didn't earn that title for the way it ranked results. Rather, AltaVista utilized its parent company Digital Equipment Corp.'s (DEC) new Alpha processor to run "a thousand [web] crawlers at once" (46), allowing AltaVista's creator, Louis Monier, to build an index of 10 million documents by 1995, which Battelle calls "the closest thing to a complete index the young Web had ever seen" (46). While AltaVista gained widespread popularity and was eventually purchased by Yahoo! to add search

functionality to their directory, innovation in sorting results came from Lycos, which Battelle cites as "the first major engine to use links to a Web site as the basis of relevance" (53). Coincidentally, Lycos's creator Michael Mauldin claims that by November 1996 Lycos's index contained "more than any other Web search engine" at the time—60 million documents.

Battelle explains that Lycos, which launched in 1994, analyzed "anchor text, or the descriptions of outbound links on a Web page" (53) to determine the meaning of the page linked to with that text. Although Battelle implies that Lycos was the first engine to incorporate analysis of anchor text, Jon Kleinberg claims that it was also utilized by "one of the oldest WWW search engines, McBryan's World Wide Worm" (18). Anchor text refers to the words that appear linked on a web page. The HTML to create a link in which the words "New York University" take the user to the NYU website looks like this: `New York University`. In this example, the words "New York University" are the anchor text and will appear as a hyperlink in a web document. Users clicking a link with the words "New York University" would expect to be taken to the NYU website, and a search engine finding such a link on a page in its index would likewise attribute the website at <http://www.nyu.edu> with a higher degree of relevance for the term "New York University."

Michael Mauldin doesn't mention the use of anchor text in Lycos's rankings in his essay on the engine's workings, but does lend other insight into its rankings. Although the Lycos web crawler did scan the full text of web documents, it made its query results more efficient by creating abstracts of the documents it crawled to store in its index instead of the full text of the original documents. The engine used what Mauldin describes as "standard information-retrieval statistical methods" to identify "the 100 keywords most related to [a] document," which comprised the abstract, along with parts of the document which the hypertext structure lent implied importance—namely the "titles, header text, and an excerpt of the first 20 lines, or 10% of the document." Using these abstracts, Lycos ranked the relevance of documents for a particular query according to the following criteria:

- How many of the query terms are contained in the document?
- How frequently are the query terms used in the document?
- How close are the query terms to each other in the document (proximity)?
- Where do the query terms occur in the document (position)?
- How closely do the query terms match the individual words? (Mauldin)

Excite launched one year after Lycos and was, according to Battelle, "the first search engine to transcend classic keyword-based searching with technology that grouped Web pages by underlying concepts" (55). Excite, which acquired WebCrawler in 1996, also "used

statistical analysis of word relationships" (55) on web pages to determine relevance.

iii. Google and PageRank

Although Pinkerton had experimented with link counting as a determinant of a website's inherent quality, Serge Brin and Lawrence Page were the first to implement a search engine that combined intrinsic rankings of all URLs in its index with relevance sorting per search term based on full-text indexing that utilized "the additional structure present in hypertext" (Brin and Page 2) to analyze a web page's content more effectively (much as Lycos had done with its prioritization of words in the title of a document over words in the body and of words in the first 10% of a document over words appearing later). Their goal, as stated in the paper in which they presented Google to the Stanford University Computer Science Department in 1998, was to create a "notion of 'relevant'" in their search engine results that "include[d] only the very best documents"—a goal notable for its addition of a guarantee of the quality of the sites their search engines would return to the rather slippery descriptor 'relevance,' which implies no assurance of or responsibility for quality. Brin and Page explain that with the "number of documents in [search engine] indices [...] increasing by many orders of magnitude," "there may be tens of thousands of slightly relevant results" for any given query, but that users "are still

only willing to look at the first few tens of results" (3)—a generous estimate considering the small percentage of user who got to the second page of WebCrawler results according to the Pinkerton stats cited above.

They describe the great difference between endeavoring to provide high quality search results on the web and in other digital environments as based upon the difference in the number and diversity of documents available online as compared to the "fairly small, well controlled collections" used for offline digital search (7). The Text Retrieval Conference, which they cite as "the primary benchmark for information retrieval" (6-7), has a "Very Large Corpus" benchmark of "only 20 GB compared to the 147 GB" that Brin and Page boasted from their then rather-limited crawl of web documents (7). They explain that the standard vector-space retrieval model (upon which the only publicly available versions of the WebCrawler engine were based) "tries to return the document that most closely approximates the query," a strategy that tends to "[return] very short documents that are the query plus a few words" when used on the web, providing the example of a search engine "return[ing] a page containing only 'Bill Clinton Sucks' and a picture from a 'Bill Clinton' query" (7).

This phenomenon results, in Brin and Page's words, from the web's status as "a vast collection of completely uncontrolled heterogeneous documents" with "virtually no control over what people can put on [it]" (7).

