

# The Relevance of Algorithms

Tarleton Gillespie

forthcoming, 2013, in *Media Technologies: Essays on Communication, Materiality, and Society*, ed. Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot. Cambridge, MA: MIT Press.

Algorithms play an increasingly important role in selecting what information is considered most relevant to us, a crucial feature of our participation in public life. Search engines help us navigate massive databases of information, or the entire web. Recommendation algorithms map our preferences against others, suggesting new or forgotten bits of culture for us to encounter. Algorithms manage our interactions on social networking sites, highlighting the 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 chatter that's on offer. Together, these algorithms not only help us find information, they also provide 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 on which we depend, 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 mathematical formulas 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 new information as the output. But these are procedures that could conceivably be done by hand -- and in fact were (Light 1999).

But as we have embraced computational tools as our primary media of expression, and have made not just mathematics but *all* information digital, we are subjecting human discourse and knowledge to these procedural logics that undergird all computation. And there are specific implications when we use algorithms to select what is most relevant from a corpus of data composed of traces of our activities, preferences, and expressions.

These algorithms, which I'll call *public relevance algorithms*, are -- by the very same mathematical procedures -- 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 its most relevant components. That we are now turning to algorithms to identify what we need to know is as momentous as having relied on credentialed experts, the scientific method, common sense, or the word of God.

What we need is an interrogation of algorithms as a key feature of our information ecosystem (Anderson 2011), and of the cultural forms emerging in their shadows (Striphas 2010), with a close attention to where and in what ways the introduction of algorithms into human knowledge practices may have political ramifications. This essay is a conceptual map to do just that. I will highlight six dimensions of public relevance algorithms that have political valence:

1. *Patterns of inclusion*: the choices behind what makes it into an index in the first place, what is excluded, and how data is made *algorithm ready*.
2. *Cycles of anticipation*: the implications of algorithm providers' attempts to thoroughly know and predict their users, and how the conclusions they draw can matter.
3. *The evaluation of relevance*: the criteria by which algorithms determine what is relevant, how those criteria are obscured from us, and how they enact political choices about appropriate and legitimate knowledge.
4. *The promise of algorithmic objectivity*: the way the technical character of the algorithm is positioned as an assurance of impartiality, and how that claim is maintained in the face of controversy.
5. *Entanglement with practice*: how users reshape their practices to suit the algorithms they depend on, and how they can turn algorithms into terrains for political contest, sometimes even to interrogate the politics of the algorithm itself.

6. *The production of calculated publics*: how the algorithmic presentation of publics back to themselves shape a public's sense of itself, and who is best positioned to benefit from that knowledge.

Considering how fast these technologies and the uses to which they are put are changing, this list must be taken as provisional, not exhaustive. But as I see it, these are the most important lines of inquiry into understanding algorithms as emerging tools of public knowledge and discourse.

It would also be seductively easy to get this wrong. In attempting to say something of substance about the way algorithms are shifting our public discourse, we must firmly resist putting the technology in the explanatory driver's seat. While recent sociological study of the Internet has labored to undo the simplistic technological determinism that plagued earlier work, that determinism remains an alluring analytical stance. A sociological analysis must not conceive of algorithms as abstract, technical achievements, but must unpack the warm human and institutional choices that lie behind these cold mechanisms. 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 that the seemingly solid algorithm is in fact a fragile accomplishment. It also should remind us that algorithms are now a communication technology; like broadcasting and publishing technologies, they are now "the scientific instruments of a society at large," (Gitelman 2006, 5) and are caught up in and are influencing the ways in which we ratify knowledge for civic life, but in ways that are more "protocological" (Galloway 2004), in other words, organized computationally, than any medium before.

## **Patterns of Inclusion**

Algorithms are inert, meaningless machines until paired with databases on which to function. 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.

For users, algorithms and databases are conceptually conjoined: users typically treat them as a single, working apparatus. And 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." (Manovich 1999, 84). Nevertheless, we can treat the two as analytically distinct: before results can be algorithmically provided, information must be collected, readied for the algorithm, and sometimes excluded or demoted.

### *Collection*

We 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 (Stalder and Mayer 2009): Google, for example, crawls the web indexing websites and their metadata. It digitizes real-world information, from library collections to satellite images to comprehensive photo records of city streets. It invites users to provide personal and social details as part of their Google+ profile. It keeps exhaustive logs of every search query entered and every result clicked. It adds local information based on each user's computer's data. It 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 spot 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 reminds us that the answer to the question, "What does this street corner look like?" has different implications for those who want to go there, those who live there, and those who believe that the answer should not be available in such a public way. But it also reveals what Google thinks of as "public," an interpretation that is being widely deployed across their service.

