The Relevance of Algorithms

Tarleton Gillespie

Algorithms are playing an increasingly important role in selecting what information is most relevant to us, a crucial feature of our participation in public life. Search engines help us navigate massive databases, or the entire web. Recommendation algorithms map our preferences against others, proposing novel information for us to encounter. Algorithms manage the complex social interactions on social networking sites, highlighting news of one friend while excluding another's. Algorithms designed to calculate what is "hot" or "trending" or "most discussed" skim the cream from the seemingly boundless information on offer. Together, these algorithms provide a means not only to find and ascertain the value of information, but a means to know what there is to know and how to know it, to participate in social and political discourse, and to familiarize ourselves with the publics in which we participate. They are now a key logic governing the flows of information we depend on, with the "power to enable and assign meaningfulness, managing how information is perceived by users, the 'distribution of the sensible.'" (Langlois 2012)

Algorithms need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations. The procedures name both a problem and the steps by which it should be solved. Instructions for navigation may be considered an algorithm, or the mathematics required to predict the movement of a celestial body across the sky. "Algorithms do things, and their syntax embodies a command structure to enable this to happen." (Goffey 2008, 17) We might think of computers, then, fundamentally as algorithm machines – designed to store and read data, apply mathematical procedures to it in a

controlled fashion, and offer the output as new information. To decrypt a message requires an algorithm, transforming codes into plain expression according to a given cipher. Sharpening a digital image requires an algorithm, designed to calculate differences in neighboring color tones an alter them accordingly. Predicting market fluctuations requires an algorithm, that can match current conditions to a database of past financial events. All of these can have both implicit politics and political implications, of course, since the data have meaning and purpose and the tools make choices about handling them: what counts as a coherent message and with what consequences; what approximation of human perception is implied in the transformation of the image; what will be done with the market predictions and why.

But as we have embraced computational forms of media as our primary media of expression, and have made not just mathematics but *all* information digital, we have subjected human discourse and knowledge to the logics that undergird computation. And there are specific implications when we use algorithms to select from a corpus of data information deemed most relevant, and the data from which it draws is us and our public discourse. It is these algorithms, which I will call *public relevance algorithms*, that are my concern here. What these algorithms are doing, by the very same kind of mathematical procedures, is producing and certifying knowledge. The algorithmic assessment of information, then, represents a particular knowledge logic, one built on specific presumptions about what knowledge is and how one should identify the most relevant of it. It is just as significant that we are now turning to algorithms to identify what we need to know as it is to rely on credentialed experts, the scientific method, common sense, or word of God.

What we need now is an interrogation of the algorithm as a key feature of our information ecosystem (Anderson 2011) and the cultural forms emerging in their shadow

(Striphas 2010), with a close attention to where and in what ways its intervention into human knowledge practices can have political valence. A sociological analysis would have to conceive of public relevance algorithms not as abstract, technical achievements, but would unpack the warm human and institutional choices that are in fact behind these cold algorithms. This essay is a conceptual map for uncovering the ways algorithms can have political valence when they are asked to evaluate knowledge.

It would also be seductively easy to get this wrong. Attempting to say something of substance about the way algorithms are shifting our public discourse must resist the temptation to put the technology in the explanatory driver's seat. While recent sociology of the Internet has worked hard to move away from the simplistic technological determinism that plagued earlier work, it remains an alluring analytical stance. I suspect that a more fruitful approach will turn as much to the sociology of knowledge as to the sociology of technology – to see how these tools are called into being by, enlisted as part of, and negotiated around collective efforts to know and be known. This might help reveal the seemingly solid algorithm as a fragile accomplishment, dependent on a series of socio-material meeting points, where the "articulation" (Slack and Wise 2006) of human, material, institutional, and discursive elements produces a functioning artifact, but one with political valence. This also should remind us that algorithms have become a communication technology; like broadcasting and publishing technologies, they are now caught up in and are influencing the ways in which we ratify knowledge for civic life. The lessons learned from the study of communication and media may further illuminate algorithms, as they too subtly reshape public discourse.

Patterns of Inclusion

Algorithms are inert, meaningless machines until paired with databases upon which to function. Before results can be algorithmically provided, information must be collected, assessed, and sometimes categorized. For users, algorithms and databases are analytically indistinct: users typically conceive of them as a single, working apparatus. In the eyes of the market, the creators of the database and the providers of the algorithm are often one and the same, or are working in economic and often ideological concert. "Together, data structures and algorithms are two halves of the ontology of the world according to a computer. The computerisation of culture involves the projection of these two fundamental parts of computer software – and of the computer's unique ontology – onto the cultural sphere." (Manovich 1999, 84).

Though they may appear to be mere storehouses of information, just waiting for the real action to occur, databases are already the result of processes of selection, categorization, curation, and repair. Decisions about what to include in a particular database are the first moments in which the construction of the algorithm as a political intervention begins. A sociological inquiry into an algorithm must always grapple with the databases to which it is wedded; failing to do so would be akin to studying what was said at a public protest, while failing to notice that some speakers had been stopped at the park gates. What is excluded from the database will never be offered up by the algorithm; what is included is fodder for the algorithm and can potentially shape its assertions of relevance.

live in a historical moment in which, more than ever before, nearly all public activity includes keeping copious records, cataloging activity, and archiving documents – and we do more and more of it on a communication network designed such that every login, every page view, and every click leaves a digital trace. Turning such traces into databases involves a

complex array of information practices (Stadler and Mayer 2009): Google crawls the web indexing websites and their metadata; digitizes real world information from library collections to satellite images to comprehensive photo records of city streets; invites users to enter personal and social data as part of their Google+ profile; keeps exhaustive logs of every search query entered and every result clicked; adds local information based on each user's computer's data; and stores the traces of web surfing practices gathered through their massive advertising networks.

Understanding what is included in such databases requires an attention to the collection policies of information services, but should also extend beyond to the actual practices involved. This is not just to identify cases of malfeasance, though there are some, but to understand how an information provider thinks about the data collection it undertakes. The political resistance to Google's StreetView project in Germany and India helped reveal what Google determines to be "public" and how that interpretation was not universally shared. The algorithmic answer to "what does this street corner look like?" has different implications for those who what to go there, those who live there, and those who believe that the answer should not be available in such a public way.

Including information in a database is not a mere act of storage, it is a rendering of that information into data. "Raw data is an oxymoron." (Gitelman and Jackson, *forthcoming*) Data is both already desiccated and remains messy. Nevertheless, there is a premeditated order necessary for algorithms to even work. More than anything, algorithms are designed to be and prized for being functionally *automatic*, to act when triggered without any regular human intervention or oversight. (Winner) This means that the information included in the database must be formalized up front, so that algorithms can act on it automatically. Data must be "imagined and enunciated against the seamlessness of phenomena." (Gitelman and Jackson, *forthcoming*) We call them up

out of an otherwise undifferentiated blur. Recognizing the ways in which data must be "cleaned up" is an important counter to the seeming automaticity of algorithms. Just as one can know something about a sculpture from studying its inverted mold, algorithms can be understood by looking closely at how information must be oriented to face them, to make it *algorithm ready*.

In the earliest database architectures, information was organized in strict and, as it turned out, inflexible hierarchies. Since the development of relational and object-oriented database architectures, information can be organized in more flexible ways, where bits of data can have multiple associations with other bits of data, categories can change over time, and data can be explored without having to navigate or even understand the hierarchical structure by which it is archived. Nevertheless, even with these more flexible forms of databases, categorization remains vitally important to database management.

Categorization is a powerful semantic and political intervention: what the categories are, what belongs in a category, and who decides are all powerful assertions about how things are and are supposed to be (Bowker and Star 2000). And a category, once instituted, draws a demarcation that will be treated with reverence by an approaching algorithm. A useful example here is the #amazonfail incident. In 2009, more than fifty-seven thousand gay-friendly books disappeared in an instant from Amazon's sales lists, because they were accidentally categorized as "adult." Complex information systems are open to error. But this error also revealed that Amazon's algorithm that calculates the "best sellers" is instructed to ignore books designated as adult. Even when mistakes are not made, whatever criteria Amazon uses to determine adult-ness are being reified in the process – visible only in the unexplained absence of some books and the presence of others.