The heterogeneity of web documents alone provides an obvious impetus for assigning web pages with intrinsic rankings regardless of their correspondence with search terms—Brin and Page give the example that a search engine should treat "a major homepage like Yahoo's" and "an obscure historical article which might receive one view every ten years" very differently in its results, even if they contain the same search terms with the same degree of ad-hoc relevance based on a vector-space model. Furthermore, Brin and Page explain, the "flexibility to publish anything" online combined with "the enormous influence of search engines to route traffic" gives birth to "companies which deliberately manipul[at] search engines for profit" (7). Inherent ranking to pages based on their citation by other pages provides a mechanism to minimize the impact of sites designed solely to manipulate search engines, which presumably would not be linked from other sites.

Brin and Page called their ranking of web pages "PageRank" and explained it as "an objective measure of [a web page's] citation importance that corresponds well with people's subjective idea of importance" (4). Essentially, they explain, they applied "academic citation literature" to the web "largely by counting citations or backlinks to a given page" as an "approximation of a page's importance or quality." The PageRank rating does "not [count] links from all pages equally, and normaliz[es] by the number of links on a page" (4). Unlike Pinkerton, Brin and Page do provide

the exact equation used to define PageRank. Their intention, like that of the wide accessibility of the RFCs that define the internet's protocol and the open membership of the organization from which those documents originate, is transparency, but the algorithm remains inscrutable to those lacking a technical or mathematical background, making their authority to state its quality in determining the rank of web pages decidedly operational. Most end users of Google, who rely upon PageRank (among other factors) to sort their results, must trust its objectivity based on their trust of Brin and Page's understanding.

Not wanting to be the sort of guardians of reason whom Kant despises for making the people who trust in them afraid to use their own sense, Brin and Page also attempt to diminish this feeling that we must blindly trust their algorithm an "intuitive justification" for PageRank (5). They claim that it acts as "a model of user behavior" in the sense that a page's PageRank is the probability that "a 'random surfer' who is given a web page at random and keeps on clicking links, never hitting 'back' but eventually get[ting] bored and start[ing] on another random page" will come across the page in question (5), but without mathematical knowledge our assurance of the truth of this claim remains tied to our trust in Brin and Page. They also justify their algorithm by explaining that "a page can have a high PageRank if many pages [...] point to it or if [...] some pages [...] point to it and have a high PageRank," elaborating that "pages that are

well cited from many places around the web are worth looking at" and that "pages that have perhaps only one citation from something like the Yahoo! homepage are worth looking at" because "if a page was not high quality, or was a broken link, it is quite likely that Yahoo's homepage would not link to it." PageRank, Brin and Page explain, "recursively propagat[es] weights through the link structure of the web" (5).

Although Google uses PageRank to determine the intrinsic ranking to URLs in its index, many other factors contribute to where a given web page ends up in the sorting of results for a given query, and Brin and Page claim to have "designed [their] ranking function so that no particular factor can have too much influence" (15). They specifically cite the user of anchor text, as described above in the context of the Lycos engine. Though Brin and Page do not specify why anchor text contributes to quality search results, the implication is that the creator of a web page will likely provide a more accurate and objective description of the pages that she links to than of her own page because, in an attempt to provide her users with an ideal browsing experience, she will most likely try to anchor her links with highly specific and accurate text. Google also "keeps track of some visual presentation details such as font size of words," attributing "words in a larger or bolder font" with more relevance than other words on a page (6). Notably, however, Google normalizes this valuation by interpreting font size "relative to the rest of the document" in order to not "rank otherwise

identical documents differently just because one is in a larger font" (11). The final sorting factor specified in the initial Google paper is that, for multi-word searches, "hits occurring close together in a document are weighted higher than hits occurring far apart" (15).

iv. Kleinberg

In some senses Google's algorithm, featuring PageRank, might be considered the final word in how search engine results are sorted—Google still dominates the market for online search due to a general consensus that its engine delivers the most relevant results for most, though certainly not all, queries. But Google certainly has its critics, and the enormous corporation built around the engine continuously tweaks the algorithm that runs it to counteract the success of spammers in gaming the algorithm—or, as critics suggest, to increase advertisers' dependency upon AdSense, which is Google's main source of revenue, or to otherwise censor results.

One alternative search result ranking mechanism deserves investigation: Jon Kleinberg's never-implemented model described in his paper, "Authoritative Sources in a Hyperlinked Environment." Writing contemporaneously to Brin and Page, Kleinberg also saw "the hyperlink structure among WWW pages" as a possible source for attributing authority to web pages (2). "Hyperlinks," Kleinberg explains, "encode a considerable amount of latent human judgment," which he specifies as follows: "the creator of page p , by including a link to page q , has in some measure

conferred authority on q " (2). Kleinberg admits that this formulation cannot be relied upon entirely because "links are created for a variety of reasons, many of which have nothing to do with the conferral of authority" (2-3). He also clarifies that "a universally popular page" like the Yahoo! or Netscape homepage should not be considered "authoritative with respect to any query string that it contained" (3), which would be the result of an overly simplistic interpretation of links as conferrals of authority.

