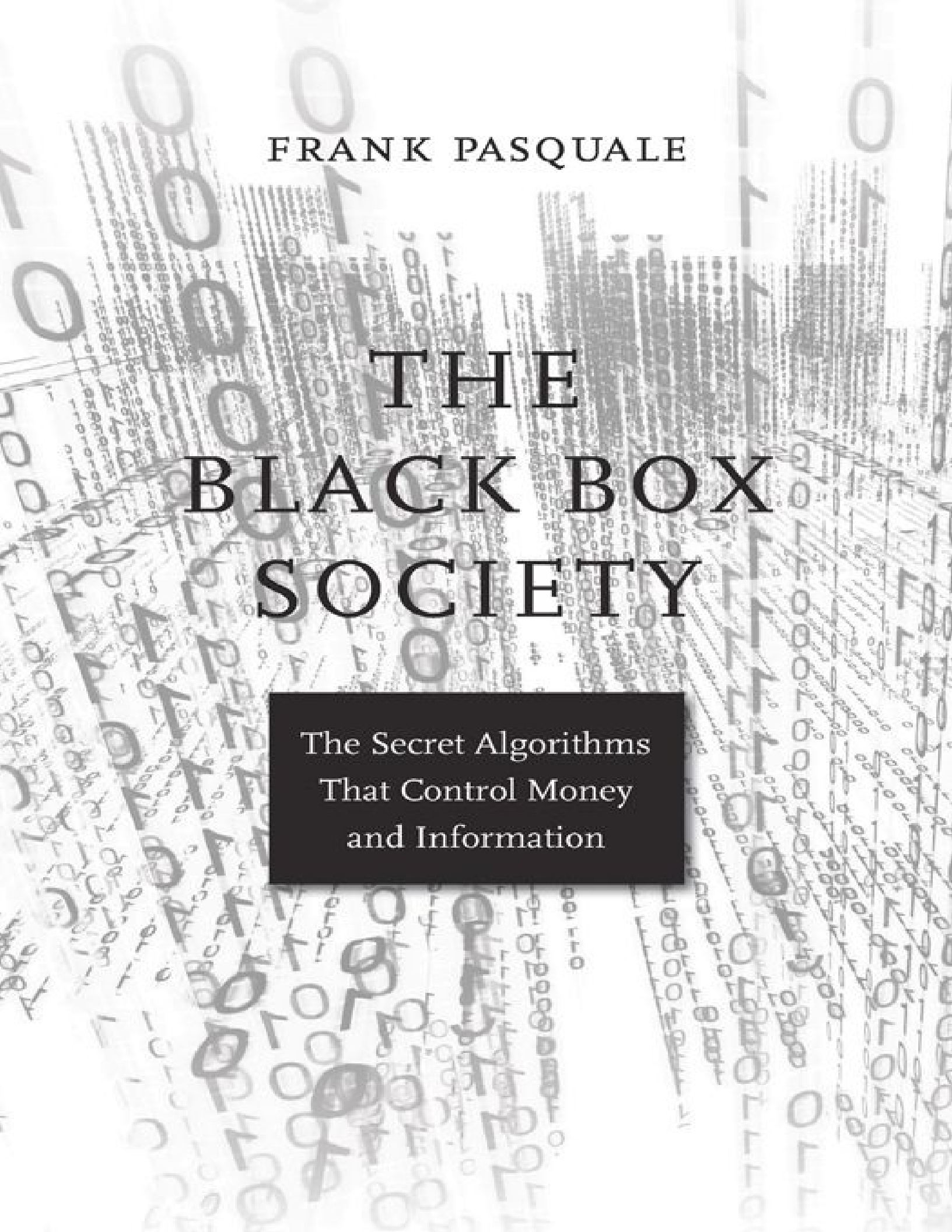
FRANK PASQUALE
THE BLACK BOX SOCIETY
The Secret Algorithms That Control Money and Information

3

THE HIDDEN LOGICS OF SEARCH

Search, in the view of economic sociologist David Stark, is "the watchword of the information age."[1] Though most people associate the "search space" with Google, search is a far more general concept. Whether looking for information or entertainment, products or soulmates, we are relying more on dynamic searches than on stable sources. Search pervasively affects our view of the Internet and, increasingly, of "real life."[2]

Search engines host billions of queries per day. They "answer" more and more of them without the asker ever having to click through to another site. They keep track of our friends, real and virtual. They find our entertainment. They rank and rate everything for us, from movies to doctors to hotels. Search engines can be general, specialized, or social.[3] There are mammoth ones and tiny ones, public ones and encrypted ones, and the array is becoming more varied and more important as content offerings proliferate.[4]

These new masters of media are more than just conveniences. Thanks both to their competence and our inertia, they often determine what possibilities reach our awareness at all.[5] They are guides; they influence, sometimes quite profoundly, our decisions about what we do and think and buy (and what we don't). They are revolutionaries; Apple's and Amazon's portals have definitively reshaped commerce.[6] They are our agents: search for and "friend" a few dozen people on Facebook or follow them on Twitter, and the platforms deliver up a steady stream of content.

Search is a leveler. It lets us, the scrutinized, turn the tables and check out everyone else. It is our entrée to the pool of reputational data to which we all willy-nilly contribute, and at its best it lets us keep tabs on the "digital selves" that so often stand in for us at fateful junctures with bosses, bankers, and other decision makers.

Search gives anyone with a computer or a nearby public library access to resources that were once out of reach of all but the very few with unlimited funds and leisure time. It has the power to give each of us a perfect little world of our own, a world tailored so exquisitely to our individual interests and preferences that it is different from the world as seen by anyone else.

But like everything else in the digital age, search has a dark side, and that dark side has to do with trust. How does a platform decide on the coverage given a third-party mayoral candidate? Or how long to let a meme like Obama's leaden debate performance or Romney's 47 percent speech dominate campaign coverage? New media giants can tame information overload by personalizing coverage for us.[7] But how do those neat and compact presentations of a messy and sprawling world occur? Was a story selected for its statistical prominence among news organs, or because a personalization

algorithm picked it out for us? If the selection was based on statistics, then *which* statistics—the number of mentions of the story, the authority of the news outlets promoting it, or something else entirely?

Businesses large and small worry over such matters daily. Hotels appear to be paying more or less stealthily for premium placement on Google's map and travel services.[8] How can we know whether news outlets or political campaigns are engaged in subtler manipulations, like routing readers and volunteers to Google+ to increase their salience in Google Search? At least with a dead-tree newspaper we know that everybody looking at it sees the same thing, and there are editors to write to when something doesn't smell right. But the decisions at the Googleplex are made behind closed doors or, as we'll see, within black boxes. How far can we trust the people who make them?

The power to include, exclude, and rank is the power to ensure which public impressions become permanent and which remain fleeting.[9] That is why search services, social and not, are "must-have" properties for advertisers as well as users. As such, they have made very deep inroads indeed into the sphere of cultural, economic, and political influence that was once dominated by broadcast networks, radio stations, and newspapers. But their dominance is so complete, and their technology so complex, that they have escaped pressures for transparency and accountability that kept traditional media answerable to the public.

There's a lot that we don't know about these services to which we hand over so much of our lives.[10] Despite their claims of objectivity and neutrality, they are constantly making value-laden, controversial decisions. They help create the world they claim to merely "show" us. I will explore four areas in which the behavior of the great search companies raises pressing issues of trust: transparency, competition, compensation, and control.

Search and Transparency

"Better user experience" is the reason the major Internet companies give for almost everything they do. But surely their interests must conflict with ours sometimes—and then what?[11] Disputes over bias and abuse of power have embroiled most of the important Internet platforms, despite the aura of neutrality they cultivate so carefully. It would be reassuring to have clear answers about when conflicts happen and how they're handled. But the huge companies resist meaningful disclosure, and hide important decisions behind technology, and boilerplate contracts. What happens, happens out of our sight.[12]

Sex and Politics in the Apple Store. Apple remade the world of online music by designing a simple interface, cutting a Gordian knot of copyright conflicts, and providing instant access.[13] iTunes, iPod, and iPad unleashed a whole new ecosystem of music options and compensation.[14] The power of a well-maintained and popular platform like that is enormous.[15] Common standards let people share, cooperate, and play. As Amar Bhidé, finance expert and professor at Tufts University's Fletcher School of Law and Diplomacy, has put it, those "innovations that sustain modern prosperity … are developed and used through a massively multiplayer, multilevel, and multiperiod game."[16]

But the rules of Apple's game can be pretty ambiguous. The company's business practices are notoriously secretive—so much so that legal scholars like Jonathan Zittrain and Tim Wu have worried that too much central control might be constraining the creativity of app developers.[17] More to my own point, users have sometimes had occasion to worry that all that invisible control is constraining *us,* as when Apple excludes popular programs from its app store, or prevents them from running on its products. Here are three disconcerting cases.

EUCALYPTUS. In 2012, developers were submitting about 10,000 apps per week to Apple. Quite a lot featured sexual subject matter. Apple's response has been pragmatic and efficient: an antiporn policy that purportedly reflects user demand and deflects spam.[18] The policy also allows Apple to process the flood of new apps efficiently.