### *Readied for the algorithm*

"Raw data is an oxymoron" (Gitelman and Jackson 2013). Data is both already desiccated and persistently 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 1977). This means that the information included in the database must be rendered into data,

formalized so that algorithms can act on it automatically. Data must be "imagined and enunciated against the seamlessness of phenomena" (Gitelman and Jackson 2013). 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 sculptures from studying their inverted molds, algorithms can be understood by looking closely at how information must be oriented to face them, how it is made *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. The sociological implications of database design has largely been overlooked; the genres of databases themselves have inscribed politics, as well as making algorithms essential information tools. As Rieder (2012) notes, with the widespread uptake of relational databases comes a "relational ontology" that understands data as atomized, "regular, uniform, and only loosely connected objects that can be ordered in a potentially unlimited number of ways at the time of retrieval," thereby shifting expressive power from the structural design of the database to the query.

Even with these more flexible forms of databases, categorization remains vitally important to database design and management. Categorization is a powerful semantic and political intervention: what the categories are, what belongs in a category, and who decides how to implement these categories in practice, are all powerful assertions about how things are and are supposed to be (Bowker and Star 2000). Once instituted, a category 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 had been accidentally categorized as "adult." Naturally, complex information systems are prone to error. But this particular error also revealed that Amazon's algorithm calculating "sales rank" is instructed to ignore books designated as adult. Even when mistakes are not made, whatever criteria Amazon uses to determine adult-ness are being applied and reified -- apparent only in the unexplained absence of some books and the presence of others.

### *Exclusion and demotion*

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 2006, 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, themselves, refuse to allow data collectors (like search engines) to index their sites. Elmer (2009) 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. But beyond self-exclusion, some information initially collected is subsequently removed before an algorithm ever gets to it. Though large-scale information services pride themselves on being comprehensive, these sites are and always must be censors as well. Indexes are culled of spam and viruses, patrolled for copyright infringement and pornography, and scrubbed of the obscene, the objectionable, or the politically contentious (Gillespie forthcoming).

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 home page generated 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, are reminiscent of twentieth century debates (Tushnet 2008) about the ways choices made by commercial media about who is systematically left out and what categories of speech simply don't qualify can shape the diversity and character of public discourse. Whether enacted by a newspaper editor or by a search engine's indexing tools, these choices help establish and confirm standards of viable debate, legitimacy, and

decorum. But here, the algorithms can be touted as automatic, while it is the patterns of inclusion that predetermine what will or will not appear among their results.

## **Cycles of Anticipation**

Search algorithms determine what to serve up based on input from the user. But most platforms now make it their business to know much, much more about the user than the query she just entered. Sites hope to anticipate the user at the moment the algorithm is called upon, which requires knowledge of that user gleaned at that instant, knowledge of that user already gathered, and knowledge of users estimated to be statistically and demographically like them (Beer 2009) - drawing together what Stalder and Mayer (2009) call the "second index." If broadcasters were providing not just content to audiences but also audiences to advertisers (Smythe 2001), digital providers are not just providing information to users, but also 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.

Much of the scholarship about the data collection and tracking practices of contemporary information providers has focused on the significant privacy concerns they provoke. Zimmer (2008) argues that search engines now not only aspire to relentlessly index the web but also to develop "perfect recall" of all of their users. To do this, information providers must not just track their users, they must also build technical infrastructures and business models that link individual sites into a suite of services (like Google's many tools and services) or an even broader ecosystem (as with Facebook's "social graph" and its "like" buttons scattered across the web), and then create incentives for users to remain within it. This allows the provider to be "passive-aggressive" (Berry 2012) in how it assembles information gathered across many sites into a coherent and increasingly comprehensive profile. Providers also take advantage of the increasingly participatory ethos of the web, where users are powerfully encouraged to volunteer all sorts of information about themselves, and encouraged to feel powerful doing so. As our micro-practices migrate more and more to these platforms, 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 have little or no means to challenge this arrangement even if they do (Turow 2012).

Yet privacy is not the only politically relevant concern. In these cycles of anticipation, it is the bits of information that are most legible to the algorithm, and thus tend to stand in for those users. What Facebook knows about its users is a great deal; but still, it knows only what it is able to know. The most knowable information (geo-location, 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 rendering of that user, a "digital dossier" (Solove 2004) or "algorithmic identity" (Cheney-Lippold 2011) that is imperfect but sufficient. What is less legible or cannot be known about users falls away or is bluntly approximated. As Balka (2011) described it, information systems produce "shadow bodies" by emphasizing some aspects of their subjects and overlooking others. 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 politically productive.