Instead, Kleinberg proposes "a link-based model for the conferral of authority" "based on the relationship that exists between authorities for a topic and those pages that link to many related authorities" (3), referring to these types of pages as authorities and hubs, respectively. These two types of web pages "exhibit what could be called a *mutually reinforcing relationship*: a good *hub* is a page that points to many good authorities; a good *authority* is a page that is pointed to by many good hubs" (8). Later in his paper, Kleinberg explains how greatly this structure differs from that of academic citation, in which "highly authoritative journals on a common topic reference one another extensively" (16). Like Brin and Page, Kleinberg cites the heterogeneity of the web relative to more traditional material as necessitating a divergence from classic models of determining authoritativeness, explaining that, online "the strongest authorities consciously do not link to one another" and "can only be connected by an

intermediate layer of relatively anonymous hub pages, which link in a correlated way to a thematically related set of authorities."

Thus a model of determining authoritativeness that takes into account the fact that "hubs [...] may not know of one another's existence" and "authorities [...] may not wish to acknowledge one another's existence" (16) becomes necessary—and although Kleinberg does not say this explicitly, his feeling that a hub like Yahoo! should not be considered relevant for every string it contains seems to imply that his hypothetical search engine leverages the link structures of such hubs to weight different authoritative sites, then excludes them algorithmically from the results that list those authorities. Kleinberg claims to identify "a certain natural type of equilibrium [...] between hubs and authorities," which he "exploit[s ...] to develop an algorithm that identifies both types of pages simultaneous" (3). He uses "the output of a text-based WWW search engine" to "construct [...] small collections of pages likely to contain the most authoritative pages for a given topic" (3).

Although in theory Kleinberg advocates "the integration of textual and link-based analysis" in sorting the results of a search engine, he aims in his paper to demonstrate "that a considerable amount can be accomplished through essentially a "pure" analysis of link structure" that "ignore[s] the textual content of pages" (12). Kleinberg also uses his hubs-and-authorities model to successfully locate "similar" pages to a given

page. He notes that "if p is a highly referenced page," an attempt to locate similar pages by analyzing link structure without differentiating between hubs and authorities "will implicitly represent an enormous number of independent opinions about the relation of p to other pages" (13). By instead searching the link structure around page p for only the strongest authorities, however, Kleinberg finds that he can locate intuitive results for "similar" pages to both the Honda Motor Company homepage and the New York Stock Exchange homepage without returning generally authoritative sites for either query (13-14).

Kleinberg discusses one other important extension of his basic hubs-and-authorities model: clustering. For some search terms, he explains, "the relevant documents can be naturally grouped into several clusters" (21). This could occur because "the query string [...] may have several very different meanings" (Kleinberg provides the example of 'jaguar'), because it "may arise as a term in the context of multiple technical communities" (his example here is 'randomized algorithms,' but non-technical terms may have multiple contexts as well), and because it "may refer to a highly polarized issue, involving groups that are not likely to link to one another," such as 'abortion' (20-21). Kleinberg further analyzes the results of his algorithm to obtain multiple sets of hubs and authorities for query terms could be divided by meaning, context, or opinion.

Although Kleinberg's idea of differentiating between hub sites and authoritative sites in search engine results has not been implemented in a commercially available engine, Ask.com's description of its ranking system sounds like it may be based on similar ideas (Ask does not cite Kleinberg as an inspiration). Ask launched in 1997 as AskJeeves, a 'natural language' search engine ("Ask: A brief history of the Ask search engine") but eventually reinvented itself as a more typical keyword search engine. According to Ask.com's about page, its search rankings are now based on technology from Teoma, a startup it acquired in 2001, which it claims as "the first, and is still the only, major search technology based upon the clustering concept of subject-specific popularity." Ask refers to its ranking as "ExpertRank" and elaborates that "Our ExpertRank algorithm goes beyond mere link popularity (which ranks pages based on the sheer volume of links pointing to a particular page) to determine popularity among pages considered to be experts on the topic of your search," a process that "requires many additional calculations that other search engines do not perform." Ask's promise that its results provide expertise on a users' query topic makes its promise of authoritative results more explicit than most search engines': Expertise on a topic gives a site's author the noetic authority on that makes us accept his influence over the information we believe to be true about the topic at hand.

