Google AdWords as a Network of Grey Surveillance

Harold M. Roberts

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Science and Technology Studies

Janet Abbate, Chair

Matthew H. Wisnioski

John Palfrey

January 26, 2010

Falls Church, Virginia

keywords: media, advertising, google, adwords, surveillance, privacy

Google AdWords as a Network of Grey Surveillance

Harold M. Roberts

ABSTRACT

Google's AdWords processes information about what sorts of content users are browsing for about a quarter of all web site visits. The significance of AdWords' use of this vast amount of personal data lies not in its use for such obviously authoritarian purposes but instead as a network of grey surveillance with Google acting as the hub and the various publishers, advertisers, and users watching (and controlling) each other in distinct ways. Google's model of collective intelligence in its search and ad ranking systems has so deeply intertwined itself into user experiences online (and offline) that it acts as a shared nervous system. AdWords' use of specific words to target simple ads directly connects advertising topics with the content supported by the advertising, encouraging the content to do more of the work of assigning social meaning traditionally done by the ads themselves. And the AdWords pay-per-click ad auction system greatly increases the level of mechanization within the advertising and content production system, replacing the historical human bureaucracy of the advertising industry with the mechanical bureaucracy that is much more difficult to predict or understand. That mechanical bureaucracy shapes, in constitutive but unpredictable ways, the relationship between content and ads that drives the what content is published online and how advertisers and users interact with that content.

Acknowledgements

Thanks first to my advisor Janet Abbate, who not only provided much helpful guidance for this paper proper but also patiently guided me through the process of restarting my degree program after a long hiatus. Thanks to Matthew Wisnioski for close and careful readings of multiple drafts of the paper.

Much of the work for this paper was performed for and at the Berkman Center for Internet & Society at Harvard University under a grant by the MacArthur Foundation for research into Internet surveillance. Thanks to the Berkman Center and to the MacArthur Foundation. Thanks also to the many folks at the Berkman Center who have helped support and guide my career there, including Rob Faris, Colin Maclay, Jonathan Zittrain, and Catherine Bracy. Thanks to John Palfrey for helpful comments on multiple drafts of this paper, for sponsoring this research, and for jump starting my research career at the Berkman Center. Thanks to my research partner Ethan Zuckerman who also helped place me on the path of research at the Berkman Center and whose challenges served as the inspiration for a big part of this paper. Thanks to folks at Berkman who have provided feedback on drafts of this paper, including Fernando Bermejo, Aaron Shaw, Wendy Seltzer, Lewis Hyde, Carolina Rossini, and David Ardia.

Finally thanks to my girls Angela and Lucy and to my wife Kristine, who provided not only love and support but her enormous writing and editing talents.

Contents

Figures

1. Introduction

Google AdWords sits at a nexus of the web infrastructure, processing detailed data on approximately one quarter of all visits to web sites. The amount and intrusiveness of the data involved raises the specter of surveillance, but Google does not use this data to do any of the sorts of things typically associated with surveillance – it does not jail, fire, or kill any of the subjects of this data or deny them health insurance or mortgages. However, Google's use of this data has fundamentally important impacts on how users interact with content and ads online, on the relationship between content and the ads that support the content, and on the mechanisms for valuing some ad topics (and therefore content topics) over others.

First, Google's model of collective intelligence in its search and ad ranking systems has so deeply intertwined itself into user experiences online (and offline) that this model acts as a shared nervous system. This nervous system is both less than an artificial intelligence in that it requires constant user input to operate and more than an artificial intelligence in that it continuously integrates the thoughts and interests of its network of users. Google paradoxically has had to devolve most of its power over this nervous system to the publishers, advertisers, and users at the edges of the network because that is where this nervous system generates its power.

Second, AdWords' use of specific words to target simple ads contrasts with the widespread practice of using creatively produced ad content to connect social meanings to consumer products. Instead, AdWords directly connects advertising topics with the content supported by the advertising, encouraging the associated content to do more of the work of assigning social meaning traditionally done by the ads themselves. This new relationship between ads and content increases the pressure on publishers to produce content that gives meaning to consumer products.

Finally, the AdWords pay-per-click ad auction system greatly increases the level of mechanization within the process of valuing ads and content online, replacing the human bureaucracy of the advertising industry that has traditionally performed this valuing task. AdWords uses instead the mechanical bureaucracy of an automated market that is much more difficult to predict or understand. That unpredictable and therefore unaccountable mechanical bureaucracy pressures publishers to produce certain kinds of content over others.

All of these processes – the integration of the Google nervous system into its users' lives, the increasing role of online content in providing social meaning to consumer products, and the increasingly mechanized valuation of advertising and content – rely on complex feedback networks that generate their power by mining the collective knowledge of all of the involved actors. One helpful way to analyze those processes is as networks of grey surveillance – using existing frameworks of surveillance but applying them to all of the ways that the various groups of actors (Google, publishers, advertisers, and users) watch one another. This approach, rather than looking only at Google as a totalitarian surveillor, helps to tease out points of control within the networks. Indeed, this line of analysis shows that Google has some limited control over the network but is bound by the dilemma that the network is only powerful to the degree that Google allows the network to operate with minimal interference.

The following chapter will explore the theoretical groundings for treating Google's use of this data as a network of grey surveillance and provide an overview of how AdWords works both as a technical system and as a network of grey surveillance. The remaining chapters will examine specific parts of the AdWords system as individual networks of grey surveillance. Chapter 3 will argue that Google's search ranking system acts as a shared Google nervous system that relies on the deep interaction between Google, publishers, advertisers, and users, and Chapter 4 will argue that the AdWords ad ranking system works in a similar way but with more central control by Google. Chapter 5 will argue that AdWords's model for attaching advertising terms with content changes the relationship between advertising, content, and social discourse. Chapter 6 will argue that AdWord's model for assigning value to ads through the network of advertisers, publishers, and users creates a new sort of unpredictable, mechanized bureaucracy.

2. Grey Surveillance

Google serves over seventy percent of all ad traffic on the Internet, about half of that through its recently acquired DoubleClick subsidiary and about half through its homegrown AdWords system. (Attributor Corporation 2008) It is hard to overstate the amount and intrusiveness of personal data processed by Google through these two advertising platforms. Google uses its online advertising market share to processes data about the majority of all web site visits. As users have integrated the web into their lives in increasingly casual and intimate ways, that stream of data about online activity has grown increasingly personal. Google processes data about users' social networks, health problems, political and religious beliefs, shopping habits, and favorite sports teams, among many other topics. The sum total of this data for each individual paints an incredibly detailed (though in ways dangerously incomplete) picture of the monitored individuals.

Even though much of this data is not directly tied to specific users through their names, it is tied to IP addresses which can be used to identify users through their Internet Service Providers, and Google in any case can use search data to identify specific users through the search terms. In 2006 AOL released a portion of its search logs with the IP addresses removed but replaced by random unique identifiers. Researchers, journalists, and bloggers were able to track down the identities of many searchers merely through the search queries themselves. For example, the New York Times figured out that user 4417749 was Thelma Arnold from Lilburn, Georgia based on searches for "'landscapers in Lilburn, Ga,' several people with the last name Arnold and 'homes sold in shadow lake subdivision gwinnett county georgia.'" (Barbaro 1996) Arnold's search history also included highly personal queries such as "numb fingers," "60 single men," and "dog that urinates on everything." Google processes this sort of personal data for virtually every web user, including those who do not use its search engine.

Google's use of this huge flow of highly personal data is intuitively very troubling but does not fit a model in which a powerful central actor uses surveillance as a lever of physical violence to control a population. Google does not directly control the subjects of the data in any important way, let alone with physical violence. This is particularly true for the AdWords system, since Google does not use AdWords to store the data about which users are browsing which subjects. Instead, it uses the data in real time to target ads without ever actually storing it (even though it collects and stores other sorts of data about users in the process, as detailed in the following chapters).[1]

Google's use of this data fits a model of surveillance through networks, crowds, and social discourse rather than the model of surveillance through direct exertion of power over individuals. Google's use of personal data in its AdWords system seeds a network of grey surveillance that helps constitute the discourse of the larger society. This network includes not only Google's

1 Google does not use AdWords itself to store data about user web histories, but it does collect information about which ads users click on, how they navigate landing pages after clicking on ads, and other data about user behavior associated with AdWords ads. Google has also begun experimenting with using browsing history data collected through its DoubleClick system to target AdWords ads. Detailed discussion of exactly what sorts of data Google collects through AdWords and how it uses that data is in the following chapters.

activities but also the activities of all the other participants – publishers, advertisers, and users – within the network. Google seeds this network rather than creates or controls it in the sense that its biggest role is to create the conditions helpful for the network to thrive and to prune the growth of the network into certain directions. But Google must ultimately rely on the network's own mechanisms to grow itself, and in that reliance Google gives up a great deal of control. The control and impact of this network depend on all of the participants' various watching activities, rather than just Google's. The watching activities range from clearly surveillance to merely watching, but all of them contribute to the control of the larger network, if not to the direct control of the watched subject. Even the larger network of grey surveillance seeded by AdWords does not directly put anyone in jail, but it does drive social discourse – which is the intermediate goal of most surveillance, including the kinds of surveillance that put dissidents in jail.

From Orwell's Big Brother to Foucault's Networks

The archetype for modern state surveillance is Big Brother in George Orwell's *1984*. Orwell describes a world in which everyone is watched always by a combination of two way televisions (that watch as well as display video) and the suspicious eyes of fellow citizens, all strongly directed by the figure of Big Brother. In *1984*, the mechanism and result of control for surveillance is obvious. The Ministry of Love arrests, tortures, and executes anyone discovered acting against Big Brother, and so there is very little dissent. Even in this clear, top down, physically enforced example of surveillance, there are a few interesting complications: the technology of surveillance is embedded everywhere locally, rather than in a big eye in the sky; the surveillance system relies as much on community participation (snitching) as on technology; and the Ministry of Truth and its role of rewriting social discourse by rewriting history matter as much as the physical violence of the Ministry of Love. But the overriding theme is the use of surveillance by a single central actor to exert clear and direct control over all individuals in society.

There are many historical and current examples of surveillance with these sorts of clear mechanisms of control, for example the use of surveillance to support slavery. The mechanisms of control of slave surveillance were obvious – slaves found to be out of the enforced norm (escaping, lack of productivity, etc) were subject to brutal violence by both individuals and the state. The forms of surveillance included obvious forms of law enforcement that look like later mainstream police techniques, such as interrogation. (Hadden 2001, p. 219) Slave masters also used forms of surveillance similar to current forms of workplace surveillance and even inventories reminiscent of the consumer product centered surveillance of AdWords (Raboteau 2004, p. 53; Parenti 2004, p. 15) This is not to drive any sort of equivalence between slave surveillance and Google AdWords surveillance. To the contrary, it is to point out the obvious difference – no one gets whipped or shackled as a direct result of AdWords' use of personal data.

There are likewise many current cases of the use of Internet surveillance to exert direct control over the subjects of surveillance, for example the surveillance of U.S. Internet backbones by the National Security Agency (NSA) and the use of Yahoo registration data to arrest a Chinese dissident. Thanks to documents leaked by an AT&T engineer and subsequent investigations by the New York Times, we know that the data mining equipment was (and presumably still is)

being used by the NSA to watch the Internet traffic flowing over the backbone network connections carrying most U.S. Internet traffic. (Markoff and Shane 2006) This story found public traction precisely because the NSA has the power to whisk people away in the middle of the night. And in 2005, a Chinese office of Yahoo knowingly turned over information about the email account of a Chinese journalist named Shi Tao that contributed to Shi Tao's ten year prison term for disclosing state secrets. (MacKinnon 2007) In both of these cases as in *1984*, a screen both watches and is watched by citizens and leads to police raids on dissidents. And as with the slave surveillance examples, the modes of surveillance in both of the NSA and Yahoo examples resemble the kinds of the surveillance undertaken by a broad range of different companies on the Internet: by companies monitoring their employees, by governments monitoring their citizens, by schools monitoring their students and teachers, and so on.

There is certainly a danger that data processed by Google will end up in the hands of governmental or other directly powerful (corporate, criminal, etc) hands. There is a significant risk that the data processed through AdWords will be stolen from Google. Google has not been the source of any known large data breaches to date, but major corporate data breaches are very common.[2] There is also the risk that any data collected by Google will be subject to government access through warrant, subpoena, or court orders. Google's stated policy is to comply with such government requests for data, and it does so constantly. For example, in 2008 Google handed to Viacom the complete logs of every video viewed by every user on YouTube to comply with a court order in a suit pressed by Viacom against Google. (Helft 2008)

But the argument that we should care about AdWord's use of personal data because Google might give the data away is insufficient for two reasons. The AdWords system is designed not to capture the core flow of data – the search and content terms used to target ads at users. Unlike the DoubleClick system, the AdWords system uses this information in real time to target ads and then drops the data. The search term data is stored by Google's search system, and ad clicks and other sorts of data are stored by AdWords, but the core set of data about user browsing topics cannot be disclosed because Google does not store it. More importantly, without discounting the danger of such potential leaks of data, the question remains whether we should care about Google's use of data on its own terms. Merely arguing that we should care about Google's use of the data because someone else might use it for bad purposes seems disproportional to the amount and intrusiveness of the data being processed.

Michel Foucault's model of panoptic surveillance provides an alternative model to Orwell's Big Brother that is a helpful guide to making sense of the AdWords use of data. Foucault argues that modern forms of surveillance exert control over individuals through the setting and application of social norms rather than through direct application of force. (Foucault 1977) Foucault uses the metaphor of the panopticon to explain his model. The panopticon is an eighteenth century prison design by Jeremy Bentham that allowed an unseen central guard to see every prisoner. The panopticon works because the prisoner knows that he must always behave for fear that he might be watched at any given time. Foucault uses the panopticon as a metaphor for contemporary technologies of institutional surveillance in prisons, hospitals, schools, and

2 In 2008 there were over six hundred data breaches including over thirty-four million records of personal data from government, corporate, and other entities. (ITRC 2008)

factories, all of which inculcate norms into their subjects by making them feel as if they are being judged on the norms of the institutions at all times. Monarchies used the physical threats (á la Big Brother) of spectacular executions to scare their populations into compliance. Modern institutions use the social threats of constant surveillance to make individuals enforce institutional norms on themselves. Any actor (state or non-state) can use the panoptic mode of surveillance because it is based on the subject's willingness to enforce norms on herself, albeit backed up by some threat through an existing source of control (loss of job, loss of benefits, loss of health, loss of schooling, etc.).

This simple but powerful insight – that any actor can use surveillance to exert control through self-enforced norms rather than through the direct use of physical violence – is the basis for most modern social science research on surveillance. Governments use surveillance of welfare recipients to encourage the recipients to strictly control the intimate details of their lives in accordance with welfare agency policies (Gilliom 2001; Eubanks 2006). Companies use surveillance to enforce specific forms of behavior on employees, not only in Foucault's factories but also in hospitals, call centers, and all kinds of modern workplaces. (Fisher 2006; Ball 2002) And entities of all sorts mine the vast sets of personal data in the corporate / governmental datasphere to make constitutive decisions about the subjects of the data: decisions about employment, health care, insurance, finance, and so on. (Solove 2004; Gandy 1993; Lyon 2003) But this insight that surveillance can act through broad impacts on society is not sufficient for determining the social impact of the AdWords system. Google has no obvious agenda nor the power to enforce an agenda on the subjects of its surveillance, so it is not obvious either how Google is trying to make its users behave or how it would enforce the behavior in any case.

The final picture that Foucault draws is of surveillance as a "multiple network of divergent elements" in battle with one another, with social discourse as a key battlefield. Within that argument lie several key points useful for the AdWords case. Under panoptic surveillance, everyone becomes at least a potential surveillor of everyone else. If anyone can do surveillance, anyone can do surveillance, and if the user is responsible for watching herself, the user is responsible for watching herself. By investing power in the crowd to watch themselves, the controlling entity necessarily grants some amount of its own power to the crowd:

> It [surveillance] was also organized as a multiple, automatic and anonymous power; for although surveillance rests on individuals, its functioning is that of a network of relations from top to bottom, but also to a certain extent from bottom to top and laterally; this network 'holds' the whole together and traverses it in its entirety with effects of power that derive from one another; supervisor perpetually supervised. (Foucault 1977, p. 176)

Foucault further argues that the institutions that are the sites of surveillance (prisons, hospital, schools, etc) have replaced the authority of the king (or Big Brother) with the authority of a social discourse that is embedded in and so largely derived from the crowds. Institutions define science as separate from the controlling institutions in order to invest more legitimacy in scientific findings for the crowd. Factory managers rely on scientific management to justify control of their workers. Prosecutors gain power by using scientific experts to condemn the accused rather than merely asserting their guilt through the power of the king. The jury accepts

the judgment of the scientific expert because the expert is separate from the prosecutor. But that same scientific expert relies on the judgment of the jury for legitimacy. By basing the power of social science on its legitimacy with the crowds, the crowds gain some power to control social sciences: "the formation of knowledge and the increase of power regularly reinforce one another in a circular process." (Foucault 1977, p. 224). This point is a key, sometimes overlooked, part of Foucault's argument and is key to understanding how AdWords' use of data impacts society. The impact of Google's use of AdWords data is not just in the targeting of topical ads to users, but also, as described in detail below, in the ability of users themselves to play a central role in determining which ads get targeted to which users.

Foucault concludes *Discipline and Punish* by arguing that the influence of the king has been replaced by a machinery of surveillance that relies on the participation of the community and on a complex feedback loop with the larger social discourse. And so their power is only somewhat controllable by those at the top of society:

> at the centre of this city, and as if to hold it in place, there is not the "centre of power," not a network of forces, but a multiple network of diverse elements ... ultimately what presides over all these mechanisms is not the unitary functioning of an apparatus or an institution, but the necessity of combat and the rules of strategy. ... In this central and centralized humanity, the effect and instrument of complex power relations, bodies and forces subjects by multiple mechanisms of "incarceration," objects for discourses that are in themselves elements for this strategy, we must hear the distant roar of battle. (Foucault 1997, p. 308)

This model of surveillance as a "multiple network of diverse elements" brings to the analysis not just the influence of the subjects of surveillance but also of the publishers, advertisers, and users involved in the various watching activities surrounding Google's own activities.

But Foucault misses the ability of the crowd to exert direct control over this process on equal footing with the institutions. Foucault's model certainly allows for including not only Google but also publishers and advertisers as watchers in the analysis of the AdWords system, but he does not provide much support for including the users themselves as equal actors, helping to control not only themselves but also the larger network (including Google, publishers, and advertisers). Thomas Mathiesen likewise argues for a bottom up version of the panopticon he calls the "synopticon" in which traditional mass media (primarily newspapers and television) act as the primary tool for the crowds to watch the powerful few who run the institutions. (Mathiesen 1997, p. 219) But Mathiesen's core argument is that the synopticon gives more control to institutions by causing the crowd to further strengthen the self-enforced norms described by Foucault. And Roy Boyne extends this argument that the synopticon reinforces existing norms: "The daily television news, the quality newspapers, the blockbuster novel, Coronation Street and the output of Hollywood, few escape some degree of self-identification and self-understanding through repeated exposure to one or more of these and other similar forms" (Boyne 2000, p. 301)

None of these notions of bottom-up surveillance address the community-based production enabled by the current form of the Internet (Mathiesen in 1997 actually predicts a dystopic Internet as yet another mass broadcast system similar to the early, walled garden version of

AOL). The scope and speed of mass user interaction on the Internet have altered the balance of power between those controlling the large institutions and the crowd by enabling the crowd to insert its own norms directly into the larger social discourse. (Benkler 2007; Shirky 2008) As argued extensively in the following chapters, users within the various AdWords systems play at least an equal role (as compared to Google, publishers, and advertisers) in driving the decisions the network makes through and about its participants.

Philip Agre proposes a new model of data collection as "capture" rather than "surveillance" that helps to start to make sense of mass participation in surveillance networks. (Agre 2004) Agre's key insight is to focus on the impacts of the systems of measurement (the capture) rather than on the watcher's use of the data. He argues that the simple act of capturing certain types of information in certain ways has constitutive impacts on the subjects of the capture:

> As human activities become intertwined with the mechanisms of computerized tracking, the notion of human interactions with a "computer" – understood as a discrete, physically localized entity – begins to lose its fore; in its place we encounter activity-systems that are thoroughly integrated with distributed computational processes. (Agre 2004, p. 743)

Agre is arguing that the mechanism of input into the capture / surveillance system is powerful in itself. Because processing data about users (and publishers and advertisers) through a system of realtime feedback is critical to Google's success, it builds its tools specifically as data processing and feedback tools as much as content and advertising tools. This focus leads to the kinds of networks of grey surveillance described below in which, for example, it is difficult to tell where the users (and publishers and advertisers) end and where the Google nervous system begins. The mode of interaction as continuous, real time feedback is as important as what Google ultimately does with the data itself. Google does not merely collect data, analyze it, and spit out results consumed by users (or publishers or advertisers). Google's model of capture is inherently reflexive. Users view the AdWords ads as ranked by Google; Google watches the users watching the ads and changes the ad rankings as a result; users view those re-ranked ads; Google views the new reaction to the new rankings; and so on. This model of capturing user (and publisher and advertiser) behavior through feedback systems is a main theme of the following sections.

Networks of Grey Surveillance

A network of the grey surveillance is the answer to the question “who watches who, how, and why?” Each of the resulting watching activities may or may not qualify as clear surveillance individually, but each acts as a point of control in the larger network, which has its own emergent impacts beyond those of the individual activities. This network model blurs black and white distinctions between who is a surveillor and who is a subject of surveillance, between which watching activities qualify as surveillance and which do not, and between the individual and aggregate impacts of any given activity in the network. A common feature of all of these activities is that they function with a network of complex feedbacks, and so they have to be understood both individually as a part of that larger network.

Three core questions help to analyze each of the individual activities on both the micro and macro levels: Does the activity grant the watcher (or allow the watcher to exercise) constitutive

control over the subject? Does the surveillor fail to get meaningful *consent* from the subject? Does the activity use data out of *context* or within the context of other personal data to alter the equations of control and consent? An individual watching activity is clearly surveillance if it is used directly to exert constitutive control over the subject of surveillance using data collected without the subject's consent, taking into account the context in which the data is used. On the micro level, the answers to these questions identify which individual activities are clearly surveillance (those with control and lack of consent), which are weak surveillance (those with control or lack of consent), and which are merely watching (those with neither control nor lack of consent).

This simple test is subjective, but it helps determine where the network is driven by surveillance activities with clear control and lack of consent and where it is driven by activities that in themselves are merely watching. Two core hypotheses of this paper are that the larger network is driven as much by the merely watching activities as by the clearly surveillance activities and that individual watching activities matter according to their impact on their larger network rather than according to whether they qualify individually as surveillance. Even if the judgments about some individual activities are off, the important conclusion is that there are a range of activities between clearly surveillance and merely watching.

These questions will tease out, one by one, the many points of control within the larger network. The fact that Google controls advertisers by enforcing a laundry list of rules for AdWords ads is an interesting case of clear individual surveillance in its own right. But this particular activity plays a more important role in the larger network of activities (Google watching advertisers to enforce rules, advertisers watching Google to learn the rules, Google watching users to rank ads, users watching advertisers to read and follow ads) that determine how AdWords ads are ranked. The exercise of making difficult and often subjective judgements about the control, consent, and context of each activity within the network helps to make those conclusions about the larger network more clear.

The guiding question of control is whether the watcher is using the collected data to exert constitutive control over the subject of surveillance, borrowing the sense of constitutive impact from Paul Starr's treatise on the creation of modern media: “By *constitutive* choices I mean those that create the material and institutional framework of fields of human activity.” (Starr 2004, 1) Constitutive decisions do not necessarily rise to the level of the totalitarian decisions that Jed Rubenfeld has argued should bound privacy issues. (Rubenfeld 1988, p. 787) Think of the influence of the kindergarten school teacher. Her disciplinary instruction is likely to have constitutive effects on the lives of her students but is unlikely to determine their lives totally. At the same time, this framing rules out shallow effects, such as the control Google exerts over whether a given user sees a given mortgage ad in a given list of ads.

This notion of constitutive control weakens the label of surveillance as applied to cases of totalitarian control. To describe the monitoring of call center workers as surveillance is to risk minimizing the harm done through surveillance of slaves. But in a world of panoptic surveillance, individual acts of surveillance are largely self enforced and so are limited in the control they can exert. But all of these individual acts of self enforced surveillance strongly

influence the larger social and political order of the state. A danger of requiring that the surveillance label only be applied to totalitarian surveillance is the same as the danger of buying into the assumption that surveillance should be judged as a balance between security and liberty. Security is more than security from being bombed, and liberty is more than liberty from imprisonment (Monahan 2006).

For the question of consent, the guiding principle is to identify whether the consent is meaningful, in the sense that the subject of surveillance is aware of what data is being collected and how it is being used, has agreed to the collection and use of the data, and has the realistic option to prevent the collection and use of the data. As pointed out by Gary Marx, many people are heavily dependent on the subway to go to work, to go to school, to buy food, and so on. To claim that these people have meaningfully consented to surveillance simply by riding the subway does not hold water because the consent is structured in a way that opting out of the surveillance has wildly disproportionate costs. (Marx 2006) In the digital world, it is likewise meaningless to claim that cell phone users consent to the monitoring of their calls because use of a cell phone is a basic utility of modern life for most people in the developed (and many in the developing) world.