But algorithms are not always about exhaustive prediction; sometimes they are about sufficient approximation. Perhaps just as important as the surveillance of users are the conclusions providers are willing to draw based on relatively little information about them. Hunch.com, a content recommendation service, boasted that they could know a user's preferences with 80-85 percent accuracy based on the answers to just five questions. While this radically boils down the complexity of a person to five points on a graph, what's important is that this is sufficient accuracy for their purposes.<sup>1</sup> Because such sites are comfortable catering to these user caricatures, the questions that appear to 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 these knowable categories. When we encounter these providers, we are encouraged to choose from the menus they offer, so as to be correctly anticipated by the system and provided 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. Designers hope to anticipate users' psycho-physiological capabilities and tendencies, not just specific users' preferences and habits. But in these anticipations, too, implicit and



sometimes political valences can be inscribed in the technology (Sterne 2003): the perceptual or interpretive habits of some users are taken to be universal, contemporary habits are imagined to be timeless, particular computational goals are 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, which is itself reshaping the political landscape. Regardless of their techniques, information providers who amass this data, third-party industries who gather and purchase user data as a commodity for them, and those who traffic in user data for other reasons (that is, credit card companies), have a stronger voice because of it, in both 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 microtargeted campaign ad or the next pop phenomenon.

### **The Evaluation of Relevance**

When users click "Search," or load their Facebook News Feed, or ask for recommendations from Netflix, algorithms must instantly and automatically identify which of the trillions of bits of information best meets the criteria at hand, and will best satisfy a specific user and his presumed aims. While these calculations have never been simple, they have grown more complex as the public use of these services has matured. Search algorithms, for example, once based on simply tallying how often the actual search terms appear in the indexed web pages, now incorporate contextual information about the sites and their hosts, consider how often the site is linked to by others and in what ways, and enlist natural language processing techniques to better "understand" both the query and the resources that the algorithm might return in response. According to Google, its search algorithm examines over two hundred signals for every query.<sup>2</sup>

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 some of the evaluative terms media scholars have already unpacked, 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, or make changes based on evidence from their users, treating quick clicks and no follow-up searches as an approximation, not of relevance exactly, but of satisfaction. To accuse an algorithm of bias implies that there exists an unbiased judgment of relevance available, to which the tool is failing to hew. Since no such measure is available, disputes over algorithmic evaluations have no solid ground upon which to fall back.

### *Criteria*

To be able to say that a particular algorithm makes evaluative assumptions, the kind that have consequences for human knowledge endeavors, might call for a critical analysis of the algorithm to interrogate its underlying criteria. But in nearly all cases, such 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, even leaves the definition of "trending" unspecified. The criteria they use to assess 'trendiness' are only described in general terms: the velocity of a certain term's surge, whether it has appeared in Twitter's Trend list before, and whether it circulates within or spans across clusters of users. 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 that of the designers (Gillespie 2012).

An information provider like Twitter cannot be much more explicit or precise about its algorithm's workings. To do so would give competitors an easy means of duplicating and surpassing its service. It would also require a more technical explanation than most users are prepared for. It would hamper their ability to change their criteria as they need. But most of all, it would hand those who hope to "game the system" a road map for getting their sites to the top of the search results or their hashtags to appear on the Trends list. While some collaborative recommendation sites like Reddit have made public their algorithms for ranking stories and user comments, these sites must constantly seek out and correct instances of organized downvoting, and these tactics cannot be made public. With a few exceptions, the tendency is strongly toward being oblique.<sup>3</sup>

### *Commercial aims*

A second approach might entail a careful consideration of the economic and the cultural contexts from which the algorithm came. Any knowledge system emerges amidst the economic and political aims of information provision, and will be shaped by the aims and strategies of those powerful institutions looking to capitalize on it (Hesmondhalgh 2006). The pressures 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.

For example, early critics worried 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 that providers wanted us to see for financial reasons, to be intermingled with content that the provider had algorithmically selected. But the concern 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 map this complex terrain requires a deep understanding of the economic relationships and social assumptions it represents.

### *Epistemological premises*

Finally, we must consider if the evaluative criteria of the algorithm are structured by specific political or organizational principles that themselves have political ramifications. This is not just whether an algorithm might be partial to this or that provider or might favor its own commercial interests over others. It is a question of whether the philosophical presumptions about relevant knowledge on which the algorithm is founded matter. Some early scholarship looking at the biases of search engines (in order of publication, Introna and Nissenbaum 2000; Halavais 2008; Rogers 2009; Granka 2010) noted some structural tendencies toward what's already popular, toward English-language sites, and toward commercial information providers. Legal scholars debating what it would mean to require neutrality in search results (Grimmelmann 2010;

Pasquale and Bracha 2008) have meant more than just the inability to tip results toward a commercial partner.

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, and so on. 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 a particular logic, one that assumed that pages that include the queried term were likely to be 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, is seen as "ratified" by other users, and is more likely to be relevant to this user as well. By preferring incoming links from sites themselves perceived to be of high-quality, Finkelstein notes, Google had shifted from a more populist approach to a "shareholder democracy:" "One link is not one vote, but it has influence proportional to the relative power (in terms of popularity) of the voter. Because blocks of common interests, or social factions, can affect the results of a search to a degree depending on their relative weight in the network, the results of the algorithmic calculation by a search engine come to reflect political struggles in society" (Finkelstein 2008, 107). When a news discussion site decides what ratio of negative complaints to number of views is sufficient to justify automatically hiding a comment thread, it represents their assessment of the proper volatility of public discourse, or at least the volatility they prefer, for the user community they think they have (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.