IV. Search Engines' Protocological/Proprietary Construction of Authority

Michael Mauldin wrote in 1997 that "since 1994, most people start their Internet surfing at either a search service or a directory service, and these have become the roadmaps to the Internet." In the past ten years, search engines have far surpassed directories in usage, coming to constitute our primary entry point to the web. Battelle cites a 2004 study by the Pew Internet & American Life Project concluding that "of all Americans who use the Internet, about 85 percent use search engines" (25). The same project's November 2005 study increased that number to 90% (Rainie and Shermak 2). This transition, as established above, stems primarily from the exponential growth in the web's scale, the increasing comfort level of its user base, and the improving quality of search results, and gives search engines' interpretation of users' queries and their ordering of the results they deem relevant a protocological control over the construction of authoritativeness online.

Protocol, as Galloway describes it, "defines the landscape of the network," determining "who is connected to whom" (12). Search brings order to the seeming disorganization of the internet, allowing users to find the information they're looking for despite the lack of intrinsic structure. I believe that even Kant with his distrust of "mechanical aids to [... our] natural gifts" would agree that, in the face of so much information, this particular mechanical aid is wholly necessary for a user attempting to

efficiently navigate the internet—and even more so for a publisher looking to disseminate information. Even though nearly anyone may publish content online without submitting to the sort of authority necessary to publish print media, they must submit to the protocol of search to effectively *distribute* what they publish. In this sense, search controls the location and distribution of information online in much the same way that the protocols that define the internet's most basic operation control its functioning. Galloway writes that "if one chooses to ignore a certain protocol, then it becomes impossible to communicate on that particular channel" (167), and indeed to ignore search engines would make it nearly impossible to find information online, and challenging and inefficient to distribute it.

A. Functional vs. Stated Attribution of Authority

Although the papers that explain the basis of search engine rankings do refer to their algorithms as measures of authoritativeness, the user interfaces of the active engines that have grown out of those papers—like Google—and whose ranking methodologies have been inspired by them—like Yahoo!, and AskJeeves—usually label their results with the more ambiguous phrase "sorted by relevance." Only one major search engine explicitly claims to order its results by "authority": the exclusively blog-indexing Technorati. Technorati, which WebCrawler's Brian Pinkerton

now works on, takes a link-based measure of authoritativeness to its logical extreme. When users view Technorati results sorted "by authoritativeness" (which is the default when using Technorati), all blogs matching their search terms are displayed in order according to their Technorati rank, which works like a non-recursive PageRank with an expiration date. Technorati weights all incoming links equally, regardless of the rank of the link's source, but only counts links within the past 180 days and assigns a numerical rank to each blog by comparing the number of unique incoming links in the past 180 days for a given blog to the number of unique incoming links in the past 180 days counted for all the other blogs in its index (Pinkerton, "Making Sense of Link Counts").

Even when search engines decline to explicitly frame their results as necessarily the most authoritative sources available on a given topic, or as necessarily listed in the order in which a user should trust in their content, their widespread use constructs for their results a very powerful functional authority—especially those listed first for common or important queries. The increasing casualness with which we approach the internet breeds a desire to obtain information as quickly as possible. The impatience of users with long pages of search results is widely acknowledged anecdotally, and confirmed by statistics demonstrating the much smaller number of users who even view the second page of a search engine's results, such as those cited by Pinkerton in his paper on

WebCrawler. Furthermore, as the subjectively perceived quality of search engines' results for *most* queries has improved, particularly with the implementation of rankings based on incoming links as pioneered in theory by Pinkerton and in practice by Brin and Page, our trust in the relevance of their results for *all* queries has likewise increased.

A recent online incident dramatically exemplified the casualness with which careless users might mistakenly attribute sites with authoritativeness simply because they are highly ranked in search engine results. The Rock the Vote organization—which, interestingly enough, is dedicated to encouraging apathetic young people to make educated political decisions—created a feature on its website for Martin Luther King Day that linked to a hate site that makes outlandish claims against King and demands the discontinuation of the holiday that commemorates his birthday. After their mistake was pointed out, Rock the Vote made a blog post explaining: "our webmaster searched Google and chose one of the top results" (Reimer). This would seem to be an extreme example of someone relying to heavily on Google's rankings without thinking for himself at all—precisely the behavior that Kant claims will arise from the use of "rules and formulas." The fact that the hate site is still ranked fourth in a Google search for "Martin Luther King" nearly five months later, however, indicates that other website owners have made the same mistake as Rock the Vote, since that rank is solidified at least in part by incoming links.

Even when we exercise enough independent review of the sites that search engines provide to us that our trust in their relevance does not translate to an automatic assumption of their authority, and even in fact when that trust in their relevance does not exist to begin with, the vast majority of search engine users will necessarily give more attention to web pages that rank highly for whatever query they happen to make. This attention is an *opportunity* for the web pages' creators to influence visitors, and when that opportunity comes to fruition, be it due to "real" authoritativeness of the information a website provides or the perceived authoritativeness of potentially biased or misleading information, these site creators achieve effective authority to influence the understanding and beliefs of their visitors.

i. Multiplicity

The authority provided by search engines conspicuously lacks exclusivity. Although search engines do attribute an order to their results, they do not presume to select the site they find to be the most authoritative for a given query and send the user directly there. Engines' provision of multiple results for a query implies that all applicable information available online can likely not be found from any one source and returns at least part of the burden of responsibility for sifting through the results to the user. Though engines proclaim their sorting to be based

on "relevance"—and sometimes even on "authority"—they do not by any means guarantee the accuracy of the information on the sites they list.