Though all database producers share an appetite for gathering information, they are made distinctive more by what they choose to exclude. "The archive, by remembering all and only a certain set of facts / discoveries / observations, consistently and actively engages in the forgetting of other sets… The archive's jussive force, then, operates through being invisibly exclusionary. The invisibility is an important feature here: the archive presents itself as being the set of all possible statements, rather than the law of what can be said." (Bowker 2008, 12-14) Even in the current conditions of digital abundance (Keane 1999), in which it is cheaper and easier to err on the side of keeping information rather than not, there is always a remainder.

Sites can, of course, refuse to allow data collectors (like search engines) to index their sites. Elmer (2008) reveals that robot.txt, a bit of code that prevents search engines from indexing a page or site, though designed initially as a tool for preserving the privacy of individual creators, has since been used by government institutions to "redact" otherwise public documents from public scrutiny. If only for this reason, we must firmly reject the common presumption that these algorithms take into account the entirety of a category of information (all websites, all online video, all tweets), that this is anything other than a *construction* of knowledge.

But beyond self-exclusion, some information initially collected are deliberately removed from algorithmic evaluation by the platform. Though large-scale information services pride themselves on being comprehensive, these sites are and always must be curators as well. Indexes are culled of spam and viruses, patrolled for copyright infringement and pornography, scrubbed of the obscene, the objectionable, or the politically contentious. (Gillespie *forthcoming*) Where these lines are drawn differs somewhat by information provider, based on their sense of their audience, their legal obligations, and the demands of their chosen revenue model.

Offending content can simply be removed from the index, or an account suspended, before it ever reaches another user. But, in tandem with an algorithm, problematic content can be handled in more subtle ways. YouTube "algorithmically demotes" suggestive videos, so they do not appear on lists of the most watched or on the front page for new users. Twitter does not censor profanity from public tweets, but it does remove it from their algorithmic evaluation of which terms are Trending.

The particular patterns whereby information is either excluded from a database, or included and then managed in particular ways, parallel the politics of inclusion in public discourse. The politics of media representation begin with who is systematically left out, what categories of speech simply don't count. This extends 20th century debates (Tushnet 2008) about how the choices made by commercial media, about what and who to include, can shape the diversity and character of public discourse. Standards of inclusion, whether enacted by a newspaper editor or by Google's indexing tools, help to reify standards of viable debate, legitimacy, and decorum. And, algorithms can be touted as impartial and automatic processes, while the patterns of inclusion prefigure what will and will not appear among their results.

Cycles of Anticipation

Search algorithms determine what to serve up based on input from the user – beginning with something as simple as the query entered into a search engine or the request of a particular product at an e-commerce site. But most platforms now make it their business to know much more about the user than what they've explicitly asked for. Sites hope to anticipate the user at the moment the algorithm is called upon, requiring knowledge of that user at that instant, knowledge of that user previously gathered, and knowledge of users estimated to be statistically and

demographically similar (Beer 2009). If broadcasters were providing not just content to audiences but also audiences to advertisers (Smythe 2001), digital platforms are not just providing information to users, they are also providing users to their algorithms. And algorithms are made and remade in every instance of their use because every click, every query changes the tool incrementally. So ever more trace data must be collected, and algorithms must be designed to "learn" ever more about their users, producing what Stadler and Mayer (2010) call the "second index."

Much of the scholarship about the data collection and tracking practices of contemporary information providers focuses on the significant privacy concerns they raise. Zimmer (2008) notes that search engines now aspire to not only relentlessly index of the web but also to develop "perfect recall" of all of its users. As the economics of online information providers shifts towards effectively anticipating users, information providers are rewarded for building technical infrastructures and business models that draw user activity ever more into their "social graph" or inside a provider's ecosystem of services, so they may assemble information gathered at many sites into a coherent and increasingly comprehensive profile. also take advantage of the increasingly participatory ethos of the web, where users are powerfully encouraged to share all sorts of information about themselves, and encouraged to feel powerful doing so. As our micro-practices migrate more and more the platform, it is seductive (though not obligatory) for information providers to both track and commodify that activity in a variety of ways (Gillespie and Postigo 2012). Moreover, users may be unaware that their activity across the web is being tracked by the biggest online advertisers, and they are in little position to challenge this arrangement.

In these cycles of anticipation, it is the bits of information that can be most easily known, that are most legible to the algorithm, that tend to stand in for those users. What Facebook knows about its users is a great deal; but it knows only what it is able to know. This knowable information (geolocation, computing platform, profile information, friends, status updates, links followed on the site, time on the site, activity on other sites that host "like" buttons or cookies) is a rendition of that user that is imperfect but sufficient; what is less legible or cannot be known about users falls away or is approximated. As Star and Balka described it, information systems produce "shadow bodies" by emphasizing some aspects of their subjects and overlooking others (Balka 2011). These shadow bodies persist and proliferate through information systems, and the slippage between the anticipated user and the user themselves that they represent can be either politically problematic or productive.

But algorithms are not always about exhaustive prediction; sometimes they are about sufficient approximation. Perhaps just as important as the anxious surveillance of users (Turow 2012) are the kinds and depths of conclusions providers are willing to draw based on relatively little information about their users. Hunch.com boasts that, based on the answers to just five questions, they can know a user's preferences with 80-85% accuracy. While this radically boils down the complexity of a person to five points on a graph, what is important is that this is considered a sufficient accuracy for their purposes (Zuckerman 2011). Such sites are comfortable catering to these user-caricatures, and the questions that sort us most sufficiently, particularly around our consumer preferences, are likely to grow in significance as public measures. And to some degree, we are invited to formalize ourselves into the knowable categories. We are encouraged to choose from the provided menu, so as to be correctly anticipated by the system providing us the right information, the right recommendations, the right people.

Beyond knowing the personal and the demographic details about each user, information providers conduct a great deal of research trying to understand, and then operationalize, how humans habitually seek, engage with, and digest information. Most notably in the study of human-computer interaction (HCI), the understanding of human psychology and perception is brought to bear on the design of algorithms and the ways in which their results should be represented. This is an anticipation of users' psycho-physiological capabilities and tendencies, rather than of specific users' actions. But in these anticipations, too, implicit and sometimes politically valences can inscribed in the technology (Sterne 2008): the perceptual or interpretive habits of some users taken to be universal, contemporary habits taken presumed to be timeless, particular computational goals assumed to be self-evident.

We are also witnessing a new kind of information power, gathered in these enormous databases of user activity and preference, that is reshaping the political landscape. Algorithmic information providers who amass this data along the way, third party industries who gather and purchase user data as a commodity for them, and those that traffic in user data for other reasons (like credit card companies), have a stronger voice than ever in the marketplace and in the halls of legislative power, and are increasingly involving themselves in political debates about consumer safeguards and digital rights. We are seeing the deployment of data mining in the arenas of political organizing (Howard 2005), journalism (Anderson 2011), and publishing (Striphas 2009), where the secrets drawn from massive amounts of user data are taken as compelling guidelines for future content production, be it the next micro-targeted campaign ad or the next pop phenomenon.