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" (Latour 1987; Pinch and Bijker, 1984) 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 as they did yesterday, while the evaluations going on beneath them have been thoroughly remade. The black box metaphor fails us here, as the workings of the algorithm are

both obscured *and* malleable, "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,<sup>4</sup> presenting different rankings to different subsets of users to gain on-the-fly data on speed and customer satisfaction, then incorporating the adjustments preferred by users in a subsequent upgrade.

Each algorithm is premised on both an assumption about the proper assessment of relevance, and an instantiation of that assumption into a technique for (computational) evaluation. There may be implicit premises built into a site's idea of relevance, there may be shortcuts built into its technical instantiation of that idea, and there may be friction between the two.

### **The Promise of Algorithmic Objectivity**

More than mere tools, algorithms are also stabilizers of trust, practical and symbolic assurances that their evaluations are fair and accurate, and 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 carefully crafted fiction. "Search engines pride themselves on being automated, except when they aren't." (Grimmelmann 2008, 950) 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 speech, and it will not appear in France if it promotes Nazism. Yet it's very important for the providers of these algorithms that they seem hands-off. The legitimacy of these functioning mechanisms must be performed alongside the provision of information itself.

The *articulations* offered by the algorithm provider alongside its tool are meant to provide what Pfaffenberger (1992) calls "logonomic control," to define the tool within the practices of users, to bestow the tool with a legitimacy that then carries to the information provided and, by proxy, the provider. The careful articulation of an algorithm as impartial (even when that characterization is more obfuscation than explanation) certifies it as a reliable sociotechnical actor, lends its results relevance and credibility, and maintains the provider's

apparent neutrality in the face of the millions of evaluations it makes. This articulation of the algorithm is just as crucial to its social life as its material design and its economic obligations.

It is largely up to the provider to describe its 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, sometimes in a variety of ways: an algorithm can be defended as a tool for impartial evaluation to those critical of its results, and at the same time be promised as a tool for selective promotion to potential advertisers (Gillespie 2010). As Mackenzie (2005) observes, this process requires more than a single, full-throated description: 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 its algorithm must be repaired both discursively and technically. And users are complicit in this: "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 articulation happens first in the presentation of the tool, in its 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 to what it should be understood as measuring. An equally important part of this discursive work comes in the form of describing how the algorithm works. Even what may seem like a clear explanation of a behind-the-scenes process is not an actually glimpse of a real backstage process, but 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 rendition of its mathematical workings), but was subsequently mythologized -- as the defining feature of the tool, as the central element 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 hundreds of other criteria.

Above all else, the providers of information algorithms must assert that their algorithms are impartial. The performance of *algorithmic objectivity* has become fundamental to the maintenance of these tools as legitimate brokers of relevant knowledge. No provider has been

more adamant about the neutrality of its algorithm than Google, which regularly responds to requests to alter its 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 Google had complied with China's rules before, and they may have pulled out rather than admitting that it was losing to its Chinese competitors. Despite Google's stance, it did alter its search results when complaints arose about a racist Photoshopped image of Michelle Obama at the top of the Image search results; Google also provides a SafeSearch mechanism for keeping profanity and sexual images from minors; and the provider refuses to autocomplete search queries that specify torrent file-trading services. Yet Google regularly claims that it does not alter its index or manipulate its results. Morozov (2011) believes that this 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 algorithmic objectivity plays in many ways an equivalent role to the norm of objectivity in Western journalism. Like search engines, 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, a goal that they admit is messier to pursue than they might appear, and a principle that helps set aside but does not eradicate value judgments and personal politics. 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), to lend public legitimacy to knowledge production tactics that are inherently precarious. "Establishing jurisdiction over the ability to objectively parse reality is a claim to a special kind of authority" (Schudson and Anderson 2009, 96).

Journalist and algorithmic objectivities are by no means the same. 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; their choices represent a careful expertise backed by a deeply infused, philosophical and professional commitment to set aside their own biases and political beliefs. The promise of the algorithm leans much less on institutional norms and trained expertise, and more on a technologically inflected promise of mechanical neutrality. Whatever

choices are made are presented both as distant from the intervention of human hands, and as 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 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 corporate Kool-Aid. It is a deeply ingrained understanding of the public character of Google's information service, one that both influences and legitimizes many of its technical and commercial undertakings, and helps obscure the messier reality of the service it provides.

Still, these claims must compete in the public dialogue with other articulations, which 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 groups of actors who also 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 astute 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 to embody populist, meritocratic ideals, how it comes to appear that way depends not just on its design but also on the mundane realities of news cycles, press releases, tech blogs, fan discussion, user rebellion, and the machinations of the algorithm provider's competitors.

