

Who Controls the Public Sphere in an Era of Algorithms

05.16.2016 - Workshop Summary

DATA & SOCIETY RESEARCH INSTITUTE

Event Overview

On February 26, 2016, Data & Society hosted a one-day workshop that brought together a group of stakeholders whose work address themes of power, control, and values in the shaping of the public sphere as algorithms play an increasingly significant role in political and public life. Workshop participants included scholars from diverse academic fields, as well as technologists and civil society representatives who contributed perspectives on policy and engineering to the discussion. An underlying objective throughout the day's discussion was to move beyond academic debates towards engaging in translation work across fields and sectors.

In her opening remarks, danah boyd began by highlighting concerns and tensions raised by critics around the degree of power U.S. companies have in controlling information flows through sites like Facebook and Google. She urged participants to consider what questions arise out of the gaps in knowledge about how these platforms affect political participation and public discourse. Moreover, while a commonality of values is often assumed, she stressed the importance of drawing out the different commitments among participants. For example, while most hold a strong beliefs in democracy, there is not necessarily a consensus on the *boundaries or constraints* of freedom of expression and First Amendment principles. She argued that understanding conflicting values and commitments are just as important as grappling with harms and social analyses.

Ethan Zuckerman, director of the MIT Civic Media Lab, continued to set the frame for the day with a historical lecture on the function the postal service once served in political participation, and asked

participants to consider the effect algorithms are having on current civic discourse. Zuckerman asked participants to consider the difference between potential and real risks, and to think deeply about what and whose values we want guiding policymaking decisions.

Nick Couldry, professor at London School of Economics, and a remote participant in the workshop, sent over a page-length provocation for the day's events. In this piece Couldry urged participants to consider whether "algorithms" or even the "public sphere" are the right framings for this debate, arguing that what we are really analyzing are social processes which become embodied and embedded within computational processes. Building off of some of the work of his colleague (and another remote workshop participant, Jose Van Dijk), Couldry noted that algorithms are "only the last stage of a layered process whereby social life today becomes 'platformed.'" Couldry urged participants to consider as a research question, how algorithms and computed sociality introduce new realities, and how these differences or disruptions should be evaluated (and by whom).

The workshop day was divided into three sessions. The first session asked participants to share their concerns about the function algorithms are playing in public spheres, as they relate to wider values and assumptions about the role media plays within democracies. The second session focused on small group discussions addressing interventions that have been posed thus far. In the last session of the day, participants self-selected their groups to dive deeper into three types of solutions proposed in the second session: (i) professional ethics and association; (ii) regulation and direct rule-making; and (iii) how users/individuals are shaped by and shape algorithmic public spheres.

Conversations throughout the day focused on surfacing participants' concerns as they related to algorithms and publics, highlighting the hidden values that undergird these concerns, and questioning to what degree these anxieties are caused by new developments regarding the use of algorithms in media, or whether they relate more to historical tensions between media and publics that are continuing to persist.

Defining Key Terms:

As a first exercise to ground the day's discussion, workshop participants were asked to define several key terms. The purpose behind the activity was to assess how keywords can be made more accessible within more public conversations. These six terms were: (1) Algorithmic accountability, (2) Transparency, (3) Digital/algorithmic gatekeeping, (4) Values-in-design, (5) Networked publics, and (6) Automated media.

Transparency is a term frequently used by researchers and policymakers seeking more powerful mechanisms for the oversight of algorithms and data. One participant defined the term as a "normative position" to aid accountability, while others defined it more in terms of a set of procedures, stating that transparency is "visible and auditable decision-making" as it relates to "people affected by decisions." Transparency was also defined as necessary for purposes of replication, or the process of

“understanding the inputs, processing procedures, and outputs of the systems,” and “the right amount of information to replicate, audit and understand the source material and outcome.” One tied this more general process to normative ideals, saying transparency is “showing whatever needs to be shown so that those external to a process can evaluate whether it needs ethical standards.” A related idea, **algorithmic accountability**, was also defined by participants in terms of its relation to the *public* or normative ideals, holding that code “whether public or private” should answer not only to its makers but to the public interest.

Values-in-Design brought forward more academic perspectives, due to its more recent ties to media studies and computer science curricula. Participants stressed the political dimensions of information systems, discussing the ways in which technical artifacts “embed normative values through decision choices.” Some stressed that technology is not neutral and that design should explicitly include or exclude certain values. Others defined it more in terms of methodological considerations, such as “a way to trace assumptions and priorities of systems,” and an orientation to guide the design process as a “philosophical commitment engineers might make to soberly consider the impact their design choices might have in real use contexts.”

Digital/algorithmic gatekeeping was defined in terms of how hidden values and workflows affect processes of selection or curation in information and media content flows. This process entails allowing some information at the expense of other content. For instance, one participant defined it as “controlling access, and mediating access to something in such a way that multiple equivalent alternatives don’t exist.” The participants also considered a related term, **automated media**. One participant said it was a “buzzword to fade in less than 6 months.” A number of other participants stressed how automated processes were *taking* over or entering into spaces that had been traditionally been governed by human beings and not machines. One participant said it was “media with social processes rendered mechanical.”

Participants defined **networked publics** in terms of relationships between individuals and groups “created via networks,” “who share overlapping concerns around political/social issues,” where people can have “multiple allegiances” that are “interlinked.” Another participant defined it as “spheres of discourse” which are not *static*, “[such] that moving between one and another poses few barriers.”

All told, as attendees grappled with these terms, it became clear that there were both commonalities within the community and differences of opinion. It also became apparent that certain assumptions and beliefs framed not just the terms that are being deployed but the value ascribed to them.