When considering questions of control and consent, the issue of context must be considered as well, both whether the data is being used out of context to misrepresent the subject and whether the data is being combined with other, contextual data to create a more detailed (and therefore less consensual) picture of the subject. The use of data out of context is one of the biggest dangers of the use of private data and is the justification for Warren and Brandeis' principle of the "right to be let alone," which serves as a foundation of U.S privacy law. (Warren and Brandeis 1890) Jeffrey Rosen has expanded on Warren and Brandeis to argue that invading people's private thoughts leads to a break down of personal relationships, which cannot flourish without the time and space to develop the deep social relationships that give context to isolated bits of personal data. (Rosen 2000)

Data can also be combined with other, contextual data to create a much more invasive aggregate data collection. For example, both AOL and Netflix released large collections of personal data (AOL released search terms, Netflix movie ratings) keyed to unique but random user tokens. Researchers quickly identified many of the supposedly anonymous users by combining individual bits of data. An "anonymous" user who searches for her home address, employer, and social security number (as many users do) is not long anonymous. (Narayanan and Shmatikov 2006; Electronic Frontier Foundation 2006) Issues of context are particularly important when thinking about Google because of the vast width and depth of the data collected by Google through its many different services. A user thinking about the privacy of her search terms on Google must think not only about the privacy of the search engine, but also how that data could be combined with any data she has left in GMail, in Google Calendar, in Google Health, in Google Docs, and on and on.

It may be jarring to group Google's use of data about who is visiting sports web sites and even who is visiting porn web sits with the use of on- and off-line surveillance by totalitarian states to jail and kill dissidents. But Google's activities matter to the extent that they affect the larger

social discourse, and driving social discourse is a key goal of most kinds of surveillance. There are some examples of top down control through extensive, micro surveillance which directly control the actions of all of its subjects. Call center monitoring of employee is a good example of this sort of surveillance -- there are few enough workers and extensive enough surveillance that the employer can directly monitor essentially every action of every employee every moment and correct as needed. In these cases, there is no need for the surveillance to drive some larger discourse within the workplace because of the comprehensive nature of the surveillance (to be sure, even call center workers have methods of resistance). In most cases of surveillance, however, the goal is to use surveillance as a lever to drive changes in the larger society precisely because surveillance cannot be comprehensive. This is the core of Foucault's argument: in industrial factories where it wasn't possible to watch every action of every worker, employers controlled workers merely by watching a few of them and convincing the rest that they might be watched as well.

Even the dramatic results of prototypical, dictatorial surveillance, such as the arrest and imprisonment of dissidents in totalitarian states, primarily lever larger social discourse to impact society. The most significant impact of imprisoning a dissident is usually not to enforce some sort of ideal justice on the dissident but rather to disrupt the dissident's own efforts to transform his society by discouraging others from engaging in similar work. This is not at all to discount the real and horrible suffering that is caused by the use of dictatorial power to jail, kill, and otherwise summarily control people. But the biggest social impact of almost any surveillance activity is its use as a social lever, and any such social levering in modern societies is necessarily intertwined with widely dispersed and difficult to control changes to the larger social discourse. Even though less dramatic and less individually hideous, grey surveillance activities like those in Google AdWords matter because of the tremendous power they have to shape the same social discourse as police state surveillance.

AdWords Ranking, Terming, and Valuing

To appreciate Google's power to shape discourse requires understanding how AdWords works, both technically and socially. The rest of this chapter provides a roadmap for the rest of the paper, describing AdWords as three related networks of grey surveillance: the ranking part that drives its search and ad lists, the terming part that drives its association of ads with content, and the valuing part that drives its valuation of ads. But each of these networks relies heavily on the others, and so it is necessary to understand how the whole system works together before delving into the individual parts. This chapter will start with a high level description of what AdWords does and then will describe briefly how the watching activities within each of the parts of AdWords act as points of control within a network of grey surveillance.

Google's AdWords is an ad brokerage system. It collects payments from advertisers and distributes them to publishers who host the ads. Instead of the traditional model of displaying ads on manually chosen sites, AdWords displays the ads according to the content of the hosting web page ("travel," "new york giants," "perfume"), and advertisers pay the host each time a user clicks on an ad. Google makes money from the system both by hosting ads on its own search and other sites and by collecting a commission for all ads hosted on other sites.

For example, the following ad for smartertravel.com appears in the AdWords box underneath the headlines on cnn.com:[3]

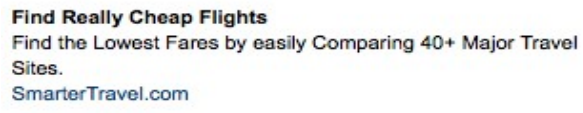

Figure 2.1: AdWords ad for smartertravel.com on cnn.com

Smartertravel.com bid enough, about $1.48, for Google to display its ad first in ad listings alongside content and searches about "smart travel."[4] Google's AdWords engine thinks the cnn.com home page was about "smart travel" in some sense, so Google includes the smartertravel.com ad in the AdWords list on cnn.com. Every time a user clicks on that smartertravel.com ad on cnn.com, smartertravel.com pays cnn.com $1.48, minus Google's brokerage fee of about twenty percent. The same advertiser also appears first under the Google search for "smart travel:"

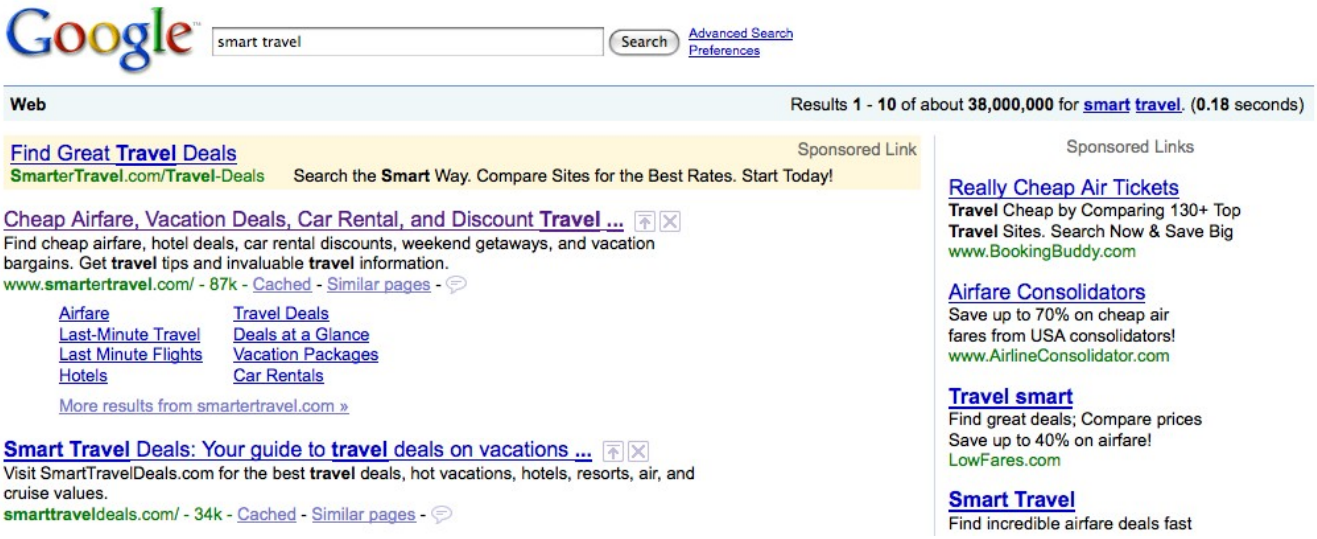

Figure 2.2: Google search results for "smart travel"

Every time someone clicks on the smartertravel.com Google ad, smartertravel.com pays Google the full $1.48. Smartertravel.com also appears first under the Google search results for "smart travel," but the process that lists smartertravel.com in the search results is separate from the AdWords ad placement system, and smartertravel.com does not pay Google anything for clicks on the search result link.

---

3 The smartertravel.com appeared on cnn.com on December 5, 2008 along with two other ads.

4 $1.48 bid price is the estimated average cost per click (CPC) for the words-as-topic "smart travel" returned on April 16, 2009 by the Google Keyword Tool (https://adwords.google.com/select/KeywordToolExternal). The price required to gain the top ad spot on pages about "smart travel" is probably somewhat higher than that $1.48, though as discussed at length below, ad ranking encompasses many factors and so the exact bid price for a given advertiser is only known by Google and that advertiser.

The core data that drives this process is the topic of the web page currently being viewed by the user. But as described above, that data is not simply used by Google in a top down fashion to control the user browsing the web pages; the data is not even used by Google to control fully the ad listings. The data is instead used to drive a set of systems that list the smartertravel.com site in "smart travel" search and ad listings, that associate the smartertravel.com ad and the cnn.com content with the words "smart travel," and that value the ad at $1.48 per user click. Google has specifically designed those processes to distribute much of the control to the participating publishers (like cnn.com), the participating advertisers (like smartertravel.com), and the users who read the content and the ads, within some basic parameters set by Google (including that the system is commercial in nature and is therefore largely driven by consumer products among other things). Together, the activities of all of the participants constitute a network of grey surveillance, seeded but not solely controlled by Google, that determines what sorts of ads appear within AdWords and what sort of content those ads support.

AdWords technically refers to only one of several sub-systems (the one that attaches the smartertravel.com *ad* to the *word* "smart travel"*)* that constitute the larger AdWords system, along with Google's search and AdWords ad ranking systems and the AdWords pay-per-click / ad auction payment system. The Google search and ad ranking systems are responsible for placing the smartertravel.com link, respectively, first in its search results for "smart travel" and first in the ads that appear alongside "smart travel" content. Google's search tool is likely the biggest host of smartertravel.com's AdWords ads and as such is the biggest source of AdWords clicks for smartertravel.com. The Google search ranking system also drives traffic directly to the smartertravel.com site through its search results link.

This traffic is used by smartertravel.com among other ways to host its own AdWords ads in its role as a publishers as well as advertiser. The core of Google's search ranking is the PageRank system, which mines the collective intelligence of publishers and users to determine which sites are relevant to "smart travel." Smartertravel.com ranks highest for the "smart travel" search according to the PageRank model because more sites link to it as a site about "smart travel" than any other site and because over time users have confirmed that ranking by following its search result link after searching for "smart travel." Google also uses a ranking system to decide to place smartertravel.com at the top of the ad listings for "smart travel," combining the bid amount from smartertravel.com with the number of times users click on the smartertravel.com ad and the ad's adherence to Google's extensive set of rules and guidelines for AdWords ads.

There are a variety of different watching activities involved in the ranking of the smartertravel.com search and ad links, some of which look like surveillance and some of which do not. The following figure maps out the various ways that the various actors watch one another in order to rank Google's search and ad results:

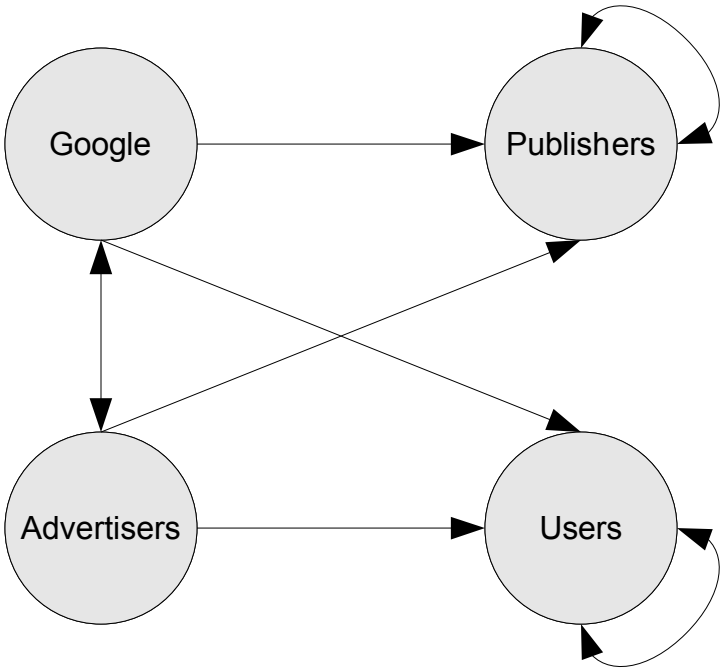

Figure 2.3: Google ranking network

Each of the watching activities in the above map represent a point of control in the network of grey surveillance that powers the ranking systems. To generate its search rankings, Google watches millions of publishers both to collect information about which publishers are linking to which pages and to enforce its own manual rules for search results. The publishers in turn watch themselves both by linking to one another and by self-enforcing Google's rules. Google also watches billions of users to measure how well PageRank and its various other ranking signals are working as a user simulation. This search-as-user-simulation model puts users in the role of watching themselves by turning every Google search into further feedback for the system to refine its user simulation. Similarly, to generate its ad rankings, Google watches millions of advertisers to enforce its much larger set of rules for the format, content, and links of AdWords ads. The advertisers watch both Google and themselves to carefully interpret and self-enforce these rules for ads. And users watch the advertisers by reading the ads and in the process providing data to Google about which ads are most interesting to users.

The collective scope and speed of each of these activities and their feedback mechanisms make the network a virtual Google nervous system that continuously reconfigures itself according to the interests of the various actors by strengthening some links and weakening others. Google is only one actor out of many in this network, and its control over the network is limited by the fact that the network derives its power from harvesting the wisdom of the participating publishers, advertisers, and users. Google does exercise its veto power over the network in various ways, but it does so knowing that every exertion of this power weakens the power of the network as a whole (and therefore weakens the power that Google wields through the network).

The search and ad ranking systems only place the smartertravel.com link at the top of the results once the AdWords system has determined that the link is relevant to “smart travel.” Google uses

its AdWords and AdSense terming systems to determine that the smartertravel.com ad should be displayed (in the order determined by the ranking system) alongside pages about “smart travel.” The AdWords system proper (the part of the larger system that Google calls “AdWords”) connects each ad to a specific word or phrase (thus the “words” in “AdWords”). The complementary AdSense system derives a relevant set of specific ad words from the content on each partner publisher site (and so allows the content to be connected to the ads associated with those words). This model of attaching ads to specific words-as-topics allows Google to profit from its search results without losing the integrity of its core search results (displaying the ads separately as “sponsored links”). By generating those ad words from the “sense” of arbitrary pages of content, Google is able to include those pages in its AdWords ad brokerage system just as it would include a search engine with explicit search terms for each page.

The following figure maps out the watching activities that combine to attach words to ads within AdWords:

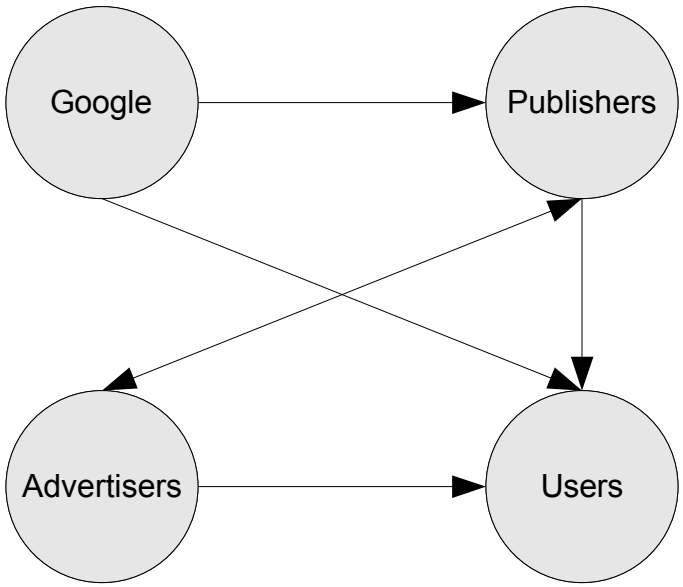

Figure 2.4: AdWords terming network

As with the search and ad ranking network, each of the activities represents a point of control in the larger network. Google watches the topics of users as they browse web pages and uses that data to target ads to users reading about specific words-as-topics. Google also watches the content of publishers participating in AdSense to generate the words-as-topics for the publishers' content. Advertisers watch publishers through Google tools that allow them to exclude their advertising from specific sorts of content, like war and famine. And publishers watch advertisers through Google tools that allow them to determine the most profitable words-as-topics. All of these activities form a network whose effect is not only to encourage the production of content that is relevant to specific, profitable topics but also to encourage that content to do the work of giving social meaning to those ads.

The system that determines that “smart travel” content is worth $1.48 per click is Google's

AdWords ad auction. The auction allows advertisers to bid on the price they will pay to a publisher each time a user clicks on an ad hosted by that publisher. The auction mechanizes the process of valuing an advertising topic, allowing the community of advertisers set the value of a given words-as-topic and thereby allowing Google to scale the AdWords system to a vast number of advertisers bidding on a vast number of advertising topics appearing on a vast number of publishers with little manual (but plenty of technology) work. A map of the watching activities within the AdWords pay-per-click ad auction system looks like this:

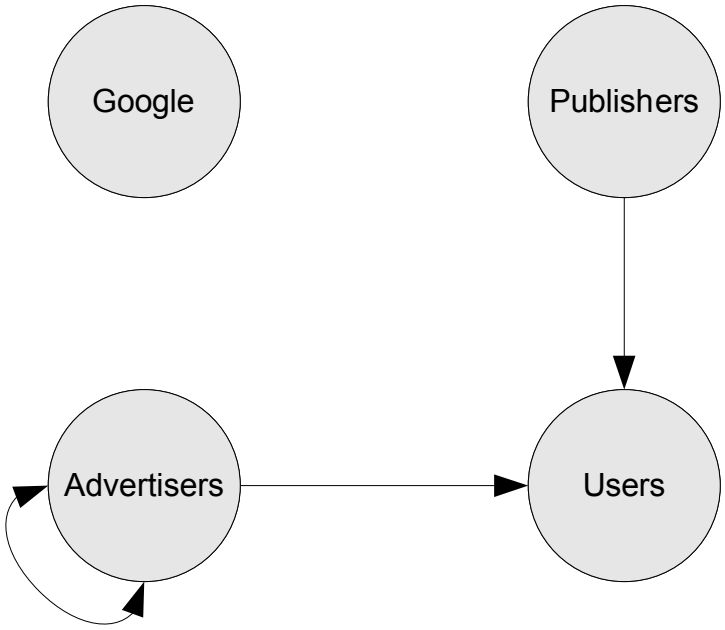

Figure 2.5: AdWords valuing network

Google's great innovation in the valuation of ads is that it plays no role other than to design the model itself, thus Google's isolation in the figure above. Google instead relies on the auction mechanism to scale the ad valuation process up to millions of advertisers and publishers. The auction serves as a form of watching in this system, allowing the advertisers to watch the value that other advertisers place on a given words-as-topic simply by participating in the auction. To make informed decisions about the values of ads within the auction, advertisers and publishers both watch users as well to track per-click payments (and, through the per-click payments, the ultimate effectiveness of the ads they buy or sell). Where in the past advertising has relied heavily on the magic of creative advertising agencies both to create advertising content and to value ads, AdWords relies on the spectacular accuracy of a machine that makes emergent decisions based on the inputs of a vast crowd of users, publishers, and advertisers watching each other.

The goal of the rest of the paper is to understand how the networks of grey surveillance that make up these systems decide that smartertravel.com is the most relevant search and ad link for

“smart travel,” that cnn.com's home page means “smart travel,” and that the act of a user clicking on “smart travel” is worth $1.48. The scope of the paper is limited to a specific company's implementation of a specific advertising model. AdWords' large market share makes it a major influence on the online advertising and publishing markets, but there are certainly other advertising models online, including Google's DoubleClick system as well as a cadre of smaller competitors. And even though Internet advertising now produces as much revenue as either cable or broadcast television advertising, it remains only a small portion of the entire advertising system. (Internet Advertising Bureau 2008) And a large portion of content created online – notably on Wikipedia but also on a vast number of ad-less blogs, educational and government sites, and other non-profit oriented content sites – is not directly influenced by any advertising model at all. The point of this paper is not to argue that the influence of AdWords is greater than or less than other models of advertising or that all publishing is driven by AdWords. The point is rather to explore the effects of this one model so that we can better understand both the implications of its growing influence and how grey surveillance works as a mechanism of control.

This paper is also limited by considering Google, publishers, advertisers, and users mostly as discrete units of analysis. It is not the case that these groups are totally or maybe even mostly distinct from one another – Google uses its own products to publish, advertise, and consume content on the web; publishers advertise their own content and acts as users when performing searches; and so on. It is also not the case that Google, publishers, advertisers, or users are homogeneous groups that act in lockstep. Google is a large company made up of many employee, executives, shareholders, and other stakeholders with diverse goals and actions. Google's users, publishes, and advertisers are almost as diverse as the entire population of Internet users (indeed one of the core benefits of AdWords has been to increase the classes of people that can publish and advertise through it). With some exceptions, this paper treats these classes of actors as distinct and consistent to make it possible to explore at least from a high level how they interact with one another to create a set of network effects. Certainly one fruitful line of inquiry following this paper will be to tease apart what it means, for example, to say that “Google” does anything as a discrete actor in this space. Likewise, further inquiries might explore which specific users are clicking on what sorts of ads and how the resulting differences in classes of users will drive the larger network of grey surveillance.

3. Ranking Search

Google commonly refers to its search engine as an artificial intelligence, but the search engine is both less and more than an artificial intelligence. It is less than an artificial intelligence because it does nothing when disconnected from the crowd of millions and billions of publishers, advertisers, and users from which it draws its intelligence. It is more than an artificial intelligence because it does not merely exist on its own, but is instead intimately integrated into the lives and thoughts of the crowd that powers it. That network acts as a Google nervous system – continually collecting, processing, and adjusting to the feedback of users, publishers, advertisers, and Google itself. Calling the network a nervous system instead of an artificial intelligence better captures this sense that it operates in the closest possible integration with the crowd.

The core work of the search ranking system happens in real time – in the equivalent of short term memory. It relies heavily on vast, immediate collection of publisher, advertiser, and user feedback; accepts the collective input continuously; and adjusts its behavior immediately in response. Google maintains a level of control over the system through a combination of manual editing, explicit rules for publishers and advertisers, and the design of the feedback model itself. But Google's degree of control is limited by the nature of the feedback system itself, which must be largely driven by publishers, advertisers, and users to maintain its power. The power of the system derives not only from the ability of the system to react instantly to and even anticipate the needs of the publishers, advertisers, and users who drive it, but also from the deep integration with the various actors that its collective nature fosters.

Library card catalogs, to provide just one example, have long provided indices on which people rely to sift through vast collections of information. So why talk about a Google nervous system but not a Dewey Decimal nervous system? Bruno Latour asked precisely this question in a talk about "Virtual Society:"

> Now what is more disembodied, abstracted, de-localised than the good old paperwork offices of the 1950's? That is, a large part of the ancestry of Virtual Society has to do not only with geography and maps, but has to do with the whole practice of bureaucracy, folders, files, and it's not an easy question to answer. What does the computerisation, the digitality add to these offices? (Latour 1998)

Latour further questions the ability of digital systems to capture the notion of "society," without getting lost in the small details (indeed, breaking down continuous reality into tiny discrete bits is the definition of being digital). Latour describes a commuter's dilemma: commuters feed the details of a traffic jam to a computer and are in return fed back the information about the traffic jam, but what use is this information for those stuck in the traffic? Google's answer to these questions is to accelerate the collectivity of the user experience. That collectivity emerges into a larger social experience, accelerating communication enough to allow deep participation of the community beyond even Latour's traffic example: by improving the index to the point that it recursively generates ever more data and therefore an ever better index, and by integrating the resulting multitude of small details into the larger life and thought processes of the digital

participants.

Google's search ranking system relies on extensive monitoring of users, but not monitoring in the traditional Big Brother or even factory foreman sense. The key piece of information the search ranking system gathers is what sorts of content the publishers and users themselves are watching. This information is continuously fed back into the system, making the system fundamentally a collective one, and the feedback system is understandable only as a network. The output is the result of the flow of data between the various different surveillance activities rather than the observations of one actor. As such, the various actors within the network collectively have more power to control the results that come out of the network than does Google. Google may have absolute veto power over the results of the network (exercised for example by manually filtering search results in response to government fiat), but the more Google uses that veto power the more it lessens the value of the collective input that powers the overall network.

Ranking Search: Google Watching Publishers

Google claims that it uses more than two hundred different measurements to rank search results, but the core of its search ranking is its PageRank algorithm. Google says that it counts each link to a given site as a vote for the importance of that site, then repeats the process iteratively, giving greater votes to sites that have themselves received lots of votes (Google Inc. Technology Overview 2008) So Yahoo has a very high page rank because other sites vote for it by linking to it. In return, Yahoo linked sites get more votes within PageRank because Yahoo itself has a high PageRank. This system makes intuitive sense: the sites that are most interesting are the ones to which other sites link the most. And a site linked to by an interesting site like Yahoo is generally interesting even if Yahoo is the only site linking to it.