But although the "objectionable content" guidelines at Apple are well publicized, the way they are applied is not. Take the veto of an app called Eucalyptus, which was intended for formatting and downloading public domain texts. Apple rejected Eucalyptus on the grounds that it could be used to access "a Victorian-era, text-only version of the *Kama Sutra.*"[19] Yet Apple had previously approved apps that do precisely the same thing, and the *Kama Sutra* could be found on Apple's own Safari browser in illustrated (including some truly pornographic) editions. Until *Ars Technica*'s Chris Foresman highlighted this absurdity in a scathing column, Eucalyptus's creator knocked in vain against a "mysterious black box." Press coverage finally spurred Apple into action, and Eucalyptus's fate was reversed by higher-ups.[20]

In this case, a well-placed story provoked corrective action and a quick apology. But how many apps never attract the attention of journalists? We don't know. There's no census of app developers to

poll, and Apple's not telling.

DRONES +. Eucalyptus seems to have been a victim of incompetent or arbitrary decision making.[21] Other rejections look less benign. NYU graduate student John Begley developed Drones+ as U.S. drone warfare expanded. It aggregates news stories on drone targets, maps them, and delivers a pop-up notification whenever a new strike is reported. Begley included the real-time alerts to help users keep track of an underreported military initiative.[22]

Apple rejected Drones+ twice. The first reason given was that it was "not useful."[23] (Apple has, however, approved an app that does nothing but display a flame on the screen.) A second rejection letter called the app's content "objectionable and crude," a violation of the App Store Review Guidelines. But the content of Begley's app was news stories, quoted and plotted on a map.[24] Apple has approved plenty of apps that describe *and depict* the destruction reported in the news, so that rationale is hard to swallow.[25] Despite national publicity criticizing the decision, Apple held firm for two years.[26] After five rejections, Begley finally got the app included in the store in 2014 by removing the word "drone" from its name and description, rechristening it Metadata+.[27] Whether those interested in tracking drone strikes can find his app without its using the term "drone" is anyone's guess.

IN A PERMANENT SAVE STATE. Artist Benjamin Poynter submitted his In a Permanent Save State as a "persuasive gaming" app, a form of combined entertainment, provocation, and instruction.[28] It offered an interactive narrative inspired by the suicides of workers at Apple supplier Foxconn's plant, which had taken an enormous public relations toll on Apple the year before.[29] Poynter intended Permanent Save State to highlight the dark contrast between Apple's dream machines and nightmarish conditions in its supply chain.

Apple did not say why it removed the app shortly after it first appeared. It might have been Guideline 16.1, the catchall ban on "objectionable content," or 15.3, which forbids depictions of "a real government or corporation, or any other real entity." Or the topic might have just menaced the company's famous "reality distortion field."[30] Political speech is especially protected under the First Amendment, but Apple isn't bound by the Bill of Rights.[31]

Zittrain anticipated opportunistic behavior like this in his 2008 book *The Future of the Internet—And How to Stop It*. His work is a complex and nuanced call for technology companies to reflect public values in their decisions about what apps to make accessible. Technology scholar Rob Frieden has gone further, challenging the need for app approval at all. When we buy desktop computers we don't have to "phone home" for the manufacturer's permission before we can run a program on it.[32] Why does Apple insist on such control? Wouldn't free access to apps work better?[33]

In Apple's defense, some control may be necessary to ensure the smooth operation of their phones. Buggy, slow, or spammy apps do hurt its customer base. But Drones+? Since it clearly provides information that people want, why should Apple care? At the very least, it could tell users clearly which apps have been rejected, and why.[34]

Google as the "Universal" Index. Google is perhaps the most instructive case of how the black box culture developed, and why it matters. Before Google, web navigation for consumers often meant cluttered portals, garish ads, and spam galore. Google took over the field by delivering clear, clean, and relevant results in fractions of a second. Even Silicon Valley skeptics credit Google with bringing order to chaos. For the skilled searcher, Google is a godsend, a dynamic Alexandrian Library of digital content. But commercial success has given the company almost inconceivable power, not least over what we find online.[35]

Google does not reveal the details of its ranking methods. It has explained their broad outlines, and the process sounds reassuringly straightforward. It rates sites on *relevance* and on *importance.* The more web pages link to a given page, the more authoritative Google deems it. (For those who need to connect to a page but don't want to promote it, Google promises not to count links that include a "rel:nofollow" tag.) The voting is weighted; web pages that are themselves linked to by many other pages have more authority than unconnected ones. This is the core of the patented "PageRank" method behind Google's success.[36] PageRank's hybrid of egalitarianism (anyone can link) and elitism (some links count more than others) both reflected and inspired powerful modes of ordering web content.[37]

It also caused new problems. The more Google revealed about its ranking algorithms, the easier it was to manipulate them.[38] Thus began the endless cat-and-mouse game of "search engine optimization," and with it the rush to methodological secrecy that makes search the black box business that it is. The original PageRank patent, open for all to see, clandestinely accumulated a thick crust of tweaks and adjustments intended to combat web baddies: the "link farms" (sites that link to other sites only to goose their Google rankings), the "splogs" (spam blogs, which farm links in the more dynamic weblog format); and the "content farms" (which rapidly and clumsily aggregate content based on trending Google searches, so as to appear at the top of search engine result pages, or SERPs). Beneath the façade of sleek interfaces and neatly ordered results, guerrilla war simmers between the search engineers and the spammers.[39]

The war with legitimate content providers is just as real, if colder. Search engine optimizers parse speeches from Google the way Kremlinologists used to pore over the communiqués of Soviet premiers, looking for ways to improve their showing without provoking the "Google Death Penalty" that de-indexes sites caught gaming the system. And just as wartime gives governments reasons (and excuses) to hide their plans from the public, Google has used the endless battle against spam and manipulation to justify its refusal to account for controversial ranking decisions.[40]

Google is an ambitious company. Its stated goal, as cultural theorist Siva Vaidhyanathan noted in his thoughtful 2010 book *The Googlization of Everything,* is to "organize the world's information."[41] But faced with shareholder demands for ever-rising profits, it is also angling for new sources of growth.[42] It is positioning Google Books and Google Shopping to rival Amazon and eBay as marketplaces. It has made YouTube a critical hub in the entertainment industry. To shake up travel, Google acquired Zagat, the famed restaurant reviewer, and Waze, a leading traffic app.[43] As of 2013, it has been acquiring at least a company a month, often in spaces adjacent to its core search business.[44]

Many welcome this expansiveness. Google brings user-friendly design and scale to areas that sorely need them—in its free Gmail and map services, for example. But it also gives cause for concern about what Google's immensity means, both for us as searchers and for the economy at large.

Google, for instance, has become a double-edged sword as web organizer and archivist.[45] Yes, its index dwarfs anyone else's. But that is precisely why it can no longer be relied upon as the "indexer of last resort." Virtually any needle can be "disappeared" into a haystack of that size; it is just too easy for the company to hide content it would rather we didn't see. Furthermore, pressing questions have arisen about whether Google is using its dominance in general purpose search to leverage undue power elsewhere. It cloaks its answers in layers of bureaucratic, technical, and contractual obscurity.

We pay no money for Google's services. But *someone* pays for its thousands of engineers, and that someone is advertisers. Nearly all the company's revenue comes from marketers eager to reach the targeted audiences that Google delivers so abundantly. We pay with our attention and with our data, the raw material of marketing. (You are not Google's client, Senator Al Franken once warned users of the World Wide Web. "You are its product."[46]) Sometimes we invest time and effort in a Google service (like arranging blog feeds in Google Reader), only to find the plug pulled abruptly when it isn't profitable enough.[47] We also pay in our ignorance of how the company operates, how it guides us through the web, and how it uses the data it collects on our activities there.

Secret algorithmic rules for organizing information, and wars against those who would defeat them, exist at Facebook and Twitter, too. Apple and Amazon have their own opaque technologies, leaving users in the dark as to exactly why an app, story, or book is featured at a particular time or in a particular place. The secrecy is understandable as a business strategy, but it devastates our ability to understand the social world Silicon Valley is creating.[48] Moreover, behind the technical inscrutability, there's plenty of room for opportunistic, exploitative, and just plain careless conduct to hide.

Search, Transparency and Fairness. We trust our search engines to play straight with us: to show us what's there; to put the best suggestions on top so that we don't have to click through thousands of pages to find them; and to rank by relevance unless they tell us otherwise. But do they?

Foundem is a UK-based firm that provides specialized "vertical search" for price comparisons. It is run by a team of husband and wife engineers with formidable CVs and a track record of innovation. Leading consumer and technology organs in the UK ranked Foundem extremely high in comparative studies of its niche.

But Foundem has not been able to convert this critical acclaim into a mass user base, and it blames Google. Less than six months after Foundem launched, Google appeared to block it from the front pages of its *organic* (that is, unpaid) search results when users queried for price comparisons.[49] The reason, according to Google, was that Foundem was a "low-quality" site, composed mainly of links to other sites. Downranking it could have been a direct result of Google's algorithmic procedure for protecting users from spammers and link farms.

But sometimes there's a legitimate reason for a site to sample other sites—in fact, that's exactly

what search engines do, including Google. Google acknowledges this. So, it says, it distinguishes among such sites by downgrading any whose guesses about what a searcher wants are inferior to its own. But, it says, it allows *good* finding tools to make it into the top search results.[50]

Foundem favors another explanation. If Google has no interest in an area, it will let an upstart be. But once it enters (or plans to enter) the market of a smaller finding service, it downranks that service to assure the prominence of its own offerings. (Major incumbents are not displaced lest their users revolt, so they usually retain their access to prime real estate.)