In some senses this plurality of results for a given query is a departure from the construction of authority in print information. Although serious research using print materials would certainly necessitate consultation of more than one source, a casual offline inquiry could easily begin and end with a single publication that wields a large amount of general authority such as the *Encyclopædia Britannica* or the *O.E.D.* Although these edifices of traditional noetic authority cite their sources and often suggest further reading on given topics, they offer self-contained information perceived to hold great authority for a wide variety of topics.

Search engines, which function as the online analogues of these staples of noetic authority, point users to external sources rather than providing their own authoritative information. Engines are authorities on the authoritativeness of other documents, rather than authorities on the specific information that their users seek. Furthermore, the engines insist on providing multiple sources. Even when users fail to take advantage of this diversity of information—as research into the clickstreams of searchers indicates is frequently the case—the message that more than one source may be consulted is clear. The volume of material returned by search engines for most queries is so large that, as Battelle relates, "in the early

days of Google, a popular sport among search watchers was to find a query that had exactly one result,"—a game called GoogleWhacking (28).

ii. Homeostasis

The relatively low barrier to web publishing as compared to print publishing means that virtually any of a search engine's users might also have websites of their own. If users with their own sites believe they have found authoritative, relevant, or otherwise interesting information in one of the sites ranked highly enough in a search engine's results for them to access it in the first place, they may provide a link from their own site to the site in question. Indeed, even if a user has a negative impression of a website, he may link to it as part of commentary critical to the site in question. Regardless of the context of this link, it will add to the intrinsic ranking of the page linked to in search engines using a link-based rating system, as most now do.³ If the user creates this link with descriptive anchor text, the site linked to will furthermore receive a higher relevance rating for the terms in that text.

³ There have been a few humorous attempts to mine search engines for context to references made online, such as <http://sucks-rocks.com>, which uses Yahoo!'s application programming interface (API) to compare the number of instances where a given search term can be found in a negative context to the number of instances where it can be found in a positive context. Although the site is meant to be a joke it drives home the point that evaluating anything based on the volume of references it receives may be misleading without evaluating the context of the references.

This pattern provides a mechanism whereby a site initially ranking highly for a given search, being resultantly linked by a large enough number of other sites, may gain a generally high ranking for any query that might be applicable based on textual and hypertextual analysis of the site's content. Thus, the practical authority that a site gains from appearing near the top of search engine results for a given search can be leveraged to automatically reinforce its authority as provided by search engine results for that topic and to create search engine-provided authority for any content that the site's creator may choose to include in the future. As Clive Thompson writes in an article on the barrier to gaining popularity in the world of blogs, "this pattern is called "homeostasis"—the tendency of networked systems to become self-reinforcing" (3). The phenomenon is not limited to the internet; Kevin Kelly describes a similar occurrence in the decision-making process of a swarm of bees deciding where to build a nest. According to Kelly, a few bees "scout ahead to check possible hive locations in hollow trees or wall cavities" and "report back to the resting swarm by dancing on its contracting surface" (7). The decision-making process continues as follows:

...the more theatrically a scout dances, the better the site she is championing. Deputy bees then check out the competing sites according to the intensity of the dances, and will concur with the scout by joining in the scout's twirling. That induces more followers to check out the lead prospects and join the ruckus when they return by leaping into the performance of their choice. [...] By compounding emphasis, the favorite sites get more visitors, thus

increasing further visitors. As per the law of increasing returns, them that has get more votes, the have-nots get less. (7)

As Thompson's article demonstrates, search engines aren't alone in perpetuating the importance of popularity in internet authority. Thompson claims that "inbound links are an 80 percent-accurate predictor of traffic" (2) for websites, and his explanation of how the social dynamics of blogs reinforce popularity sounds much like how the algorithms of search engines reinforce rank and, by extension, authority. He cites internet theorist Clay Shirky as finding that "A very small number of blogs enjoy hundreds and hundreds of inbound links" while "almost all others have very few sites pointing to them" (2). This phenomenon, which Shirky calls a "power-law distribution," seems intrinsic to human nature: "When we are asked to decide among a dizzying array of options," Thompson explains, we do not "[weigh] each option on its own merits." Instead, in Shirky's words, we "base [our] decisions on what other people are doing" (2). When a website's creator bases her decision to link to another site on what other people are linking to, "a feedback loop emerges." This means that "the few sites lucky enough to acquire the first linkages grow rapidly off their early success, acquiring more and more visitors in a cascade of popularity" (2). Social bookmarking sites like del.icio.us further reinforce this pattern. Users can post their favorite links on these sites without the hassle of creating their own. The more people who post a given link, the more likely that link is to appear on del.icio.us's highly trafficked homepage

and thus get even more links and more visitors—and a higher ranking in search engine results.