Google uses PageRank to determine only the general importance of each site in relation to other sites without regard to topic. To determine the topic of the site, Google watches both the text on the page (including the prominence of the text through font sizes and other visual cues) and critically the text within links to the site from other sites. The initial, surprisingly successful strategy for Google's search ranking was simply to search for sites with text matching the query (where the text includes the text of the links to the site) and to rank the results by the PageRank score. (Brin and Page 1998) Google closely guards the other criteria used for ranking search results as trade secrets and as protection against gaming its ranking algorithms. But Google includes some measurement of which search results users click on in the aggregate. Google considers continuous experimentation on its users through this click data with new ways of sorting and displaying search results to be a core philosophy of the company. (Huffman 2008)

Google's monitoring of publishers through their web sites to generate its PageRank score clearly qualifies as surveillance. Google monitors publisher web sites closely and uses the collected data to exert direct, constitutive control over publishers without their meaningful consent. Google's page indexing system collects the content of almost every public web page.[5] Google uses the content of a given page to determine what the page is about and uses the outgoing links

5 Google allows publishers to opt out of its search indexing system. Google also cannot index some pages that require form submissions or use non-html formats, though it has been steadily increasing its ability to index both of these sorts of pages.

of the page to determine the importance and topics of the linked pages. The topic analysis of the page content is controlled completely by Google. Google directly determines for instance how to weigh text in the title, text in the body, highlighted text, text with search terms in close proximity, and so on.

Google eliminates from its index entirely pages that it determines, through direct analysis of the content, are violating certain web page quality guidelines. Google says that these guidelines are intended to prevent sites from artificially boosting their search rankings, to discourage users from creating spam content solely for the purpose of generating ad click or affiliate revenue, and to prevent the spread of malware. For instance, Google removes from its index sites that it decides are incorporating blocks of hidden text meant only to improve search rankings, are including different (search ranking friendly) content for the Google indexing bot than for users, or are automatically generating "low quality" pages solely for the sake of generating affiliate links. (Google Inc. Webmaster Guidelines 2008)

Google directly uses the content topic analysis to determine which monitored sites appear under searches for which terms. This use is a good example of traditional use of surveillance power to control a surveilled subject. Google collects data from the subject, the web site publisher, and then uses that data directly to control the subject, by either changing the site's search ranking or removing it from the search results altogether. The ultimate impact of Google's topic analysis is unclear because Google only discloses the broadest information about the calculus of its search rankings. However, the impact of eliminating non-conforming sites from Google's index is obvious; it is the online equivalent of a jail sentence for any site that depends on Google searches to generate traffic.

Two of the most well known lawsuits over Google's search ranking (SearchKing v. Google and Kinderstart v. Google) involve Google's use of this power to control search rankings. In the SearchKing case, Google removed the site from its index because SearchKing was selling outgoing links for the express purpose of increasing its clients' search rankings, in clear violation of Google's webmaster guidelines. In the Kinderstart case, casual examination of kinderstart.com shows that it is predominantly a list of links to other sites. So it is likely that Google judged the site to be in violation of the guideline against sites that merely aggregate existing content to generate ad revenue. In both cases, Google made the decision to remove the site either through a manual judgment or through some sort of analysis of the content of the site. The court dismissed each case in Google's favor, in the Search King case on free speech grounds among others. The harm claimed in both cases was that Google was unfairly undermining the will of the community by manually overriding the link-as-vote PageRank model. (Bracha and Pasquale 2008; Goldman 2006)

Google also has the ability to control search rankings by manually editing the results, and it exercises this power often, primarily by complying with government requests to filter its search results in various ways. Among many other examples, Google has agreed to work with the Chinese government since 2006 to filter search results on google.cn (its Chinese language search site) such that, for example, a search for pages from the site for Human Rights in China (http://hrichina.org) returns 4,380 pages on U.S. Google (http://google.com) but no results on

Chinese Google (http://google.cn).[6] (OpenNet Initiative 2007) Since at least 2002, Google has been removing a variety of content, mostly neo-Nazi sites, from its index on its French and German search engines. (Zittrain and Edelman 2004) Google has been obeying court orders to filter search results about a large number of celebrities and public officials in Argentina. (Valle and Soghoian 2008) And Google complies with the Digital Media Copyright Act in the U.S. by removing links from their search results in response to copyright violation complaints. (Google Inc. Digital Millennium 2008)

Sites that rely heavily on search referrals for survival have no meaningful choice about whether or not to participate in the Google indexing system, because Google hosts over sixty percent of all searches. (comScore Inc. 2008) Google allows sites to opt out of indexing through technical means, including the web standard robots.txt file and Google's own sitemap system. But opting out of indexing is not a meaningful option for sites that depend on traffic for survival (whether for the generation of direct ad revenue or for motivation to continue producing the site). Sites can technically opt out of Google's indexing, but exercising that option would cut the traffic of most sites fatally.

Google collects so much information about so many publishers in so many ways through its various tools and services that this potential for aggregating its data alone raises concerns about the context of the data. In the digital world, Google *is* context. The collection and analysis of a variety of large sets of data is its reason for existence. Because of the vast amounts of data collected by Google through its wide variety of different services (search, advertising, mail, office tools, calendar, and on and on), Google always has at least the potential to integrate data about a surveillance subject from any one service into the larger context of data about the same subject from a variety of other sources. And while Google has not disclosed to what degree it is currently sharing data between its various services, it does reserve the right in its umbrella privacy policy to "combine personal information collected from you with information from other Google services or third parties to provide a better user experience." (Google, Inc. Privacy Policy 2008)

In this specific case, Google possesses a vast store of information about publisher web sites through its various services – including links to sites in Gmail and Google Calendar; advertisements run on and by sites through AdWords and DoubleClick; and traffic statistics through Google search, Google Analytics, and other tools. It is not clear whether and how Google uses this data in its search rankings because Google closely guards the details of its search ranking methods. But we know that Google has the ability to access this rich store of data about publishers, so it is at least possible that Google is using some of the contextual information about publishers from its other services.

All of the elements of surveillance – control, consent, and context – combine to make Google's monitoring of publishers look like clearly surveillance. But Google's ultimate control over the rankings is limited by its reliance on the collective nature of its PageRank algorithm. Google's watching of publishers acts as the seed of the larger network of grey surveillance that produces

6 Searches performed by author on google.com and google.cn for "site:hrichina.org" (the url of a large, U.S. based Chinese dissident organization) on January 23, 2009.

search results. The power of the system ultimately comes from the collective intelligence of publishers harvested by the PageRank algorithm. Indeed, the design of the AdWords system is a direct outgrowth of Google's need to display advertiser supported links without tainting the collective nature of its ranking algorithm.

Ranking Search: Publishers Watching Publishers

Google's PageRank model consists not so much of watching publishers as of watching the publishers watch each other. As such, the PageRank gives at least as much control over the ranking to the crowd of publishers as to Google itself. The basis of the Google PageRank mechanism is harnessing the fact that publishers are always watching themselves and that links between sites are an indication of who is watching whom. This is the sense in which a link to another site is a vote for that site. A link indicates that another site is worth watching, not in an abstract sense, but rather in the practical sense that a link is an encouragement by a publisher for its users to look at another site.

This watching of publishers by publishers looks like weak surveillance because even though some individual publishers exert a high degree of control over other individual publishers through the choice of whether to link (and thereby increase the ranking of the linked site), it is hard to argue that there is a lack of consent. Publishers watch by linking, and in almost all cases, sites want other sites to link to them. There is a low level of contextual use of data in general, because publishers do not generally have any more data about the sites to which they link than other actors.

Publishers are actively and knowingly participating in the network of surveillance around Google's search ranking. Most publishers are well aware that linking serves not only to direct users to another site, but also to increase the linked site's Google search ranking. It is common practice, for example, for bloggers not to link directly to a site or article with which they strongly disagree to avoid increasing the search ranking of that site. And publishers that have sites with a high PageRank can exert strong control over the search rankings for specific terms, especially uncommon ones. A single link for a site with a very high PageRank can move a link to the top of the search results for any but the most popular search terms.

There is an active underground economy in the sale of outgoing links known as Search Engine Optimization (SEO). Many SEO companies offer legitimate advice about how to increase a site's search ranking (including relevant terms in page titles, link to important sites in the hope that they will discover and return the link, etc). But many other SEO companies offer services that clearly game Google's ranking model. For example, the first Google search result for “buy PageRank” is a site called IncreasePR that allows web site owners to buy increased PageRank by adding a variety of incoming links, including “Reciprocal Link Buidling [sic],” “Directory Submissions,” “One-Way Linking,” “Article Publishing,” and “Blog Submissions.” (IncreasePR 2008) Any sale of outgoing links for the purpose of increasing a site's PageRank is specifically disallowed by Google's webmaster guidelines, but in addition to companies that explicitly offer to directly sell links, many sites like IncreasePR couch the sale of links within more acceptable language such as “directory submissions.” The large number of such sites in the face of Google's active attempts to quash them (as in the SearchKing case) indicates that publishers are well

aware of the active role they play in Google's search ranking system.

Publishers acting together have manipulated the search ranking using a technique called Google bombing: encouraging web publishers to use a specific phrase to link to a specific site to make that site appear at the top of the search results for the phrase. During the 2004 U.S. presidential election, political bloggers led Google bombing campaigns to make a search for "waffle" return the John Kerry campaign page and a search for "miserable failure" return the George Bush campaign page:

Figure 3.1: Google bombing on "miserable failure"

Google changed its ranking algorithm in 2007 to make this sort of manipulation more difficult, so "miserable failure" searches returned pages about the well known "miserable failure" Google bombing as a phenomenon rather than the link directly to George Bush. (Cohen 2007) Even though Google was typically closed about the exact change to the ranking algorithm, Google presumably tweaked the ranking model to give more weight in some cases to the text within the linked page and less to the text in the link to the page. Google specifically claims not to have manually changed any results but to have relied purely on a change to its model. (Moulton and Carattini 2007)

Liberal bloggers attempted to Google bomb the term "john mccain" during the 2008 election to increase the ranking of a video depicting John McCain as a flip-flopper. (Bowers 2008) Shortly after the election, the Google bombed video had risen to the fifth result in a search for "john mccain":

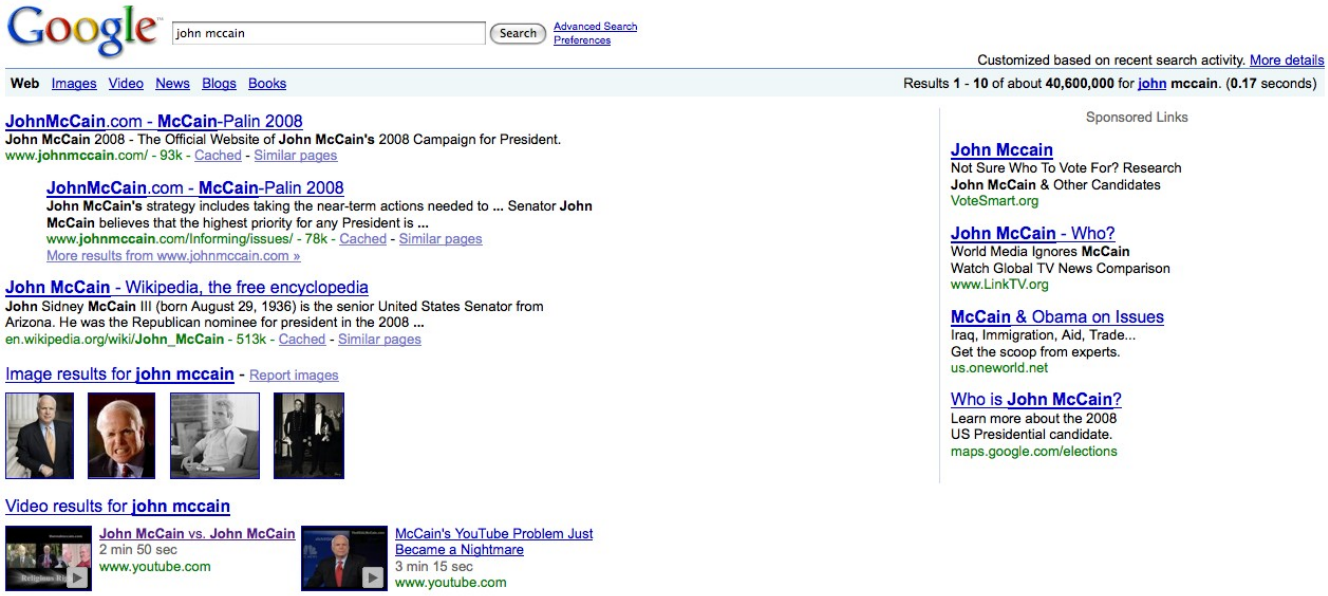

Figure 3.2: Google bombing on "john mccain"

Google bombers find it much more difficult to effect the rankings of a very popular search term (such as "john mccain" during the election) because there is a much larger pool of existing links for that topic that serve as ballast. It is impossible to determine whether the high ranking of the video was related to the Google bombing attempt or arose out of the community through some combination of the video's own merits and promotion by the Google bombing campaign. It is difficult to tell the difference between "true" community interest and Google bombing in any case: what if the result of lots of links in blog posts was not only to increase the search ranking in its own right but also to encourage users to watch the video? This ambiguity around the effectiveness of the McCain video Google bombing demonstrates well how the threads of control over PageRank are hopelessly tangled between Google and the community of publishers.

This is the same confusion that Foucault describes as the impact on social discourse of the modern jury process. Consider the network of actors that Foucault describes in the modern jury trial – a judge watching the process by ceding judgment to the jury, the jury watching the defendant but also the various scientific experts testifying about the meaning of the evidence, the scientific expert watching the evidence but also the impact of her testimony on the jury, and potential defendants (which is to say everyone) watching the whole process to determine how guilt is defined. The process of presenting the scientific evidence to a jury is an inherently social one, so the judgment (and surveillance) of the jury is a social act as much as a scientific one.

And the need for presenting the evidence socially in turn impacts how the scientific evidence is generated, creating a feedback loop between science and society that no one in particular controls. This messy process described by Foucault parallels the messy process between Google and publishers that produces search rankings.

Jill Fisher's exploration of workplace surveillance in hospitals provides a case in point for how modern forms of surveillance rely on the participation of their subjects. Fisher describes two different cases of the use of RFID tags for tracking in hospitals – one to track the location of nurses' equipment and the other to track the location of the nurses and doctors themselves. (Fisher 2006) Surprisingly, the first project fails, even though it is seemingly less intrusive than the second. The second project succeeds mostly because it includes the nurses who are its subjects in the design process and includes features in the design of the system that allow the nurses some measure of control over their own privacy. Nurses in the first case subverted the system by removing tags from and hiding scarce equipment, whereas nurses in the second case willingly used the tracking data to monitor their own movements (reminding themselves to visit patients on more regular schedules) and to track down previously difficult-to-locate doctors (who were also tracked). Fisher's study shows that the subjects of surveillance can use the surveillance to increase their power within the institution if given the right sorts of support (and in fact will often resist if not given the right support) and that some forms of surveillance are made more powerful by the active participation of the subjects.

Ranking Search: Google Watching Users

Google actively tracks the users of its search engine by recording both which terms users search for and which result links they follow. Every time a user performs a search on its site, Google stores data about the request, including the search terms, the date and time, and the IP address of the request. This is the same basic data that every web server collects (or has the potential to collect) for every web page visited on the given web server. Google's collection of this particular data is particularly interesting and important because of the combination of the enormous popularity of Google's search engine and the deeply intrusive nature of many web searches (especially when viewed together as a history for each user). Google currently keeps all of this data about individual requests for nine months, when it replaces part of the IP address with a new, unique value.[7] After eighteen months, Google removes a part of the IP address altogether, aggregating most users into blocks of a couple hundred members. Google also stores a cookie on each user's browser, enabling it to identify that a user's searches are all coming from the same user (or at least computer / browser). For users logged onto a registered Google account (for

7 Google's policy for storage of its search logs is under active negotiation, primarily with the European Union to comply with the requirements of the EU privacy directive. Google has recently shrunk the time it maintains individual IP addresses from 18 months to 9 months. Google has not stated clearly what it does with search log entries after nine months, but the best guess is that Google is hashing the final 8 bits into a unique number that cannot be used to generate the original IP address. Hashing the last bit of the IP address still allows Google to identify a set of queries as coming from a single IP address, but it cannot directly trace those entries back to a particular IP address. It can still trace the entries back to a 256 address block of IP addresses, through, which could be used by an ISP to identify the neighborhood a user lives in. And as addressed in the text of the paper, the search terms themselves can often be used to identify a unique individual in the logs even without the IP address.

example, users who are logged into a Gmail account), Google also associates all of the user's searches (and clicks as described below) with the personal information in the user account. Finally, Google collects data about which links users click on within search results.

These monitoring activities qualify as weak surveillance, because even though Google does not have meaningful consent gather this data, there is no mechanism for direct control. Google controls the ranking of its search results through this user data by controlling the underlying algorithm and data and exercising its ability to filter results manually. Google does use this click data to personalize search results, but it is hard to argue that such a change represents constitutive control of a user. This is the core difficulty of making sense of Google's collection of the personal data of its users as an act of surveillance. Google is not using the data it collects on individual users to do anything life changing to those users directly.

Google has published a series of videos on YouTube that minimize the privacy implications of the data collected by its search engine, implying that it does not need consent to collect the data:

> To improve our search results as well as maintain security and prevent fraud we remember some basic information about searches. ... We're able to [replace "carss" with "cars"] because we've analyzed search queries in our logs and found that when people type in "carss" they really mean "cars." ... What a cookie doesn't tell Google is personal stuff about you like where you live and what your phone number is. ... As you can see, logs don't contain any truly personal information about you. (Google, Inc. 2007)

Google's explanation is true only in a narrowly technical sense. Google is using "personal" in an industry accepted sense of data that directly connects to an offline identity: for example a name, phone number, or street address. But the context of any data determines its personal nature as much as the data itself. So Google's assertion that the it does not collect "personal stuff about you like where you live" is only true in the same sense that your driver's license number does not tell the police where you live. Even though the cookie itself is just a random string of gibberish letters, it can indeed be used to look up personal information "like where you live." For users logged in to a Google account, Google directly connects all of the collected search data with the personal data from the account, which includes the profile information from the account (usually including a full offline name). For users searching without being logged in, an IP address can be used (and often is by law enforcement) to track a user down to a specific offline identity with the help of the user's Internet Service Provider (ISP). Even without the participation of an ISP, the cookie and IP address each connect all searches performed by a single person, at least reasonably well. Google may or may not choose to do the relatively easy work necessary to translate its collection of search data into a database of personally identifiable data, but it does have the data and the ability to query personal data out of the collection at any time if it chooses.

Google also collects data about which links users click within its search results, which, unlike the basic data collection described above, is not typical practice for web servers. Every search result link is actually a link back to Google that redirects the user to the end site. So a search for "new york times" returns http://nytimes.com, labelled as "The New York Times," as its first result.

When the user hovers the mouse cursor over the result link, the browser displays the nytimes.com url. But when she clicks on the link, the browser actually goes to an invisible Google page with a url something like http://google.com/redirect/http://nytimes.com.[8] The sole purpose of that page is to allow Google to record the user's click on http://nytimes.com after searching for "new york times" and then redirect the user to http://nytimes.com. Google casually but actively hides the fact that it is recording these clicks by adding a snippet of javascript code to its search results web page that only converts the search result links to the Google redirected version when clicked.[9]

The generous and plausibly true explanation of Google's obscuring of the redirection of search result links is that the Google redirected link is confusing for users because they are most interested in (and expecting to see) the ultimate destination rather than the Google tracking mechanism. Google is certainly not trying to hide its tracking from careful observation – there are more sophisticated methods it could use to try to hide its tracking of user clicks from even close inspection. But the end effect of this model is to hide the fact that Google is tracking search result clicks.

In addition to contextual questions about the aggregation of anonymized data into identifiable, personal data, any data collected by Google raises serious questions about its contextual use because of the sheer amount and variety of data that Google collects and because Google reserves in its privacy policy the right to combine the data from its various services. It is not clear whether Google is doing so now, but its privacy policy reserves the right to use search histories to target ads within Gmail or to mine the content of documents written in Google Docs to adjust minimum bids within AdWords. All of this contextualized information is also potentially available to government disclosure through warrants and subpoenas, so Google might be forced to disclose not only a user's search terms and clicks, but also that same user's Gmail emails, Google Calendar events, and Google Health records.

Google uses its record of searches and search result clicks to influence search result rankings for all users by, for example, privileging search results that are clicked on more often and search

8 The actual url in this case is: "http://www.google.com/url?sa=t&source=web&ct=res&cd=1&url=http%3A%2F%2Fwww.nytimes.com%2F&ei=WmRaSaX7MuHAtgePstDmBg&usg=AFQjCNEtLodOdxWZSGdJpL7WJaEeUJVlnw&sig2=JFEk3Z3sZA-4IakO0ZTMZA". The important thing to notice is that the url is actually pointing to google.com and that the nytimes.com url is embedded within the google.com url.

9 The relevant javascript and html for the nytimes.com link are:

```
        window.rwt=function(b,d,e,g,h,f,i){var a=encodeURIComponent||
escape,c=b.href.split("#");b.href="/url?sa=t\x26source\x3dweb"+(d?"&oi="+a(d):"")+
(e?"&cad="+a(e):"")+"&ct="+a(g)+"&cd="+a(h)+"&url="+a(c[0]).replace(/\+/g,"%2B")
+"&ei=WmRaSaX7MuHAtgePstDmBg"+(f?"&usg="+f:"")+i+
(c[1]?"#"+c[1]:"");b.onmousedown="";return true};

        [...]

        <a href="http://www.nytimes.com/" class=l onmousedown="return
rwt(this,'','','res','1','AFQjCNEtLodOdxWZSGdJpL7WJaEeUJVlnw','&amp;sig2=JFEk3Z3sZA-
4IakO0ZTMZA')">
```

results that do not generate follow-up, clarifying searches. But more importantly, it continually evaluates changes to its search ranking model with this data. The reliance of the search ranking model on the enormous set of input from the community and the emergent, unpredictable nature of its results makes the effects of even small changes to the model very hard to predict. And there are no absolute set of criteria to determine whether the model is working. The only way to test whether a change improves the model is to test its effect on users. In addition to testing changes on trained human evaluators, Google performs tests on its users continuously by deploying experimental changes to small subsets of users (without their knowledge) and analyzing the effects of those changes on the search and click data of those users. (Huffman 2008)

Ranking Search: Users Watching Users

Most current descriptions of PageRank, including Google's own, use an election metaphor to describe the algorithm, as I have done in places above. (Google Inc. Technology Overview 2008) But in their 1998 paper describing the PageRank algorithm, Sergey Brin and Larry Page instead use a metaphor of user simulation to describe PageRank:

> PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank. (Brin and Page 1998)

According to this description, Google is modeling the intent of its users, not the interests of the publishers; the publisher-links-as-votes model only serves as an approximation of what users are interested in. Google quotes Larry Page in the first sentence of the "Technology Overview" page on its web site saying that the "perfect search engine" is something that "understands exactly what you mean and gives you back exactly what you want." (Google Inc. Technology Overview 2008) There is no publisher-links-as-votes democracy in that statement, but instead a model that strives for simulation of user intent.

Google uses the publisher-links-as-votes model only because publisher links are the most readily available metric for determining the intents of users. Even though these publisher-links-as-votes are a key input to the search ranking system, the success of the system is measured primarily by user feedback. It is probably impossible even for Google to answer precisely how it weighs the user feedback against publisher links in its ranking because comparing the two is not simply a matter of weighing different signals within a given model. Instead, Google evaluates and adjusts the working of its publisher (and other) inputs based on this user data.

As with the publisher based PageRank model, the underlying power of the mechanism comes from the edges of the network watching themselves through a feedback loop setup by Google. Google provides search results. Users evaluate those results by acting on them. For instance, users may consistently click on a lower ranked site first for a given search, indicating that the lower ranked site is more relevant query. Or they may consistently replace one search with another (for instance "pritney spears" with "britney spears"), indicating that users who search for

the first query really want results for the second query. Google adjusts the search results to reflect this user feedback, and its users provide feedback for those adjusted search results by interacting with them. Google and its users process and react to further changes, and so on.

The power of this feedback comes not from Google as an active watcher as much as from the users watching each other. Google only decides how to setup a model that can scale to capture the decisions of its billions of users and whether to override any of the specific results (as described in the previous section). Certainly this act of users watching each other looks only like merely watching. There are legitimate questions about whether users are giving meaningful consent to participate in this feedback loop, since it is not at all clear to most users that Google is watching this activity, let alone the degree to which Google mines the resulting data to power its search engine. But almost no individual user is exerting significant control over any other individual user by participating in this feedback loop.

William Bogard has written about a kind of surveillance through simulation in which surveillance consists not just of watching subjects as they act, but of generating simulated profiles which are themselves closely watched and used to watch the subjects. (Bogard 1996) The Google search ranking system creates a simulations of users ("assume there is a 'random surfer' ...") that is used to predict the intent of users as they enter search terms. The goal is to correct search results based on the reactions of the simulated user so that the real user does not need to. Google has repeatedly stated that it is interested in what users mean, not what they say when they search. Not only does the user provide feedback to improve the user simulation, but the user simulation provides feedback to "improve" the user.

Google's auto-correction feature is a concrete representation of this model of search ranking as user simulation. The auto-correction system looks through Google's log of user search clicks to find cases in which users consistently replaces one search with another similar one (for instance "pritney spears" with "britney spears"). The search engine either suggests the corrected search at the top of the search results or actually returns the corrected search result above the literally requested results. So when the user corrects a search result, Google takes this behavior as a sign that its user simulation has failed and correspondingly adjusts the simulation. And the resulting simulation in turn corrects the users who make the same mistake.