There is a fundamental paradox in the articulation of algorithms. Algorithmic objectivity is an important claim for a provider, particularly for algorithms that serve up vital and volatile information for public consumption. Articulating the algorithm as a distinctly technical intervention helps an information provider answer 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 web-wide collective opinion of a particular website lends it authority. And there is commercial value in claiming that the algorithm returns "better" results than its provider's competitors, which posits customer satisfaction over some notion of accuracy (van Couvering 2007). In examining the articulation 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 with Practice**

Though they could be studied as abstract computational tools, 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 (Couldry 2012). This is especially true when the algorithm is the instrument of a business for which the information it delivers (or the advertisements it pairs with it) is the commodity. If users fail or refuse to fit 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 tactics of users who take them up. 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 rearticulate the algorithms they encounter; and algorithms impinge on how people seek information, how they perceive and think about the contours of knowledge, and how they understand themselves in and through public discourse.

It is important that we conceive of this entanglement not as a one-directional influence, but as a recursive loop between the calculations of the algorithm and the "calculations" 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 choices made by photographers. What people do and do not photograph is already a kind of calculation, though one that is historical, multivalent, contingent, and sociologically informed. But these were not Flickr's only design impulses; sensitivity to photographic practices had to compete with cost, technical efficiency, legal obligation, and business imperatives. And the population of Flickr

users and the types of photos they post 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: some are professional photographers looking for employment, some are seeking communities of like-minded hobbyists, some are simply proud of their work. So just as the algorithm must be sensitive to photographers, 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 invites screenwriters to write in generic ways,<sup>5</sup> the Flickr algorithm may induce subtle reorientations of photographers' practices toward its own constructed logic, that is, toward aspiring to photograph in ways adherent to certain emergent categories, or orienting their choice of subject and composition toward those things the algorithm appears to privilege. "What we leave traces of is not the way we were, but a tacit negotiation between ourselves and our imagined auditors" (Bowker 2006, 6-7).

### *Algorithmically recognizable*

This tacit negotiation consists first and foremost of the mundane, strategic reorientation of practices many users undertake, toward 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 an algorithm. A whole industry, search engine optimization (SEO), promises to boost websites to the top of search results. 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 producers of all sorts orient themselves toward algorithms. When we use hashtags in our tweets -- a user innovation that was embraced later by Twitter -- we are not just joining a conversation or hoping to be read by others, we are redesigning our expression so as to be better recognized and distributed by Twitter's search 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.<sup>6</sup> 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 might find "Britny Speers" recordings but the record industry software would not.<sup>7</sup>

Is this gaming the system? Or is it a fundamental way we, to some degree, orient ourselves toward 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 make ourselves *already algorithmically recognizable* in all sorts of ways. This is not so different than newsmakers orienting their efforts to fit the routines of the news industry: timing a press release to make the evening broadcast, or providing packaged video to a cable outlet hungry for gripping footage, are techniques for turning to face the medium that may amplify them. Now, for all of us, social networks and the web offer some analogous kind of "mediated visibility" (Thompson 2005, 49), and we gain similar benefit by turning to face these algorithms.

### *Backstage access*

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 (as with repeatedly retweeting the same message in the hopes of trending on Twitter). Search engine optimizers and spammers have just as little access, but have developed a great deal of technical skill in divining the criteria beneath the algorithm through testing and reverse-engineering. 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. 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 must not be divulged in a public venue.

Furthermore, some stakeholders are in fact granted access to the algorithm, though under controlled conditions. Advertisers are offered one kind of access to the backstage workings of that system, for bidding on preferred placement. Information providers that offer Application Programming Interfaces (API) to their commercial partners and third-party developers give them a glimpse under the hood, but bind them with 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.

### *Domestication*

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 matter for the way users find information, communicate with others, and know the world around them. But more than that, users express preferences for their favorite search engines, opine about a site's recommendations as being buggy or intuitive or spot on. Some users put great stock in a particular tool, while others come to distrust it, using it warily or not at all. Apple iPhone users swap tips on how to make its Siri search agent speak its repertoire of amusing retorts,<sup>8</sup> then share in the outrage about its answers on hot-button political issues.<sup>9</sup> Satisfied Facebook users today become critics tomorrow when the algorithm behind their news feed is altered in a way that feels economically motivated -- while through and after the uprising, they continue to post status updates. 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). While it is crucial to consider the ways algorithmic tools shape our encounters with information, we should not imply that users are under the sway of these tools. The reality is more complicated, and more intimate.

Users can also turn to these algorithms for a data-inflected reflection; many sites allow us to present ourselves to others and back to ourselves, including our public profile, the performance of our friendships, the expression of our preferences, or a record of our recent activity. Facebook's Timeline feature curates users' activities into chronological remembrances of them; the pleasure of seeing what it algorithmically selects offers a kind of delight, a delight beyond composing the photos and news posts in the first place. But algorithms can also function as a particularly compelling "technology of the self" (Foucault 1988) when they seem to independently ratify one's public visibility. It is now common practice to Google oneself: seeing

me appear as the top result in a search for my name offers a kind of assurance of my tenuous public existence. There is a sense of validation when your pet topic trends on Twitter, when Amazon recommends a book you already love, or when Apple iTunes' "Genius" function composes an appealing 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 are a powerful invitation to understand ourselves through the independent lens they promise to provide.