If the smaller engine is a potential acquisition target, Google has another interest in suppressing traffic: to discourage its hope of succeeding independently. Like Pharaoh trying to kill off the baby Moses, it denies its rival the chance to scale.[51] When a would-be purchaser controls significant access to its target's potential customer base, overtures of interest are offers that can't be refused.[52]

The downranking of Foundem drastically reduced its visibility in Google's unpaid results. When the company tried to reach users with ads, Google cut off that option too. Foundem had been bidding five pence to participate in Google ad auctions, but now Google required a minimum bid of five *pounds*. This made the cost of advertising so prohibitive that, according to Foundem, for more than a year it was effectively eliminated from the view of those searching Google for price comparison websites.[53]

In September 2007, Google relented, "whitelisting" Foundem in its *paid* search results, and lifting its penalty. But the exclusion from *organic* search persisted until the tech press began to cover the story. Finally, in December of that year, Google "manually whitelisted" Foundem, assuring its owners that the algorithms that had branded Foundem as useless or spammy web junk would no longer act to penalize (and thus hide) the site.[54]

Google insists that "the system worked" with respect to Foundem; its algorithms for detecting low-quality sites had hurt it for a while, but eventually human intervention addressed the problem.[55] As Google's engineers like to say, "Search is hard." Evaluation and ranking protocols are as potentially controversial in search as they are anywhere else, and when controversies arise, users can't expect instantaneous resolutions.

But for Foundem and its supporters in the tech press, it's more sinister than that. Google must meet Wall Street's expectations and has demanding shareholders. They expect it to grow, and to do so it must expand. It has: with e-mail (Gmail), video (YouTube), social networking (Orkut and Google+), a blog platform (Blogger), and various specialized search technologies such as Image Search and Google News. Now it is venturing into the realms of shopping, travel, advice, reviews, and price comparisons.[56] Who will Google's system "work" for next? As Metafilter has found, a rapid decline in Google traffic can be a devastating and mysterious blow to even a well-known, well-respected site.[57]

Google counters that it is under no obligation to help other companies eat into its revenue. Its antitrust lawyers insist that what may look from the outside like self-serving bias is just a consistent commitment to customer service. If engineers know that Google Product Search works, why should they expend time and effort in due diligence on every untested alternative? YouTube has dedicated

staff and an active user community that root out spam, porn, and other undesirable material. Is an upstart video service likely to be as well run as Google's own? The company frames its inexorable advance from text search into image, video, and who knows what next as a public service. That is one reason American courts have been so forgiving of Google in considering the copyright complaints against it; it has been seen as a benevolent force for order on the web.[58]

(The situation also highlights the limits of economic analysis. If competition law authorities decide to protect specialized services from domination by a general purpose search behemoth, they are effectively delineating a specialized market.[59] Their decision is not a *reflection* of market forces, but an engine shaping them.[60] The same can be said of the search engines themselves. Left to their own devices, they create the online marketplace at the same time that they participate in it.[61] There is no neutral ground here: the state either takes steps to protect the upstarts, or allows the giant platforms to swallow them. Like banks that, if allowed to grow too large, can effectively control commerce thanks to their power over its financing, massive internet platforms can similarly dominate because of their power over finding.)

Google's dominance is recognized in Europe, too, but differently. EU antitrust authorities recognize that Google is not really a competitor in numerous markets, but instead serves as a hub and kingmaker setting the terms of competition for others. To settle a longstanding antitrust investigation (requested by Foundem, among others), Google as of mid-2013 had offered to guarantee a place on its results page for at least three rival services whenever it offered a service of its own in response to a query.[62] This is a stark contrast with American antitrust authorities' minimalist approach.[63]

Was Foundem's exclusion really a side effect of Google's effort to protect searchers from spammy sites? Or was it an attempt to undermine a nascent competitor? The results are susceptible to either interpretation, but Google's "quality scoring" algorithms are so thoroughly black-boxed that we can't know which is correct. More on Google and competition shortly.

Search, Transparency, and "Murketing." "Stealth marketing" is another area of collision between search and trust. Like broadcast networks, search engines survive by offering unpaid content (in this case, *organic* search results) to sell advertising (*paid* search results).[64] As search engines developed, most of them placed ads at the top and sides of result pages, but used the center for rankings that were free of commercial influence.

American law has long required the separation of editorial and paid content.[65] At first, Google honored those requirements in spirit as well as in letter. When it was just one of many search engines scrambling for market share, this was not only wise compliance but also good business. Google's transparency about advertising delivered high quality results and gained trust.[66] Early search leaders who succumbed to the siren song of ad-disguising drove their users away with irrelevant links while Google's audience grew. As more people signed into its system, Google learned more about them and became ever better at tailoring its search results.[67] Its ad income increased as its targeting improved. This triumph of "Don't Be Evil" is still a celebrated Silicon Valley success story. Patiently gathering data, the company entrenched its privileged position between advertisers, content providers, and

audiences.[68]

But in 2012, as it moved from general purpose search into specialized fields like shopping, Google began to back away from strong separation of paid and editorial material.[69] The Federal Trade Commission strongly encourages search engines to label sponsored content,[70] and has reserved the right to file suit for unfair and deceptive practices against any search engine that fails to do so. Yet it has never actually filed such a suit. This passivity has emboldened small Internet players, and now Google itself, to weaken some of the visual distinction between paid and unpaid content.[71] Accordingly, it becomes harder to discern whether the inclusion, say, of a given hotel or florist shop in a page of search results reflects its quality or its willingness to pay for visibility.[72] And the secretiveness of Google's search ranking processes doesn't help. Even Danny Sullivan, a Silicon Valley journalist who has defended Google from many critics, was disappointed in the shift:

> For two years in a row now, Google has gone back on major promises it made about search.… In the past, Google might have explained such shifts in an attempt to maintain user trust. Now, Google either assumes it has so much user trust that explanations aren't necessary. Or, the lack of accountability might be due to its "fuzzy management" structure where no one seems in charge of the search engine.[73]

And Google is not alone in arousing watchdogs. Blogs constantly speculate about what it might take to get one of the 500,000 or so apps in Apple's store to stand out. Paid-content issues also dog those seeking attention via Facebook.[74] Facebook doesn't disclose the "EdgeRank" methods it uses to sort the items in a user's news feed into the stream of links, pictures, and information from friends that makes the site so addictive.[75] But in 2012, it offered users a chance to pay to promote certain posts. Confusion and resentment ensued almost immediately, as some nonpayers noticed their sudden obscurity and interpreted it as Facebook's way of forcing them to pony up. Without knowing exactly how EdgeRank works, it is very difficult to assess how much substance there might have been in that particular concern.[76] But anyone with a critical mass of friends can see how unwieldly Facebook's "News Feed" has become: how hard it is, say, to be sure you see all your friends' posts, even when you choose to see "Most Recent" posts rather than "Top Stories." Facebook is increasingly a kingmaker for "digital content providers," but it's entirely unclear how it's choosing which sites to promote and which to doom to obscurity.

This confusion may be to Google or Facebook's advantage, but it is not to ours. Blending paid and editorial content creates a confusing world of "murketing" (murky marketing tactics).[77] Google founders Sergey Brin and Larry Page admitted in 1998 their expectation that "advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers."[78]

This situation comes up reliably enough in the communications context that there is a long-standing solution: require *both conduits and content providers* to disclose whether they are raising the profile of those who pay them.[79] Consumers and competitors alike suffer when sub rosa payments are

permitted. Money confers an enormous advantage in the battle for mindshare, and fairness requires—at the very least—that when advantage has been bought, it be disclosed.[80] The question now is whether regulators will adopt and enforce classic rules in a digital age, or let them wither into desuetude.

Search, Transparency, and Judgment. More complex trust issues come up in the ways that the Silicon Valley behemoths handle other disconcerting search surprises.

Google came under fire in 2012, for example, in an awkward situation regarding a prominent German woman, Bettina Wulff.[81] Users who typed her name into the search box were likely to see "bettina wulff prostituierte" and "bettina wulff escort" appear in the "autocomplete" list underneath. Those phrases reflected unfounded rumors about Wulff, who has had to obtain more than thirty cease-and-desist orders in Germany against bloggers and journalists who mischaracterized her past salaciously. Wulff feared that users would interpret the autocompletes (which Google offers as a convenience to users) as a judgment on her character rather than as an artifact of her prolonged and victorious legal battles against slanderers.[82]

Google's help pages say that the autocompletes are "algorithmically determined" and usually reflect "the search activity of users and the content of web pages indexed by Google."[83] The company maintained that Bettina Wulff's wrongful association complaint was none of its business—that it is the obligation of users to appraise the validity of what they read. Yet Google's own behavior refutes that position. The company is not generally indifferent to what its users think; on the contrary, it is constantly trying to educate us, to discern our intent, to give us "the right answer" in ever more contexts. It even corrects our spelling. Type in "lock ness monster" and we see the results for "*loch* ness monster," along with a small offer to "Search instead for 'lock ness monster' " underneath. Google makes at least *some* provisional judgments about what searchers are looking for. Given its interpretive activism about misspellings, one might think that it would lend a hand to a person defamed online, or otherwise dogged by unrepresentative and demeaning material.[84]

Not only autosuggestions, but also search results, can seem inappropriate or unfair. Consider what happened when politician Rick Santorum irked activist Dan Savage. Santorum had compared gay marriage to bestiality, and Savage led an outraged network of bloggers in retaliation. They linked so enthusiastically to a site associating *santorum* with anal sex that soon that site was the first result for most Google searches on the candidate's name. The online comeuppance of the ultraconservative candidate delighted many. Santorum supporters complained to Google to no avail. Only after he made a surprisingly strong showing in three GOP primaries in early 2012 did the anal sex association fade from the very top of the search results.[85]

In its public statements about such controversies, Google mostly characterized them as a reflection of the zeitgeist. Its defenders worried that Google would be "opening the floodgates" to political lobbying if it were to override its search algorithms in Santorum's favor. Google itself pointed out its great efficiency and speed are due to its automated search process; to call in human reviewers would likely slow response times. (An outsider might be forgiven for wondering whether it might also

depress profit margins.) Above all, Google said, an override would contradict the culture of the company, which was committed to organizing and presenting information based on math, rules, and facts, not on opinion, values, or judgment.[86]

But Google has surrendered its "objectivity" position from time to time.[87] In 2004, anti-Semites boosted a Holocaust-denial site called "Jewwatch" into the top ten results for the query "Jew."[88] (Ironically, some of those horrified by the site may have helped by linking to it in order to criticize it; PageRank by and large looks only to linking itself, and not the reasons behind it, to determine a site's prominence.)[89] The Anti-Defamation League complained. Google added a headline at the top of the page entitled "An explanation of our search results."[90] A web page linked to the headline explained why the offensive site appeared so high in the relevant rankings, thereby distancing Google from the results.[91] It might want to consider doing the same at YouTube, where (according to a noted author) watching a few videos of old speeches on the Federal Reserve can quickly provoke a rabbit hole of anti-Semitic "suggested videos" on financial conspiracy theories.