For instance, a search for "chevrolet magnum" returns a few results for the literal query but then inserts results for "dodge magnum" in the query because the system has discovered that users who search for chevrolet magnum usually correct the query to "dodge magnum" and click on those results:

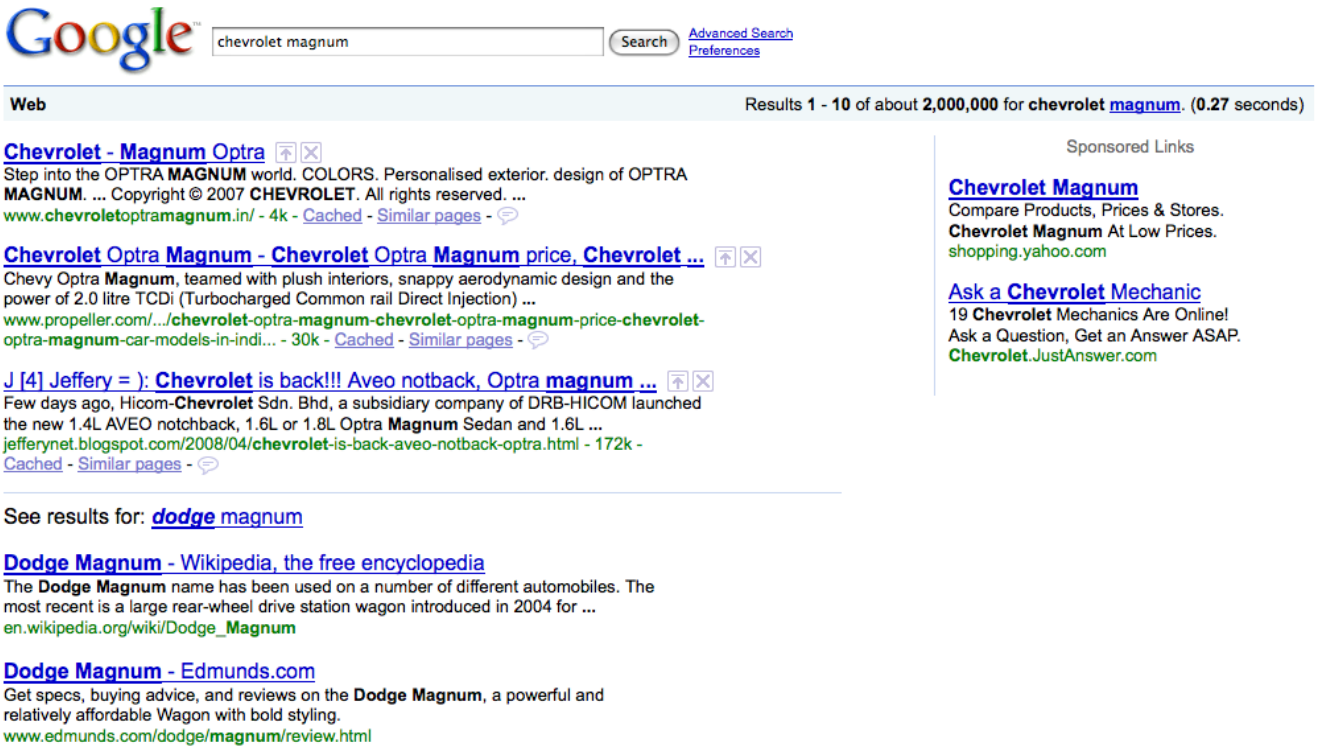

Figure 3.3: Google search correction for "chevrolet magnum"

These spelling and clear factual corrections are only a concrete representation of the work that his reflexive user simulation does. In most cases, the issue is not the literal truth of the results versus user intent, but rather determining user intent for underspecified terms:

> ... understanding what a user really wants when they type a query – the query's "intent" – can be very difficult. For highly navigational queries like [ebay] or [orbitz], we can guess that most users want to navigate to the respective sites. But how about [olympics]? Does the user want news, medal counts from the recent Beijing games, the IOC's homepage, historical information about the games, ... ? This same exact question, of course, is faced by our ranking and search UI teams. Evaluation is the other side of that coin. (Huffman 2008, brackets included by author to indicate search terms)

Google does not return some objectively relevant set of search results (whether determined by publishers or some other metric) but rather what its collective engine thinks each user means by any given term. This sort of surveillance as self correcting simulation looks like a lot like Foucault's description of the Panopticon as a laboratory – a place of active experimentation where surveilling actor is always adjusting the monitoring equipment according to the current behavior of the subjects: "the Panopticon was also a laboratory; it could be used as a machine to carry out experiments, to alter behavior, to train or correct individuals." (Foucault 1975, 203)

Ranking Search: The Google Nervous System

Google's founders consider its search system to be a budding artificial intelligence:

> Artificial intelligence would be the ultimate version of Google. So we have the ultimate search engine that would understand everything on the Web. It would understand exactly what you wanted, and it would give you the right thing. (Page and Brin 2000)

This artificial intelligence is embodied by the simulated user sitting over the user's shoulder while she is searching telling her she really meant to type “chevrolet magnum” instead of “dodge magnum.” But this behavior looks more like a nervous system shared between Google and its users than like an independently conscious being. Unlike the archetypical AI computer HAL, Google's system only works through continuous interaction with each user. Not only is it capable of action only in reaction to a user request, but it derives its “intelligence” through the constant feedback of its users. This Google nervous system is not so much an independent artificial intelligence as a global feedback mechanism that connects users (and publishers and advertisers) to one another through a collective user simulation, continuously interacting with, adjusting to, and adjusting each of its users.

Tracking user behavior is not merely an act of watching by Google, because the core actor in this case is the user herself. The user is watching as well as being watched by the search ranking system. Google's search ranking system serves as not only as an index to web content but also as a model for the capture of information to further improve the model, in the sense that Agre uses “capture.” Google has designed the system to be fundamentally a laboratory tool for the capture of data about its users. It uses captured data to create a reflexive system whose purpose is not to return objectively relevant results but rather the results that user simulation thinks are relevant. Google has designed the model of search as a shared Google nervous system, but the users' (and publishers' and advertisers') interaction with the system provides the system with the power of its intelligence. The control rests wholly with neither Google nor with the search users or publishers or advertisers, but in the interaction of all three.

4. Rankings Ads

Just as it does with its search results, Google applies a complex set of criteria to sort AdWords ads. The amount bid by advertisers is an important criteria, but a number of other criteria are taken into account, including Google's determination of the quality and relevance of the landing page (the page the ad links to) and the click through rate for the ad (the proportion of users viewing the ad who click on the link). Google refers to these non-bid price criteria as the ad's "quality score." So an ad that attracts lots of clicks from users may appear at the top of the list even though it pays less per click than the other ads in the list. Google offers a couple of explanations for why they factor the "quality score" into the ad ranking. The more straightforward explanation is that the total revenue collected by Google for an ad is a function of the cost per click and the number of clicks. An ad that pays Google $0.10 per click and gets 100 clicks pays Google more than an ad that pays $1.00 per click but only gets 1 click. But Google also argues that factoring in the quality of the ad improves the experience of the user by only presenting relevant, interesting ads. Ads that get very few clicks, argues Google, discourage users from reading and clicking on ads in general. (Google Inc. Quality Scores 2008)

Together these various factors create a network of grey surveillance that drives the ranking of AdWord ads. This network operates much like the network that drives Google's search rankings, with users in particular driving the ranking of ads by clicking (or not) on them, creating an analogue Google nervous system that determines which ads get shown in what order. The key difference between this network and and search ranking network is that Google retains considerably more control over the ad ranking by enforcing an extensive set of rules about both the format and content of the ads.

Ranking Ads: Google Watching Advertisers

Google analyzes the text of both the AdWords ads themselves and of the ads' landing pages to adjust the ranking of the ads. The analysis is based on a detailed set of rules published on the AdWords site and is directly controlled by Google, without significant moderation by the community. The enforcement of these rules is clearly surveillance on Google's part. Google uses the rules to exert a high level of control over advertisers, enforcing a very specific view of what an ad should look like by decreasing the ranking of or even removing altogether ads that break these rules. And Google so dominates the contextual advertising market that most advertisers do not have a meaningful choice about whether to obey Google's rules. The enforcement of these ad rules acts as a major point of control over the ad ranking system.

The list of AdWords rules is much longer than the list of webmaster guidelines for the search engine, which includes only a handful of general rules about relevance, originality, and transparency. The rules are divided into guidelines and policies. The guidelines only threaten to degrade the ranking of an ad, but the policies threaten to remove the ad from display altogether. The policies are much more detailed, broken down into formatting, link, and content rules. The formatting rules require that the ads themselves adhere to about ten different criteria, such as proper grammar, spelling, and punctuation; length limits; relevance to the keywords; and even bans on certain "unacceptable phrases" like "click here" that Google has found to be a waste of advertising space. The link rules require that the ad links be functional, transparent, and non-

malicious in various ways. Finally, Google disallows content advertising a wide variety of different types of products, roughly broken into rules limiting Google's liability under various laws (for example against advertising hate speech sites) and rules disallowing categories of sites based merely on Google's own non-legal considerations (for example violations of its own webmaster guidelines). (Google Inc. Text Ads 2008)

All together, the guidelines and rules, shortened to a few words apiece, look like this:

be accurate	*don't sell hard alcohol*	*don't advertise mobile scams*
use capitalization	*don't advertise spam*	*don't use personal attacks*
be short	*don't advertise pirating*	*license pharmacies*
use correct grammar	*don't advertise counterfeits*	*don't advertise phishing*
use correct spelling	*don't advertise ad data entry*	*don't solicit without approval*
don't claim affiliation	*don't advertise drugs*	*don't advertise ad templates*
don't curse	*don't advertise mushrooms*	*don't advertise radar sensors*
don't use names as keywords	*don't advertise salvia*	*don't advertise weapons*
use punctuation	*don't advertise e-gold*	*don't break these rules*
don't repeat yourself	*don't advertise endangered*	*don't break the back button*
don't use superlatives	*don't advertise prostitution*	*don't link to a broken url*
approve trademarks	*don't advertise fake ids*	*don't use a false display url*
don't say "click here"	*don't advertise fireworks*	*don't use popups*
don't advertise steroids	*don't advertise gambling*	*use https for collecting data*
don't advertise violence	*don't advertise hacking*	*don't be malware*
don't advertise cheating	*don't advertise miracle cures*	

Google's watching of advertisers to influence ad ranking is clearly surveillance. Google exerts a high degree of control over the formatting, tone, and topics of both the ads and the landing pages to which the ads link. Advertisers not following Google's rules will either find their ads blocked entirely or find that they have to pay much more for their ads to appear. Even though use of AdWords is technically voluntary, Google is by far the largest player in the market. Especially for advertisers who want to advertise on many small sites as allowed by the AdSense component of AdWords, there is no real choice about whether to use Google's AdWords. Compounding the issues of control and consent, Google actively combines a variety of contextual data collected about both the surveilled advertiser and other advertisers through this and other watching activities. For instance, Google uses data collected through its indexing engine to determine whether a given landing page is original content, whether it is hosting malware, whether it is an affiliate site, and so on.

Ranking Ads: Advertisers Watching Google

Google publishes the AdWords guidelines and policies openly and in considerable detail. This transparency is not a necessary or obvious approach to its ad ranking system. Google is considerably more closed about the details of its search ranking system, referring opaquely to the two hundred metrics used in addition to PageRank, on the grounds of trade secret protection and protecting against gaming of the rankings. In contrast, the AdWords quality score guidelines and policies are much more open and detailed. Google is vague whether it actively enforces the

guidelines or the guidelines are merely best practices. For example, the following text implies that Google itself will enforce the guidelines by adjusting the quality score, but it also emphasize that advertisers should follow the guidelines for their own sakes:

> We've found that when our advertisers' sites reflect these guidelines, two important things happen:
>
> - The money you spend on AdWords ads will be more likely to turn into paying customers.
> - Users develop a trust in the positive experience provided after clicking on AdWords ads (and this turns in to additional targeted leads for you).
>
> Furthermore, following our site guidelines will help improve your landing page quality score. As a component of your keywords' overall Quality Scores, a high landing page quality score can affect your AdWords account in three ways:
>
> - Decrease your keywords' cost-per-clicks (CPCs)
> - Increase your keyword-targeted ads' position on the content network
> - Improve the chances that your placement-targeted ads will win a position on your targeted placements
>
> (Google Inc. Landing Page 2008)

The overall quality score for an ad is some unstated combination of Google's direct ad and landing page analysis and the community based click through rate. Conveying the sense that the rules are just reflecting the will of the community minimizes Google's control over the process. And so Google deflects criticism of its outsize role in the content of online advertising and more importantly makes its guidelines seem like a natural mirror of the user interests that drive advertising (on- and off-line). By publishing the guidelines and describing them as self-justifying but not specifying how Google enforces them, Google encourages advertisers to treat the guidelines as self-enforced norms. Advertisers have to watch Google very carefully to figure out exactly what the rules are for ad ranking and how best to optimize their ads to meet those rules. But as with any panopticon, investing the power to watch in the advertisers necessarily grants some control to the advertisers – to the degree Google does not enforce all of the rules all of the time, the advertisers gain some control back over the system.

Ranking Ads: Google Watching Users

Google tracks AdWords users in two different ways – by tracking the topics of browsed web pages and by tracking clicks on ads. When users view pages that host AdWords ads, the AdWords system processes but does not store data about which users are browsing which topics. When a user performs a search on google.com, Google processes that user search data as described above, additionally using the search terms to target ads to users. For ads displayed on partner content sites, the Google servers only collect the content of the page and use that content to produce a list of targeted ads – they do not have direct access to the identify of the browsing

user.[10] When users click on ads, Google stores data about the ad clicks. That ad click data generates feedback about how effectively the system is mapping ads, topics, and users to each other.

When a user clicks on an ad, Google uses a redirect page to record the click. Google has to know when a user clicks on a given ad in order to charge advertisers each time a user clicks on an ad. Google stores at least the cookie itself along with the identity of the ad clicked on, but it deletes each cookie every 30 days. (Google Inc. Advertising and Privacy 2008) This redirect page gives Google access to the IP addresses of users as they click on ads, but Google does not state whether it stores the IP addresses. For participating advertisers, Google also uses this cookie for "conversion tracking." Conversion tracking tells advertisers which users reach a specific page beyond the landing page within an advertised site. Advertisers generally use a purchase page for conversion tracking to track how many users coming from an AdWords ad they "convert" into buyers. Google reports this conversion data to advertisers as aggregate data – for example that one hundred of two hundred users clicking on an ad subsequently made it to a purchase page.

Google's tracking of user browsing habits and clicks looks like weak surveillance due to a lack of consent for Google to monitor AdWords users and Google's universally high use of contextual data. The lack of a clear opt in mechanism, especially for the ad click tracking, constitutes a high lack of consent for Google to monitor AdWords users. Google does not actively hide the redirecting urls in the ad links as it does in the search result links, but there is no reason for a casual user to suspect that a link immediately above the text "smartertravel.com" will actually lead to Google, and there is no indication other than the url in the status bar that the ad links go to Google before going to the advertised site. Google discloses its use of these redirecting links in its privacy policy, but it is a stretch at best to expect a user at cnn.com to have found, read, and agreed to the privacy policy for AdWords on google.com. And as with all questions about Google's watching activities, Google uses or reserves the right to mine the combination of extensive contextual information about its users from its myriad services, including data collected about ad clicks.

But as with search data, Google does not use the ad click data to control users directly, making it unclear how it would qualify as surveillance. Helen Nissenbaum argues for privacy as a sort of contextual integrity that takes into account "not only whether information is appropriate or inappropriate for a given context, but whether its distribution, or flow, respects contextual norms of information flow." (Nissenbaum 2004, p. 123) According to Nissenbaum, previous approaches to privacy fail to account for many modern cases of surveillance, including the widespread collection of consumer data, because they look for a strict separation between public

10 The AdWords hosting web page executes a block of javascript code that is fetched from Google each time the page is loaded. The javascript code calls a script on the Google servers that presumably looks up the content of the hosting web page as periodically spidered by Google. Even though the code does not directly have the ability to associate browsing data with individual users by IP address or cookie, it does have the ability to collect such information by, for example, causing the user's browser to load an extra, blank image on the Google servers. This potential for the AdWords code to collect personal data is worth considering because the code is used on millions of hosting sites and can be updated immediately by Google to do whatever Google wants at any time. Any change without notice would of course put Google risk at of being discovered.

and private data, rather than for the context of the data and its flow. To judge whether to protect the privacy of a given piece of data, we should ask whether existing norms deem it appropriate to collect and to distribute that data:

> Although the online bookseller Amazon.com maintains and analyzes customer records electronically, using this information as a basis for marketing to those same customers seems not to be a significant departure from entrenched norms of appropriateness and flow. By contrast, the grocer who bombards shoppers with questions about other lifestyle choices – e.g., where they vacationed, what movies they recently viewed, what books they read, where their children attend school or college, and so on – does breach norms of appropriateness. The grocer who provides information about grocery purchases to vendors of magazine subscriptions or information brokers like Seisint and Axciom is responsible not only for breaches of norms of appropriateness but also norms of flow. (Nissenbaum 2004, p. 135)

The problem with Nissenbaum's approach for the AdWords case is that the answer begs the question. She leaves aside the key question of who decides the existing norms of appropriateness and flow. Nissenbaum walks through some of the different ways that these norms can be evaluated, but the list reads as a typical list of individual privacy invasions (informational harms, information unequality, autonomy and freedom, human relationships, and democracy and other values) that do not directly apply to Google's use of browsing data to target ads.

Daniel Solove struggles with this same problem of how to create a standard for privacy when the definition of privacy itself is not clear. Walking through an exhaustive catalogue of scholarship about privacy, Solove argues that we have to consider privacy an amorphous term á la Wittgenstein and that the only thread connecting all of the notions of privacy is that they all consist of one or more of sixteen different activities. (Solove 2008) He argues that privacy is best conceived as the bundle of all of the related problems caused by these different activities, rather than as one semantically coherent entity. For Solove, then, analyzing concerns about a new technology is a matter of determining which of the sixteen types of activities it includes (surveillance? aggregation? secondary use? decisional interference?) and addressing the problems that those activities might cause. It is hard to see what direct problems the activities related with the AdWords system cause, though, since the collected data is neither published nor used to make constitutive decisions about the subjects surveilled.

In an earlier book specifically on the use of highly detailed “digital dossiers” by government and corporate actors, Solove argues that the danger of consumer (and other personal) data collection is that it empowers faceless bureaucracies to make constitutive decisions about the people represented by the dossiers – for instance, about whether a given person will be granted a mortgage. But AdWords is not making decisions about mortgages; it is just displaying ads, mostly for products not constitutively important to the lives of users. Even ads for products like mortgages are only ads. The ads may have some impact on the user's awareness of mortgage possibilities, but using personal data to determine whether to display a mortgage offer is not comparable to using personal data directly to decide whether to grant a mortgage.

Google's monitoring of AdWords users exerts control not by direct control of the AdWords users but rather by establishing and directing a model of stateless market research. Google uses search and content terms to connect ads with the real time interests and intents of AdWords users. Google could target ads based on historical browsing interests, so users who tend to search for information about golf and fancy beers would see ads about golf, fancy beers, and maybe expensive cars when using the Google search engine to search for anything; users who tend to search for information about a specific health problem would see ads about related drugs on Google regardless of the specific search; and users who have been searching for information about digital cameras would consistently see digital camera ads on Google. Google could technically use its javascript AdWords inclusion code to collect information about the browsing histories of its users on all participating AdWords sites, including the vast majority of visits in which users do not click on an AdWords ad.[11] Such a service would amount to stateful market research, meaning that it would be market research that relies on storing "state," or historical data, about a given user's interests over time.

Google explicitly chose not to sell this kind of service on its search engine, developing instead the AdWords system of targeting mostly on the current search terms. Google made this choice at least partially because these sorts of services consistently provoke a strong public reaction from privacy advocates about the obvious privacy risks (imagine a woman who spends a few weeks searching about fertility and is bombarded by ads for fertility drugs on all of her favorite sites for the next year, including when viewing those sites in public areas). There are a host of different companies providing exactly this sort of stateful market research, albeit from much smaller sets of data, and they have generally provoked some level of public reaction. Most notably, Google's DoubleClick uses this model of targeting, but it uses only data collected from its own DoubleClick ads through third party cookies, and it comes with its own company history as a recent purchase of Google.[12]

Google currently chooses instead to sell only real time data about the intent of the user in the immediate present for its AdWords system.[13] Instead of selling the stateful information that a

11 Google could for instance include in the AdWords javascript inclusion code a single line to load a "web bug" (a tiny transparent image that exists solely to transmit data about the client to a third party). The AdWords include code is generally included remotely from Google with every page load, so it could make this change instantly on virtually all sites that serve AdWords ads. Whether and how quickly this change would be detected by the community is another question.

12 A third party cookie is a cookie that originates from a entity other than the primary host of a given web site. DoubleClick serves image ads from its own ad servers. Anytime a user requests anything from a server, including an image, that server can store a cookie on the user's machine. Every time a user loads a page with a DoubleClick image ad, the user requests the image from DoubleClick and in the process stores a cookie from DoubleClick on her machine. DoubleClick serves ads on a wide swatch of different sites (amounting for about 35% of all advertising web site visits) and uses the same cookie ID when it displays an ad to user on any of those sites, so it can identify that single cookie ID from a given user with all of the DoubleClick advertising sites the user has ever visited.

13 Google announced on March 11, 2009 in a post titled "Making ads more interesting" on its official blog that it has begun beta testing a version of AdWords among a few advertisers that combines the current page targeting strategy of the current AdWords system with the historical interest data from its Double-Click system. It is not yet clear whether or quickly this new hybrid system will be rolled out to all advertisers or precisely how Google

given user has been interested in infertility for the past month, Google sells the stateless information that a user is actively looking for (or at least looking at) infertility content right now. This current intent approach has the first benefit of provoking much less concern among users. Certainly it is less intrusive in the obvious sense that it is less intrusive to sell one individual bit of data (this user just now performed a search about infertility) than a whole collection of data (this user has performed several searches about infertility, digital cameras, and movies over the past month). Google's active public outreach through blogs, videos, and other such media indicates that Google recognizes that concerns about obvious privacy risks pose a risk to its market position, especially its dominance of Internet search. (Google Inc. Google Search Privacy 2007)

This data about the real time intent of the user also allows AdWords to adjust its topic / ad mappings in real time according to the feedback provided by users as they indicate interest in targeted ads by clicking on them. An historically based system like that used for DoubleClick can (and does) track ad clicks and adjust ad displays accordingly, but the historical nature of the model by definition slows down any response. AdWords responds always and immediately to what Google thinks the users are doing right now. An ad launches as soon as it wins an auction for placement, users click or do not click on the ad immediately, the ad moves up or down in the rankings (and potential for exposure to users) immediately, encouraging or discouraging more clicks, which move the ad up or down, and so on.

This real time feedback loop between Google's systems and the collective participation of individual users is the defining characteristic of the Google nervous system. Marshall McLuhan and Lewis Lapham write grandly that:

> During the mechanical ages we had extended our bodies in space. Today, after more than a century of electric technology, we have extended our central nervous system itself in a global embrace, abolishing both space and time as far as our planet is concerned. Rapidly, we approach the final phase of the extensions of man - the technological simulation of consciousness, when the creative process of knowing will be collectively and corporately extended to the whole of human society, much as we have already extended our senses and our nerves by the various media. (McLuhan and Lapham 1994)

McLuhan and Lapham highlight the two characteristics of media technology that are key to understanding the role and impact of the Google nervous system: that it works in real time and that it operates collectively. The real time nature of the process reduces the time and space required to moderate the results served by Google. The feedback between users and the AdWords ranking system and between individual users of the system is immediate.

This real time feedback loop integrates with users' offline lives as well as their online lives. Compare ads shown in AdWords to shopping done at a store. The same bit of knowledge displayed in an AdWords ad in an instant can take much longer to act on offline – compare the time required to get into a car, drive to a store, and purchase a bottle of perfume after seeing an

is combining the stateless and stateful forms of user data.

ad in a magazine with the time required to click on a link in an AdWords ad and buy a bottle of perfume online. Certainly people make impulse purchases offline as well as online, and people have the ability to consider online purchases as well as offline, but the convenience (and encouragement) of instant purchasing linked directly from online ads at least potentially makes instant purchasing much more common.

That extra time required for offline shopping serves as a buffer, providing space and time for the user to reflect on and moderate the knowledge before importing it into her own brain. The much quicker, broader, and more accurate data provided by mining the community greatly shrinks that buffer. The buying process has the potential to happen within the short term memory of the feedback system between Google, publishers, advertisers, and users, reducing the opportunity for users to mull over decisions offline, away from the shared nervous system. And, as we will see in subsequent chapters, Google provides powerful tools to help advertisers and publishers optimize their advertising and content for precisely this process, "converting" users into buyers immediately after clicking on an ad.