Thompson balances his implication that this reiterative linking creates a barrier to new blogs hoping to obtain influence by claiming that "without fresh postings every day [...] even the most well-linked blog will quickly lose its audience" (5). This might hold true for direct readership, the links to such a blog—or any other type of website—will continue to provide high ranking in search engine results (with the exception of Technorati, which expires the authority attributed by incoming links after 180 days as explained above). For popular or contentious topics that return a large number of search engine results, this means that once a site gains a substantial amount of link-based authority, it will be slow to lose its status, and will set a consequentially higher bar for rival sites to meet. But the necessity of gaining links from other sites to obtain high search engine rankings and the functional authority that comes with them certainly does not amount to a barrier to an exogenous barrier to distributing content online. The mechanisms of search engine rankings and the network structure of the internet make inbound links the protocol of spreading information online. Links are the language of communicating authority on the internet, and by adopting and succeeding in this protocol, new sites can and do gain widespread influence in the network these links enable.

Google's dominance as the most authoritative of all search engines has been even more self-reinforcing than that of the pages it ranks highly. Brin and Page's initial breakthrough in providing intrinsic ratings to the URLs it indexed based on incoming links provided much higher quality results than other engines. As its reputation for high quality results spread, Google's usage grew, as did the subjective and practical authority conferred to sites ranking highly in Google's results compared to those of other engines. Furthermore, Google's success spurred competing engines to create their algorithms along similar lines, and, although Google itself does continually tweak its algorithm to thwart spammers attempting to game its results, the basis in PageRank has been a constant due again to its enormously positive reception.

AOL, which, according to the Piper Jaffray statistics that Battelle cites, held 5% of the global search market share in Q1 of 2005 (Battelle 30), even gave up on developing its own search model in favor of a partnership with Google. AOL's search homepage now reads "AOL search, enhanced by Google" and prominently features the Google logo. Its about page even clarifies that "enhanced by" is an understatement, explaining that "Web results are Websites listed in order of relevance [...] these listings are administered, sorted, and maintained by Google" and directs users looking for more information to Google's about page. This raises Google's 51% market dominance according to the Piper Jaffray study to

56%, even without taking into account the fact that all of its competitors, including Yahoo! with a 24% share; MSN with a 13% share; and AskJeeves with a 5% share, have ranking models that seem to have been inspired by Google's based on the limited information those sites provide about how their rankings work. The Google model for ranking search results has gained a decided dominance and authoritativeness over users' web experiences—in the Google engine and in its competitors' engines as well—to the exclusion of different models such as Kleinberg's.⁴

B. Proprietary Protocol

The control of search results by increasingly gargantuan corporations significantly complicates its function as a protocol of information distribution. According to Galloway, "if technology is proprietary, it ceases to be protocological" (172). If protocol "operate[s] according to certain pre-agreed 'scientific' rules of the system" (38), the power structure changes dramatically when a corporation (or even several corporations) have exclusive control over those rules instead of the users

⁴ Although Kleinberg's model does bear much similarity to the Google algorithm, it differentiates itself by distinguishing between hub sites and authoritative sites in order to efficiently cluster authoritative sites in its results by topic when the search term is vague and to avoid mistakenly attributing hubs with a high ranking for any search term they contain. One could speculate that prevalence of the Google model, which, unlike Kleinberg's, rewards incoming links without possibly counting too many outgoing links against a site, has given rise to the tendency of all types of sites—not just hubs—to provide many links to other sites.

who make up the system. In fact, Galloway believes that "proprietary or otherwise commercial interests [...] represent a grave threat to and failure of protocol" (120) because they "[coopt] protocol's [...] explosive architecture" (121). Indeed, the closed, commercial nature of search engines was one of Brin and Page's main complaints in their initial paper on Google: "Up until now most search engine development has gone on at companies with little publication of technical details. This causes search engine technology to remain largely a black art and to be advertising oriented" (3).

Brin and Page aimed to "push more development and understanding into the academic realm" (3), elaborating that "the predominating business model for commercial search engines," advertising, is "inherently biased towards the advertisers and away from the needs of consumers" (24). They posit that "the better the search engine is, the fewer advertisements will be needed for the consumer to find what they want"—a claim that leads Brin and Page to conclude that improving the quality of search engine results would not be financially wise for a company relying on advertising dollars and that "it is crucial to have a competitive search engine that is transparent and in the academic realm" (25).