Algorithms are not just what their 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 the algorithms as truly our own.

### *Knowledge logics*

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 also lead users to internalize their norms and priorities: Bucher (2012) argues 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 recognize that gestures of affinity (such as commenting on a friends' photo) are a key criteria in Facebook's algorithm. Longford (2005) argues that the code of commercial platform "habituates" us, through incessant requests and carefully designed default settings, toward 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 such muscular theories of ideological domination to suggest that algorithms designed to offer relevant knowledge also offer ways of knowing -- and that as they become more pervasive and trusted, their logics are self-affirming. Google's search engine, amidst its 200 signals, does presume that relevant knowledge is assured largely by public ratification, adjusted to weigh heavily the opinions of those who are themselves publicly ratified.

This blend of the wisdom of crowds and collectively certified authorities is Google's 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, Google's solution is operationalized into a tool that billions of people use every day, most of whom experience it as something that simply, and unproblematically, "works." To some degree, Google and its algorithm help assert and normalize this knowledge logic as "right," as right as its results appear to be.

### **The Production of Calculated Publics**

Ito, boyd, and others have recently introduced the term "networked publics" (boyd 2010; Ito 2008; Varnelis 2008) to highlight both the communities of users that can assemble through social media, and the way the technologies structure how these publics can form, interact, and sometimes fall 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). To the extent that algorithms are a key 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 choose to navigate them, could undermine our efforts to be involved citizens. The ability to personalize search results and online news 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 precisely tailored to the user's preferences (by the user, or the provider) such that, in practice, the stories presented 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" (ibid.), where we find only 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, they also traffic in *calculated publics* that 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 invited to feel an affinity -- though the population on which it bases these recommendations is not transparent, and is certainly not coterminous with its entire customer base. When Facebook offers as a privacy setting that a user's information be seen by "friends, and friends of friends," it transforms 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 algorithmically generated 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. I have argued elsewhere that Twitter's Trends algorithm promises users a glimpse of what a particular public (national or regional) is talking about at that moment, but that it is a constructed public, shaped by Twitter's specific, and largely unspecified criteria (Gillespie 2012). Klout promises to measure users' influence across the various social media platforms. Their measures are intuitive in their definition, but completely opaque in their mechanisms. 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 can be the entire user base, sometimes hundreds of millions of people (but only that user base the algorithm provider has access to). Instead, the central problem here is that the intention behind these calculated representations of the public is by no means actuarial. Algorithms that purport to identify what is "hot" engage in a calculated approximation of a public through its participants' traceable activity, then report back to them what they have talked about most. But behind this, we can ask, What is the gain for providers in making such characterizations, and how does that shape what they're measuring? Who is being chosen to be measured in order to produce this representation, and who is left out of the calculation? And perhaps most important, how do these technologies, now not just technologies of evaluation but of representation, help to constitute and codify the 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 "Americans" or what Amazon says teens read are forms of authoritative knowledge that can and will 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, has proven alluring for scholars as well. Social science has turned eagerly toward computational techniques, or the study of human sociality through "big data," (Lazer et. al. 2009; for a critique, see 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 all this data. The approach is seductive: having millions of data points lends a great deal of legitimacy, and the way algorithms seem to spot patterns that researchers 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 a researcher's data has 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, they tell powerful but often mythological stories -- usually in the service of the gods.

Finally, when the data is us, what should we make of the associations that algorithms claim to identify about us as a society -- that we do not know, or perhaps do not want to know? In Ananny's (2011) uncanny example, he noticed the Android Market recommending a sex-offender location app to users who downloaded Grindr, a location-based social networking tool for gay men. He speculates how the Android Market algorithms could have made this association -- one 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 description? 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, between homosexuality and predatory behavior? Zimmer (2007) notes a similar case: where (until Google changed the results) a search for the phrase, "she invented," would return the query, "did you mean 'he invented'?" That is, it did so until Google changed the results. While unsettling in its gender politics, Google's response was completely "correct," 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-- and mistakenly presumed it meant the search query "she invented" was merely a typographical error. Google, here, proves much less sexist than we are. In a response to Ananny's example, Gray has suggested that, just as we must examine algorithms that make associations such as these, we might also inquire into the "cultural algorithms" that these associations represent, (that is, systematically associating homosexuality with sexual predation) across a massive, distributed set of "data points" -- us.

## **Conclusion**

Understanding algorithms and their impact on public discourse, then, requires thinking not simply about how they work, where they are deployed, or what animates them financially. This is not simply a call to unveil their inner workings and spotlight their implicit criteria. It is a sociological inquiry that does not interest the providers of these algorithms, who are not always in the best position to even ask. It requires examining why algorithms are being looked to as a credible knowledge logic, how they fall apart and are repaired when they come in contact with the ebb and flow of public discourse, and where political assumptions might not only be etched into their design, but also constitutive of their widespread use and legitimacy.