There are principled grounds for a large Internet firm like Google to leave the Santorum results alone, while aggressively intervening to stop the spread of virulent discrimination. But we need to know more about how such decisions are made, given the power of large Internet firms, and the much harder issues on the horizon. A psychologist has conducted experiments suggesting that a "dominant search engine could alter perceptions of candidates in close elections."[92] Jonathan Zittrain spells out how known technology at a dominant social network could have an even more insidious effect:

> Consider a hypothetical, hotly contested future election. Suppose that Mark Zuckerberg personally favors whichever candidate you don't like. He arranges for a voting prompt to appear within the newsfeeds of tens of millions of active Facebook users.... Zuckerberg makes use of the fact that Facebook "likes" can predict political views and party affiliation, even beyond the many users who proudly advertise those affiliations directly. With that knowledge, our hypothetical Zuck chooses not to spice the feeds of users unsympathetic to his views.[93]

When Facebook tried the "vote prompt" in 2010, 0.39 percent more users notified by it voted—well more than enough to swing the outcome in contests like the 2000 U.S. presidential election. Note that Facebook is neither obliged by current law, nor by its terms of service, to announce such interventions.

Are tech titans' political preferences skewed enough to make such a plot tempting? Many Republicans have complained that Google[94] skews search results to mock or marginalize the right;[95] columnist Michelle Malkin charged that websites like hers weren't appearing in Google News results.[96] Later, after George W. Bush and Barack Obama were both subjected to "google bombs"[97] that linked their names to the words "miserable failure," Fox News reported conservative discontent that the manipulation involving Obama was resolved quickly, but it took Google almost four years to address the issue with respect to Bush.[98] Certainly its responses in these varied cases don't present a

picture of a clear policy.

Moreover, Google *did* defuse the Bush and Obama g-bombs, although at different speeds. Why did they rate an override and Santorum didn't? Did the company learn enough from the response to the Bush prank to somehow respond faster when it was Obama's turn?[99] Did the difference reflect more years of practical experience? A new policy? Political views? We don't know. It's an odd thing to trust a search engine so much when we have no way of ascertaining whether or not it acts on a political agenda, or to what extent it will allow clear manipulation to go unchallenged.

Limited "rights of reply" would constitute one way of adding information to a digital platform; annotations could be permitted in certain instances of express or implied defamation, for example.[100] Google continues to maintain that it doesn't want human judgment blurring the autonomy of its algorithms. But even spelling suggestions depend on human judgment, and in fact Google developed that feature not only by means of algorithms, but also through a painstaking, iterative interplay between computer science experts and human beta testers who report on their satisfaction with various results configurations.[101] It's true that the policy for alternative spellings can be applied generally and automatically once the testing is over, while every situation like Wulff's or Santorum's would require a fresh independent judgment. Perhaps Google fears that reputational micromanagers would overwhelm it with requests. But would it really be so hard for the search engine to turn off autocomplete when it's causing unnecessary harm?

Google's repeated refusals even to entertain such reform proposals suggest that the companies' executives believe they've found one best way of ordering the web, outside input be damned. That is an ironic stance for a company that once accused critics (in the context of an FTC antitrust investigation) of a naïve, outdated, and overly rigid conception of search results as "ten blue links."[102] Google argued successfully at that time that certain prerogatives of malleability were due a company that has to make rapid and dramatic changes in its "product." Don't those prerogatives come with responsibilities, too?[103]

Unfortunately, technology firms tend to resist accountability. Consider how America's leading microblogging platform, Twitter, deflected concerns about its algorithms. Twitter hosts whatever short bursts of content *(tweets)* its users contribute. Their message varies widely: from the banal (@KimKardashian) to the profound (@SorenKQuotes), from networking to gibberish to satire (@KimKierkegaard). It can function as either a broadcaster or a narrowcaster, according to the predilections of individual users. It has also become a crowd-sourced democratic search engine for news and conversation. By putting a hashtag (#) in front of a term, users form an automatic "real-time" community around it; anyone who clicks on the term will see items tweeted about it in the past few seconds, hours, or days.[104]

The hashtag also serves to nominate some terms as "trending"—that is, interesting enough to be recommended *generally* rather than simply to the *followers* who subscribe to one's own tweets.[105] Trending topics are listed on Twitter's Home, Discover, and Search pages. Users tend to understand them as hot, fun, or particularly interesting news, and activists use the Trending Topics lists to assess their success in engaging a mass audience.[106]

In late September of 2011, Occupy Wall Street was starting to gain media attention. But although #OWS and #occupy seemed to be collecting more tweets than other terms on the official Trending Topics list, Twitter didn't show them there. Organizers and sympathizers began to accuse Twitter of overriding its trending topics algorithm to suppress those terms, and therefore of censoring their politically controversial movement.[107] @TheNewDeal (@ identifies a Twitter username) put it bluntly on October 1: "It is Official, @ witter is Censoring #OccupyWallStreet There is No Way in Hell That it is Not the #1 Trending Topic in America."[108]

The response from the company was swift: no censorship was occurring. Sean Garrett, head of communications at Twitter, replied to @TheNewDeal that "Twitter is not blocking #OccupyWallStreet from trending. Trends are based on velocity not popularity." Twitter also pointed to a similar situation in 2010, when people had been complaining that #wikileaks did not appear prominently enough in Trending Topics. At that time, the company explained:

> Twitter Trends are automatically generated by an algorithm that … captures the hottest emerging topics, not just what's most popular. Put another way, Twitter favors novelty over popularity.…
>
> Topics break into the Trends list when the volume of Tweets about that topic at a given moment dramatically increases.… Sometimes, popular terms don't make the Trends list because the velocity of conversation isn't increasing quickly enough, relative to the baseline level of conversation happening on an average day; this is what happened with #wikileaks this week.[109]

The #wikileaks and #occupy controversies died down quickly after Twitter offered these explanations. But when a site called Thunderclap attempted to hold a trending topic in reserve until it could unleash its followers all at once, timing all their tweets for maximum impact, Twitter suspended Thunderclap's access to its API.[110]

Media studies scholar Tarleton Gillespie analyzed the company's position in a widely shared blog post titled "Can an Algorithm Be Wrong?" He observed that "as more and more of our online public discourse takes place on a select set of private content platforms and communication networks, and these providers turn to complex algorithms to manage, curate, and organize these massive collections, there is an important tension emerging between what we expect these algorithms to be, and what they in fact are."[111] For Gillespie, the problem is less one of fair platform practices than of media literacy. People were misunderstanding Trending Topics.[112]

But at what point does a platform have to start taking responsibility for what its algorithms do, and how their results are used? These new technologies affect not only how we are understood, but also how we understand. Shouldn't we know when they're working for us, against us, or for unseen interests with undisclosed motives?

Dizzying shifts in the ways Internet platforms characterize themselves amount to a form of regulatory arbitrage, evading the spirit of classic legal obligations.[113] When faced with copyright and

defamation lawsuits, they claim not to be media companies (that is, producers of content), but only conduits (that is, pipelines for content).[114] A conduit does not enjoy the most robust First Amendment protection, but it gains freedom from liability in cases of defamation.[115] (For example, the phone company can't refuse to serve me on First Amendment grounds, but it also can't be sued by someone I defame using the phone.) Thus Google can argue that the very idea of suing it for its autocompletes is as nonsensical as a lawsuit against the phone company for enabling a slanderer to spread lies over its network.

But in other cases, Google has also maintained that its services *are* content and that it *is* entitled to the media's fullest First Amendment protections, which include not only the right to free expression but also the right not to be forced to express opinions not its own.[116] Expansive interpretations of the First Amendment could leave Google nearly unregulable. Fortunately, there is also plenty of legal doctrine suggesting the limits of opportunistic civil libertarianism.[117]

Search, Transparency, and Personalization. The secret workings of our search engines deeply inform our views of the world. That truth comes as a real shock to many of us. I don't know how often I've heard someone say, "I'm the top Google result for my name!" But if *I* searched for your name, would I see the same thing? Only Google knows, but very likely not. We can only guess at how our Google-mediated worlds differ.