The Evaluation of Relevance

When a user clicks "Search" or loads their Facebook News Feed or asks for a recommendation from Netflix, their algorithm must instantly and automatically identify which of the trillions of bits of information best meets the criteria at hand, and will best satisfy this user and their presumed aims. While this has never been a simple calculation, the evaluations involved have grown more complex as the public use of these services has matured. Search results, for example, once based on tallying how often the actual search terms appear in the indexed web pages, now incorporates contextual information about the sites and their hosts, considers how often the site is linked to by others and in what way, and incorporates natural language processing techniques to better "understand" both the query and the resources that the algorithm returns in response. According to Google, its search algorithm examines over 200 signals for every query.[1]

These signals are the means by which the algorithm approximates "relevance." But here is where sociologists of algorithms must firmly plant their feet: "relevant" is a fluid and loaded judgment, as open to interpretation as the evaluative terms media scholars have had to unpack, like "newsworthy" or "popular." As there is no independent metric for what "actually" are the most relevant search results for any given query, engineers must decide what results look "right" and tweak their algorithm to attain that result. To accuse an algorithm of bias implies that there is an unbiased judgment of relevance available that the tool is failing to hew to. Yet there is no such measure available, which means disputes over algorithmic evaluations have no solid ground to fall back upon.

The catch is that their evaluative criteria are hidden, and must remain so. Twitter's Trends algorithm, which reports to the user what terms are "trending" at that moment in their area, leaves the definition of "trending" unspecified. Though in their corporate blog they identify some

of the criteria they use to assess 'trendiness,' these criteria are only described in general terms: the velocity of a certain term's surge, whether it has appeared in the Trend list before, whether it remains within or spans across clusters of users (Gillespie 2011). What is unstated is how these criteria are measured, how they are weighed against one another, what other criteria have also been incorporated, and when if ever these criteria will be overridden. This leaves algorithms perennially open to user suspicion: that their criteria skew to the provider's commercial or political benefit, or incorporate embedded, unexamined assumptions that act below the level of awareness, even of the designers.

Publicly, an information provider like Twitter cannot be much more explicit or precise about how its algorithm works. Being plain about their algorithm's workings might give competitors an easy means of duplicated and surpassing their service. It would require a more technical explanation than most users are prepared for. Information providers want to retain the ability to change their criteria as they need. But most of all, those who hope to "game the system," to get their sites to the top of the search results or their hashtags to appear on the Trends list, would be able to take advantage of more specific knowledge of the algorithm's workings.[2]

Most users do not dwell on algorithmic criteria, tending to treat them as unproblematic tools in the service of a larger activity: finding an answer, solving a problem, being entertained. However, while the technology may be "black boxed" (Pinch and Bijker, 1984; Latour 1987) by designers and users alike, that should not lead us to believe that it remains stable. In fact, algorithms can be easily, instantly, radically, and invisibly changed. While major upgrades may happen only on occasion, algorithms are regularly being "tweaked". Changes can occur without the interface to the algorithm changing in the slightest: the Facebook news feed and search bar may look the same, while the evaluations made beneath have been thoroughly remade. So they

are "likely so dynamic that a snapshot of them would give us little chance of assessing their biases." (Pasquale 2009) In fact, what we might refer to as an algorithm is often not one algorithm but many. Search engines like Google regularly engage in "A/B" testing,[3] presenting different rankings to different subsets of users to gain on-the-fly data on speed and customer satisfaction, incorporating the adjustments preferred by users in a subsequent upgrade.

To say that a particular algorithm makes evaluative assumptions, the kind that have consequences for human knowledge endeavors, requires not just an understanding of the algorithm and its underlying criteria, but a careful consideration of the economic and the cultural contexts from which the algorithm came. Any knowledge system emerges amidst economic and political structures of information provision, and will be shaped by the aims and strategies of powerful institutions looking to capitalize on it. (Hesmondhalgh 2006) The economic and political demands faced by search engines, content platforms, and information providers can subtly shape the design of the algorithm itself and the presentation of its results. (Vaidhyanathan 2011) As the algorithm comes to stand as a legitimate knowledge logic, new commercial endeavors are fitted to it (for instance, search engine optimization), reifying choices made and forcing additional ones.

The earliest and easiest version of this was the worry that search engines would offer up advertisements in the form of links or featured content, presented as the product of algorithmic calculations. The rapid and clear public rejection of this ploy demonstrated how strong our trust in these algorithms is: users did not wish the content sites wanted us to see for financial reasons intermingled with content that had been algorithmically selected. But the issue is now multidimensional: the landscape of the Facebook News Feed, for example, can no longer be described as two distinct territories, social and commercial; rather, it interweaves the results of

algorithmic calculations (what status updates and other activities of friends should be listed in the Feed, what links will be recommended to this user, which friends are actively on the site at the moment) structural elements (tools for contributing a status update, commenting on an information element, links to groups and pages) and elements placed there based on a sponsorship relationship (banner ads, apps from third party sites). To understand this complex terrain requires both careful description and a deep understanding of the economic relationships and social assumptions it represents.

A more difficult question is whether the evaluative criteria themselves contain political assumptions, over and above how they might be tweaked towards some specific commercial benefit for the provider. Some early scholarship in the field examined the biases of search engines (Introna and Nissenbaum 2000; Halavais 2008; Granka 2010), noting structural tendencies towards what's already popular, towards English-speaking sites, and towards commercial information providers. Legal scholars have debated what it would mean to require neutrality in search results (Pasquale and Bracha 2008; Grimmelmann 2010).

The criteria public information algorithms take into account are myriad; each is fitted with a threshold for what will push something up in the results, position one result above another, etc. So evaluations performed by algorithms always depend on inscribed assumptions about what matters, and how what matters can be identified. When a primitive search engine counted the number of appearances of a search term on the web pages it had indexed, it reified the following logic: pages that include the queried term often are likely to be "about" or highly relevant to someone interested in that term. When Google developed PageRank, factoring in incoming links to a page as evidence of its value, it built in a different logic: a page with many incoming links, from high quality sites, has already been "ratified" by other users as relevant to

their own purposes, so it is more likely to be relevant to this user's purposes as well. Each of these algorithms reveals both a worldview about the assessment of relevance, and an instantiation of that worldview into a technique for (computational) evaluation. There may be assumptions built into a site's idea of relevance, there may be shortcuts built into its technical instantiation, and there may be slippage between the two. When a news discussion site decides what ratio of negative complaints to number of views is sufficient to justify automatically deleting a comment thread, it represents their assessment of the proper volatility of public discourse, or at least the volatility they prefer. (Braun 2011) A great deal of expertise and judgment can be embedded in these cognitive artifacts, (Hutchins 1995; Latour 1986) but it is judgment that is then submerged and automated.

Algorithmic Objectivity

 "Search engines pride themselves on being automated, except when they aren't."
(Grimmelmann 2008, 950) Algorithms are stabilizers of trust, assurances that their evaluation of relevance is a fair and accurate one, free from subjectivity, error, or attempted influence. But, though algorithms may appear to be automatic and untarnished by the interventions of their providers, this is a deliberate fiction. It is very important for the providers of these algorithms that it seem so. The legitimacy of these functioning mechanisms must be performed along side the provision of information itself. It is this careful "figuration" of an algorithm (even when that characterization is more obfuscation than explanation) that certifies it as a reliable socio-technical actor, that lends its results relevance and credibility, that maintains its apparent neutrality in the face of all the millions of evaluations it makes. This figuring of the algorithm is, in my mind, just as crucial to its social life as its material design.

It is largely up to the providers to describe their algorithm as being of a particular shape, having therefore a certain set of values, and thus conferring to it some kind of legitimacy. This includes carefully characterizing the tool and its value to a variety of audiences: an algorithm can be presented as neutral to critics of its results, while being presented as powerful tool for leverage and visibility to potential advertisers (Gillespie 2010). As Mackenzie (2005) notes, this process requires more than a single, full-throated articulation: it depends both on "repetition and citation," (81) and at the same time requires "the 'covering over' of the 'authoritative set of practices' that lend it force." (82) When an information provider finds itself criticized for the results it provides, the legitimacy of their algorithm must be repaired both discursively as well as technically. And, users are complicit in this figuration: "A society that obsesses over the top Google News results has made those results important, and we are ill-advised to assume the reverse (that the results are obsessed over because they are important) without some narrative account of why the algorithm is superior to, say, the "news judgment" of editors at traditional media." (Pasquale 2009)

This figuration happens first in the very presentation of the tool, in its very deployment within a broader information service. Calling them "results" or "best" or "top stories" or "trends" speaks not only to what the algorithm is actually measuring, but what it should be seen as measuring. An equally important part of this discursive work comes in the form of describing how the algorithm works. What can seem like a clear explanation of a behind-the-scenes process is in fact a "performed backstage" (Hilgartner 2000), carefully crafted to further legitimize the process and its results. The description of Google's PageRank system, the earliest component of its complex search algorithm, was published first as a technical paper (already a crafted performance of its mathematical workings), but it was subsequently mythologized – as the

defining feature of the tool, as the central feature that made Google stand out above its then competitors, and as a fundamentally democratic computational logic – even as the algorithm was being redesigned to take into account dozens of other criteria.