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, can only take it for what it's worth. Every public medium previous to this has faced this challenge, from town criers to newspapers to broadcasting. In each, when we turn over the provision of knowledge to others, we are left vulnerable to their choices, methods, and subjectivities. Sometimes this is a positive, providing expertise, editorial acumen, refined taste. But we are also wary of the intervention, of human failings and vested interests, and find ourselves with only secondary mechanisms of social trust by which to vouch for what is true and relevant (Shapin 1995). Their procedures are largely unavailable to us. Their procedures are unavoidably 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 commercial or 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, who are themselves made and authorized through institutional processes of training and certification, or validated by the public through the mechanisms of the market. The algorithmic logic, by contrast, depends on the proceduralized choices of a machine, designed by human operators to automate some proxy of human judgment or unearth 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, 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 grapples with one or both techniques and their pitfalls.

A sociological inquiry into algorithms should 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 algorithms. 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 perhaps more than that, 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 this kind of faith (Vaidhyathan 2011) renders it difficult to soberly recognize their flaws and fragilities.

So in many ways, algorithms remain outside 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 we may also need to prepare ourselves for more and more encounters with the unexpected and ineffable associations they will sometimes draw for us, the fundamental uncertainty about who

we are speaking to or hearing, and the palpable but opaque undercurrents that move quietly beneath knowledge when it is managed by algorithms.

### **Acknowledgements**

I want to thank my colleagues at Culture Digitally for their help and advice on this essay, and the generous support of the Collegium de Lyon and The European Institutes for Advanced Study (EURIAS) Fellowship Programme.

## References

- Ananny, Mike. 2011. "The curious connection between apps for gay men and sex offenders." *The Atlantic*, April 14. Available at <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, OH.
- Beer, David. 2009. "Power through the algorithm? Participatory Web cultures and the technological unconscious." *New Media and Society* 11 (6): 985-1002.
- Berry, David. 2012. "Introduction." In *Life in Code and Software: Mediated Life in a Complex Computational Ecology*, ed. David Berry. London: Open Humanities Press.  
[http://www.livingbooksaboutlife.org/books/Life\\_in\\_Code\\_and\\_Software/Introduction](http://www.livingbooksaboutlife.org/books/Life_in_Code_and_Software/Introduction)
- Bijker, Wiebe E. 1997. *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press.
- Bowker, Geoffrey. 2008. *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.
- Bowker, Geoffrey, and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: 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*, ed. Zizi Papacharissi, 39-58. New York: Routledge.
- boyd, danah, and Kate Crawford. 2012. "Critical questions for big data." *Information, Communication, and 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). Available at <http://firstmonday.org/ojs/index.php/fm/article/view/3493/2955>.
- Bucher, Taina. 2012. "Want to be on the top? Algorithmic power and the threat of invisibility on Facebook." *New Media & Society* 14 (7): 1164-1180.

Cheney-Lippold, John. 2011. "A new algorithmic identity: Soft biopolitics and the modulation of control." *Theory, Culture and Society* 28 (6): 164-181.

Couldry, Nick. 2012. *Media, Society, World: Sociasl Theory and Digital Media Practice*. Cambridge UK: Polity Press.

Elmer, Greg. 2009. "Exclusionary rules? The politics of protocols." In *Routledge Handbook of Internet Politics*, ed. Andrew Chadwick and Philip N. Howard, 376-383. London: Routledge.

Finkelstein, Seth. 2008. "Google, links, and popularity versus authority." In *The Hyperlinked Society: Questioning Connections in the Digital Age*, ed. Joseph Turow and Lokman Tsui, 104-120. Ann Arbor: University of Michigan Press.

Foucault, Michel. 1988. "Technologies of the self." In *Technologies of the Self: A Seminar with Michel Foucault*, ed. 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. 2010. "The politics of 'platforms'." *New Media and Society* 12 (3): 1-18.

---. 2012. "Can an algorithm be wrong?" *Limn* 2. Available at <http://limn.it/can-an-algorithm-be-wrong/>.

---. Forthcoming. *The Politics of Platforms*. New Haven, CT: Yale University Press.

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

Gitelman, Lisa. 2006. *Always Already New: Media, History, and the Data of Culture*. Cambridge, MA: MIT Press.

Gitelman, Lisa, and Virginia Jackson. 2013. "Introduction." In *"Raw Data" Is an Oxymoron*, ed. Lisa Gitelman, 1-14. Cambridge, MA: MIT Press.

Goffey, Andrew. 2008. "Algorithm." In *Software studies: A Lexicon*, ed. 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.

---. 2010. "Some skepticism About search neutrality." In *The Next Digital Decade: Essays on the Future of the Internet*, ed. Berin Szoka and Adam Marcus, 435-459. Washington DC: TechFreedom.