We know that what we do while signed into Google services (like Gmail) will be reflected in our search results. This has been true for a long time. As far back as 2007, Google was investing heavily in customization technology.[118] By late 2009, it had changed its algorithms to deliver "personalized search" to all web users. Our locations, our search histories, our computers—all of these and more influence Google Search results, and therefore our view of the world.[119]

The basic outlines of similar processes are clearer on Facebook and Twitter, where users curate continuously scrolling feeds. But even there, judgments have to be made about what to do with, say, a sudden burst of content from one source, or the flagging of potentially "objectionable" content.

Personalization lets us hide annoying relatives on our Facebook feeds, list our favorite microbloggers, and get updates from crucial RSS feeds. It means that Google News might give pride of place to baseball, music, or left-wing politics according to the reputations we establish. It means that Google Search orders its results according to searches we've made before, through clicks collected by the Google-owned ad network DoubleClick and through activity on other Google-related accounts.

Personalization makes for digital magic. Let's say that you've lost a favorite earring and want to replace it. And that when you first found the pair many years ago, you took a picture of it and sent it in an e-mail to your sister. When you next search Google Images for earrings, you may find an exact match at the very top. You wouldn't know that the critical data point was the picture in your e-mail; you don't even have to remember that there ever *was* a picture or an e-mail at all. This is just what happens when you've got a search engine (as aggressive about data aggregation as Google) attached to your own e-mail account. Multiply that experience by years of people, e-mail, and search—that's

how powerful the dominant platforms really are as artificial intelligence aids for virtually any tasks we undertake.[120] They have unmatched abilities to advance users' data-dependent interests.

But personalization has unnerving effects, too. Google results have become so very particular that it is increasingly difficult to assess how much of any given subject or controversy any of us actually sees. We see what we have trained Google to show us and what Google gradually conditions us to expect. Entrepreneur Eli Pariser calls this phenomenon "the filter bubble" and worries that all this personalization has serious side effects, namely increased insularity and reinforced prejudice.[121] So intense is the personalization of search results, for instance, that when British Petroleum's (BP) massive oil spill was dominating cable news in the summer of 2010, searches for "BP" on Google led some users to fierce denunciations of the company's environmental track record, and others to investment opportunities in the company.[122] Only the search engineers at the Googleplex can reliably track who's seeing what and why. And they are bound by nondisclosure agreements not to tell us.[123]

Personalization means vulnerability as well as power. If a social network knows you love poker, it can prioritize posts about casinos. But it might also get you included on a "sucker's list" of problem gamblers for casino advertisers.[124] The same platforms on which Arab Spring protesters virtually assembled to overthrow corrupt rulers also generate intelligence for autocrats.[125] Data deployed to serve users one moment can be repurposed to disadvantage them the next. In contemporary American policy debates, these concerns are often framed as privacy issues. But they are equally concerns about search. Who are the men behind the curtain, and how are their black boxes sorting and reporting our world?

Shaping it, too. Personalization is critical to both buying and selling, and that is why reputation and search go hand in hand in the digital economy. How we are seen by websites in turn affects the choices they present to us. Businesses want to know how we search precisely so they can shape our view of the marketplace. We shape the marketplace too, in our search for the best prices and the widest choice. Accurately attuned search results attract users, and accurately targeted users attract advertisers. The most lucrative ads are those "narrowcast" on search result pages, because they reach niche audiences who have already volunteered information about what they want.[126] A florist is likely to pay more to advertise to people searching for "roses" than to any random group of computer users.[127] But it's better still when Google can tell it not only how often its searchers query "roses," but also the sites they go to in response. And what goes for Google is increasingly true of Facebook, Twitter, and so on.

As usual, there's danger here. The advantages of this sort of pinpointing are leading advertisers to abandon traditional, and even not-so-traditional, publishers in favor of the huge Internet platforms. Why? Because nobody else can approach either the granularity or the comprehensiveness of their data. The result is a revolution-in-process about who can afford to keep publishing, and concomitant alarm about the concentration of media clout into fewer and fewer hands.

Search, Trust, and Competition

Neoclassical economists envision a direct and positive relationship on the Internet between privacy and competition. If a large online company is abusing its position, market-oriented scholars say, economic forces will solve the problem.[128] Can't find something on Google? There's always Bing. Don't like the new version of iTunes? Subscribe to Rhapsody. Google not private enough? Try Duck-Duck-Go.[129] Users can select for a preferred level of privacy the way car buyers select for miles per gallon.[130] And if they choose services that *don't* provide privacy protection? Well, that just reveals the place of privacy in their priorities.[131]

It would be great if market forces really were directly promoting optimal levels of privacy. It would also be splendid if antitrust law were promoting them indirectly, by assuring that a diverse range of firms could compete to offer them.[132] But the plausibility of these desiderata is fading. Competitive striving can do as much to trample privacy as to protect it.[133] In an era where Big Data is the key to maximizing profit, *every* business has an incentive to be nosy.[134] What the search industry blandly calls "competition" for users and "consent" to data collection looks increasingly like monopoly and coercion.

Silicon Valley is no longer a wide-open realm of opportunity. The start-ups of today may be able to sell their bright ideas to the existing web giants. They may get rich doing so. But they're not likely to become web giants themselves. Silicon Valley promulgates a myth of constant "disruption"; it presents itself as a seething cauldron of creative chaos that leaves even the top-seeded players always at risk. But the truth of the great Internet firms is closer to the oligopolistic dominance of AT&T, Verizon, and Comcast.

In 2008, I testified before a congressional committee about Google's market power. Just about every representative who questioned me assumed that a clique of twenty-somethings could at that very moment be developing an alternative. They didn't know much about the Internet, but they knew that Larry Page and Sergey Brin had risen from grad students to billionaires by building a corporate colossus out of old servers and ingenuity. In their imaginations, Google's own rags-to-riches story foreshadowed its eventual displacement.[135] Even law professors who ought to know better buy into this myth. "No one's even going to care about Google in five years!" one heatedly told me. That was six years ago. Too many still believe that the digital economy is by its nature open, competitive, and subject to the disruption that it preaches for other fields.[136]

But how realistic is this? The electricity consumption of Google's data centers rivals that of Salt Lake City.[137] Technology historian and journalist Randall Stross estimated in 2008 that the company uses close to a million computers to index and map the web.[138] If he is within even an order of magnitude of the real number (a strictly protected trade secret), it's pretty hard to imagine how an alternative could be brewing in somebody's garage. Even with millions in venture capital funding, even with computing space leased from Amazon, a start-up with valuable new search technology is far, far more likely to be bought up by Google than to displace it.[139]

Well, then, maybe another giant could take Google on? So far Microsoft is losing $2.6 billion a

year on Bing.[140] The government? They tried that in Europe, but the Quaero project sputtered out, perhaps because the $450 million or so allocated to it could not compete with Google's $100 billion in annual revenue. Anyway, it's a virtual certainty that any other Goliath that could seriously squeeze Google has its own secretive and restrictive black box carapace.[141]

It's not only prohibitive infrastructure costs that keep competitors from emerging in the general search space. Innovation in search depends on access to a user base that "trains" algorithms to be more responsive.[142] But the user base belongs to Google. Innovation in analysis depends on access to large quantities of data. But the data belongs to Google. And Google isn't sharing. As long as Google's search data store remains secret, outside innovation is dead in the water. Robert Merton called this the "Matthew Effect": to those who have much, more is given.[143]

Furthermore, what if someone *did* manage to come up with a terrific alternative? They'd often have to market it through the very channels they wish to displace. If Google, Apple, Amazon, and Facebook really don't want most of their users to see something—a competitor, an alternative, whatever—they are well able to make sure it won't be seen.

Restrictive terms of service are another deterrent.[144] Every user who types in a search query agrees not to copy, modify, distribute, sell, or lease any Google service for *any* reason, or attempt to reverse engineer one.[145] Advertisers have faced other restrictions.[146]

Finally, there's the black box itself. Google's secrecy not only keeps spammers from manipulating its results but also keeps rivals from building upon its methods or even learning from them. Unlike patented procedures, which must be disclosed and whose protection eventually expires, trade secrets need never be revealed, let alone released into the public domain of free reuse.

All of these factors undermine robust competition. Silicon Valley rushes to monetize and control access to information that would better be anonymized and licensed openly as the raw material of future innovation. Quantum leaps in technology sufficient to overcome such disadvantages are unlikely. Search now is as much about personalized service as it is about principles of information organization and retrieval.[147] Many more people use search now than when Google conquered the field in the early 2000s, and they are mostly Google's. So its current advantage is likely to be self-reinforcing.[148] There have been isolated consumer boycotts, but a company so dominant can do without the business of, say, hardcore Santorum supporters. Serious complaints lodged against the company are seldom loud enough to be noticed by ordinary searchers, let alone to provoke sympathy. Users lack both the ability and the incentive to detect manipulation as long as they are getting "good enough" results.

So we're stuck. And again the question arises: With whom? The exciting and radical Internet platforms that used to feel like playmates are looking more like the airlines and cable companies that we love to hate. "Don't Be Evil" is a thing of the past; you can't form a trusting relationship with a black box. Google argues that its vast database of information and queries reveals user intentions and thus makes its search services demonstrably better than those of its whippersnapper rivals. But in doing so, it neutralizes the magic charm it has used for years to fend off regulators. "Competition is one click away," chant the Silicon Valley antitrust lawyers when someone calls out a behemoth firm

for unfair or misleading business practices.[149] It's not so. Alternatives are demonstrably worse, and likely to remain so as long as the dominant firms' self-reinforcing data advantage grows.