Ranking Ads: Users Watching Advertisers

Users watch advertisers by clicking on the ads (or not) and ultimately by buying the advertisers' products (or not), activities that look like merely watching in themselves. Users have no leverage to control advertisers individually. In nearly all cases, a single ad click has negligible effects on an ad's rating, as does the purchase of a single product. The point of advertising is to get as many users as possible to watch the ads and hopefully buy the advertised products, so there is no issue of consent. And users as a group only have access to public information about advertisers, so there is no issue with the contextual use of privileged information about the advertisers. Users watching advertisers nonetheless play an important role in ranking ads by providing collective feedback about the ads.

Advertisers have traditionally struggled to find meaningful ways to measure the impact of ads on sales. If feedback drove advertisers, spending on advertising would drive sales, but Michael Schudson argues that at most companies the opposite is true – increased sales leads to increased advertising. (Schudson 1986) A 1977 survey that found that 75% of U.K. companies based their advertising budgets on the previous year's sales revenues. (Gilligan 1977) And other studies have shown that many companies spend more than the optimal amount on advertising and even that reducing advertising spending can in some cases increase sales. (Aaker and Carman 1982). The most widely used metric for advertising has been recall – the proportion of users who can recall something about a product some period of time after and ad for the product was shown – but there is no necessary relationship between recall and product purchases. (Schudson 1986) This struggle to measure the connection between advertising and sales reflects an inherent disconnect between offline advertising and sales. That same buffer that gives buyers the opportunity to better consider whether to buy a product after reading an offline ad also makes the relationship between offline ads and sales opaque to advertisers.

AdWords makes this connection between online ads and sales much more transparent by watching two kinds of user feedback – clicks on ads to indicate attention paid to the ad and, for participating advertisers, conversion into purchasing customers. The latter especially is a near

holy grail for advertisers in comparison to previous metrics. The AdWords conversion metric is certainly not perfect, but an advertiser that was basing its advertising budget on last year's sales and surveys of recall can now base its budget and advertising strategy on a reasonable measure of how much revenue a given ad generates. And both of these sorts of feedback happen in real time, so an advertiser who finds a particular AdWords ad leading to direct sales can instantly increase the budget for and prominence of that particular ad (or decrease the budget for an ad that is not leading to sales). Because AdWords is itself an automated system, those changes take place immediately. The core strength of the Google nervous system is measuring the kinds of interactions that have traditionally taken place in the buffer of people's offline brains. By moving the relationships between users, ads, and products into the online (and offline) world of the Google nervous system, those relationships become measurable and therefore become a lever of control for the larger network.

This ability to measure the relationship between users, ads, and products certainly gives advertisers more control over the results of the advertising and over users as a group. But it also gives users a great deal of control over advertisers in the sense that they are collectively driving the direction of advertising campaigns. The collective input by users takes two forms corresponding to the two metrics used by AdWords. Users contribute ad clicks as collective votes for which ads are the most interesting and they contribute product purchases as collective votes for which ads lead to sales of (usually offline) products. The first form of collectivity is that of a community talking openly, with room for contribution from everyone. Yochai Benkler argues that the Internet has fostered a new kind of collective participation in the public sphere. No longer are conversations beamed into the community by broadcast mass media and only interpreted by small, segregated listening groups of users:

> The Internet allows individuals to abandon the idea of the public sphere as primarily constructed of finished statements uttered by a small set of actors socially understood to be “the media” (whether state owned or commercial) and separated from society, and to move toward a set of social practices that see individuals as participating in a debate. (Benkler 2007, 180)

Obvious examples of this sort of new collective public sphere are the successes of Wikipedia and blogs in general. The participation of users in ranking AdWords ads does not rise to the same level of open participation because much of the cycle of production is hidden and controlled by other entities, as described in the sections above. However, the basic mechanism of allowing large numbers of people to participate in the conversation is the same for the AdWords users. Users in the traditional mode of advertising production are merely an audience that at best remember what they are being told, with only relatively weak and slow mechanisms for influencing ads through, for example, focus groups run by advertising agencies. Users in the AdWords system actively, constantly, and publicly vote for which ads are the most interesting and privately vote for which ads are the most effective at converting them into buyers. The mere availability of that data forces advertisers to pay attention and thereby gives users a place in the conversation that has been lacking.

User feedback within AdWords is also collective in the sense that it melds on- and offline worlds

by converting the reading of webs page online into the purchase of products used offline. John Palfrey and Urs Gasser argue that this is a defining attribute of the new generation of "digital natives" (those born after the advent of "social digital technologies" and with the access and skill to use them):

> They live much of their lives online, without truly distinguishing between the online and the offline. Instead of thinking of their digital identity and their real-space identity as separate things, they just have an identity (with representations in two, or three, or more different spaces). (Palfrey and Gasser 2008)

danah boyd describes how this melding of on- and offline worlds works in the specific context of shopping:

> Using Google and a variety of online shopping sites, Mary researched dresses online, getting a sense for what styles she liked and reading information about what was considered stylish that year. Next, Mary and her friends went to the local department store as a small group, toting along their digital cameras (even though they're banned). They tried on the dresses, taking pictures of each other in the ones that fit. Upon returning home, Mary uploaded the photos to her Facebook and asked her broader group of friends to comment on which they liked the best. (boyd 2008)

The integration of the users and advertisers extends into the offline world, especially among the growing cohort of digital natives. This integration of on- and offline worlds is true of individual queries, where the user is performing a series of on- and offline queries, integrating them, processing them, and repeating. It is also true of the input to the system as a whole, in which AdWords users collectively contribute conversion statistics that help shape the products that, although sold online, are mostly delivered and used offline.

Ranking Ads: Decoding the Google Nervous System

Judith Williamson talks about advertising as a system of signification: ads connect consumer products to social meanings (ads give "special k cereal" the meaning "healthy family") and then rely on consumers using those meanings as indices for the products (people who want a "healthy family" think of "special k" when buying cereal). (Williamson 1978) The changing roles of ads and content in doing this social work is the subject of chapter 5, but this frame is also useful when thinking about the role that the Google search and ad ranking systems themselves play in the consumption of online content. The Google ranking systems assign meaning to individual pieces of the vast hordes of online content in the same way that Williamson argues that ads assign meanings to products. Just as ads do the work of assigning the meaning "healthy family" to "special k" cereal, the ranking algorithms assign the meaning "new york times" to the site nytimes.com. As with the ad meanings, this association works in reverse as well. A user who looks to ads may think the way to get "healthy family" is "special K." And a user who looks to Google search or ad results may think the way to get "new york times" is nytimes.com. Indeed, if the user asks the wrong question ("chevrolet magnum"), Google will even tell her that she should have asked another question ("dodge magnum").

Google presents itself – through its intentionally simple interface, through its self-justifying ad and content guidelines, through its claims to know what users mean – as merely a mirror, avoiding the appearance that it is actively making particular sorts of judgments through particular mechanisms about what content corresponds to a given search or advertising topic. This presentation encourages users to think interchangeably of the Google nervous system and their own brains. In fact, Google's search and ad ranking systems are actively and always creating that mirror through the network of grey surveillance described in detail above. The point of this highly complex process of surveillance and simulation is to decipher "what a user means, not what he says." But every Google search or list of AdWords ads is the result of a particularly designed system. Paradoxically, Google has modeled that network not to be under its direct control but instead to be constantly contested by Google, publishers, advertisers, and users, growing the power of the overall network by necessarily limiting its own role.

5. Terming Ads

AdWords uses a set of words to attach a given ad to a given piece of content. That set of words defines what topic the user is currently reading for advertisers. In the case of search page ads, the words Google uses are simply the search terms entered by the user. In the case of content, Google has to convert the content into some small set of words that can be sold to advertisers as ad topics. Google's AdSense is the system that pulls a set of ad words-as-topics from an arbitrary block of content by reading the content of each page, analyzing it using Google's set of proprietary algorithms for determining whether the content is about "smart travel" or "cereal" or "obama," and then retrieving AdWords ads associated with the resulting ad words-as-topics.

The resulting AdWords ads look more like classifieds than like traditional print display ads, partially because Google requires classified length text, partially because Google requires that the ads be directly relevant to the associated ad words-as-topics, and partially because AdWords associates the ads with content related to each user's current browsing topics. Just as a user looking for information in classifieds searches specific sections of classifieds ("used foreign cars"), a user viewing AdWords ads ideally sees ads directly related to her current, active interests. A user searching for "used foreign cars" sees AdWords ads from companies selling used foreign cars (or, in fact, sites hosting classifieds for used foreign cars). The difference between classified and AdWords ads is that the classifieds user actively seeks out the classifieds, whereas the AdWords user is generally primarily browsing the related content and is only incidentally (for the user) displayed the ads. As a result, classifieds are not generally associated with content at all, or only very loosely, whereas AdWords attaches ads to specific pieces of content (either a list of relevant search results or directly to a relevant page of content).

This AdWord ad-as-classified model differs from other forms of advertising in that the ad count is both simple and directly tied to the specific content that the user is currently reading, whereas most mass media ads are usually only loosely tied to the associated content. Commercials alongside a televised golf tournament feature golf clubs but also luxury cars, financial services, and hair implants. Television advertisers expect that golf viewers will generally be interested in a specific array of other products, but this association is only loose. And even given the correct audience, the ads still have to do the social work of selling the products. AdWords ads associated with "golf" feature only golf equipment, and they neither have the space nor the editorial leeway to do much social work through the ad content.

The following chapter will argue that AdWords encourages the content linked to its ads to do more of the social work of selling products than do other kinds of mass media advertising. Google's guidelines force the tiny AdWords ad content to refer only to literal, relevant products, strongly limiting the amount of social work the ad content can do directly. Certainly the consumer products the ads are promoting, the words-as-topics used to connect the ads to particular pieces of content, and the basic market mechanisms of competition are deeply tied to the larger social framework, and so the ads must work within that social framework to have meaning. It is impossible to make sense of the act of selling, for example, perfume without referring to the social meaning of perfume for people who buy it, even if the perfume ad only advertises prices.

But Google has structured AdWords ads so that perfume ads refer overwhelmingly to simple qualities like price and variety, in sharp contrast to the prototypical glossy magazine and television ads for perfume that feature beautiful, visceral images meant to reinforce particular social meanings for perfume. Even notions used by the AdWords ad content like "sale" and "variety" are deeply social, but only in the sense that they work through the larger framework of capitalism. Typical mass media ads, in contrast, do the additional social work of giving a specific consumer product a specific meaning for consumers (golf clubs = power / perfume = beauty). The AdWords ads rely on the existing social meanings of perfume but do not attempt to reinforce or edit them (other than perhaps to make it, and everything, more embedded in the larger system of capitalism and more of a generic commodity). But the magic of the AdWords system, which connects content to ads with matching words-as-topics, puts the content in the position of itself assigning social meaning to consumer products.

Terming Ads: Google Watching Users

As discussed in chapter 4, AdWords' real time watching of user browsing and search topics amounts to a form of stateless market research. This activity looks like weak surveillance because, even though there are legitimate questions about the consent granted to Google to process that data, it is difficult to argue that Google is using the data to constitutively control individual users. But Google does use this data as the basis of its ad targeting system, which allows it to connect its advertisers' consumer products to its publishers' content through real time interests and simple reason-why ads rather than through demographics and jingle ads.

The topics that Google watches are particular words rather than merely semantic topics – they are the actual words used by people to search for particular results. An advertiser who wants to advertise a travel service, for instance, cannot just buy space alongside a broad "travel" category but instead must buy ads alongside specific variations of words or phrases related to travel. For instance, here are the most common "travel" related searches according to Google Insights, each of which is auctioned separately through AdWords:

Search terms related to travel

Top searches		Rising searches	
1. travel insurance	100	1. travel supermarket	+1850%
2. travel inn	90	2. premier travel inn	+450%
3. travel lodge	85	3. premier travel	+300%
4. cheap travel	75	4. travel republic	+250%
5. air travel	75	5. bbc travel	+110%
6. flights	55	6. sky travel	+60%
7. travel agents	50	7. first travel	+60%
8. travel agency	50	8. aa travel	+60%
9. london travel	50	9. travel zoo	+50%
10. world travel	45	10. yahoo travel	+40%

Figure 5.1: Google search terms related to "travel"

Google's use of specific words-as-topics biases the system toward specific obtainable things

rather then general concepts ("cereal" rather than "healthy family"). Even though people are broadly interested in large topics like "love" and "happiness," they tend to search online using words that take them to actual things that are obtainable online. Accordingly, the most popular single search topic on Google from the U.S. and worldwide is "lyrics" by a fair margin, and more generic terms like "love" and "happiness" fall far below:

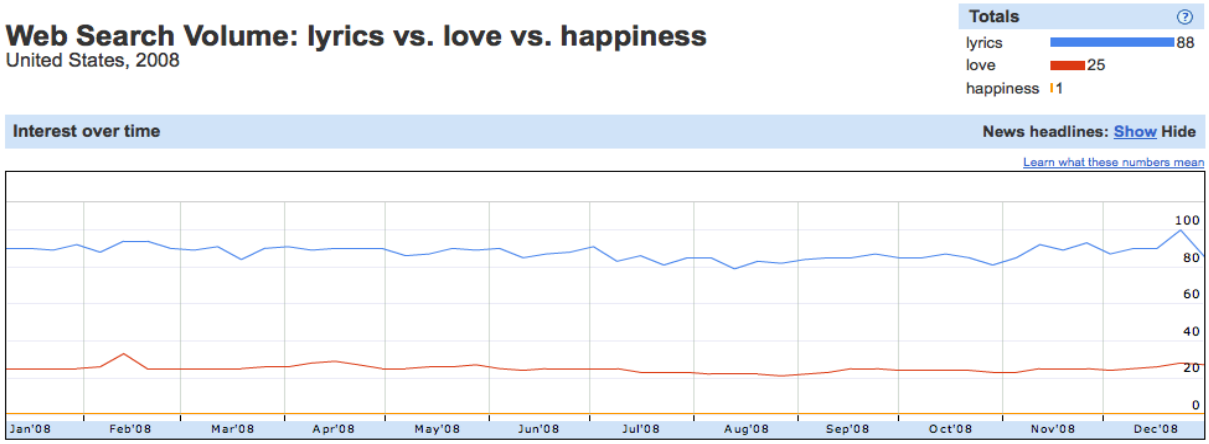

Figure 5.2: Google search volume: "lyrics" vs. "love" vs. "happiness"

For advertisers, this use of words-as-topics means that ads are much more successful when attached to these specific words-as-topics rather than to the larger semantic categories ("love," "beauty," "happiness," "power," "sex") with which advertisers have historically tried to associate their products. And so in AdWords "health family" becomes a specific obtainable thing rather than an aspiration, as shown by the following AdWords for a "healthy family" search:

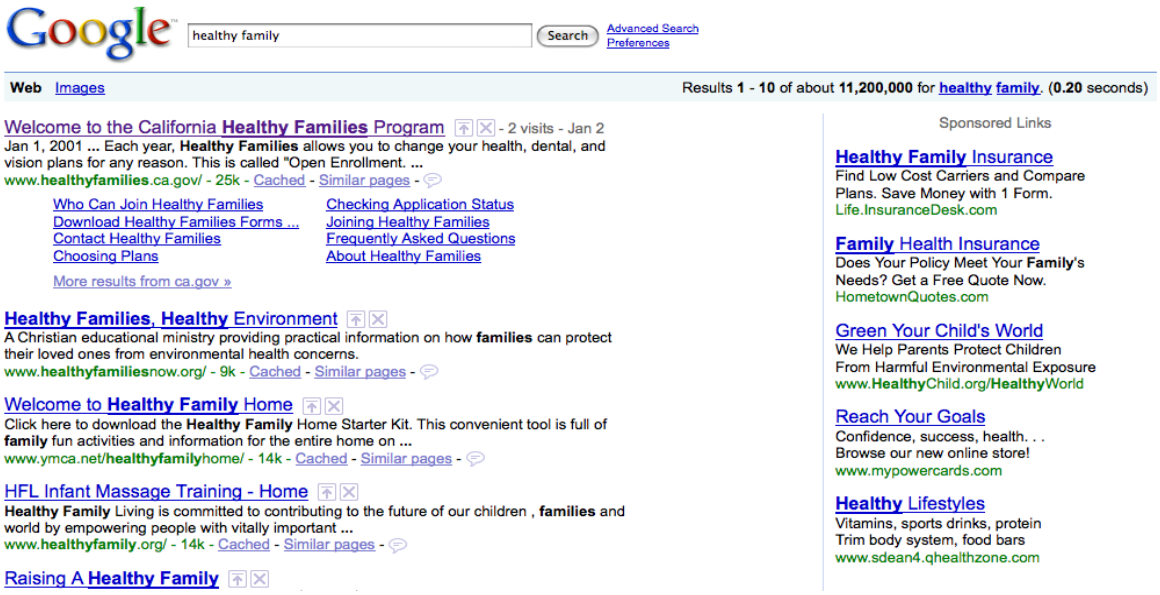

Figure 5.3: Google search for "healthy family"

Within the five ads Google has associated with "healthy family," two are related to health

insurance for families, one is for a public health campaign, one is for a motivational product promoting good health, and one is for a health food store. None of these ads use the sort of associative advertising described by Williamson, attempting to place a commodity within a particular social meaning (even though these ads certainly work within existing social systems, for instance by using "trim body system" as a reference for health). These AdWords ads are all literally offering ways to buy (or learn about) a healthier family, albeit a healthy family as already socially defined. To a degree, people who buy Special K are trying to improve the health of their families, but "healthy family" almost certainly would not be the primary function attached to Special K by most people. In contrast, it is likely that most people would assign "family health insurance" a primary function of a "healthy family." Google has structured AdWords to encourage and enforce this sort of direct, literal association, rather than the one-degree-removed association of Special K with "healthy family."

A search for "love" displays an even starker example of the classified mode of AdWords ads:

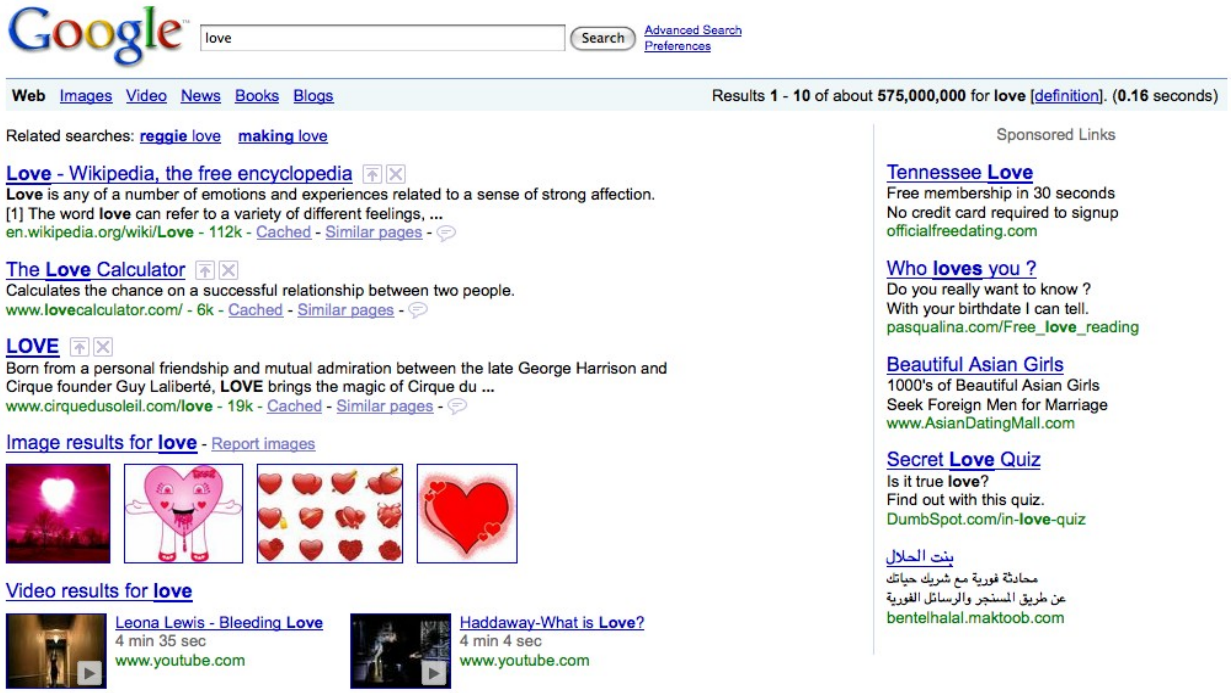

Figure 5.4: Google search for "love"

In thee results, "love" is not a general social aspiration but instead a specific word-as-commodity that can be obtained through online dating, self help quizzes, mail order brides, and "the largest arab islamic marriage site on the Internet." This model of words-as-topics is partially due to Google's extensive ad rules and guidelines that require ads to be short and directly relevant. But it is also due to the immediate, word based nature of the stateless ad targeting system and to the tendency of people to search or concretely obtainable things. The user targeted by a "healthy family" AdWords ad may or may not be the kind of person who aspires to a "healthy family," but she is definitely looking for something called a "healthy family" right now. AdWords treats the user as someone browsing through classifieds in the particular section identified by the current

words-as-topics. AdWords assumes that the user wants the specific obtainable thing called a "healthy family" through family health insurance, multivitamins, and family diet plans, presumably because people tend to search online in this concrete mode.

Stephen Fox argues that advertising has bounced back and forth over time from classified-like reason-why advertising that features simple, direct descriptions of the functions and prices of products and jingle advertising that makes more creative attempts to attach specific social meanings to products. This vacillation between modes of advertising has been driven by the social nature of advertising. Consumers grow resistant to either approach, and advertising agencies follow strong individual personalities subject to changes in fashion. (Fox 1997) AdWords ads fall on the far end of that reason-why spectrum, selling specific products through direct pitches about price, availability, and quality rather than through creative attempts to give meaning to the products. For example, Fox includes the following pre-radio print jingle used to sell cereal:

> Jim Dumps a little girl possessed,
> Whom loss of appetite distressed;
> "I des tan't eat!" the child would scream;
> Jim fixed a dish of Force and cream –
> She tasted it — then joy for him –
> She begged for more from "Sunny Jim." (Fox 1997)

By comparison, here are the AdWords ads for "cereal:"

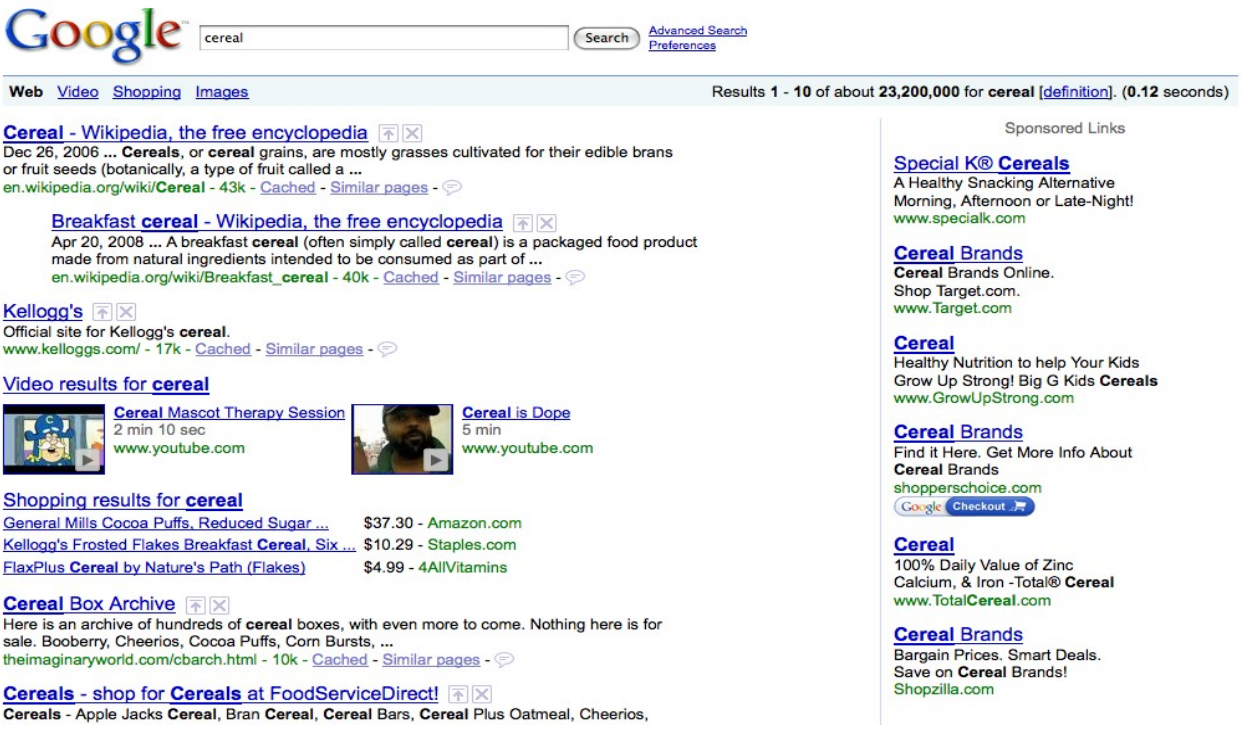

Figure 5.5: Google search for "cereal"

There is some social work happening in these AdWords ads: notice the words “healthy” or “kids” in two of the six ads. But even that work is closer to straightforward argument for how the cereal will make you healthier than the highly social work that the Sunny Jim ad does to convince the consumer that buying “Sunny Jim” cereal would make not only the consumer but his family both happier and healthier. The Target.com ad does not even bother to try to make a selling point at all: “Cereal Brands Online. Shop Target.com.” The comparison with this particular jingle is particularly interesting because the Jim Dumps ad is both text based and relatively short and so points to the possibility of this mode of advertising within the text-format and length restrictions of AdWords.