- Halavais, Alexander. 2008. *Search Engine Society*. Cambridge UK: Polity Press.
- Hesmondhalgh, David. 2006. "Bourdieu, the media and cultural production." *Media, Culture and 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, MA: 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*, ed. Kazys Varnelis, 1-14. Cambridge, MA: MIT Press.
- Keane, John. 1999. "Public life in the era of communicative abundance." *Canadian Journal of Communication* 24: 165-178.
- Langlois, Ganaele. 2013. Participatory culture and the new governance of communication: The paradox of participatory media. *Television & New Media* 14 (2): 91-105
- Latour, Bruno. 1986. "Visualization and cognition: Thinking with eyes and hands." *Knowledge and Society* 6: 1-40.
- . 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational social science." *Science* 323: 721-723.
- Light, Jennifer. 1999. "When computers were women." *Technology and Culture* 40 (3): 455-483.
- 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 and Society* 22 (1): 71-92.
- Mager, Astrid. 2012. "Algorithmic ideology: How capitalist society shapes search engines." *Information, Communication, and Society* 15 (5): 769-787.
- 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. Available at

<http://www.newrepublic.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, and Oren Bracha. 2008. "Federal Search Commission? Access, fairness and accountability in the law of search." *Cornell Law Review* 93 (6): 1149-1210. Available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1002453](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1002453).

Pasquale, Frank. 2009. "Assessing algorithmic authority." *Madisonian.net*, November 18. Available at <http://madisonian.net/2009/11/18/assessing-algorithmic-authority/>.

Pfaffenberger, Bryan. 1992. "Technological dramas." *Science, Technology, and 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.

Rieder, Bernhard. 2012. "ORDER BY column\_name: The relational database as pervasive cultural form." Presented at The Lived Logics of Database Machinery conference, London.

Rogers, Richard. 2009. "The Googlization question, and the inculpable engine." In *Deep Search: The Politics of Search Engines beyond Google*, ed. Konrad Becker and Felix Stalder, 173-184. Edison, NJ: Transaction Publishers.

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*, ed. 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*, ed. Ruth Ellen Bulger, Elizabeth Meyer Bobby, and Harvey Fineberg, 388-408. Washington, DC: National Academies Press.

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

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.

Solove, Daniel. 2004. *The Digital Person: Technology and Privacy in the Information Age*. New York: New York University Press.

Stalder, Felix, and Christine Mayer. 2009. "The second index: Search engines, personalization and surveillance." In *Deep Search: The Politics of Search beyond Google*, ed. Konrad Becker and Felix Stalder, 98-115. London: Transaction Publishers.

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

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

---. 2010. "How to have culture in an algorithmic age." *Differences and Repetitions*, July 13. Available at <http://www.diffandrep.org/2010/07/13/how-to-have-culture-in-an-algorithmic-age/>.

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

Thompson, John. 2005. "The new visibility." *Theory, Culture and Society* 22 (6): 31-51.

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: University of California Press.

Van Couvering, Elizabeth. 2007. "Is relevance relevant? Market, science, and war: Discourses of search engine quality." *Journal of Computer-Mediated Communication* 12 (3). Available at <http://jcmc.indiana.edu/vol12/issue3/vancouvering.html>.

---. (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: MIT Press.

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

Zimmer, Michael. May 9, 2007. "Google: "Did you mean: 'He invented'?" Available at <http://www.michaelzimmer.org/2007/05/09/google-did-you-mean-he-invented/>.

---. 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). Available at <http://firstmonday.org/ojs/index.php/fm/article/view/2136/1944>.



---

<sup>1</sup> Ethan Zuckerman, "Eli Pariser talks about the filter bubble." *The Boston Phoenix*, May 26, 2011.

<http://thePhoenix.com/Boston/arts/121405-eli-pariser-talks-about-the-filter-bubble/>.

<sup>2</sup> Google, "Facts about Google and competition," <http://www.google.com/competition/howgooglesearchworks.html>.

Google and Bing have since engaged in a little competitive "signals" war, first when Bing announced that it uses 1000 signals, and Google following that its 200 signals have as many as 50 variations, bringing their total nearer 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>.

<sup>3</sup> Foregoing the possibility of a perfectly transparent algorithm, there is 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 providing more careful site documentation.

<sup>4</sup> 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/business/2012/04/ff\\_abtesting/](http://www.wired.com/business/2012/04/ff_abtesting/).

<sup>5</sup> Christian Sandvig, personal communication.

<sup>6</sup> danah boyd, personal communication.

<sup>7</sup> ABC News, "Napster faced with big list, trick names," March 5, 2001

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

<sup>8</sup> <http://siri-sayings.tumblr.com/>

<sup>9</sup> Jenna Wortham, "Apple says Siri's abortion answers are a glitch." *New York Times*, November 30, 2011.

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