Above all else, the providers of information algorithms must assert that their algorithm is impartial. The performance of "algorithmic objectivity," more than any other, has become fundamental to the maintenance of these tools as legitimate brokers of relevant knowledge. In fact, no information service can be completely hands-off in its delivery of information: though an algorithm may evaluate any site as most relevant to your query, that result will not appear if it is child pornography, it will not appear in China if it is dissident political commentary, it will not appear in France if it promotes Nazism.

No provider has been more adamant about the neutrality of its algorithm than Google, who regularly responds to requests to alter their search results with the assertion that the algorithm must not be tampered with. Google famously pulled out of China in 2010 entirely rather than censor its results, though they had complied with China's rules before. But despite their stance, they nevertheless altered their search results when complaints arose about a racist Photoshopped image of Michelle Obama made it to the top of the Image search results; they provide a SafeSearch mechanism for keeping profanity and sexual images from minors; they refuse to autocomplete search terms that specify torrent file-trading services. Yet despite all this, Google regularly claims that it does not alter its index or manipulate its results. Morozov (2011) believes that figuring the algorithm as neutral is a way to deflect responsibility: "Google's spiritual deferral to "algorithmic neutrality" betrays the company's growing unease with being the world's most important information gatekeeper. Its founders prefer to treat technology as an autonomous and fully objective force rather than spending sleepless nights worrying about

inherent biases in how their systems – systems that have grown so complex that no Google engineer fully understands them – operate."

This assertion of and faith in the objectivity of algorithms is in many ways equivalent to the norm of objectivity in Western journalism. Journalists have developed tactics for determining what is most relevant, how to report it, and how to assure its relevance – a set of practices that are relatively invisible to their audience, that they will admit are messier to deal with than they might seem, that try to set aside but do not eradicate value judgments and politics, but that nevertheless accomplish something in the form of assessing information's relevance and organizing it accordingly. These institutionalized practices are animated by a conceptual promise that, in the discourse of journalism, is regularly articulated (or overstated) as a kind of totem. Journalists use the norm of objectivity as a "strategic ritual" (Tuchman 1972), aimed at achieving public legitimacy for knowledge production tactics that are inherently precarious. "U.S. journalism's claim to objectivity - i.e., the particular method by which this information is collected, processed, and presented - gives it its unique jurisdictional focus by claiming to possess a certain form of expertise or intellectual discipline. Establishing jurisdiction over the ability to objectively parse reality is a claim to a special kind of authority." (Schudson and Anderson 2009, 96)

Journalistic objectivity depends on an institutional promise of due diligence, built into and conveyed via a set of norms journalists learned in training and on the job; the choices made represent a careful expertise backed by a philosophical commitment to set aside subjectivity, a commitment built deep into the profession itself. The promise of the algorithm leans much less on institutional norms and trained expertise, more on a technologically-inflected promise of mechanical neutrality. Whatever choices are made are presented both as distant from the

intervention human hands, and submerged inside of the cold workings of the machine. But in both, legitimacy depends on accumulated guidelines for the *proceduralization* of information selection.

The discourses and practices of objectivity have come to serve as a constitutive rule of journalism (Ryfe 2006). Objectivity is part of how journalists understand themselves and what it means to be a journalist. It is part of how their work is evaluated, by editors, colleagues, and by their readers. It is a defining signal by which journalists even recognize what counts as journalism. The promise of algorithmic objectivity, too, has been palpably incorporated into the working practices of algorithm providers, constitutively defining the function and purpose of the information service. When Google includes in its "Ten Things We Know to Be True" manifesto that "Our users trust our objectivity and no short-term gain could ever justify breaching that trust," this is neither spin nor is it corporate Kool-Aid. It is a deeply ingrained understanding of their information service, which both shapes and champions many of their technical and commercial efforts, and helps obscure the messier reality of the service they provide.

The figurations offered by the algorithm provider alongside their tool are meant to provide what Pfaffenberger (1992) calls "logonomic control," to define their tool into the practices of users, to bestow the tool with a legitimacy that then carries to the information provided and, by proxy, the provider. But they must compete in the public dialogue with other figurations, that may or may not be so friendly to the economic and ideological aims of the stakeholders. Bijker (1997) calls these competing "technological frames," the discursive characterizations of a technology made by those groups of actors who have a stake in that technology's operation, meaning, and social value. What users of an information algorithm take it to be, and whether they are correct or ignorant, matters. How the press portrays such tools will

strengthen or undermine the providers' careful discursive efforts. This means that, while the algorithm itself may seem to possess an aura of technological neutrality, or embody populist, meritocratic premises, how it comes to appear that way depends on the mundane realities of news cycles, press releases, tech blogs, fan discussion, user rebellion, and the machinations of their competitors.

At its base, there is a fundamental tension in the figuration of algorithms. Algorithmic objectivity is an important claim for a provider, particularly for algorithms that serve up vital and volatile information for public consumption. Figuring the algorithm as a distinctly technical intervention helps an information provider avoid charges of bias, error, and manipulation. At the same time, as can be seen with Google's PageRank, there is a sociopolitical value in highlighting the populism of the criteria the algorithm uses. To claim that an algorithm is a democratic proxy for the collective opinion of a particular website lends it a kind of authority. Further, there is commercial value in claiming that the algorithm returns the best results, better than its competitors, which works against the claim of neutrality. For this, it is imperative to offer not a neutral tool, but the best neutral tool, the tool that finds the relevant content its competitors cannot. The claim that an algorithm simulates high-quality editorial judgment offers another kind of authority, but a different one. To suggest that it can ascertain patterns from human activity data that cannot be perceived by human observers offers still another kind of authority. In examining the figuration of an algorithm, we should pay particular attention to how this tension between technically-assured neutrality and the social flavor of the assessment being made is managed – and, sometimes, where it breaks down.

Entanglement

Though they could be studied as abstract, computational code, algorithms are built to be embedded into practice: in the lived world that produces the information they process, and in the lived world of their users. This is especially true when the algorithm is the key instrument for a business for whom the information delivered is the commodity, or is the means to deliver users to the advertisements it pairs with it. If users fail or refuse to build that tool into their practices, to make it meaningful, that algorithm will fail. This means we must consider not their "effect" on people, but a multidimensional "entanglement" between algorithms put into practice and the social life of their users. This relationship is of course a moving target, because algorithms change, and the user populations and activities they encounter change as well. Still, this should not imply that there is no relationship. As these algorithms have nestled into people's daily lives and mundane information practices, users shape and refigure the algorithms they encounter; and algorithms may impinge on how people seek information, how they perceive and think about the contours of knowledge, and how they understand themselves in and through information.

First, we need to conceive of this entanglement not as a one-directional influence, but as a recursive loop between the algorithmic calculations of the tool and the procedural activities of people. The algorithm that helps users navigate Flickr's photo archive is built upon the archive of photos posted, which means it is designed to apprehend and reflect back the lived practices of photographers. What people do and do not photograph is already a kind of calculation, though one that is historical, multivalent, contingent, and sociologically informed. The Flickr algorithm was designed to be aware of and responsive to the "algorithms" of photographers. But these were not the only design impulses; sensitivity to the dynamics of photographic practice had to compete with cost, technical efficiency, legal obligation, and business imperatives. And, the population of Flickr users and the types of photos they posted changed as the site grew in

popularity, was forced to compete with Facebook, introduced tiered pricing, was bought by Yahoo, and so forth.