Search and Compensation

At the 2013 Governing Algorithms conference at New York University, a data scientist gave a dazzling presentation of how her company maximized ad revenue for its clients. She mapped out information exchanges among networks, advertisers, publishers, and the other stars of the Internet universe, emphasizing how computers are taught by skilled programmers like herself to find unexpected correlations in click-through activity. To some extent the algorithms were machines that would go of themselves, freed from supervision. "That gives me more time to ride my horses," she joked.[150]

Intrigued by the idea of machines learning, one listener asked, "At what point do the algorithms do *your* job?" In other words, when does the computing process itself reach the third level of sophistication and start determining for itself which metrics are the best metrics for measuring past metrics, and recommending further iterations for testing?[151] The presenter brushed off the question. *She* remains indispensable, even as machine learning methods are said to render millions of other jobs obsolete.[152]

Maybe she's right. But to know, we'd need expert access to the interactions between humans and machines in her firm, and we don't have it. So some of us will keep wondering about the extraordinary returns that top CEOs, managers, and investors are deriving from the Big Data economy. Compensation, like competition, raises major legal and moral issues. The first step in approaching them is awareness, especially since the black box aspect of Internet infrastructure has been so notably successful in keeping its economic arrangements out of the public eye.[153]

There are two intertwined issues here. One has to do with concern about appropriate levels of compensation for executives, intermediaries, and investors. These questions do not apply uniquely to search firms; on the contrary, they are very common in other fields. They were central, for instance, in the struggle over the Affordable Care Act, which aimed to keep insurance premiums from being siphoned disproportionately out of health care proper and into insurer profits and CEO compensation. They will come up acutely in the next chapter, on Wall Street. They haunt other corners of the information world—for instance, the cable and telephone companies that benefit along with Silicon Valley firms from the massive increase in traffic engendered by the world of search. These companies have also been accused of capturing an unfair share of revenues. These are not new questions, but it's time to ask them in our new context.

The second issue has to do with appropriate recompense not for search firms and their investors, but for the innumerable contributors to the Internet who make search worthwhile. I will start with the second, and then circle back to the first.

If there were nothing on the net, no one would be looking for it. In their book *Unjust Deserts,* Lew Daly and Gar Alperovitz document the centuries of past endeavor on which today's technical progress rests. The top dogs of Webs 2.0 and 3.0 are enriched as surely by the millions of searchers who improve their services and attract their advertisers as they are by their own ingenuity. They are further enriched by the army of creative people without whom the web would be contentless. And

they are enriched by all the old technologies that contribute to new ones. Without the communication and computing of the nineteenth and twentieth centuries, for example, search would not exist at all. Yet the revenue generated online goes more and more to the masters of search infrastructure, and less and less to support the culture that makes the infrastructure possible and meaningful.[154]

The retail dominance of Walmart offers a cautionary tale here. Walmart grew to be the largest retailer in the United States by attracting consumers and squeezing suppliers. As its customer base expanded, it forced its suppliers to accept ever smaller margins. Consumers had little loyalty to the sources of their shampoo, socks, and dog food; they were pleased to accept Walmart as the place to find ultracheap everything.[155]

Firms like Google and Apple are the Walmarts of the information economy.[156] They aggressively scheme to restrict their own workers' wages.[157] They squeeze content producers (for whom making it on a big platform may mean everything), and habituate users to value the finding service itself over the sources of the things found. The content contributors—the writers, musicians, filmmakers, artists, historians, scholars, photographers, programmers, journalists, activists, cooks, sailors, manufacturers, yoga teachers, knitting gurus, auto mechanics, dog trainers, financial advisers, Lego architects, and muckrakers in quest of whose output people use major internet platforms—may receive no share at all of the revenues that that vast user base occasions. The ones that do are often obliged under contract not to reveal what their share is.[158] That is an ingenious way for the platform to cripple any opportunity for them to unite to organize for better terms.[159]

Even some progressive voices trivialize the value of ordinary Internet users' work and play. When one gadfly called Google out as a parasite extracting value created by others, law professor and digital rights activist Lawrence Lessig answered: "In the same sense you could say that all of the value in the *Mona Lisa* comes from the paint, that Leonardo da Vinci was just a 'parasite' upon the hard work of the paint makers. That statement is true in the sense that but for the paint, there would be no *Mona Lisa.* But it is false if it suggests that da Vinci wasn't responsible for the great value the *Mona Lisa* is."[160]

This is a provocative but very puzzling metaphor. Is Lessig really implying that Google's organization of the web by query does for it what da Vinci did for some pots of paint? That it is not the content, but the *index,* that gives the web meaning? After all, the new economy preaches that "information" is just another commodity. From Google's perspective, content, data, and information are basically 1's and 0's and the ad payouts they generate. But to most of us, the value of a website lies in its meaning, not its salience. And real careers, real incomes, and real achievements are won and lost in the struggle for salience that platforms host daily.

This brings us back to our equestrienne presenter, to the lords of the cloud, and to the question that is really the theme of this chapter. Who are these people and these companies that wield so much power in our lives? What do we owe them? Are they really the Gandalfs of the digital world, wizards selflessly guiding us through digital brambles? Or is it time to reconsider some conventional views about technology, labor, and value in the information economy?[161]

Silicon Valley's top managers are well educated and technically skilled, but they are not great

sages. They hide behind corporate operations so covert that their actual contributions are hard to assess, and it's hard not to wonder whether other firms or other individuals might make more constructive use of their data than they do. If not, why all the secrecy? Certainly they are beneficiaries of what is for them a wonderfully virtuous cycle. Thanks to the ingenuity and luck of company founders, they have acquired an audience. This allows them to offer data-driven targeting to advertisers, with whose handsome payments they can buy content, apps, and other enticements (the fruits of *other* people's ingenuity) that draw a bigger audience still, and so on. The well-realized technological vision that attracts the initial user base deserves recompense. But it does not entitle present corporate leaders to endlessly leverage past success into future dominance. What Thomas Piketty said of unlimited capital accumulation applies as well to untrammeled tech giants: "the past devours the future."[162]

The data advantage of the Silicon Valley giants may owe as much to fortuitous timing as to anything inherent in the firms themselves. Social theorist David Grewal has explained the "network power" of English as a lingua franca*;* it's not "better" than other languages; it's not easier to learn, or any more expressive. It just happened to be the language of an imperial power during an important period of globalization, and that of the world's dominant economic power from 1945 on. So it serves well now as a common standard for the communications of far flung elites. To have been prominent at a critical point in Internet development was a similar piece of luck. Google or Facebook were once in the right place at the right time. It's not clear whether they are still better than anyone else at online data science, or whether their prominence is such that they've become the permanent "default."

We also have to ask whether data *science* is still key here, or just the data itself. When intermediaries like Google and Facebook leverage their enormous databases of personalized information to target advertising, how much value do they add in the process? This is a matter of some dispute. Every so often we see an old-style advertising genius come up with a brilliant angle for introducing a new product to an unfamiliar audience. But that's not what Google and Facebook do. The frenzy of ad-matching described in books like Joseph Turow's *The Daily You* is not a triumph of creative ingenuity.[163] Much depends on a store of personal and demographic information: who has the best list of single white women between 25 and 35; or wealthy, exurban gun-owning households. The matching game may simply rest on a catalog of crude correlations: who has the biggest set of past data on what X group of people (say, fathers under 30) does when Y appears (say, a Mother's Day ad for flowers). Some algorithmic expertise may be needed to infer telling characteristics from the websurfing habits of a particular IP address. But in some ways, the new media giants, for all their glamour, are glorified phone books, connecting message senders with message receivers. They just present businesses with a yellow pages of *people,* organized into audiences.

For all these reasons, it's time to recast the black box search culture as an occasion for skepticism, not for deference, adulation, or more fawning tech press profiles. But even though a more realistic assessment of the relative contributions of the search giants and the content makers, and the diversion of a fairer share of intermediary revenues to the latter, are necessary first steps toward a better online landscape, they are only first steps. There are other reasons to beware the concentration of so much power and money into so few hands, and they are not all economic.[164] They include the

importance of media diversity, of independent gatekeepers, and of "distribution of communicative power and opportunities among private actors."[165] A series of laws passed over the course of the twentieth century ensures some basic ground rules for the communications infrastructure, but the new information environment raises new challenges at every turn.

Consider Google's breathtaking aspiration to scan millions of books, many still under copyright, into a searchable index of unprecedented proportions. Google Book Search has provoked storms of public controversy and private litigation.[166] The plan raises countless questions about fair compensation and transparent organization. The most highly publicized aspect of the debate centers on the rival property rights of Google and the owners of the copyrights of the books it wishes to scan and index.[167] But there are others just as important.

Journalistic narratives largely portray the Book Search project as an unalloyed advance in public access to knowledge, and Google has indeed established alliances with some of the leading libraries of the world. Its 2013 fair use victory also paves the way (in principle) for rival book search engines to arise. But here, again, competition may be illusory: it's hard to see the rationale (or investor or public enthusiasm) for subjecting millions of volumes (many of them delicate) to another round of scanning. Once again, Google reigns by default. The question now is whether its dictatorship will be benign.