The direct connection of AdWords ads to words-as-topics (in addition to Google's guidelines requiring short, directly relevant ads) strongly encourage reason-why over jingle ads. The social work of AdWords ads happens not in the ads themselves but rather in the connection between the ads, the content, and the words. This process ideally (according to Google) creates a perfect classified system in which users see helpful information only about products they are currently, actively interested in and so do not need to be persuaded about the relevancy. There is less need to convince a consumer about the larger social value of cereal if she is already searching for cereal in the first place. Or rather the persuasion happens somewhere between the position of the ad beside the content and the content itself.

Terming Ads: Google Watching Publishers

The beauty of the search engine as an advertising platform is that the user explicitly tells the system her current interests every time she searches. So attaching ads to those search terms is an obvious way to make money. Google's primary innovation in AdWords was not merely that it targeted ads to search terms but rather that it applied (through its AdSense program) the same ad brokerage system to general content by automatically turning any block of content into a set of potential words-as-topics. Google's monitoring of web pages, to perform this content to words-as-topics conversion, looks like clearly surveillance. Many publishers have no meaningful choice about whether to host AdWords ads because of AdWords' dominant market share. And Google exercises a great deal of power over participating publishers through this act of assigning particular words-as-topics to particular web pages. Web pages to which Google assigns valuable words-as-topics generate much more revenue than those it assigns less profitable words.

Google has not published details about the implementation of this process of conversation from content to ad words-as-topics. But rule-based, not crowd-based, algorithms likely drive the process, and Google controls these algorithms. Google's control is not total – the pool of possible words is determined by the pay-per-click ad auction described in chapter 6. And Google presumably uses some measure of profitability to bias the selection of words-as-topics toward profitable ones. But there is no obvious mechanisms for a level of community input like ad ranking click-through rates or search ranking link counting and user experimentation. Finally, Google has the ability, as with all of its monitoring, to use the data from all of its various other content monitoring system (including its search index) to contextualize further the content it

analyzes within AdSense, making it look more like surveillance.

Advertisers buy ads within the New York Times because of its general reader demographics and because of the reputation of the paper as a whole. Users also tend to read offline newspapers and magazines in more of a general fashion than their online versions, paging through much of the paper (or at least interesting sections) to discover stories of interest. Online, users are more likely to dive directly to a story of interest. With AdSense, instead of attaching content to the paper as a whole in the hope that readers will see it as they browse through the whole newspaper, AdWords always attaches the content to one page, which may be the only page that a user reads in the publication. Indeed, many or most users come to an online story from a Google search or from an external link, rather than from the front page. Users sometimes pick up print magazines and newspapers to read specific stories (or even cut out stories to share with friends), and print publications certainly produce specific content to run alongside (and support) specific advertisements (fashion magazines, for example, are almost entirely constructed of such content). But with AdWords, ads run alongside all sorts of content, even the sort of hard news stories that publishers have ideally tried to keep separate from specific advertising pressure. And the pressure exerted by the AdWords ads on all stories to provide meaning for those ads is necessary, direct, and much more easily measured than with offline ads.

Judith Williamson argues that advertising gives social meaning to consumer products by associating products with specific social aspirations. (Williamson 1978) For instance, here's a perfume ad that explicitly associates both perfume in general and this specific perfume with the aspiration of being “beautiful:”

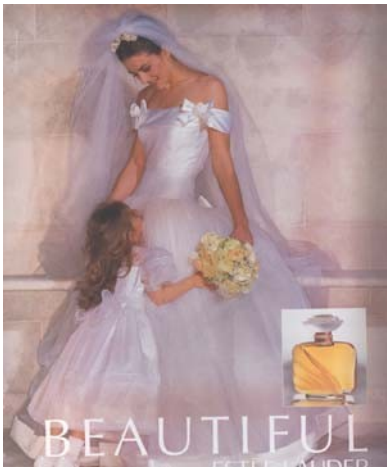

Figure 5.6: glossy magazine perfume ad

A consumer is unlikely to see the above ad and be consciously persuaded that the perfume will make her beautiful. But advertisers hope that the repetition of these ads gives the tangible perfume product a social meaning (“beauty”) that the consumer will use at some point in making a purchase decision. AdWords both directly connects ads with specific pieces of content and restricts ad content from doing this sort of aspirational social work. That combination puts pressure on the specific pieces of content to do the work of connecting products to social meanings that jingle ads have traditionally done. For comparison, here are the AdWords ads for

"perfume":

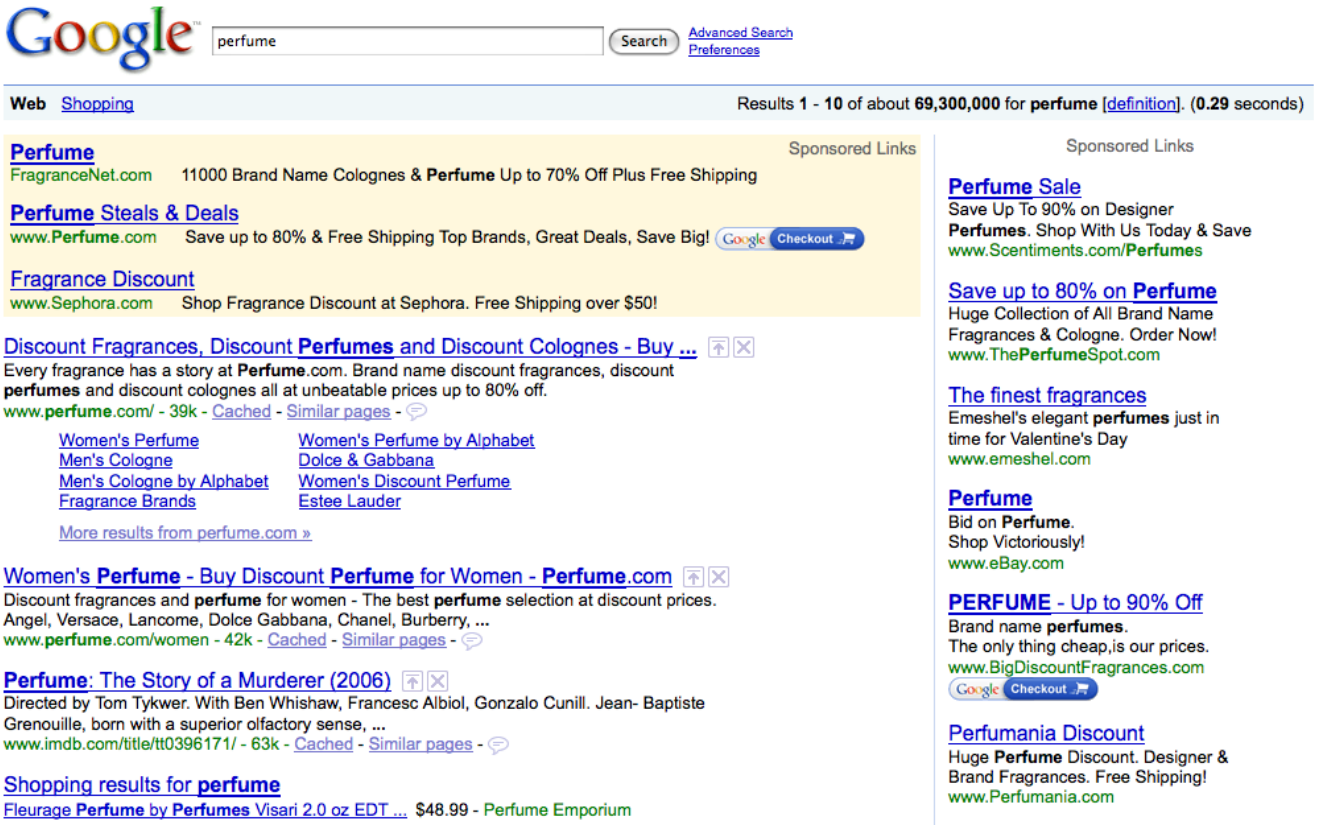

Figure 5.7: Google search for "perfume"

As with the cereal example above, the ads here work more like classifieds than traditional glossy advertising. The third ad on the right tries do a little bit of social work with the word "elegant," but all of the other ads focus on some combination of the selection and prices of perfumes offered. There are no mentions of beauty of any sort within the ads. For further comparison, here are the AdWords ads for "beautiful":

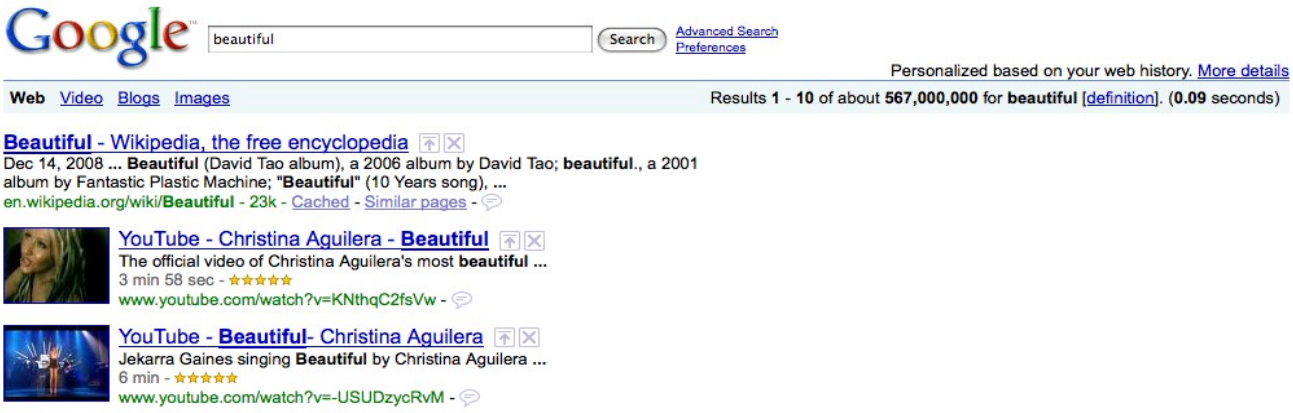

Figure 5.8: Google search for "beautiful"

As with the “love” search, AdWords tries to convert the term into a concrete product but in this case fails. According to the words-as-topics nature of the AdWords and AdSense systems, there is no thing called “beautiful” to be sold, and the nature of the AdWords system doesn't allow the ads to do the social work necessary to connect “beautiful” to an actual consumer product. So the AdWords system simply is not able to sell “beautiful.” What this means for publishers is that there is little direct profit from AdWords in writing about the concept of beauty in the larger social sense. This total inability of AdWords to sell “beautiful” is a dramatic difference from the offline magazine advertising industry, which supports the entire sector of fashion magazines on the basis of connecting “beautiful” to a wide slate of consumer products.

As a coda, here is the original perfume ad with the attached AdWords as found on a site that sells ad images:

Figure 5.9: AdWords ads for perfume magazine ad

The AdSense system determined that the most profitable word to target above is something related to advertising in general through a couple of advertising services at the bottom and a combination of “advertising” related ads and “magazine” related ads at the top, rather than the mention of “beautiful” in the image description. This preference for “advertising” over “beautiful” contrasts sharply with the comparative size of the offline advertising industry and

consumer fashion magazine sectors. The preference for concretely salable products encouraged by AdWords is a fundamental shift in the social role of advertising.

Terming Ads: Advertisers watching Publishers

In addition to the kind of highly abstracted watching of publishers described above, Google allows each individual advertiser to specify a lists of individual publishers to exclude from hosting its ads. Some advertisers use this feature to make sure that their ads do not appear on sites that might offend their customers. This individual exclusion feature is hampered, however, by the unpredictability of the millions of publishers within AdWords that might serve a particular ad. So Google also provides a category exclusion feature that allows publishers to exclude specific topics and page types from their ad runs. As shown below, this system allows publishers to surveil the topics and page types of publishers and decide whether to advertise on those topics and page types.

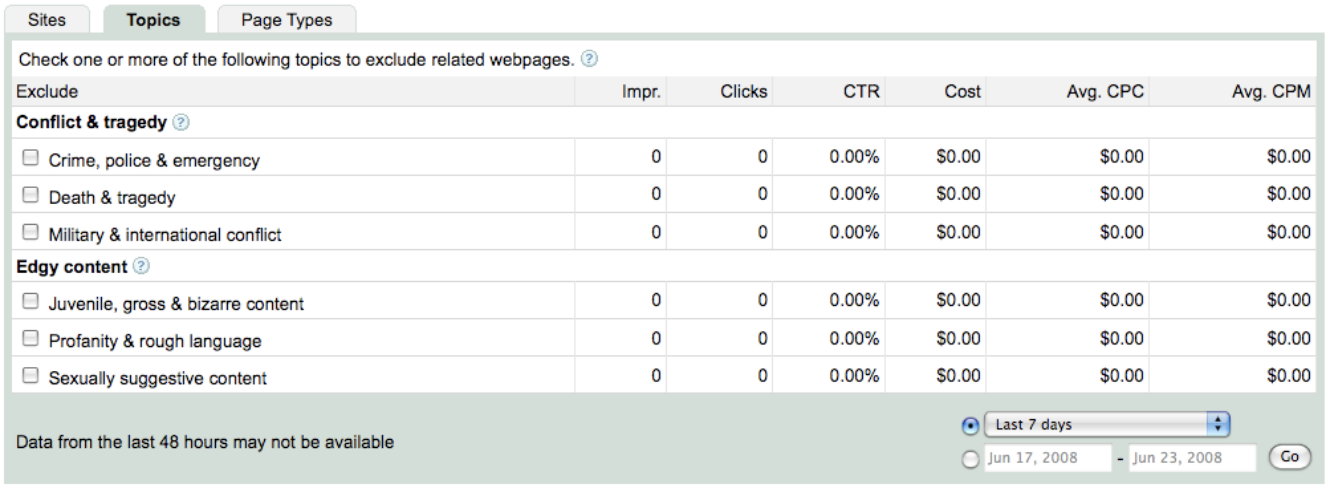

Figure 5.10: AdWords topic exclusion tool

Excludable page types include domain squatting pages that Google allows to run ads. More importantly, the system allows advertisers to exclude various kinds of "user-generated content," including discussion forums, social networks, and image and video sharing sites. Presumably Google's intent is to placate advertisers who are worried that an ad will appear alongside offensive user generated content. But this exclusion category potentially introduces a significant bias toward content produced by a few mass media actors, like content published by the New York Times, and away from content produced by the community, like content published by Flickr or YouTube.

The excludable topics include the expected "edgy content" categories of juvenile content, profanity, and porn, but they also include the following descriptions of the "conflict & tragedy" topics:

- Crime, police & emergency: Police blotters, news stories on fires, and emergency services resources
- Death & tragedy: Obituaries, bereavement services, accounts of natural disasters, and

accidents
- Military & international Conflict: News about war, terrorism, and sensitive international relations (Roberts 2008)

There is an obvious, built in bias introduced by this topic exclusion feature that encourages publishers to exclude certain kinds of content (crime, death, tragedy, war, international conflict). The resulting system looks like a dystopia ruled by the tyranny of the majority, in which the public gets the sorts of trivial content, with no death or tragedy, that it likes. But the crowds already drive advertising through the various mechanisms above. If no users click on ads about death & tragedy related stories, then those ads will fail to register ad clicks or conversions, and advertisers will not bid on them. In fact, there is a demand for precisely this sort of content, though it comes overwhelmingly from various relief and international advocacy organizations, which makes the support for war and famine stories look an interesting hybrid of commercial advertising and public funding:

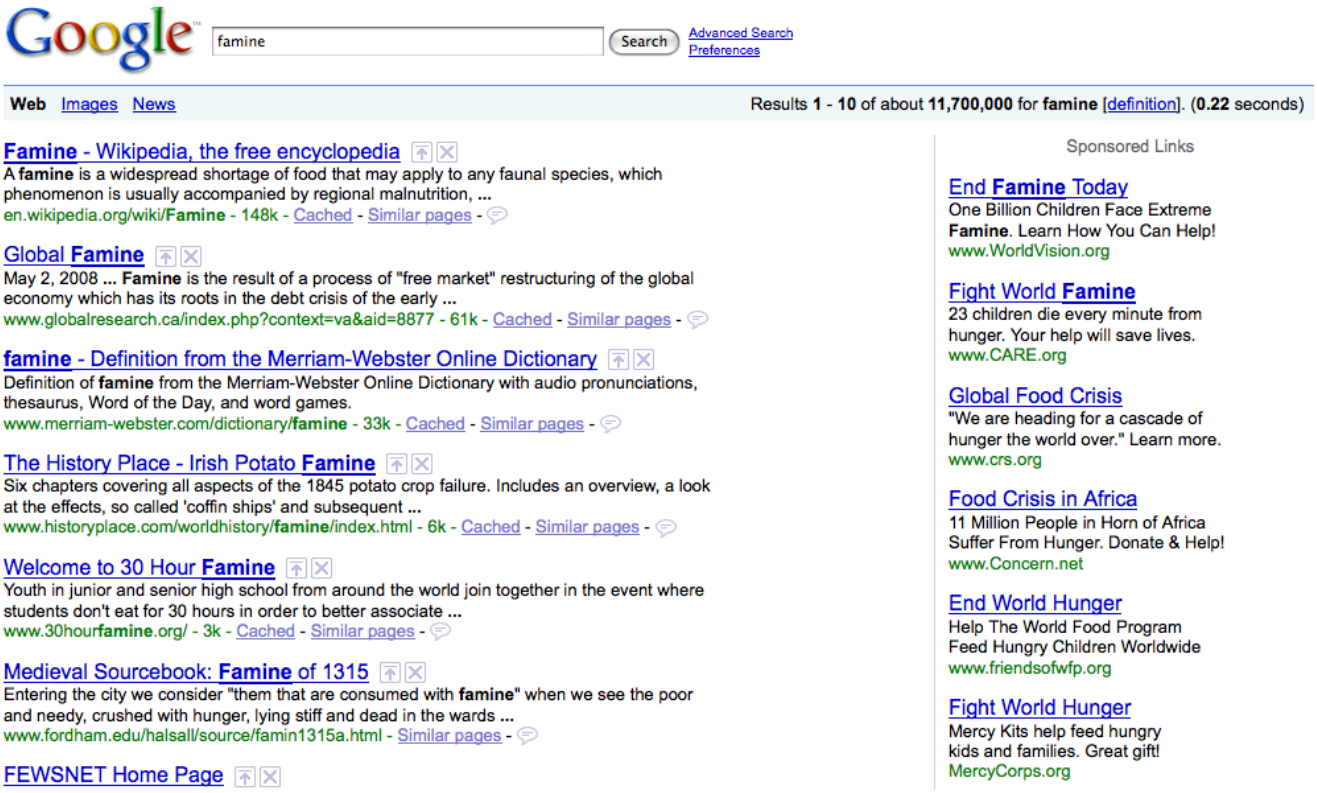

Figure 5.11: Google search for "famine"

There is some justification for this layer of filtering in the imperfection of the system as a whole (a given article may mention perfume and death or tragedy in the same article, so a perfume advertisers may find her ad placed next to a story on famine). But the need for this system also shows that the market for media does not perfectly follow user interest. The problem is not simply the tyranny of the majority but instead the complex interactions described throughout this paper that determine what sorts of content get produced. This topic exclusion only looks like

weak surveillance because no one advertiser generally has much power over the excluded publishers. But the overall effect of site and especially category exclusion is to bias the system against specific kinds of content, including many of the sorts of content that have generally been viewed as the public mission of the news media.

Terming Ads: Advertisers Watching Users

In addition to allowing advertisers to monitor (and exclude) publishers, Google freely provides a slate of some of the Internet's most powerful tools for monitoring Internet traffic in general and AdWords topics and traffic specifically. The purpose of these tools is to allow advertisers to optimize ad campaigns based on the popularity and profitability of keywords. The tools include the Keywords tool, which allows the lookup of current search frequency and average pay-per-click bid amounts for each keyword; the Search Based Keyword tool, which suggests to publishers under which keywords to advertise their sites; the Ad Planner tool, which displays the top five hundred sites according to a variety of geographic, demographic, and topical categories; and the Insights and Trends tools, which allow advertisers to see detailed comparisons of the popularity of search terms, including breakdowns by and comparisons of countries.

The use of these tools by advertisers looks like merely watching because Google aggregates all of the data before publishing it through any of these tools. So it is not possible for a given advertiser to identify, let alone directly control, a specific user through the data made available through the tools. There are, however, questions about the consent used to gather this data, not the least because Google does not fully disclose what data backs the tools. At the least, Google is using data collected through its search and ad click logs, both of which fail to gain meaningful consent for the user data they collect, as described above. The power of these tools lies in Google's ability freely to combine contextual data from its many different services, generating a data set that together is much more powerful that the individual sources. taking advantage of Google's vast potential for the combined use of its many different sets of user, publisher, and advertiser data.

It is hard to overstate how powerful these tools are, not only for advertisers but for any one interested for any reason in aggregated statistics about how people use the Internet. The Ad Planner tool, for example, replaces the Alexa data as the best free source of traffic data on the Internet. The Alexa data was based on the very skewed data from the very small portion of Internet users who use the Alexa toolbar. Google Ad Planner, in contrast, uses the data from its own search engine, which processes a large chunk of the Internet's traffic every day, as well as its own toolbar (much more popular than the Alexa toolbar), and proprietary market research companies, which likely means companies like ComScore, who pays millions of people to install highly intrusive software that records every bit of the user's network traffic, encrypted or not (contributing as well to the high lack of consent of these Google tools). (Olsen 2004) Likewise, the Keyword Tool allows advertisers to find which specific search terms (“song lyrics”) are related to which topics (“lyrics”) along with the volume of searches for those terms, the estimated cost per click of advertising on the terms, and the amount of advertiser competition for the terms. Advertisers interested in products related to lyrics can find out, for instance, that 6,120,000 people searched for “song lyrics” on Google in July 2009, that they will have to pay an average of $0.30 per click to advertise on the first page of search results, and that there is

heavy advertiser competition for advertising on "song lyrics."

These tools are all generally available to the public and are incredibly valuable to researchers and others interested in how the Internet is used. But Google built them for the use of advertisers to understand what sorts of sites and words-as-topics people are reading online. The impact of the tools is to significantly increase the knowledge of advertisers about the sites and words-as-topics users are browsing. Advertisers can use the tools to precisely translate broad product topics into specific words-as-topics with which to associate their ads. This ability gives advertisers a strong lever of control over the connection between their products and the content used to advertise the products. Google could charge very high fees for the use of the tools (as do market research companies with smaller but comparable data sets), but Google evidently thinks that it benefits more by empowering the advertisers at the edges of the network (and thereby increasing the power of the network as a whole) than by collecting revenue directly for itself at the center.

Terming Ads: Publishers watching Advertisers and Users

Google provides no tools specifically for publishers to optimize content for AdWords, presumably for fear of encouraging publishers to game the system. But publishers have access to the same tools as advertisers. Given the strong connection between ads and content established by AdWords, these tools are just as useful for publishers as for advertisers. Publishers can use the tools to watch both advertisers (to determine the value of ads) and users (to determine their interests and traffic patterns). As with advertiser use of these tools, publisher use of these tools looks like merely watching rather than surveillance, because all of the data is aggregated and so cannot be used directly to control the behavior of advertisers or users. But publisher access to the tool gives the publishers the ability to exert control over words-as-topics terming of content by optimizing their content for profitable topics.