Many Flickr users post photos with the express purpose of having them be seen by others: some are professional photographers looking for employment, some are seeking communities of like-minded hobbyists, some are simply proud of the work. So while the algorithm should be sensitive to photography, photographers have an interest in being sensitive to the algorithm, aware that being delivered in response to the right search might put their photo in front of the right people. Just as Hollywood's emphasis on specific genres calls screenwriters to write in generic ways,[4] the Flickr algorithm may induce subtle reorientations of photographers' practices, towards its own constructed logic: they may aspire to photograph in ways adherent to certain emergent categories, to orient their choice of subject and composition towards those things the algorithm appears to recognize and privilege. "What we leave traces of is not the way we were, but a tacit negotiation between ourselves and our imagined auditors." (Bowker 2009, 6-7)

This tacit negotiation consists first and foremost of the mundane, strategic reorientation of practices many users undertake, towards a tool that they know could amplify their efforts. There is a powerful and understandable impulse for producers of information to make their content, and themselves, recognizable to the algorithm. A whole industry, search engine optimization (SEO), promises to give websites the edge to make it to the top of the major search engines; information providers, whose reputation and profit depend on the perceived neutrality of their algorithmic offerings, engage in an arms race against these efforts; the optimizers work to "crack the code" of each new upgrade. But we might think of optimization (deliberate, professional) as just the leading edge of a much more varied, organic, and complex process by

which content providers of all sorts orient themselves towards the algorithms they hope to find value in. When we use hashtags in our tweets, we are not just joining a conversation or hoping to be read by others, we are designing our expression so as to be better recognized and distributed by Twitter's search algorithm. When we time our Facebook status updates, we are working with or against the rhythms, as we understand them, of Facebook's algorithm. Some may work to be noticed by the algorithm (teens have been known to tag their status updates with unrelated brand names, in the hopes that Facebook will privilege those updates in their friends' feeds[5]), others may work to evade an algorithm (Napster and P2P users sharing infringing copyrighted music were known to slightly misspell the artists' names, so users could find "Britny Speers" recordings but the record industry software would not[6]).

Is this gaming the system, like the search engine optimizers? Or is it a fundamental way we, to some degree, orient ourselves towards the means of distribution through which we hope to speak? Based on the criteria of the algorithm in question (or by our best estimate of its workings) we in all sorts of ways make ourselves *already algorithmically recognizable*. This is not so far from the tendency for newsmakers to orient their efforts to best fit the routines and rhythms of the news industry: timing a press release to make the nightly broadcast, or providing packaged video to a cable outlet hungry for gripping footage, are techniques for turning to face the media that may amplify you. Now, for all of us, social networks and the web offer some analogous kind of amplification, and we gain similar benefit by understanding and orienting our communication towards the workings of algorithms.

But who is best positioned to understand and operate the public algorithms that matter so much to the public circulation of knowledge? Insight into the workings of information algorithms is a form of power: vital to participating in public discourse, essential to achieving

visibility online, constitutive of credibility and the opportunities that follow. As mentioned

before, the criteria and code of algorithms are generally obscured – but not equally or from

everyone. For most users, their understanding of these algorithms may be vague, simplistic,

sometimes mistaken; they may attempt to nudge the algorithm in ways that are either simply

considered best practices (hashtags, metadata) or that fundamentally misunderstand the

algorithm's criteria (repeatedly retweeting the same message in the hopes of Trending on

Twitter). Search engine optimizers, advertisers, and spammers have just as little access, but have

developed a great deal of technical skill, hoping to divine the criteria beneath the algorithm

through testing and reverse engineering, so as to outfit their clients' content with the features that

will push them to the top of result lists. Communities of technology enthusiasts and critics

engage in similar attempts to uncover the workings of these systems, whether for fun, insight,

personal advantage, or determined disruption. And legislators, who have only just begun to ask

questions about the implications of algorithms for fair commerce or political discourse, have thus

far been given only the most general of explanations; information providers often contend that

their algorithms are trade secrets that cannot be divulged in a public venue.

On the other hand, some stakeholders are granted access to the algorithm, though under

controlled conditions. Advertisers are offered one kind of access to the inner workings of that

system, given backstage access to an algorithm so as to bid on placement. Information providers

that provide APIs to their commercial partners and third party developers offer those developers

a glimpse under the hood, though they are often bound by contracts and nondisclosure

agreements in the very same moment. Access to, understanding of, and rights regarding the

algorithms that play a crucial role in public discourse and knowledge will likely change, for

different stakeholders and under specific circumstances – changing also the power to build for,

navigate through, and regulate these algorithms available to these stakeholders and those they represent.

As much as these tools may urge us to make ourselves legible to them, we also take them into our practices, shifting their meaning and sometimes even their design along the way. Silverstone (1994) has suggested that, once technologies are offered to the public, they undergo a process of "domestication:" literally, these technologies enter the home, but also figuratively, users make them their own, embedding them in their routines, imbuing them with additional meanings that the technology provider could not have anticipated. Public information algorithms certainly come to play important roles in the way each user finds information and entertainment, attends to and communicates with others, and knows themselves and the world around them. Users express preferences for their favorite search engines, opine about a site's recommendation system as being buggy or intuitive or spot on. Apple iPhone users swap tips on how to make its Siri voice-activated search agent speak its repertoire of amusing retorts,[7] then share in the outrage when it doesn't return the expected answers on hot button political issues.[8] Facebook users one day become immediate critics and back-seat engineers when the algorithm behind their news feed is altered, or what felt to them like a neutral delivery of information suddenly begins to reveal a corporate calculus; through and after the uprising, they continue to post status updates and add photos. Users, faced with the power asymmetries of data collection and online surveillance, have developed an array of tactics of "obfuscation" to evade or pollute the algorithmic attempts to know them (Brunton and Nissenbaum 2011).

Algorithms are not just what designers make of them, or what they make of the information they process, they are also what we make of them day in and day out – but with this caveat: because the logic, maintenance, and redesign of these algorithms remain in the hands of

the information providers, they are in a distinctly privileged position to rewrite our understanding of them, or to engender a lingering uncertainty about their criteria that makes it difficult for us to treat them as truly our own.

Users can also turn to these algorithms as ways to make meaning of themselves; many information providers allow us to present ourselves to others and back to ourselves, including our public profile, our performance of our friendships, the expression of our preferences, or as a record of our recent activity. Facebook's Timeline curates a user's activity into a chronological remembrance of them; the pleasure of seeing what it algorithmically selects and juxtaposes offers a kind of delight, a delight beyond composing the series of photos and news posts oneself. But algorithms may also function as a particularly compelling "technology of the self" (Foucault 1988) in that they can independently ratify one's public self. It is now common practice to Google oneself: seeing myself appear as the top result in a search for my own name offers a kind of assurance of my tenuous public existence. There is a ratification when your pet topic Trends on Twitter, or when Amazon recommends a book you already love, or when Apple iTunes' "Genius" function composes a playlist from your library of songs. Whether we actually tailor our Amazon purchases so as to appear well-read (just as Nielsen ratings families used to over-report watching PBS and C-Span) or we simply enjoy when the algorithm confirms our sense of self, algorithms offer a powerful invitation to understand ourselves through the independent lens they promise to provide.