Does Google intend Book Search to promote widespread public access, or is it envisioning finely tiered access to content, granted (and withheld) in opaque ways?[168] Will Google grant open access to search results on its platform, so experts in library science and information retrieval can understand (and critique) its orderings of results?[169] Finally, where will the profits go from this immense cooperative project? Will they be distributed fairly among contributors, or will this be another instance in which the aggregator of content captures an unfair share of revenues from well-established dynamics of content digitization? If the Internet is to prosper, *all* who provide content—its critical source of value—must share in the riches now enjoyed mainly by the megafirms that organize it.[170] And to the extent that Google, Amazon, or any other major search engine limits access to an index of books, its archiving projects are suspect, whatever public-spirited slogans it may adduce in defense of them.[171]

Philosopher Iris Murdoch once said, "Man is a creature who makes pictures of himself and then comes to resemble the picture. This is the process which moral philosophy must attempt to describe and analyse."[172] The large Internet firms make pictures of us and our world and enforce the resemblances between them. But they downplay the moral implications of their work, and the legal ones, too. In the next section, I will look back to earlier times when robust regulation was still being brought to bear on these processes.

What if one of the big electric companies bought out Whirlpool and thereafter doubled its electricity rates for anyone using a different brand of refrigerator or washing machine?[173] I imagine there would be mass protest and a slew of lawsuits. The very possibility seems antique, the fever dream of a robber baron. But in the digital realm, monopolistic cable firms are angling to impose a similar arrangement: to make Internet access cheap if paired with their own content, and pricier if used to access others' work. Similarly, firms like Google and Amazon are in prime position to make money off both sides of a two-sided market: monetizing our data and purchases, while promoting to us their own products and services, or those of "partners" who let the larger platform share in their profits.

That's one reason we need to look back to the legal principles that animated Populists and Progressives in response to America's first Gilded Age. The great Internet companies and the physical networks that enable them are not the first private enterprises to achieve near monopolistic power over a key service, and to leverage that power into windfall profits and influence.[174] It happened in the nineteenth century with railroads and telegraphs.[175] Like today's search and cable companies, those firms controlled essential junctions of an emerging economic order. They were private businesses, but they controlled vital resources and enjoyed a power similar to that of a public authority.

Social, political, and legal conflicts arose around the exercise of this power, and demands to restrain it mounted. The most common and important grievances against these companies had to do with "discrimination," meaning both inequitable and unequal treatment of individuals, and complete refusals to serve.[176]

Litigants turned first to the ancient section of the common law that governed bridges, innkeepers, and other common carriers, and developed it into a comprehensive framework for governing the new entities that corporate industrialism had produced.[177] In a second stage, when court-based supervision alone proved insufficient, a statutory and administrative framework for regulation was gradually created. This became the foundation of the modern regulatory system, which over generations has established well-tested guidelines about how essential utilities can use their power.[178]

The telephone company, for instance, cannot oblige a business to pay rising shares of its revenue for service lest it be cut off. Telephone rates (or "tariffs") have to be publicly posted, and are often regulated. Utility firms may not discriminate: universal service rules keep carriers from connecting only to lucrative urban areas and ignoring others. This complex regulatory history profoundly shaped the U.S. communication landscape. The requirement that tariffs be fair and nondiscriminatory balanced the carriers' drive for profit against customers' need to be protected against exclusion or exploitation by a "must-have" service.[179] The requirement that networks include everyone established a level playing field among the different regions of the United States. And there are strict limits on the degree to which these essential companies can use their privileged access to communications for their own commercial advantage.[180]

Every time a new kind of infrastructure becomes critical to everyday life, regulators are

challenged to strike the fairest balance they can between public and private good. It's time to situate the giants of Internet search and networking in this tradition. Time-honored principles underlie the regulatory framework of our other utilities.

Admittedly, these are complex issues. Even if we had a Federal Search Commission, we couldn't just transfer the current Federal Communications Commission Rules over to it.[181] A well-established rubric of accountability like the one for carriers does not yet exist for search technology. But the carriers' rules did not spring forth, fully formed, like Athena from the head of Zeus. They were crafted over decades, and we should commit ourselves to a similar project in the world of search.

One of the most enduring principles of communications regulation has been transparency. That's needed now more than ever. In the instantaneous and fluid world of apps and search engines, it's much harder to tell what actually goes on behind the scenes. Discrimination used to be as simple as flipping a switch and denying access to a network; everybody knew it was happening, and when, and where. But an ISP or search engine can slow down transmission speed or reduce a website's ranking in nearly undetectable ways.[182] Moreover, there are many points of control for both desktop and mobile Internet users.[183] Even when something suspicious is happening, it's easy for one player to shift responsibility to others.

Many communications mavens are ready to throw up their hands at the complexity, and hope that market pressures and bad press will deter bad behavior. But as we have seen, Big Data giants entrench their dominance over time.[184] They gain power in Washington and state capitals, too, and may well influence regulation in self-serving ways. It does not follow, however, that doing nothing is the preferable option. We need to revive regulation, not give up on it. Internet service providers and major platforms alike will be a major part of our informational environment for the foreseeable future. The normative concerns associated with their unique position of power are here to stay. A properly designed regulatory approach may do much to clarify and contain the situation; without one, will deteriorate.

Content, Conduits, and Search: The Emerging Co-opetition

Once upon a time, we could imagine that scrappy Internet firms—Google among them—were doing battle on behalf of their users against old-line oligopolists like the record labels and cable companies. Silicon Valley firms fought for net neutrality and opened up troves of content. Business analysts hoped Google might even expand into "dark fiber" nationwide, to shake up the moribund Internet service market. But as Google has consolidated its own power, it is now more inclined to make common cause with these older giants than to resist them.[185] The implications are sobering. Competition is muted; cooperation accelerates; and the hoped-for dynamism of Internet economics is congealing into a static combination of the two, "co-opetition."[186]

Strange Bedfellows. The lifecycle of YouTube is a relatively straightforward example. Founded by a pair of young entrepreneurs, it grew explosively in the mid-2000s as a cornucopia of unauthorized videos: old films that had been MIA for decades; obscure gems of musical performance; early animations; political speeches. (Cats, too.) Users uploaded millions of hours of their own content, and community members helped each other organize the material, developing a tagging "folksonomy" so clever that searchers could find even the most obscure content.[187]

The sale of YouTube to Google for over a billion dollars in 2006 was cheered as another of the great tech success stories. But YouTube was not universally adored. To many leading copyrightholders, it was an unrepentant enabler of infringement.

The Digital Millennium Copyright Act (DMCA) of 1998, while increasing the penalties for copyright infringement on the Internet, had immunized some providers of online services from direct responsibility for the content posted by their users. YouTube thus maintained that it was as innocent of infringement as, say, the phone company would be if one of its customers played a copyrighted recording over its lines. But the DMCA also suggested that a video search engine did have some responsibility for screening out pirated content. For example, an "information location service" could be liable for *secondary,* if not direct, copyright infringement if it ignored obvious red flags indicating illicit behavior.[188] And so the battle was joined.

Clearly, an account advertising "!!!Bootleg Movie Releases!!!" would be one of those obvious red flags. But what about a music video that is unavailable even to would-be purchasers? Or a three-minute clip from a two-hour film? These are issues that can be extensively litigated, and rulings on "fair uses" of copyrighted material come down on both sides.

Thus major content owners tolerate many questionably legal uses, but try to crack down on users who engage in many unauthorized downloads and uploads.[189] That uneasy truce sparked a business opportunity: a video or music search engine could grab a mass audience, as long most of its users only uploaded a few pieces of infringing content. YouTube grew to prominence on the back of the pirated content of millions of users. But as it consolidated its position as the dominant video search engine online, it began cleaning up its act.[190] It struck deals with major labels and independent artists, sharing ad revenue with them based in part on how many viewers and listeners they attracted. We can

only know "in part" what the revenue share is, because Google keeps the terms of the contracts strictly under wraps.[191] But the basic industrial organization is pretty clear: like cable companies positioning themselves between subscribers and content providers, Google wants YouTube to be a broker, taking its cut of the ad revenue ultimately generated by the content it hosts.[192] And that ambition is reflected in its search results.

The recording industry has been targeting music-sharing sites for years.[193] Infringers pop up at a new address each time an old one is seized, a digital game of whack-a-mole. Content owners complained for years about Google's role in enabling infringement, especially after it bought YouTube. The search giant took its usual position with regard to most complaints: not our problem. Copyright holders could litigate against the offending sites themselves, but Google would not do more work than the DMCA required it to. This did not satisfy the copyright holders, who continued to demand that important search engines address the problem by automating punishment of the worst intellectual property scofflaws.[194]

In 2012, Google creatively capitulated to this demand. A comprehensive search engine makes it a cinch to find pirated materials—unless, of course, the search engine is trying to conceal them. Google decided to do so, agreeing to adjust its algorithm and systematically demote sites that collect multiple complaints of copyright infringement. Google's famously stubborn engineers acceded to Hollywood's demands.[195] Now that it is making serious money from copyrighted content on YouTube, it has an interest in assuring compensation for viewings.[196] It also has a brand (worth tens of billions of dollars, by some Wall Street estimates) and a business model to protect. Copyright-holders brought ad revenue to YouTube; Google had to return the favor with some takedowns of pirate havens and demotions of alleged infringers.[197]

In its public statements, Google denied that demoting sites for copyright infringement was a significant departure from existing policy. Like everything else at Google, it was framed as just another way of making results better.[198] But while it certainly did make for a change in user experience, the change was not, in many users' views, an improvement. Furthermore, Google justice was swift, secret, and arbitrary. Due process did not apply. Once a critical mass of copyright complaints accrued against a site, it just started to sink in the rankings.[199] Google didn't de-index it. But in an information environment where searches often result in thousands of results, being demoted to the ninety-ninth page of listings is tantamount to the same thing. And the demoted site might not even know that it had been demoted. If it looked for itself from its own IP address it might well appear near the top of the results, its own personalized signals for salience having locally overwhelmed the signals for demotion.