For example, the Search Based Keyword tool allows publishers to determine the best way to advertise their sites. But the same tool can be used by the publisher to explore how his content is generating traffic for AdWords ads on his own site. Consider the results from the Search Based Keyword tool for "nytimes.com":

Keyword	Monthly searches ↓	Competition	Sugg. bid	Ad/Search share	Extracted from webpage
Keywords related to nytimes.com (100)					
lg dp781	2,300		USD0.47	--	LG DP781 Portable 8" DVD Player - NYTime
rm6750	660		USD0.46	--	Polk Audio RM 6750 - home theater speaker
im510	540		USD0.32	--	Altec Lansing inMotion iM510 speakers for S
primera lx810	350		USD2.23	--	Primera LX810 Color Label Printer - label pri
rebel k2	190		USD0.37	--	Canon EOS Rebel K2 - SLR camera - 35mm
jericho playstation 3	190		USD0.28	--	Clive Barker's Jericho (PlayStation 3) screen
aluratek admpf110	155		USD0.48	--	Aluratek ADMPF110 - digital photo frame - N
d4996t	155		USD0.42	--	HP Pavilion Ultimate D4996t Customizable D
hwba54g	155		USD1.07	--	Hawking Wireless-G Multi-Function AP/Bridg
admpf110	125		USD0.37	--	Aluratek ADMPF110 - digital photo frame - N

Figure 5.12: Search Based Keyword Tool results for nytimes.com

According to these results, the most effective advertising strategy for the New York Times would

be to advertise itself mostly in association with specific consumer electronics products. Given the dire state of the newspaper industry, with a constant stream of news about bankruptcies and emergency loans, how can this insight not strongly pressure the New York Times? The New York Times is probably not going to become an electronics magazine, if for no other reason than that it likely garners readers for its electronics writing largely through its reputation as the paper of record. To garner that reputation as a paper of record, it has to continue to write about hard news, but even if we consider that a strong focus of the New York times will always be its hard news writing, the results from the Search Based Keyword tool are interesting. Consider the most profitably searched keywords that include "obama":

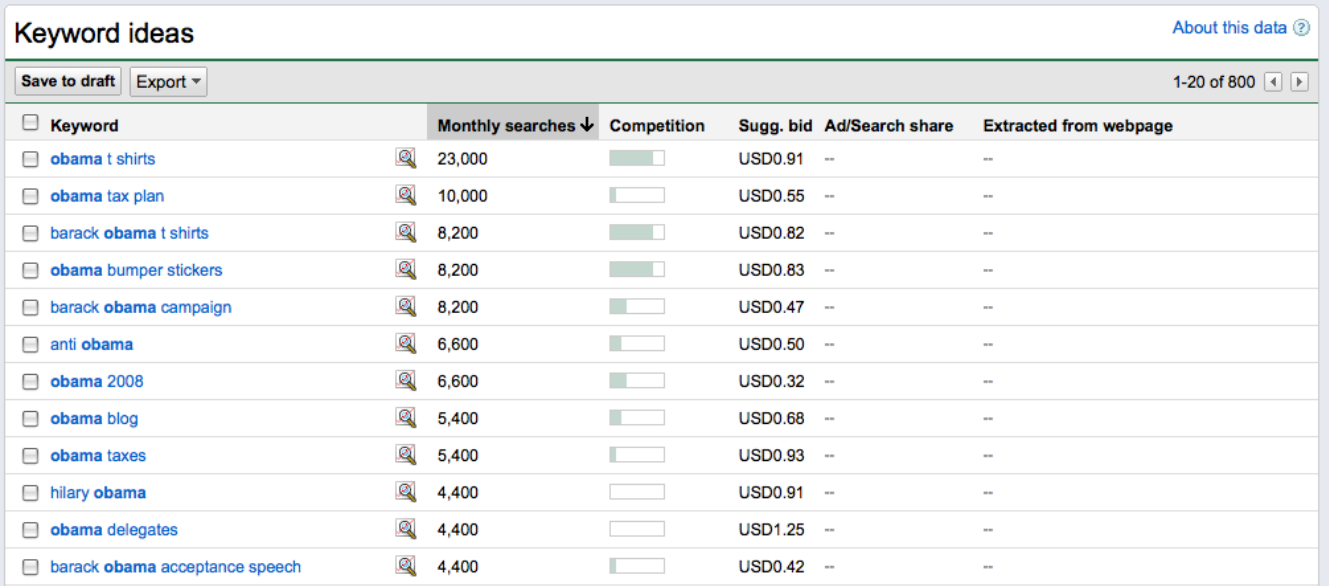

Figure 5.13: Search Based Keyword Tool results for "obama"

Overwhelmingly the most popular keywords for "obama" are about t-shirts and bumper stickers. These results mean that the most profitable way to advertise a site about Obama is as an Obama t-shirt and bumper sticker site. The second entry in the results is for "obama tax plan," which seems like the kind of result that would indicate a healthy market for substantive content about Obama. But a search for "obama tax plan" displays the following ads:

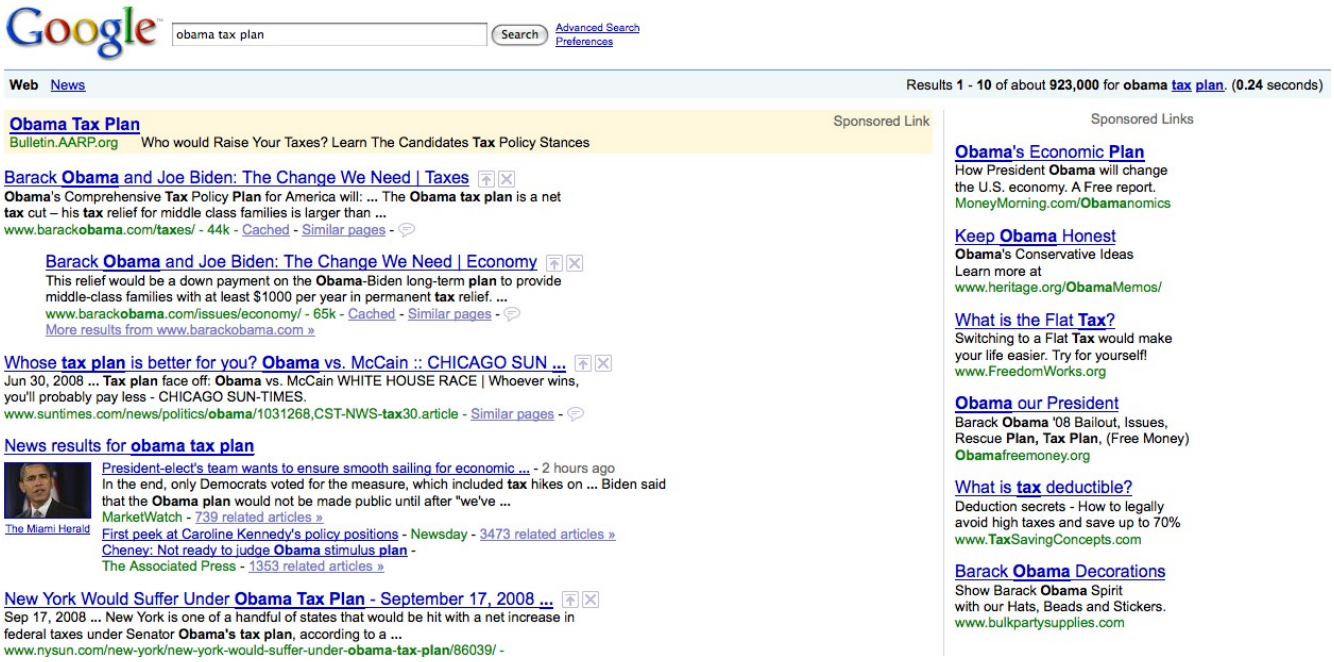

Figure 5.14: Google search for "obama tax plan"

Of the seven ads for "obama tax plan," three (aarp.org, heritage.org, freedomworks.org) are for sites with substantive information about Obama's tax plan, two (moneymorning.com, obamafreemoney.org) are borderline scams, one (taxsavingsconcepts.com) is an ad for tax software that looks vaguely scam-like ("Instant Pay Raise from the Government"), and the last is for party decorations. One response to this list is to conclude that the AdWords system simply does not work very well for content not directly connected to obvious consumer products. But the vast amount of revenue generated by the system and the huge number of sites hosting these ads suggest that it is working at least in the sense of processing lots of paying clicks from users. So if we take the system seriously, we have to consider what it means to the New York Times and other publishers that the AdWords ads supporting content about Obama's tax plans are mostly for scam-like sites and party supplies. In one sense, AdWords behavior is obvious and non-puzzling – what products other than t-shirts and bumper stickers could an advertiser possibly sell as "obama" products? But jingle advertising doesn't require such a direct connection and so allows publishers to support "obama" stories with ads selling perfume and cereal. The AdWords mode of ads as directly relevant classifieds requires that ads associated with "obama" content be directly relevant to "obama" and so forces publishers to support those stories with t-shirts and bumper stickers.

The New York Times maintains its own ad department, through which it sells big banner ads on

its site (as well as ads in the paper version), and it collects subscriptions for its paper as well, so AdWords represents only one relatively small part of the New York Times revenue. These other sources of revenue rely more broadly on the reputation and audience size of the paper as a whole, ideally encouraging the New York Times to write content that garners the trust of the largest number of readers. But even the New York Times places AdWords ads on its stories, just after the end of each story. This placement seems at a glance to minimize the importance of the AdWords ads, compared to the big banner ads at the top of each story. But this placement allows the AdWords ads to benefit from the context provided by the content of the story. The user reads the Adwords ads just after the end of the story in the context of the story she has just read, giving them meaning beyond the simple, short copy enforced by Google.

For example, consider a set of AdWords ads from the New York Times attached to a story on the 2008 stock market crash and economic crisis:

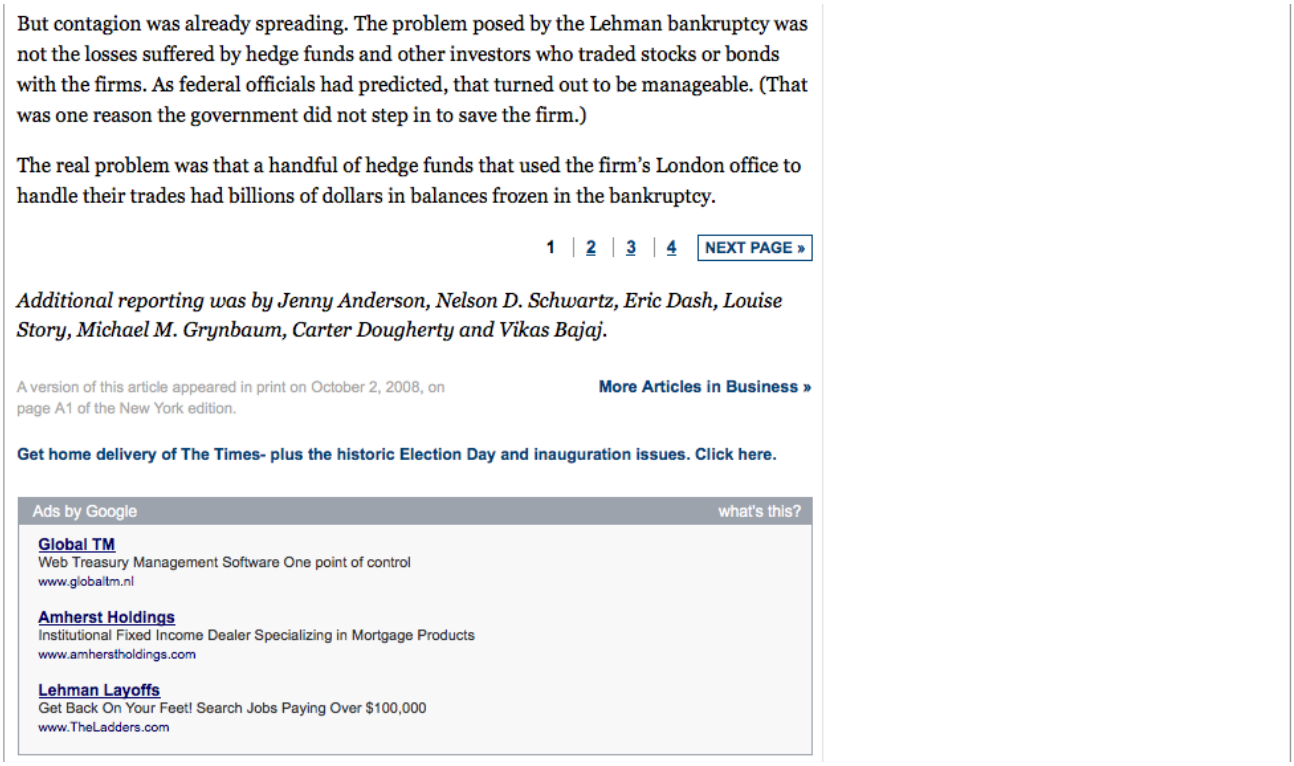

Figure 5.15: AdWords ad on New York Times story about the 2008 economic crisis

This story is doing heavy social work for the ads by helping the reader make sense of these finance and employment products. And what ad copy could provide better context for an executive recruiting company than a New York Times article on the banking crisis? This is not to say that the New York Times has been reduced to the role of a shill for the finance industry or that it is altering the content of its stories specifically to sell ads. The point is that the AdWords words-as-topics terming system automatically puts the New York Times content into this role of providing meaning for the most profitable set of ads. That positioning of contextualizing content with ads on particular topics creates a new sort of financial pressure to drive users to ads on

topics specifically related to the content of a given story. Bucking a long history of publishers bowing to the pressures of advertising, the New York Times may choose to resist that pressure and write simply to improve its larger reputation and audience size. But given the pressure exerted and the tools available for the New York Times to easily discern those pressures, resistance must be active at the very least.

Terming Ads: Decoding Ad Words

In their seminal law review article on privacy in 1890, Louis Warren and Samuel Brandeis argue that the evolution of society to increasingly concern itself with "the intellectual and emotional life" require a recognition of a "right to be let alone" that protects the liberty of "private and domestic life" as well as physical liberty. (Warren and Brandeis 1890) Warren and Brandeis were concerned particularly with the combination of new, potentially intrusive technology like instant photography and the growing circulation and intrusiveness of the press. They argued that gossip widely distributed through the press not only puts the private lives of the subjects in an unfair light, but also harms the larger social discourse by occupying the public with petty gossip:

> no one can be surprised that [gossip] usurps the place of interest in brains capable of other things. Triviality destroys at once robustness of thought and delicacy of feeling. No enthusiasm can flourish, no generous impulse can survive under its blighting influence. (Warren and Brandeis 1890)

This argument – about the effect of privacy invasion on social discourse – is the closest Warren and Brandeis come to addressing the core impact of the AdWords system. Arguably because of the legal context of the article, however, Warren and Brandeis only touch on this point about the larger social discourse and focus mostly on the harm done to the individual by this new form of public gossip. The largest influence of their article was to establish a philosophical and legal basis for this individual right to be let alone from publication of the details of one's private life. That individual right is not particularly helpful in the AdWords case simply because no personal details are actually published, certainly not in the sense of press publication that Warren ad Brandeis addressed. Since the AdWords data is not published and no individual is harmed by public gossip, Warren and Brandeis' right to be let alone does not help much in making sense of the AdWords case.

But Warren and Brandeis' broader right to be let alone provides a helpful framework for thinking about the larger effect of the AdWords model. By requiring publisher content to do much of the social work that ad content has historically done, Adwords subjects publishers to a bias for the production of stories about a certain set of literal, directly profitable topics like consumer electronics, t-shirts, bumper stickers, party supplies, scam-like services, and investment advice. Even publishers who are committed to writing about a specific domain like politics are subject to the same forces within that domain, knowing that the AdWords profit of writing even about hard news topics lies in providing the reader with a context for these words-as-topics that makes them socially meaningful.

6. Valuing Ads

AdWords uses auctions to determine how much advertisers pay for ads. Advertisers bid a pay-per-click amount against each other for the right to display an ad alongside given words-as-topics. That bid amount is used as one of several criteria to determine the order of ads displayed alongside a given words-as-topic. This auction mechanizes the valuation process, allowing AdWords to scale up to millions of ad words-as-topics bought by millions of advertisers paying millions of publishers for the clicks of billions of users.

The media industry has generally evolved in the U.S. from a state supported to a private enterprise, even though that evolution has itself always been strongly influenced by government policies. This evolution has democratized the media, exchanging explicit control by the state for control by the nexus of publishers, advertisers, and consumers, with remaining government influences. (Starr 2004) But the interaction of those forces has never created a straightforward free market in which publishers produce whatever content is most popular with consumers. The market that determines what sorts of content the media produces is distorted from this ideal by a number of factors, including the fact that the content provided by the market plays a strong role itself in informing users about their preferences. (Baker 1997) AdWords represents another branch in the evolution of media, not in the sense of bringing it closer to perfection but rather in the sense of changing it to adapt to the new environment of the Internet. In this new model, Google is a new actor within the network of publishers, advertisers, and users. And Google's primary influence is to democratize the system further by giving smaller players more influence: users in general, smaller publishers, and smaller advertisers all have more power to control the network at various points, including the pay-per-click auction model described in this chapter as well as the collective ranking and terming models described in earlier chapters.

A cost of building a machine that scales collective input to millions of publishers, advertisers, and users is that many of the watching activities look a lot like mechanized surveillance. Consider for example the motivating activity for this paper – the act of Google processing vast amounts of data about consumer browsing topics in the course of targeting AdWords ads. Analyzing this activity from within a surveillance framework is helpful because it pulls out the points of control that arise from the various different surveillance-like activities. Understanding how the different actors interact to produce content and advertising through the AdWords model requires understanding the puzzle of how those points of control come together (or not). The framework of surveillance is helpful in analyzing the issues of control. The last piece of that puzzle is understanding how the AdWords system automatically assigns values to ad words-as-topics, and through those topics to the content that hosts the ads. This chapter will argue that the Adwords action model devolves almost all of the power to the advertisers, but in the process creates a mechanized bureaucracy that is impossible to predict and therefore control.

Valuing Ads: Advertisers Watching Advertisers

The core activity driving AdWords ad auctions – advertisers watching the bids of other advertisers and bidding in accord – looks like weak surveillance. As a whole, the network of advertisers watching one another determines ad prices. Many advertisers have no meaningful

choice about whether to buy AdWords ads because of AdWords' dominant market share. In some cases, the competition around a single words-as-topic can give a single advertiser a high degree of control over other individual advertisers, for instance by bumping a given advertiser off of the list of displayed ads or by vastly increasing the required bid price. But in general, a single advertiser only contributes to the overall system of valuation and has little control over any other advertiser. So there is some control, but it is hard to argue that this activity is clearly surveillance.

The impact of the ad auction on AdWords as a whole is to mechanize the process of valuing ads. The auctions allow Google (through the advertisers) to place values on advertising terms without having to monitor the market value of the ads itself, vastly decreasing the cost of valuing each term for each advertiser and each publisher. As described in detail above, Google determines ad placements not only by auction bid amounts, but also by its ad quality score. But to the degree that the auction determines ad placements and rankings, advertisers drive the process through the auction, watching the bids of the other advertisers and responding with their own bids. Google has no direct role in pricing ads other than designing the machine that runs the auctions. Using this auction system, Google has scaled its system to sell advertising space to over a million advertisers, over a million publishers, and over a billions users. (Helft 1 Million Advertisers 2008)

The process of valuing ads has traditionally been a manual negotiation between publishers, advertising agencies, and advertisers. Advertising agencies began as re-sellers of publishers' ad space – they bought advertising space in bulk and then resold it to advertisers. Agencies eventually changed over to an "open contract" system in which they were paid by advertisers rather than publishers and therefore had a clear allegiance to the advertisers. After the move to the open contract system, agencies only gradually began to take over the role of writing advertising copy from their advertiser clients. (Fox 2007) But they continue to perform the brokerage role of finding the most appropriate publications for advertisers and negotiating ad prices with those publications.

The AdWords auction largely replaces human judgments of advertising agencies and publishers with its mechanized judgments. Certainly the majority of advertising revenue is still generated by non-Internet advertising (print, television, radio, etc), and about a quarter of Internet advertising revenue is generated from DoubleClick and other banner ad systems that act more like traditional advertising systems. (Internet Advertising Bureau 2008) And many advertisers use advertising agencies to buy AdWords ads within a larger advertising strategy that includes other, non auction based advertising systems. But the use of auctions strongly decreases the role that publishers and advertising agencies play in valuing AdWords ads. As AdWords plays an increasing role in the advertising market, the value decisions made by advertisers through its auction machine will increasingly replace the value decisions made by publishers and advertising agencies through manual negotiations.

This move – away from control by a few publishers and advertising agencies toward control by the very large crowd of advertisers – marks a further evolution of media toward democracy. As with the collective participation of users in the search and ad ranking systems, this democratization is an opening of public sphere described by Benkler, allowing many voices to

participate where only a few participated before. But the democratization of the ad valuing process is only a democracy of input, not of process. The auction that brokers the decisions of the advertisers is a specific machine with specific ways of operating. Neither the existing mode of valuing ads largely through audience sizes nor this new mode of valuing ads through the crowd of advertisers reduces merely to a poll of consumer interest. In the agency negotiated value model, saying that publishers can sell ads next to stories about Tanya Harding for more than ads about stories about Rwandan genocide is not the same as saying that people want to read about Tanya Harding more than they want to read about Rwanda genocide. (Baker 1997) In the auction negotiated value model, saying that publishers can sell ads next to stories about smart travel for more than ads about war is not the same as saying that people want to read about smart travel more than they want to read about war.

As with large human run bureaucracies that process large amounts of information, a danger of the mechanized auction is that its behavior is difficult to predict or control. Even though individual advertisers drive the process, the large number of inputs from a large number of actors makes the ad auction very difficult to predict. As with a human bureaucracy, no one entity is wholly in control of the decisions the auction makes or even really understands why it is making the decisions. Google obviously has the ability to override any one of the machine's decisions, but as with the collective ranking mechanisms, the quality of the results and its reputation in the community encourage it to act as a neutral broker of the technology.

The availability of massive amounts of consumer data has historically attracted bureaucracies to make decisions using that data. Oscar Gandy has argued that the dramatic growth in the collection of personal data about consumers creates a "panoptic sort ... a kind of high-tech, cybernetic triage through which individuals and groups of people are being sorted according to their presumed economic or political value." The danger of the panoptic sort is that existing power structures will be constantly reinforced as the wealthy and advantaged are given ever more advantages (better mortgage offers, better customer service, and so on). (Gandy 1993) David Lyon builds on Gandy's frame of the panoptic sort to argue that "social sorting" through these vast, searchable collections of personal data create "data doubles" that have profound effects on the represented individuals:

> they serve to open and close doors of opportunity and access. They affect eligibilities for credit or state benefits and they bestow credentials or generate suspicion. They make a real difference. They have ethics, politics. (Lyon 2003, p. 27)

Daniel Solove argues that the sum of data collected about a person in corporate and government databases over the course of her life constitutes a "digital dossier" and that opaque bureaucracies increasingly use these digital dossiers to make constitutive decisions about people's lives. This loss of power to these powerful bureaucracies leads to a Kafkaesque hopelessness:

> an individual's sense of helplessness, frustration, and vulnerability when a large bureaucratic organization has control over a vast dossier of details about one's life. Bureaucracy often results in a routinized and sometimes careless way of handling

information – with little to no accountability. (Solove 2004, p. 9)

Solove argues that the danger of widespread consumer data collection is not so much in the power reinforcement argues by Gandy and Lyon as in the lack of accountability inherent in the large bureaucracies required to process large collection of data. Solove compares these data processing bureaucracies to the government bureaucracies in Franz Kafka's *The Trial*, which doom the protagonist to endless redirection and paper shuffling until the day he is executed.

The AdWords auction machine risks becoming just a mechanized bureaucracy an order of magnitude more complex than those envisioned by Solove or Kafka. This auction acts as the sort of Kafkaesque (powerful, unpredictable, and unaccountable) bureaucracy described by Solove with the power to encourage the creation of some types of content and discourage the creation of other sorts of content. This auction acts as a magical, mechanical broker that makes important decisions about what sorts of content to reward but without the ability to answer why it makes those decisions, dooming those who care about what sorts of content publishers create online to follow endless redirections and shuffles trying to understand the emergent decisions of a vastly complex machine.

Valuing Ads: Advertisers and Publishers Watching Users

Each AdWords advertiser has access to the IP address of each user that clicks on the advertiser's ad, along with the specific ad topic that the user was browsing when she clicked on the page.[14] Advertisers have the ability to track users as they move around the landing site from the original landing page, tracking for instance whether and how each particular user, identified at least by an IP address and possibly also by Google's and the advertiser's own cookies, navigates to the purchase of a product. Publishers have access to less information about AdWords users as they click on ads, since Google returns only aggregate data to publishers about how many users over various time periods clicked on ads from which pages. Publishers can easily transform this aggregate data at least into data about which topics generated the most ads, giving them the ability to optimize the topics of their content for the paying ad clicks of their users.

Even though advertisers have access to more detailed information about users clicking on ads, both watching activities look like similar cases of merely watching because there are no mechanisms of direct control over users. Advertisers and publishers actively optimize their content according to the ad clicks of users, but they do so based on the aggregate data, and the impact of those optimizations happens in the aggregate as well. There is some question about the level of consent granted to advertisers and publishers to collect this data, but the data collection is the same collection that happens any time a user visits any web site (the site host has access to the user's IP address, the url requested, and usually a collection of other HTTP data such as the

14 Google does not directly pass on information to advertisers about which users were browsing which topics. But an advertiser can easily setup a distinct landing page for each topic under which it is advertising. So an advertiser who sets up http://foo.bar/ads/cereal as the landing page for an ad displayed with the words-as-topic "cereal" knows that any user who visits that page from an AdWords ad was browsing content about "cereal" at the time.

referring web page and the user agent browser identification string). Many users certainly do not understand the amount of data transmitted to a web server with each visit to a web page, but the AdWords advertisers and publishers are doing nothing different from the norm on the web (unlike Google, which is actively, if weakly, hiding its use of redirect pages to track ad and search clicks as described in earlier chapters).

Advertisers and publishers use this data to learn about not only the interests of users, but also the effectiveness of advertising (to generate product sales) and content (to generate ad sales). Michael Schudson argues that the role of advertising in directly generating sales is dubious – that its role has more to do with creating a place for consumer products within the larger society than with tricking people into buying things they do not need. Advertising is not a science aimed at convincing people to buy things but rather a sort of social magic meant to place a product (usually only after it has become successful) within the larger social realm. (Schudson 1986) T. Jackson Lears argues along similar lines that advertising, despite constant yearning for more professionalism and scientific methods, has never escaped its mystic, naturalistic, patent medicine peddler roots because advertising has always been a magical, intuitive craft. The goal of advertisers is to create the "fables of abundance" that give consumer products meaning within the larger society, that transform the act of purchasing things into an act of giving a person a meaning within society.