It is easy to theorize, but substantially more difficult to document, how users may shift their worldviews to accommodate the underlying logics and implicit presumptions of the algorithms they use regularly. There is a case to be made that the working logics of these algorithms not only shape user practices, but induce users to internalize their norms and

priorities: Bucher (2012) makes the case that the EdgeRank algorithm, used by Facebook to determine which status updates get prominently displayed on a users' news feed, encourages a "participatory subjectivity" in users, who come to appreciate that gestures of affinity (such as commenting on a friends' photo) are a criteria Facebook's algorithm considers when deciding whether their news makes it to your feed, and yours to theirs. Longford (2005) argues that the code of commercial platform "habituates" us, through incessant requests and carefully designed default settings, towards giving over more of our personal information. Mager (2012) and van Couvering (2010) both propose that the principles of capitalism are embedded in the workings of search engines.

But we need not resort to muscular theories of ideological domination to suggest that algorithms designed to identify relevant knowledge or common associations offer compelling ways of knowing, and that as they become more pervasive and trusted, these logics are self-affirming. Google's search engine, amidst its 200 signals, does presume that relevant knowledge is based largely on public ratification, adjusted to weigh more heavily the opinions of those who are also publicly ratified. This particular blend of the wisdom of crowds and the identification of collectively certified authorities is Google's own solution to the longstanding tension between expertise and common sense, in the enduring problem of how to know. It is not without precedent, and it is not a fundamentally flawed way to know, but it is a specific one, with its own emphases and myopias. Now, their solution is operationalized into a tool that billions of people use every day, most of whom experience it as simply working, as "right". To some degree, Google and its algorithm help assert and normalize this knowledge logic as "right," as right as its results appear to be.

Calculated Publics

Ito, boyd, and others have recently introduced the idea of "networked publics" (boyd 2010; Ito 2008; Varnelis 2008) to highlight both the communities of users that can come together using social media, and the way the technologies in question structure how these publics can form, interact, and sometimes come apart. "While networked publics share much in common with other types of publics, the ways in which technology structures them introduces distinct affordances that shape how people engage with these environments." (boyd 2010, 39)

Certainly, to the extent that algorithms are typically a technological component of these mediated environments, they too help structure the publics that can emerge using digital technology. Some concerns have been raised about how the workings of information algorithms, and the ways we often choose to navigate them, could undermine our efforts to be involved citizens. The ability to personalize search results and information resources was the first and perhaps best articulated of these concerns. With contemporary search engines, the results two users get to the same query can be quite different; in a news service or social network, the information offerings can be tailored to the user's preferences (by the user or the provider) such that, in practice, the stories represented as most newsworthy may be so dissimilar from user to user that no common object of public dialogue is even available. Sunstein (2001) and, more recently, Pariser (2011) have argued that, when algorithmic information services can be personalized to this degree, the diversity of public knowledge and political dialogue may be undermined. We are led, by algorithms and our own preference for the like-minded, into "filter bubbles," (Pariser 2011) populated exclusively with the news we expect and the political perspectives we already hold dear.

But algorithms not only structure our interactions with others as members of networked

publics; algorithms also traffic in calculated publics they themselves produce. When Amazon recommends a book that "customers like you" bought, it is invoking and claiming to know a public with which we are being invited to feel affinity– though the population upon which it calculates these recommendations may not necessarily be coterminous with its entire customer base. When Facebook offers as a privacy setting to allow information to be seen by "friends, and friends of friends," it has transformed a discrete set of users into an audience – it is a group that did not exist until that moment, and only Facebook knows its precise membership. These imagined publics and manufactured groups may overlap with, be an inexact approximation of, or have nothing whatsoever to do with the publics that the user sought out.

Some algorithms go further, making claims about the public they purport to know, and the users' place amidst them. Elsewhere I have argued that Twitter's Trends algorithm promises users a glimpse of what a particular public (national or regional) is talking about at that moment, though a constructed one shaped by Twitter's specific, and largely unspecified criteria (Gillespie 2012). Klout promises to measure a user's influence across the various social media platforms they use. Their measures are intuitive in their definition, but completely opaque in their mechanisms. Algorithms that purport to identify what is "hot" engage in a calculated approximation of a public through their traceable activity, then report back to them what they care about most. The friction between the 'networked publics' forged by users and the 'calculated publics' offered by algorithms further complicates the dynamics of networked sociality.

With other measures of public opinion, such as polling or surveys, the central problem is extrapolation, where a subset is presumed to stand for the entire population. With algorithms, the population is often the entire user base, which can be hundreds of millions of people (but, only

that user base the algorithm provider has access to). The central problem here is that the intention

behind such calculated representations of the public is by no means actuarial. Behind these

representations of the public or of publics, we can ask: what is the gain for the provider in

making such characterizations, and how does that shape what they're looking for? Who is being

chosen to be measured in order to produce this representation, and who is left out of the

calculation? And perhaps most importantly, how do these technologies, now not just

technologies of evaluation but of representation, help to constitute and codify publics they claim

to measure, publics that would not otherwise exist except that the algorithm called them into

existence?

These questions matter a great deal, and will matter more, to the extent that the

representations of the public produced by information algorithms get taken up, by users or by

authorities, as legitimate, and incorporated into the broader modernist project of reflexivity

(Giddens 1991). "Society is engaged in monitoring itself, scrutinizing itself, portraying itself in a

variety of ways, and feeding the resulting understandings back into organizing its activities"

(Boyer and Hannerz 2006, 9) What Twitter claims matters to "us" or what Amazon says teens

are forms of authoritative knowledge that can be invoked by institutions whose aim is to regulate

such populations.

The belief that such algorithms, combined with massive user data, are better at telling us

things about the nature of the public or the constitution of society, as proving to be an alluring

one for scholars as well. Social science has turned eagerly towards computational techniques, or

the study of human sociality through "big data," (Lazer et. al. 2009; boyd and Crawford 2012) in

the hopes of enjoying the kind of insights that the biological sciences have achieved, by

algorithmically looking for needles in the digital haystacks of these immense databases. The

approach is seductive: the millions of data points lends a great deal of legitimacy, and the way the algorithm seems to spot patterns that scholars couldn't see otherwise is exciting. "For a certain sort of social scientist, the traffic patterns of millions of e-mails look like manna from heaven." (*Nature* 2007) But this methodological approach should heed the complexities described so far, particularly when their data is generated by commercial algorithms themselves. Computational research techniques are not barometers of the social. They produce hieroglyphs: shaped by the tool by which they are carved, requiring of priestly interpretation, and telling powerful but often mythological stories, usually in the service of the gods.

When the data is us, what should we make of the associations that algorithms claim to identify about us as a society -- that we did not know, or did not want to know? Ananny (2011) offers an excellent example, when he noticed the Google Android Market recommending a sex offender location app for users who downloaded Grindr, a location-based social networking tool for gay men. He speculates how this association might have been made within the Android Market algorithms, and wonders about its possible impact on already contested cultural associations between homosexuality and predatory behavior. But he also notes the troubling ambiguity about how and why this link was made, a link even the operators of the Android Market could not easily explain. Did the algorithm make an error? Did the algorithm make too blunt an association, simply pairing apps with "sex" in the metadata? Or did the Android recommendation engine in fact identify a subtle association that, though we may not wish it so, is regularly made in our culture? Zimmer (2007) notes a similar case, where (until Google changed the results) a search for "she invented" would result in the return query "did you mean 'he invented'?" While insulting in its gender politics, the response is completely explained by the sorry fact that, over the entire corpus of the web, the word "invented" is preceded by "he" much

more often than "she." Google's algorithm recognized this disparity – and mistakenly presumed it meant the search query "she invented" was a typographical error. Google, here, proves much less sexist than we are. In a response to Ananny's case, Gray has suggested that, just as we must examine algorithms that make associations such as these, we might also inquire into the "cultural algorithm" that this association represents, i.e. systematically associating homosexuality with sexual predation, across a massive, distributed set of "data points" -- us.

Conclusion

Understanding the algorithm and its impact on the public, then, requires thinking not simply about how it works, where it is deployed, or what motivates it financially. It requires examining why the algorithm is looked to as a credible knowledge logic, how it falls apart and is repaired when it comes in contact with the ebb and flow of public discourse, and where political assumptions might be etched into its design.