Despite these noble academic intentions, Brin and Page did end up creating a corporation around Google, which indeed eventually adopted an advertising revenue model, and PageRank is now a patented algorithm and

a trademarked term. They have never, however, offered the "paid inclusion" available on other search engines, whereby site owners may purchase a place in the engine's index, and Google proudly proclaims on its "Google Technology" information page that "no one can buy a higher PageRank." Google does aspire to corporate transparency, PageRank's patented status means that more information is available about its ranking methodologies (in the form of its patent application) than about those of its competitors. But Google's technology information page itself gives a limited explanation, similar to Ask.com's (described above), and only slightly more detailed than the vague offerings from Yahoo! and MSN.⁵ The rest of its algorithm remains a company secret—and one that changes continuously.

Battelle quotes some disgruntled AdSense customers who suspect Google of tweaking its rankings to encourage dependency on AdSense instead of natural search results, but the ostensible reason that Google—and other engines—change their algorithms is to outsmart the spammers who use their understanding of how ranking works to game the results. Indeed it is difficult to imagine how even a non-commercial search engine could operate on a completely transparent algorithm without

⁵ Yahoo! says its engine "ranks results according to their relevance to a particular query by analyzing the web page text, title and description accuracy as well as its source, associated links, and other unique document characteristics"; MSN says its "ranking algorithm analyzes factors such as web page content, the number and quality of websites that link to your pages, and the relevance of your website's content to keywords," adding that "the algorithm is complex and is never human-mediated."

becoming useless due to spam-ridden results. This presents an important obstacle to the openness to which true protocol must adhere and raises some difficulties with regards to Galloway's explanation of the clash of propriety and protocol.

When Galloway describes the birth of email spam—a phenomenon not unlike search engine spam—he explains that "the openness of the network was wrenched away from its users and funneled toward a single commercial goal" (120). The first spam message, and others that follow it, are "not a failure on protocol's own terms," but an example of what happens when "protocol is not allowed to work purely on its own terms" (120). Spam exemplifies the process of proprietary interests coopting protocological systems. In the case of search engine spam, the technology is coopted not by the commercial interests of companies that maintain search engines, but by other companies that want to dominate their listings. A search engine that achieves any degree of success will not be able to prevent attempted spammings—a fact that seems to preclude the possibility of operating an engine with complete transparency.

Despite the seeming impossibility of building a transparent search engine, Jimmy Wales, the founder of the free, community-edited encyclopedia Wikipedia, recently announced that he plans to do just that, proclaiming forcefully that "search is broken." Wales elaborates, "It is broken for the same reason that proprietary software is always broken:

lack of freedom, lack of community, lack of accountability, lack of transparency." The Search Wikia project plans to leverage the enthusiastic contributions of users like those who built and continue to build Wikipedia to "generate a new kind of search engine, which relies on human intelligence to do what algorithms cannot." Indeed if any group of users could accomplish such a task as a human-edited search engine it would certainly be the community that perpetually updates Wikipedia and protects it from vandalism, but at this point Search Wikia's success remains to be seen.

Brin and Page's own early concerns about the conflicting interests of search engine quality and advertising revenue remain troubling to Google, its competitors, and its critics. The need for search makes it as powerful as protocol. Galloway writes that "opposing protocol is like opposing gravity—there is nothing that says it can't be done, but such a pursuit is misguided and in the end hasn't hurt gravity much," and an attempt to avoid the influence of search would be similarly futile for any internet users aside from the few highly skilled enough to attempt to concoct their own solution. The inner workings of search wield an operational authority over the internet—one necessary to its effective functioning—and in theory we do base our acceptance of that authority primarily on our belief in the noetic authority of those who control the engines, which is based in turn upon our satisfaction with the results that

they have provided. In the course of our daily online activities, however, we also find ourselves compelled to accept that authority in order to participate in the network of locating and distributing information for which search provides a protocological basis. The difficulty we may feel with this obligation lies in the degree to which we trust the proprietary interests behind search technology.