This move is itself a sort of magic: buying a cereal can magically transform an individual into a member of a healthy, happy family; buying a perfume can magically make a person beautiful. The tension of the advertising world is that this magic has to happen within the context of the scientifically managed corporate world that produces these (magical) products:

> A more capacious approach might acknowledge that the overall pattern of nineteenth-century change was a developing balance of tensions – within the broader society and gradually within advertising itself – between dramas of magical transformation and moralistic or managerial strategies of control. The recurring motif in the cultural history of American advertising could be characterized as the attempt to conjure up the magic of self-transformation through purchase while at the same time containing the subversive implications of a successful trick. (Lears 1994, p. 41)

Lears argues that advertising has lacked concrete metrics not because a lack of proper tools of measurement but because its fundamental job has always been a sort of cultural magic that defies measurement. The magic of AdWords is rather that it can measure how well it is working, but that to do so it has to use a complex machine that is by definition impossible to understand or predict.

AdWords transforms ads into precisely targeted and measurable classifieds whose jobs is simply to point the user toward a relevant product at the right time and to attach a value to those ads for publishers and advertisers. The magic of the AdWords model is not in the peddler creativity of advertisers, but instead in the mechanized auction that magically places certain values on certain ads based on the combined inputs of vast numbers of advertisers, publishers, and users. The AdWords model is not magical because it cannot measure its impact on culture, as Lears argues

for traditional advertising. It is magical because it measures in so many ways that the machinery of measurement takes on a life of its own. The unpredictable genius of the creative individual gives way to the unpredictable genius of the emergent machine. But that unpredictability is also the danger of the machine – the machine has the potential to devolve into a Kafkaesque bureaucracy of the sort described by Solove because by definition no on can explain or predict its decisions.

Valuing Ads: The Magical Machine

> 'Finally, he was quartered,' recounts the Gazette d'Amsterdam of 1 April 1757. 'This last operation was very long, because the horses used were not accustomed to drawing; consequently, instead of four, six were needed; and when that did not suffice, they were forced, in order to cut off the wretch's thighs, to sever the sinews and hack at the joints... (Foucault 1975, p. 75)

Foucault traces the evolution of social discipline from the spectacles of medieval executions to the panoptic surveillance of industrial institutions. Foucault argues that the spectacles were effective insofar as they represented an imposition of the king's body onto the body of the public, and at the same time provided space for the condemned to voice their frustrations against monarchy. This system of disciplining the public began to break apart as the power of the people grew, and consequently the king's symbolic body lost power relative to the body of the people. The increasingly powerful public was able to question the fairness of the terrifying executions and the one-sided prosecutions that led to them.

What eventually replaced these spectacles, Foucault argues, was the modern set of institutions whose most important impact was to embed discipline into the social fabric itself, rather than to impose discipline bodily through bloody spectacle. Prisoners, pupils, patients, and workers learned that they were being watched continuously, and that their fates were judged by a set of scientific criteria which were themselves defined by the system of watching. Instead of investing all power in the prosecutor, the modern judicial system invests its power in the jury's ability to objectively judge the truth of various pieces of witness and scientific testimony. The guilty man is condemned not because of the will of the king but because our objective system determines him to be guilty. If you do not want to be judged guilty, you have to judge yourself by these objective criteria, rather than by the arbitrary decisions of the king. And if you want to succeed in school or at work, you have to measure up to objective criteria. But those objective criteria are themselves a product of this process. Experts discipline themselves for their work in court, social institutions design the tests that determine school success, and so on.

In 1994, Scott Bukatman captured a widely held view that television and other mass media had once again made spectacle the dominant mode of social discipline. Bukatman pulls together a set of postmodern social theorists, media thinkers, and science fiction authors to describe a "terminal space" in which a flood of image, audio, and text "blips" constitute a never ending spectacle through which society defines itself. This image of society as spectacle contrasts sharply with the image drawn just twenty years before by Foucault. And the spectacle has multiplied itself many times with the explosive growth of the Internet since 1994, including the growth of Internet pornography and YouTube beatings that make Burrough's death dwarf from

his science fiction experiment *Nova Express*, via Bukatman, look prescient:

> "Images – millions of images – That what I eat – Cyclotron shit – Ever try kicking that habit with apomorphine? – Now I got all the images of sex acts and torture ever took place anywhere and I can just blast it out and control you gooks right down to the molecule – I got orgasms – I got screams – I got all the images any hick poet ever shit out – My power's coming – My power's coming – My power's coming. ... And I got millions of images of Me, Me, Me meee." (Burroughs 1994, p. 45)

Bukatman's mass media spectacle is a culmination of the advertising magic described by Lears. It is the art of the peddler multiplied and magnified through the mass media of the 1980s and 90s.

Foucault's spectacle of execution and Bukatman's (and others') spectacle of mass media each refer to a display of striking images. The public execution is striking largely because it is dramatically physical. One cannot watch a execution without a strong, physical reaction. Bukatman's flood of images (and other media) is striking partly because each image is designed for emotional impact (buy this SUV if you want to dominate the road) but mostly because of the sheer number of images. It is striking to be shown many different images at once, even if each image is just as a single solid color. Foucault's spectacle is striking because it is so strongly physical, whereas the modern media spectacle is completely virtual – it is a spectacle because of the sheer flood of input that cannot be reproduced bodily. Bukatman argues that the lack of physicality actually defines the media spectacle as such: "pure spectacle ... [is] ... a proliferation of semiotic systems and simulations which increasingly serve to replace physical human experience and interaction." (Bukatman 1993, p. 26) Most importantly, both of these spectacles are used a source of control over society, and the impact of each form of spectacle relies on the dramatic impact of the spectacle itself – this need for dramatic impact is why, for instance, modern executions by injection serve nothing like the role of the spectacular public executions that Foucault describes. In contrast to the slow diffusion of the social knowledge through institutions, spectacle derives its power from its ability to reach directly into the brain of its subjects and create an immediate reaction (who wants to make a careful argument about which textbook a school should use when you can just make a YouTube video and beam your truth directly into kids' brains?).

Both of these forms of spectacle (like Foucault's institutional discipline) involve not just being watched, but watching as well. The spectacle of execution is an application of watching onto the public: not only does the execution enact the punishment of the king on the body of the people, but the process of prosecution applies the eye of the king onto the people. By watching and judging the condemned, the king is making clear that he is watching the public, both symbolically and through his state apparatus. The watching and being watched of this process are necessarily entangled: one cannot have a public execution without a prosecution, and the prosecution has no social impact if it is not in turn watched.

The mass media spectacle provides an even more tangled relationship between watching and being watched. Much of the modern spectacle is advertising, which is about pushing images to consumers to get them to watch them. But advertisements are only useful if the advertiser knows

who is watching them. Television ratings and commercials are necessary complements, as are clicks and web advertisements. An advertisement (and any spectacle used as a social lever) is only useful insofar as its impact can be measured, and knowledge of who is watching an ad is necessary to measure that impact. In the mass media world, the television acts much in the same way that the two way television works in *1984*. Consumers watch the screen and are in turn watched to generate the ratings that drive the advertising that pays for the images on the screen.

The network of grey surveillance seeded by AdWords marks a third kind of spectacle. The difference between the mass media spectacle described by Bukatman (and many others) and the spectacle of the AdWords system is that the work of the mass media spectacle is the same work of the advertising industry described by Lears – to give social meaning to consumer products – whereas the work of the AdWords spectacle is actually to sell the products, relying on the magic of the machine that helps users find and process relevant content and ads, that associates ads with content and encourages the content to do social work for the ads, and that values the content and the ads. The magic in that machinery lies in its construction as a device of collective input, seeding a network of watching activities by and of vast crowds of different actors. What makes this system spectacular is not the death dwarf's constant flood of broadcast images but rather the constant flood of ads and content between the edges of the network. The spectacle lies in the ability of the machine to hide its own mechanisms, making it seem that it is always giving all of the involved publishers, advertisers, and users exactly what they want all of the time.

7. Conclusion

One of Google's core values for its search engine and other systems is to be data driven – to constantly use data collected through its tools to verify that any changes made to the tools are improving the system. (Huffman 2008) It has to operate this way because of the emergent nature of its systems – changes to its systems that may work anecdotally or theoretically may fall apart when plugged into the immense collective input of the live Google machine. This unpredictability is the blessing and the curse of building its system to power itself from the edges of very large networks of publishers, advertisers, and users. In the case of the AdWords auction, the emergent decisions of the network allow Google to rely on the network machine to magically make decisions for all of its users through all of its users. But the curse is that not only does Google not strongly control these decisions that are being made at the edges, but it cannot really understand them.

To be sure, part of the magic of the machine lies in how well it works. Its search and ad systems are fundamental drivers of a new content ecosystem that allows efficient access to a large subset of the vast collection of topics and pages on the Internet and that newly empowers millions of small advertisers and publishers to participate by slashing transaction costs and mining the intelligence at the edges of the network. And the Google search user simulation works well. Granted works and well are loaded, but a query for "new york times" returns the website of the New York Times, which it did not in the search engines preceding Google. It is clear that a card catalog approach, manual or automated, will not suffice for the project of organizing the web, so some sort of massive, emergent processing of data is needed to make some sort of sense of the vast amount of web content, even if Google's particular system organizes the data in particular ways that have particular effects on the control of the overall network of control, as described in detail in the chapters above. Likewise, the AdWords advertising brokerage system generates a large amount of revenue for many publishers, and AdSense in particular allows small advertisers and publishers to participate in the media ecosystem much more easily. All this is to say that Google's use of this data has strongly beneficial effects.

Jonathan Zittrain has written of this operation at the edges of the network as a blessing and curse for the Internet as a whole, arguing that the engine of the explosive growth of the Internet over the last fifteen years has been the generativity of its end points – the ability of the computers (and people) that connect to the edges of the networks to constantly create new sorts of functionality unimagined and unregulated by any central actor. (Zittrain 2008) This generativity is what has allowed Facebook, Skype, EBay, Wikipedia, and Google itself, among many others, to grow and flourish on the Internet. The whole Internet is in this sense a sort of magical machine that is generating itself as it grows. But the cost of the generativity, according to Zittrain, is that current Internet is saturated with bad actors of all sorts – spammers, thieves, malware developers and administrators, and many more – that take advantage of the generative nature of the Internet to build into the network whatever functionality they want. If anyone can operate the machine, anyone can operate the machine.

Google has built its search and AdWords systems on top of the generative power of its many participating publishers, advertisers, and users. This generative engine is spectacular in its scope and reach. Google watches millions of publishers to set guidelines for web content; millions of

publishers watch each other to rank each other's content; Google watches billions of users to create its search-nervous-system-as-user-simulation; billions of users watch this search-nervous-system-as-user-simulation, and through it each other, to judge its functionality; Google watches millions of advertisers to set rules and guidelines for AdWords ads; millions of advertisers watch Google optimize their ads according to Google's standards; Google watches billions of users click on ads to target ads and to prioritize popular ads; billions of users watch the advertisers' ads to power the collective engine that ranks the ads; Google watches the billions of users browse the web to create a system to stateless market research; Google watches millions of publishers to determine the ad words-as-topics to attach to each piece of content; millions of publishers watch millions of advertisers and billions of users to optimize their content for the ads that support it; millions of advertisers watch each other through ad auctions to determine the value of ad clicks on words-as-topics; and millions of advertisers and publishers watch billions of users click on ads to determine the popularity (and gross revenue) of the words-as-topics ads.

Whether or not anyone is directly fired from a job, refused a mortgage, or thrown in jail using the data processed by Google through AdWords, the various individual watching activities within this network – some clearly surveillance, some weak surveillance, and some merely watching – give the actors various sorts of direct control over each other: control over which content gets attention through Google's search engine, control over how the search engine's user simulation works, control over what sorts of content are supported by AdWords advertising, control over the general form and function of AdWords ads. The general shape of this network is mandated by various tools and policies developed and run by Google, but the decisions of the network as a whole are driven by these various different watching activities by the massive crowd of participants.

The mostly crowd-based, generative models that Google has established for the components of this network necessarily require that Google give up a great deal of control over the decisions of the network. At the same time, the publishers, advertisers, and users who act as the engine of this collective Google nervous system are ever increasingly integrating the decisions of that nervous system into their own lives, giving it control over their lives. Users interact with Google in a way that makes it a collective extension of their own brains rather than merely an information retrieval tool, giving the search and ad ranking systems deep impact over the ways that those users find and consume content and products on- and offline. Publishers and advertisers interact with the AdWords terming system in a way that pressures publishers to provide content that directly supports the sale of specific consumer products. And publishers, advertisers, and users interact through the AdWords to assign values to advertising topics and the associated content through a massively distributed auction that amounts to a mechanical bureaucracy that threatens to be as unaccountable as the worst offline bureaucracy.

References

Aaker, David A., and James M. Carman. “Are You Overadvertising?.” *Journal of Advertising Research* 22.4 (1982): 57-70.

Agre, Philip E. “Surveillance and Capture: Two Models of Privacy.” *The Information Society* 10.2 (1994): 101-127.

Attributor Corporation. “Get your fair share of the ad network pie.” Attributor Corporation, 2008. Web. 17 Oct 2008. <http://www.attributor.com/blog/get-your-fair-share-of-the-ad-network-pie/> accessed 17 Oct 2008.

Baker, C. Edwin. “Giving the Audience What it Wants.” *Ohio State Law Journal* 58.2 (1997): 311.

Ball, Kirstie. “Categorizing the workers: electronic surveillance and the social ordering in the call center.” *Surveillance and Social Sorting: Privacy, Risk, and Digital Discrimination*. Routledge, 2002.

Barbaro, Michael, and Tom Jr Zeller. “A Face Is Exposed for AOL Searcher No. 4417749.” *The New York Times* 9 Aug 2006.

Benkler, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, 2007.

Bogard, William. *The Simulation of Surveillance*. Cambridge University Press, 1996.

Bowers, Chris. “Googlebomb John McCain.” *Open Left* 22 Feb 2008. Web. 12 Dec 2008. <http://www.openleft.com/showDiary.do?diaryId=4098> accessed 12 Dec 2008.

boyd, danah. “Technology and the World of Consumption.” *danah boyd's blog* 10 Jan 2008. Web. 22 Dec 2008. <http://digitalyouth.ischool.berkeley.edu/blog/21>.

Boyne, Roy. “Post-Panopticism.” *Economy and Society* 29.2 (2000): 285.

Bracha, Oren, and Frank Pasquale. “Federal Search Commission? Access, Fairness, and Accountability in the Law of Search.” *Cornell Law Review* 93.6 (2008).

Brin, Sergey, and Lawrence Page. “The Anatomy of a Large-Scale Hypertextual Web Search Engine.” *Computer Networks and ISDN Systems* (1998): 107-117.

Bukatman, Scott. *Terminal Identity: The Virtual Subject in Postmodern Science Fiction*. Duke University Press, 1993.

Burroughs, William S. *Nova Express*. Grove Press, 1994.

Cohen, Noam. “Google Halts ‘Miserable Failure’ Link to President Bush.” *The New York Times* 29 Jan 2007.

comScore Inc. “comScore Releases June 2008 U.S. Search Engine Rankings.” *comscore.com* 18 Jul 208. <http://www.comscore.com/press/release.asp?press=2337> accessed 23 Jul 2008.

Electronic Frontier Foundation. “FTC complaint against AOL about 2006 release of search data.” 2006. <http://www.eff.org/Privacy/AOL/aol_ftc_complaint_final.pdf> accessed 5 Dec 2007.

Eubanks, Virginia. “Technologies of Citizenship: Surveillance and Political Learning in the Welfare System.” *Surveillance and Security: Technological Power and Politics in Everyday Life*. Routledge, 2006. 89-107.

Fisher, Jill A. “Indoor Positioning and Digital Management: Emerging Surveillance Regimes in Hospitals.” *Surveillance and Security: Technological Power and Politics in Everyday Life*. Routledge, 2006.

Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. Trans. Alan Sheridan. Allen Lane, 1977.

Fox, Stephen. *The Mirror Makers: A History of American Advertising and Its Creators*. University of Illinois Press, 1997.

Gandy, Oscar H. *The Panoptic Sort: A Political Economy of Personal Information*. Westview Pr (Short Disc), 1993.

Gilligan, Colin. “How British Advertisers Set Budgets.” *Journal of Advertising Research* 17.1 (1977): 47-49.

Gilliom, John. *Overseers of the Poor: Surveillance, Resistance, and the Limits of Privacy*. University of Chicago Press, 2001.

Goldman, Eric. “Data Mining and Attention Consumption.” *Privacy and Technologies of Identity: A Cross-Disciplinary Conversation*. Springer, 2005.

---. “Search Engine Bias and the Demise of Search Engine Utopianism.” *Yale Journal of Law & Technology* 9 (2006): 198.

Google, Inc. “Advertising and Privacy.” *Google Privacy Center* 2008. <http://www.google.com/privacy_ads.html> accessed 20 December 2008.

---. “Corporate Information - Technology Overview.” *google.com* 2008. <http://www.google.com/corporate/tech.html> accessed 20 November 2008.

---. “Google Privacy Center - Privacy Policy.” *google.com* 4 Oct 2005. <http://www.google.com/privacypolicy.html> access 7 May 2008.

---. “Google Search Privacy: Plain and Simple.” *The Google Channel* 8 Aug 2007. <http://www.youtube.com/watch?v=kLgJYBRzUXY> accessed 23 Jul 2008.

---. “Landing Page and Site Quality Guidelines - AdWords Help Center.” *adwords.google.com* 2008. <https://adwords.google.com/support/bin/answer.py?answer=46675&hl=en> accessed 2 Dec 2008.

---. “Quality scores and ad auctions.” *The Official Google Blog*. 2008. <http://googleblog.blogspot.com/2008/10/quality-scores-and-ad-auctions.html> accessed 29 Nov 2008.

---. “View all policies: Text ads.” *AdWords Help Center.* 2008. <http://adwords.google.com/support/bin/static.py?page=guidelines.cs&topic=9271&view=all> accessed 18 Dec 2008.

---. “Webmaster Guidelines - Webmaster Help Center.” *google.com* 2008. <http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=35769> accessed 11 Dec 2008.

Hadden, Sally E. *Slave Patrols: Law and Violence in Virginia and the Carolinas*. Harvard University Press, 2001.

Helft, Miguel. “Google Told to Turn Over User Data of YouTube.” *The New York Times* 4 Jul 2008.

Huffman, Scott. “Search evaluation at Google.” *Official Google Blog* 15 Sep 2008. <http://googleblog.blogspot.com/2008/09/search-evaluation-at-google.html> accessed 3 Dec 2008.

Identity Theft Research Center. “ITRC Breach Report 2008.” Identity Theft Research Center, 2008. <http://idtheftmostwanted.org/ITRC%20Breach%20Report%202008.pdf> accesed 27 Dec 2008.

IncreasePR. “Buy PageRank Text Links.” 2008. <http://www.increasepr.com/> accessed 12 Dec 2008.

Internet Advertising Bureau. “Internet Advertising Revenues in Q3 '08 at Nearly $5.9 Billion.” Internet Advertising Bureau, 2008. <http://www.iab.net/about_the_iab/recent_press_releases/press_release_archive/press_release/572194> accessed 23 Nov 2008.

Latour, Bruno. “Thought Experiments in Social Science: from the Social Contract to Virtual Society.” *1st Virtual Society Annual Public Lecture.* 1 Apr 1998. <http://www.artefaktum.hu/it/Latour.htm> accessed 14 Oct 2008.

Lears, T. Jackson. *Fables of Abundance: A Cultural History of Advertising in America*. Basic Books, 1994.

Lyon, David. *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination*. Taylor & Francis, Inc., 2003.

MacKinnon, Rebecca. “Shi Tao's case: Yahoo! knew more than they claimed.” *RConversation* 29 Jul 2007. <http://rconversation.blogs.com/rconversation/2007/07/shi-taos-case-y.html> accessed 27 Oct 2008.

Markoff, John, and Scott Shane. “Documents Show Link Between AT&T and Agency in Eavesdropping Case.” *The New York Times* 13 Apr 2006.

Marx, G. T. “Murky conceptual waters: The public and the private.” *Ethics and Information Technology* 3.3 (2001): 157-169.

Marx, Gary T. “Soft Surveillance: the Growth of Mandatory Volunteerism in Collecting personal

Information -- 'Hey Buddy Can You Spare a DNA?"' *Surveillance and Security: Technological Power and Politics in Everyday Life*. Routledge, 2006.

Mathiesen, Thomas. "The Viewer Society: Michel Foucault's 'Panopticon' Revisited." *Theoretical Criminology* 1.2 (1997): 215-234.

McLuhan, Marshall, and Lewis H. Lapham. *Understanding Media: The Extensions of Man*. The MIT Press, 1994.

McQuivey, James. "Grocery Store Loyalty Card Use is Strong Despite Privacy Concerns." Boston University College of Communication, 2004. <http://www.prnewsnow.com/Public_Release/Supermarkets/3050.html> accessed 29 Oct 2008.

Monahan, Torin. "Questioning Surveillance and Security." *Surveillance and Security: Technological Power and Politics in Everyday Life*. Routledge, 2006. 1-26.

Moulton, Ryan, and Kendra Carattini. "A quick word about Googlebombs." *Official Google Webmaster Central Blog* 25 Jan 2007. <http://googlewebmastercentral.blogspot.com/2007/01/quick-word-about-googlebombs.html> accessed 10 Dec 2008.

Narayanan, Arvind, and Vitaly Shmatikov. "How To Break Anonymity of the Netflix Prize Dataset." *arXiv.org* (2006). <http://arxiv.org/abs/cs/0610105> accessed 7 Dec 2007.

Nissenbaum, Helen. "Privacy as Contextual Integrity." *Washington Law Review* 79.1 (2004).

Olsen, Stefanie. "ComScore: Spyware or 'researchware'?" *CNET News.com* 20 Dec 2004. <http://news.cnet.com/ComScore-Spyware-or-researchware/2100-1032_3-5494004.html> accessed 21 Jul 2008.

OpenNet Initiative. "OpenNet Initiative Profile for China (including Hong Kong)." OpenNet Initiative, 2007. <http://opennet.net/research/profiles/china> accessed 23 Jan 2009.

Orwell, George. *1984*. New American Library, 1949.

Page, Larry, and Sergey Brin. "Larry Page Interview." *Academy of Achievement* 20 Oct 2008. <http://www.achievement.org/autodoc/printmember/pag0int-1> accessed 15 Dec 2008.

Palfrey, John, and Urs Gasser. *Born Digital: Understanding the First Generation of Digital Natives*. Perseus Books Group, 2008.

Parenti, Christian. *The Soft Cage: Surveillance in America, from Slavery to the War on Terror*. Basic Books, 2004.

Penenberg, Adam. "Cookie Monsters: The innocuous text files that Web surers love to hate." *Slate.com* 7 Oct 2005.

Raboteau, Albert J. *Slave Religion: The "Invisible Institution" in the Antebellum South*. Updated. Oxford University Press, USA, 2004.

Roberts, Hal. "Google Adwords Category Exclusion." *Hal Roberts: watching technology* 24 Jun

2008. <http://blogs.law.harvard.edu/hroberts/2008/06/24/google-adwords-category-exclusion/> accessed 21 Apr 2009.

Rosen, Jeffrey. *The Unwanted Gaze: The Destruction of Privacy in America*. 1st ed. Random House, 2000.

Rubenfeld, Jed. "Right to Privacy, The." *Harvard Law Review* 102 (1988): 737.

Schudson, Michael. *Advertising, the Uneasy Persuasion: Its Dubious Impact on American Society*. Basic Books, 1986.

Sen, Amartya. *Development as Freedom*. Oxford University Press, 1999.

Shirky, Clay. *Here Comes Everybody: The Power of Organizing Without Organizations*. Penguin Press HC, The, 2008.

Solove, Daniel J. *The Digital Person*. NYU Press, 2004.

---. *Understanding Privacy*. Harvard University Press, 2008.

Starr, Paul. *The Creation of the Media*. Basic Books, 2004.

Warren, Samuel D., and Louis D. Brandeis. "The Right to Privacy." *Harvard Law Review* IV.5 (1890).

Williamson, Judith. *Decoding Advertisements: Ideology and Meaning in Advertising*. Marion Boyars, London & NY, 1978.

Zittrain, Jonathan. *The Future of the Internet--And How to Stop It*. New Haven: Yale University Press, 2008.

Zureik, Elia. "Theorizing surveillance: the Workplace and Beyond." *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination* (2003).

Appendix A: Figures

within AdWords than “beautiful.”

5.10. AdWords topic exclusion tool – 52 – The AdWords topic exclusion tool allows advertisers to specifically exclude hard news topics including war and famine.

5.11. Google search for “famine” – 53 – The Google search for “famine” shows that “famine” is a salable word within AdWords.

5.12. Search Based Keyword Tool results for nytimes.com – 55 – Google's Search Based Keyword Tool says that the most profitable keywords for the New York Times to advertise itself under are all for specific consumer electronics.

5.13. Search Based Keyword Tool results for “obama” – 56 – Google's Search Based Keyword Tool says that the most profitable keywords associated with “obama” are mostly for t-shirts and bumper stickers.

5.14. Google search for “obama tax plan” – 57 – The AdWords ads associated with a Google search for “obama tax plan” show that even of the seven ads associated with this substantive search, four link off to scams of some sort.

5.15. AdWords ad on New York Times story about the 2008 economic crisis – 58 – The position of the AdWords ads under the content on the New York Times page allows the content of the story to set the social context for the ads.