I see the emergence of the algorithm as a trusted information tool as the latest response to a fundamental tension of public discourse. The means by which we produce, circulate, and consume information, in a complex society, must necessarily be handled through the division of labor: some produce and select information, and the rest of us, at least in that moment or context, can only take it for what its worth. Every public medium previous to this has taken on this challenge, from travelling bards to newspapers to broadcasting, each with its own unique solution to this division of labor. But, in each, when we turn over the provision of knowledge to others, we are left vulnerable to their choices, methods, and subjectivities. Sometimes we see this as a positive: expertise, editorial acumen, refined taste. But in other moments, we are wary of the intervention of these information producers, of their human failings and vested interests, and find

ourselves with only secondary mechanisms of social trust by which to vouch for its truth and significance (Shapin 1995). Their procedures are largely unavailable to us. By definition these procedures are selective, emphasizing some information and discarding others, and the choices may be consequential. There is the distinct possibility of error, bias, manipulation, laziness, commercial or political influence, or systemic failures. The selection process can always be an opportunity to curate for reasons other than relevance: for propriety, for institutional self-interest, or for political gain. Together this represents a fundamental vulnerability, one that we can never fully resolve; we can merely build assurances as best we can.

From this perspective, we might see algorithms not just as codes with consequences, but as the latest socially constructed and institutionally managed mechanism for assuring public acumen, a new knowledge logic. We might consider the algorithmic as posed against, and perhaps supplanting, the "editorial" as a competing logic. The editorial logic depends on the subjective choices of experts, themselves made and authorized through institutional processes of training and certification and/or validated by the public through the mechanisms of the market; the algorithmic logic depends on the proceduralized choices of a machine, designed by human operators to automate some proxy of human judgment or identify patterns across collected social traces. Both struggle with, and claim to resolve, the fundamental problem of human knowledge: how to identify relevant information crucial to the public, through unavoidably human means, but in such a way as to be free from human error, bias, or manipulation. Both the algorithmic and editorial approaches to knowledge are deeply important and deeply problematic; much of the scholarship on communication, media, technology, and publics are grappling with one or both techniques and their pitfalls.

A sociological inquiry into algorithms should, at its core, aspire to reveal the complex workings of this knowledge machine – both the process by which it chooses information for users, and the social process by which it is made into a legitimate system. But there may be something, in the end, impenetrable about them. They are designed to work without human intervention, they are deliberately obfuscated, and they work with information on a scale that is hard to comprehend (at least without other algorithmic tools). And we want relief from the duty of being skeptical about information we cannot ever assure for certain. These mechanisms by which we settle (if not resolve) this problem, then, are solutions we cannot merely rely on, but *must* believe in. But we have a tendency to then make them into mechanisms we can believe in, a kind of faith (Vaidhyanathan 2011) that renders it difficult to soberly recognize its flaws and fragilities. So in many ways, algorithms remain outside of our grasp, and they are designed to be. This is not to say that we should not aspire to illuminate their workings and impact. We should. But I think we may also need to prepare ourselves for more and more encounters with the unexpected and ineffable associations they will draw for us, the fundamental uncertainty about who were speaking to or hearing, and the palpable but opaque undercurrents that move quietly beneath knowledge managed by algorithms.

References

Ananny, Mike. 2011. "The Curious Connection Between Apps for Gay Men and Sex Offenders." *The Atlantic*, April 14. http://www.theatlantic.com/technology/archive/2011/04/the-curious-connection-between-apps-for-gay-men-and-sex-offenders/237340/

Anderson, C. W. 2011. "Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of its Public in an Age of Audience." *Journal of Communication* 5: 529-547.

Balka, Ellen. 2011. "Mapping the Body across Diverse Information Systems: Shadow Bodies and They Make Us Human." Paper presented at the annual meeting for the Society for Social Studies of Science, Cleveland, Ohio, November 2-5.

Beer, David. 2009. "Power through the Algorithm? Participatory web cultures and the technological unconscious." *New Media & Society* 11(6): 985-1002.

Berry, David. 2004. "The Contestation of Code." *Critical Discourse Studies* 1(1): 65-89.

Bijker, Wiebe E. 1997. *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA: The MIT Press.

Bowker, Geoffrey. 2008. *Memory Practices in the Sciences*. Cambridge, MA: The MIT Press.

Bowker, Geoffrey, and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press.

boyd, danah. 2010. "Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications." In *Networked Self: Identity, Community, and Culture on Social Network Sites*, edited by Zizi Papacharissi, 39-58. New York: Routledge.

boyd, danah and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication, & Society* 15(5): 662-679..

Boyer, Dominic, and Ulf Hannerz. 2006. "Introduction: Worlds of journalism." *Ethnography* 7(1): 5-17.

Braun, Josh. 2011. "Electronic Components and Human Interventions: Distributing Television News Online." PhD diss., Cornell University.

Brunton, Finn and Helen Nissenbaum. 2011. "Vernacular resistance to data collection and analysis: A political theory of obfuscation." *First Monday* 16(5) http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3493/2955

Bucher, Taina. 2012. "Want to be on the top? Algorithmic power and the threat of invisibility on Facebook." *New Media & Society* 1461444812440159*, first published on April 8, 2012* update

Collins, Harry. 2004. "Interactional Expertise as a Third Kind of Knowledge." *Phenomenology and the Cognitive Sciences* 3(2): 125-143.

Elmer, Greg. 2009. "Exclusionary Rules? The Politics of Protocols." In *Routledge Handbook of Internet Politics*, edited by Andrew Chadwick and Philip N. Howard, 376-383. London: Routledge.

Foucault, Michel. 1988. "Technologies of the Self." In *Technologies of the Self: A Seminar with Michel Foucault*, edited by Luther Martin, Huck Gutman and Patrick Hutton, 16-49. London: Tavistock.

Galloway, Alexander. 2004. *Protocol: How Control Exists after Decentralization*. Cambridge, MA: MIT Press.

Giddens, Anthony. 1991. *The Consequences of Modernity*. Stanford, CA: Stanford University Press.

Gillespie, Tarleton. *Forthcoming. The Politics of Platforms*. New Haven, CT: Yale University Press. *update*

Gillespie, Tarleton. 2012. "Can an Algorithm Be Wrong?" *Limn* 2. *http://limn.it/can-an-algorithm-be-wrong/*

Gillespie, Tarleton. 2010. "The Politics of 'Platforms'." *New Media & Society* 12(3): 1-18.

Gillespie, Tarleton and Hector Postigo. 2012. "Five More Points." *Culture Digitally,* May 4.
http://culturedigitally.org/2012/05/five-more-points/

Gitelman, Lisa, and Virginia Jackson. *forthcoming*. "Introduction." In *"Raw Data" Is an Oxymoron,* edited by Lisa Gitelman. Cambridge, MA: MIT Press.

Goffey, Andrew. 2008. "Algorithm." In *Software studies: a Lexicon*, edited by Matthew Fuller, 15-20. Cambridge, MA: MIT Press.

Granka, Laura. 2010. "The Politics of Search: A Decade Retrospective." *The Information Society* 26(5): 364-374.

Grimmelmann, James. 2008. "The Google Dilemma." *New York Law School Law Review* 53: 939-950.

Grimmelmann. James. 2010. "Some Skepticism About Search Neutrality." In *The Next Digital Decade: Essays on the Future of the Internet*, edited by Berin Szoka and Adam Marcus, 435-459. TechFreedom.

Halavais, Alexander. 2008. *Search Engine Society*. Cambridge: Polity Press.

Hesmondhalgh, David. 2006. "Bourdieu, the Media and Cultural Production." *Media, Culture & Society* 28(2): 211 -231.

Hilgartner, Stephen. 2000. *Science on Stage: Expert Advice as Public Drama*. Stanford, CA: Stanford University Press.

Howard, Philip N. 2005. *New Media Campaigns and the Managed Citizen*. Cambridge: Cambridge University Press.