Google's draconian antipiracy practices also raised questions about collateral damage. For example, what happens if a site (whose intention is not infringement) accidentally or incidentally posts pirated material and loses prominence for that reason?[200]

Google's decision to serve as enforcer for the holders of intellectual property rights left unanswered many questions that are sure to arise about the laws of its secret "Googlement." But if its behavior in the past is any guide, it will address them behind closed doors. The public won't be privy

to the considerations raised, the monetary interests involved, or the favors cut for one group or another. And as we'll all see the results through our own personalized search lenses, it will be well-nigh impossible for us to notice that a decision was even made, let alone assess the reasons or the effects.[201]

Who Can Afford to Publish? The power of the old media is waning. Traditional journalism is in crisis.[202] Some predict that investigative reporting will be sustainable only as a charity.[203] Broadcast media are in less serious financial trouble, but their political and cultural clout is declining, and their profit margins are threatened.[204] Broadcast radio too is culturally less relevant as younger listeners look online for music.[205]

All of these developments coincide with—and have in part been caused by—the rise of new media, which feature online video, text, and music. Users have abandoned old *sources* of content for new ways of *searching* for it. The huge user bases that result mean that both content providers and advertisers want to seize places at the top of Google's (or Facebook's or Apple's) users' front pages.[206] Not coincidentally, Google's U.S. advertising revenue is now greater than that of *all* newspapers.[207] If current trends continue, it will soon be larger than both newspaper and magazines combined. Current valuations of Facebook suggest it will capture 10 percent of global ad revenue by 2020.

Some web-based publishers feel empowered to use search engines and social networks to build audiences that would never have been possible in the analog world.[208] But others feel that the search intermediaries have done them ill. Microtargeted advertising by companies like Google has taken an ever-increasing share of the revenue that used to be spent directly at their sites. Google's tense relationship with many web-based political publications reveals these trends. In a provocative post titled "Has Google Destroyed the 4th Estate?," prominent progressive blogger Jane Hamsher attributes the decline of the fortunes of sites like hers to Google's rise to preeminence in key advertising markets. A *Washington Post* story confirmed that both Google and AOL played hardball during the election of 2012, negotiating portions of political campaign ad revenue that would have gone directly to sites like Hamsher's Firedoglake in past years. The ad buyers argued that it's not space on paper or pixels on a website that matters to them, but *audiences;* that's what they were looking to buy. In other words, the context of the advertisement was mere background: what really mattered was data on who was looking at the content, and Google had far more of that than anyone else. Google could connect advertisers to a precise demographic, and in an era of campaigns based on Big Data, that secret, proprietary information was the vital edge political campaigns needed.[209]

Though media is suffering now, campaigns themselves should also beware. Saving a bit now by avoiding wasted advertising may lead to huge costs down the road if data holdings further consolidate and become the key to finding undecided voters. The *Citizens United* decision is an open invitation to tech firms to escalate the prices they charge for audiences, as billionaire donors are eager to foot the bill.

Recall again Vaidhyanathan's title, *The Googlization of Everything.* For Big Data buffs,

"Googlization" is ultimately a hopeful process: systematic use of analytics to squeeze maximum effectiveness out of any decision; maximum relevance from any search; maximum risk-adjusted return from any investment. To paraphrase Jeff Jarvis, today's businesses should ask themselves, "What would Google do?" But the answer to that question is all too clear: use their data to outflank competitors and extract maximum profits from their customers.[210]

"Googlization" has an even darker meaning, too: that whole industries stand to be taken over by Google itself.[211] Walmart (Walmart!) has said that it considers Google one of its most formidable competitors. Even Apple's greatest misstep—forcing Google off its iOS in favor of an incomplete and ill-conceived maps app of its own—was an (unsuccessful) attempt to compete with Google for the locational data that Google's map services were collecting.[212] And what does "Googlization" mean to traditional publishers, booksellers, and educators, who don't have Google's opportunity to match individuals to "optimal" sources of information based on their past predilections, demonstrated abilities, and willingness to pay? That Silicon Valley engineers and managers are in charge of their fortunes.

Of course, Google isn't the only press baron on the horizon; Amazonification, Facebookization, and Twitterification also beckon. Some will further hollow out once-hallowed properties. Others will invest, as venture capitalist Marc Andreessen recommends. Though he strikes fear into publishers, Amazon's Jeff Bezos has not yet reduced writers at his newspaper (the *Washington Post*) to the status of Mechanical Turkers or warehouse pickers.[213] But we should not assume media independence as tech firms swallow more of the revenue that might have once gone to journalists. After Amazon inked a $600 million deal to provide the CIA with cloud computing services, 30,000 people petitioned the *Post* with the message "*Washington Post:* Readers Deserve Full Disclosure in Coverage of CIA."[214] Such inquiries will only become more common as Washington and Silicon Valley develop more partnerships for information dominance.

Of course, we can see why large firms want to keep their industry (and government) alliances under wraps. People want to feel like there is *someone* looking out for them. Google's decision to join forces with content industry leaders (regarding piracy) disappointed many of its users.[215] They had thought of Google as *their* agent, pushing for users' rights and a neutral, technical ordering of the Internet against the usual corporate interests' efforts to exploit it. But as Google dominates more of the search space, and as its investors' demands remain pressing, its business focus has shifted from the need to *attract more users* to the need to *monetize what the viewers see.* Google found itself needing more compelling content, and that content would only materialize for a price.[216]

These are trust issues. In a classic example of what philosopher Langdon Winner called "technological somnambulism,"[217] we have given the search sector an almost unimaginable power to determine what we see, where we spend, how we perceive. Top legal scholars have already analogized the power relationships in virtual worlds and cloud computing to medieval feudalism.[218] Technological advance goes hand-in-hand with politico-economic regression.

Toward a Digital New Deal

In the late 1990s, tech enthusiasts looked to search engines as an extraordinary democratization of the Internet. They permitted content creators from all over the world to reach far-flung audiences. Web 2.0 promised even more "democratization" by enabling self-organization of virtual communities. But recent commercial history suggests a different—even an opposite—effect. The very power that brought clarity and cooperation to the chaotic online world also spawns murketing, unfair competition, and kaleidoscopic distortions of reality.[219]

The first step toward reform is realizing the scope of the problem. Tim Wu, a prominent cyberlawyer and one of the intellectual architects of network neutrality, helps contextualize today's Internet disputes in a larger time frame. In his 2010 book *The Master Switch,* he animates a history of "industrial wars" over communications with strong moral judgments about the fairness or impropriety of the business strategies he investigates. The book is a tour de force of narrative. But it falls short, prescriptively. Wu acknowledges the coercive private power of an Apple or a Google but concludes that norms now restrain it: "Rare is the firm willing to assert an intention and a right to dominate layers of the information industry beyond its core business." However true that was then, it's outdated now: Google wants to expand to be a social network and military robot company; Facebook is not just a social network, but a kingmaker in online media; Amazon disrupts industry after industry. But Wu focuses more on the cultural and political impact of information-age giants than on the grubby economics that drives this rapid-scale expansion.[220]

I can understand why—people are far more interested in the outsized personalities of Silicon Valley than the complex money grabs that grant them their platforms. But we can't hope to reform the information economy without fundamentally changing the incentives at its core. Wu's postmaterialism would have been a good fit for the roaring nineties, when a rising tide of Internet firm profits seemed to be lifting many parts of the economy. But the economic crisis that has overtaken the United States since 2008 makes our time in many ways more similar to Franklin Roosevelt's era than Bill Clinton's. A small cadre of the lucky, the talented, and the ruthless are taking an enormous share of the revenues generated by new Internet technologies. They keep their methods strictly proprietary while reaping huge returns from content put out in the open by others.[221] Like the megafirms and CEOs that the New Deal helped bring to heel, the leaders of our largest tech firms have been very quick to misequate personal enrichment with the public good.

It is time to bring the substance and style of that era back into a progressive political economy of technology. In the first half of the twentieth century, the American lawyer, economist, and educator Robert Lee Hale studied the dominant firms of his day. Given their pervasive influence, he argued that personal freedom depended on responsible corporate conduct.[222] His theories were influential among FDR's advisers as they faced the economic catastrophe of the 1930s. Hale and Wu have both analyzed the "private coercive power" of large companies. But there are major differences between Hale's *Freedom through Law* and Wu's *The Master Switch,* and they speak volumes about changes in the American political climate over the past six decades.

Hale's work chronicles the gradual victory of democratic constraint over arbitrary and exploitative business practices. Hale discussed the "principles for determining how the wealth of the community should be distributed," patiently detailing the case law of ratemaking and taxation through the mid-twentieth century. He also made it clear that government couldn't just sit idly by as a "neutral party," in order to "avoid picking winners" in a time of technological change. If it failed to do so, there were other forces—such as finance—more than willing to step in and direct the economy. And we now see the results: monopolistic and manipulative behavior that has left many wary of a sector they once adored.

The search sector's profiteering is an effort to meet the demands of investors. Search firms may rank and rate the reputation companies that rank and rate people; but even search firms have to worry about how *they* are being rated by Wall Street. They can't keep swallowing up rivals unless investors keep betting on their enduring dominance. Opaque aspects of finance keep the leading Internet firms on their toes as surely as the Internet firms' mysterious ranking mechanisms keep everyone else alert and worried about any possible loss of standing. It is therefore to this final and most forceful aspect of the new political economy—finance—that we now turn our attention.