Works Cited

- "About website ranking." *Windows Live Help*. Microsoft Corporation. 31 March 2007
<http://help.live.com/help.aspx?project=wl_webmasters>.
- "About AOL Search." *AOL Search*. AOL, LLC. 31 March 2007
<<http://about-search.aol.com/>>.
- "Ask: A brief history of the Ask search engine." *The Web Search Workshop*. The Web Marketing Workshop Ltd. 31 March 2007
<http://www.websearchworkshop.co.uk/ask_history.php>.
- Ahrens, Frank. "So Google Is No Brand X, But What Is 'Genericide'?" 5 August 2006. *Washington Post*. 28 February 2007
<<http://www.washingtonpost.com/wp-dyn/content/article/2006/08/04/AR2006080401536.html>>.
- Baio, Andrew. "A9.com on *The O.C.*" 29 January 2005. *Waxy.org*. 28 February 2007
<http://www.waxy.org/archive/2005/01/29/a9com_on.shtml>.
- Battelle, John. *The Search: How Google and its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Penguin, 2005.
- Berners-Lee, Tim, and Mark Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. San Francisco: HarperSanFrancisco, 1999.
- Brin, Sergey and Larry Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." Stanford University, 1998. 28 March 2007
<<http://www7.scu.edu/au/programme/fullpapers/1921/com1921.htm>>.
- Broder, Andrei. "A taxonomy of web search." Working paper. *Special Interest Group on Information Retrieval Forum* 36.2 (Fall 2002). 27 February 2007
<<http://www.acm.org/sigs/sigir/forum/F2002/broder.pdf>>.
- Denton, Nick. "Chronicle writer disappears in porn clampdown." 27 December 2006. *Valleywag*. 27 February 2007
<<http://valleywag.com/tech/google/chronicle-writer-disappears-in-porn-clampdown-224670.php>>.
- Galloway, Alexander R. *Protocol: How Control Exists After Decentralization*. Cambridge: The MIT Press, 2004.
- "Googe, v²." *The Oxford English Dictionary Online*. 2007. Oxford University Press. New York University. 27 February 2007
<<http://dictionary.oed.com/>>.
- "Google Technology." Google Inc. 30 March 2007
<<http://www.google.com/technology>>.
- "How are web documents ranked?" *Yahoo! Help*. 31 March 2007
<<http://help.yahoo.com/help/us/ysearch/ranking/ranking-01.html>>.
- Kant, Immanuel. "An Answer to the Question: What is Enlightenment?"

1784.

Jardin, Xen. "Google 'disappears' sex blogs?" 27 December 2006.

boingboing. 28 February 2007

<http://www.boingboing.net/2006/12/27/google_disappears_se.html>.

Kelly, Kevin. *Out of Control: The Rise of Neo-Biological Civilization*

Reading, MA: Addison-Wesley Publishing Co., 1994.

Kleinberg, Jon M. "Authoritative Sources in a Hyperlinked Environment."

Cornell U., 1998. 30 March 2007

<<http://www.cs.cornell.edu/home/kleinber/auth.pdf>>

Lelande, John. "How to Search FTP Sites." *Smart Computing* 7.11

(November 1999). 28 March 2007

<<http://www.smartcomputing.com/editorial/article.asp?article=articles/archive/q0711/13q11/13q11.asp>>

Manschreck, Clyde L. "Authority and Skepticism: An Introductory Essay."

Manschreck, *Erosion of Authority* 7-36.

Manschreck, Clyde L., ed. *Erosion of Authority*. Nashville: Abingdon Press,

1971.

Mauldin, Michael. "Lycos: Design choices in an Internet search service."

IEEE Expert 12.1 (January/February 1997): 8-11. 29 March 2007

<<http://robot-club.com/lti/pub/ieee97.html>>.

Pinkerton, Brian. "Making Sense of Link Counts." 4 October 2006.

Technorati Weblog. 30 March 2006

<<http://technorati.com/weblog/2006/10/127.html>>.

Pinkerton, Brian. "WebCrawler: Finding What People Want." Diss. U. of

Washington, 2000. 28 March 2007

<<http://thinkpink.com/bp/Thesis/Thesis.pdf>>.

Raine, Lee and Jeremy Shermak. "Search Engine use November 2005." 20

November 2005. Pew Internet & American Life Project. 30 March 2007

<http://www.pewinternet.org/pdfs/PIP_SearchData_1105.pdf>.

Reimer, Hans. "Last night, Rock the Vote made a mistake." 17 January

2007. *Rock the Vote Blog*. 3 May 2006

<<http://blog.rockthevote.com/2007/01/last-night-rock-vote-made-mistake.html>>.

"Relevant." *The Compact Oxford English Dictionary*. 2nd ed. 1991.

Schwartz, Barry. "Google Now A Verb In The Oxford English Dictionary." 29

June 2006. *Search Engine Watch*. 27 February 2007

<<http://blog.searchenginewatch.com/blog/060629-105413>>.

Schwartz, Michael F., et al. "A Comparison of Internet Resource Discovery

Approaches." *Computing Systems* 5.4 (1992). 28 March 2007

<http://www.isi.edu/div7/publication_files/comparison_of_internet.pdf>.

Shinn, Roger L. "The Locus of Authority: Participatory Democracy in the

Age of the Expert." Manschreck, *Erosion of Authority* 92-122.

Thompson, Clive. "Blogs to Riches: The Haves and Have-Nots of the Blogging Boom." 20 February 2006. *New York Magazine*. 31 March 2007 <<http://nymag.com/news/media/15967/>>.

Wales, Jimmy. "Welcome to Wikia Search." 23 December 2006. *Search Wikia*. Wikia, Inc. 3 May 2007 <http://search.wikia.com/wiki/Search_Wikia>.