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

Introna, Lucas and Helen Nissenbaum. 2000. "Shaping the Web: Why the Politics of Search Engines Matters." *The Information Society* 16(3): 169-185.

Ito, Mizuko. 2008. "Introduction." In *Networked Publics*, edited by Kazys Varnelis, 1-14. Cambridge, MA: MIT Press: 1-14.

Keane, John. 1999. "Public Life in the Era of Communicative Abundance." *Canadian Journal of Communication* 24: 165-178.

Langlois, Ganaele. 2012. "Participatory Culture and the New Governance of Communication: The Paradox of Participatory Media." *Television & New Media*. 1527476411433519. *update*

Latour, Bruno. 1988. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1986. "Visualization and Cognition: Thinking with Eyes and Hands." *Knowledge and Society* 6: 1-40.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. "Computational Social Science." *Science* 323: 721-723.

Longford, Graham. 2005. "Pedagogies of Digital Citizenship and the Politics of Code." *Techné: Research in Philosophy and Technology* 9(1): 68-96.

Mackenzie, Adrian. 2005. "The Performativity of Code: Software and Cultures of Circulation." *Theory, Culture & Society* 22(1): 71-92.

Mager, Astrid. 2012. "Algorithmic Ideology: How Capitalist Society Shapes Search Engines." *Information, Communication, and Society,* first published on April 10, 2012 *update*

Manovich, Lev. 1999. "Database as Symbolic Form." *Convergence: The International Journal of Research into New Media Technologies*. 5(2): 80-99.

Morozov, Evgeny. 2011. "Don't Be Evil." *The New Republic*, July 13

http://www.tnr.com/article/books/magazine/91916/google-schmidt-obama-gates-technocrats

Nature. 2007. "A Matter of Trust." *Nature.* 449: 637-638.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin Press.

Pasquale, Frank A. and Oren Bracha. 2008. "Federal Search Commission? Access, Fairness and Accountability in the Law of Search." *Cornell Law Review* 93(6): 1149-1210. http://ssrn.com/abstract=1002453

Pasquale, Frank. 2009. "Assessing Algorithmic Authority," *Madisonian.net* November 18. http://madisonian.net/2009/11/18/assessing-algorithmic-authority/

Pfaffenberger, Bryan. 1992. "Technological Dramas." *Science, Technology, & Human Values* 17(3): 282-312.

Pinch, Trevor, and Wiebe Bijker. 1984. "The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other." *Social Studies of Science* 14(3): 399-411.

Ryfe, David. 2006. "The Nature of News Rules." *Political Communication* 23(2): 203-214.

Schudson, Michael, and Chris Anderson. 2009. "Objectivity, Professionalism, and Truth Seeking in Journalism." In *The Handbook of Journalism Studies*, edited by Karin Wahl-Jorgensen and Thomas Hanitzsch, 88-101. New York: Routledge.

Shapin, Steven. 1995. Trust, Honesty, and the Authority of Science. In *Society's Choices: Social and Ethical Decision Making in Biomedicine*, edited by Ruth Ellen Bulger, Elizabeth Meyer Bobby, and Harvey Fineberg, 388-408. Washington, DC.

Silverstone, Roger. 1994. *Television and Everyday Life*, New York: Routledge.

Slack, Jennifer and J. Macgregor Wise. 2006. "Cultural Studies and Communication Technology." In *The Handbook of New Media: Updated Student Edition*, edited by Leah A. Lievrouw and Sonia Livingstone, 141-162. London: Sage.

Smythe, Dallas. 2001. "On the Audience Commodity and its Work." In *Media and Cultural Studies: KeyWorks,* ed. Meenakshi Durham and Douglas Kellner, 253-279. Malden, MA: Blackwell Publishing.

Stalder, Felix, and Christine Mayer. 2009. "The Second Index. Search Engines, Personalization and Surveillance." In *Deep Search. The Politics of Search beyond Google,* edited by Konrad Becker and Felix Stalder, 98-115. London: Transaction Publishers.

Star, Susan Leigh 1999. "The Ethnography of Infrastructure." *American Behavioral Scientist* 43(3): 377-391.

Sterne, Jonathan. 2003. *The Audible Past: Cultural Origins of Sound Reproduction*. Durham, NC: Duke University Press.

Striphas, Ted. 2010. "How to Have Culture in an Algorithmic Age." *Differences and Repetitions*, July 13. http://www.diffandrep.org/2010/07/13/how-to-have-culture-in-an-algorithmic-age/

Striphas, Ted. 2009. *The Late Age of Print: Everyday Book Culture from Consumerism to Control*. New York: Columbia University Press.

Sunstein, Cass R. 2001. *Republic.com 2.0*. Princeton, NJ: Princeton University Press.

Tuchman, Gaye. 1972. "Objectivity as Strategic Ritual : An Examination of Newsmen's Notions of Objectivity." *American Journal of Sociology* 77(4): 660-679.

Turow, Joseph. 2012. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven, CT: Yale University Press.

Tushnet, Rebecca. 2008. "Power Without Responsibility: Intermediaries and the First Amendment." *The George Washington Law Review* 76(4): 986-1016.

Vaidhyanathan, Siva. 2011. *The Googlization of Everything: (And Why We Should Worry)*. Berkeley, CA: University of California Press.

Van Couvering, Elizabeth. (2010) "Search Engine Bias: The Structuration of Traffic on the World Wide Web." PhD diss., London School of Economics and Political Science.

Varnelis, Kazys, ed. 2008. *Networked Publics*. Cambridge, MA: The MIT Press..

Winner, Langdon. 1978. *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought*. Cambridge, MA: The MIT Press.

Zimmer, Michael. 2008. "The Externalities of Search 2.0: The Emerging Privacy Threats when the Drive for the Perfect Search Engine meets Web 2.0." *First Monday 13*(3). http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2136/1944

Zimmer, Michael. 2007. "Google: "Did You Mean: 'He Invented'?"" May 9. http://michaelzimmer.org/2007/05/09/google-did-you-mean-he-invented/

Zuckerman, Ethan. 2011. "Eli Pariser Talks about the Filter Bubble." *The Boston Phoenix*, May 26. http://thePhoenix.com/Boston/arts/121405-eli-pariser-talks-about-the-filter-bubble/

[1] Google, "Facts about Google and Competition,"

http://www.google.com/competition/howgooglesearchworks.html, accessed May 1, 2012.

Google and Bing have since engaged in a little competitive "signals" war, first when Bing

announced that it uses 1000 signals, and Google soon mentioning that its 200 signals have as

many as 50 variations, bringing their total to 10,000. See Danny Sullivan, "Dear Bing, We Have

10,000 Ranking Signals To Your 1,000. Love, Google," http://searchengineland.com/bing-

10000-ranking-signals-google-55473, accessed May 1, 2012.

[2] Foregoing the possibility of a perfectly transparent algorithm, there are a range of choices open

to a developer as to how straightforward to be. This can be as simple as being more forthright in

the characterization of the tool, or by providing an explanation for why certain ads were served

up with a page, or it could be more careful site documentation. But the tendency is strongly

towards being oblique about the inner workings of an algorithm.

[3] Brian Christian, "The A/B Test: Inside the Technology That's Changing the Rules of

Business," *Wired.com* April 25, 2012. http://www.wired.com/epicenter/2012/04/ff_abtesting/

[4] Christian Sandvig, personal communication.

[5] danah boyd, personal communication.

[6] ABC News, "Napster Faced With Big List, Trick Names," March 5, 2001

http://abcnews.go.com/Entertainment/story?id=108389

[7] http://siri-sayings.tumblr.com/

[8] Jenna Wortham, "Apple Says Siri's Abortion Answers Are a Glitch." *New York Times*, Nov.

30, 2011. http://bits.blogs.nytimes.com/2011/11/30/apple-says-siris-abortion-answers-are-a-

glitch/