Session 1: Identifying Concerns and Values

For the first breakout session, participants were asked to articulate the concerns they have related to the role of algorithms in the public sphere, to identify the values that underpin these concerns, and to

list other related or competing issues that come up in these discussions. What follows is an overview of major themes that emerged over the course of the session.

Algorithms as gatekeepers or editors

Participants raised concerns about private companies' ability to manipulate access to information, as well as the ability for governments to ask or pressure companies, behind closed doors, to manipulate algorithms to prioritize some content over others. Participants also raised the concern that democratic values that have been institutionalized within some media ecosystems, such as the responsibility to "inform the public," do not drive tech companies in the same way. Evaluating the goals and interests of these new gatekeepers becomes much more difficult with personalized algorithms, as each user experience is individualized and it becomes more difficult to conduct comparative research. The inaccessibility of algorithms, and the difficulties associated with comparing personalized content also leads to a lack of general awareness about how news and content is produced or filtered through algorithmic processes. This leads to concerns that, as people begin to believe that machine-to-human interactions are replacing human-to-human, this will change perceptions and beliefs about the persuasiveness of messages affecting public sphere discourse and debate. In terms of transparency, it was also noted that even if algorithms were made available to the public, having access to algorithms is not necessarily helpful for evaluating how algorithms affect us, individually or societally.

The role of engineers in embedding values

An ongoing theme was the role of engineers in building particular values into information systems, whether their position constitutes a possible point of intervention, and how present systems differ from, as Jonathan Zittrain expressed it, the less deterministic "roulette" of the pre-algorithmic era. Importantly, how do engineers' values get enacted in the systems they build, and to what extent does this matter? C.W. Anderson noted that engineers' values differ significantly from other fields like journalism, with engineers more geared towards abstract models. Krishna Gummadi seconded this, saying that engineers are driven by different values that may appear similar on the surface, such as measuring "engagement," but that this is done for other purposes such as "optimizing output." The notion of *transparency* also differs significantly for engineers and for users. Tarleton Gillespie made the point that, for engineers, the idea that an algorithm will *reveal* a pattern is a value that trumps other values. Engineers may see this as a scientific process, and not a social or constructed one.

Jonathan Zittrain said that making engineers aware of how their social and cultural values can be inserted into systems could cause additional concerns, such as engineers self-consciously inserting some values into systems and not others. Solon Barocas questioned whether this mattered significantly, noting there is no such thing as a 'neutral' value system so values become encoded into systems by engineers whether they intend this to happen or not.

danah boyd, however, pushed back on the discussion's emphasis on "values-in-design" to say that it is not really a matter of whether engineers have values or not. Rather, algorithmic systems are an example of how systems are co-constructed. Regardless of what values are inserted into the design process, these may reflect larger institutional values (such as those of the company), which in turn reflect broader social and cultural values. Algorithmic systems are also changed as a function of interaction with users who come with their own social and cultural values, and governments that also set standards and policies by which companies need to abide. Tarleton Gillespie seconded this turn towards the role of institutions and co-construction.

Gilad Lotan further underscored the complexity involved not only in co-construction between types of actors, but also within the field of engineering itself. While there are unintended consequences arising from engineers' decisions, he also cautioned against generalizing both the work of engineers and the types of systems with which they work. Instead, he suggested we consider a distinction between "supervised" and "unsupervised" machine learning. While the latter involves "letting the pattern speak for itself," the former requires taking specific data and drawing from it to achieve a specific objective. Different approaches entail different moments where human decision-making comes into play.

Which actors are accountable and to whom?

A major concern throughout the session was how the logic of algorithms and machine learning—which make choices without direct human involvement—can create the impression that "nobody is in charge," making it difficult to assign responsibility or blame for outcomes within a system. Participants lamented the lack of a critical language with which to call out the political power of algorithmic systems that are becoming infrastructural – underlying all of our communications. The gap existing between technology itself – and its impact on social processes – was noted as one obstacle to transparency and accountability. Another focus was pinpointing which institutions should be held accountable and to whom, and which actors should have the responsibility of overseeing the process.

An informal poll of the room showed that most in the group agreed that 'who' we choose to hold accountable is an important concern. However, Pranesh Prakash expressed skepticism over whether it is possible or useful to pin down and delimit a specific set of actors, particularly given the size of the field. To draw a parallel, he noted that journalism has developed clear mechanisms for accountability, but even in that field the rules are violated more often than not. Fred Turner intervened by pointing out that focusing on individuals and algorithms removes an important category of analysis, namely that of institutions that have legal and financial standing. He stressed that neglecting to think in terms of institutions would mean ignoring the existing tools at hand for holding institutions accountable. C.W. Anderson added that the task of assigning accountability has always been complicated to disentangle. There has always been an instinct to assign responsibility to a particular group. It was tentatively suggested that what is novel about the current state of affairs is an anxiety over the *diffusion* of accountability.

The challenge of separating agency across actors, both human and technical, prompted Bodó Balázs to ask what methods may be available to better understand the interactions between these different agents. Nick Diakopoulos suggested that an ethnographic approach would be effective, but is also expensive. On a practical level, Corynne McSherry pointed out that from a legal standpoint what matters is that the corporations that hire engineers are identifiable entities that are accountable to the law. Companies like Google are pressured, for example, to manipulate their search results to make it harder to find infringing content.

Robyn Caplan reminded the group that the issue at hand is not only to identify who is to be held accountable, but also *to whom*, which becomes more difficult as algorithms work to both globalize and localize particular politics and values. Pranesh, however, countered that there are power dynamics to consider, particularly in cases where the people most negatively affected by these decisions are not well positioned to pressure institutions. When it comes to pressure, moreover, Bodó Balázs made the point that the notion of responsibility needs to be distinguished from the separate task of making amends. The findings of researcher Latanya Sweeney, who found that discrimination in online ad delivery was being driven by Google users' own biases, is an illustration of this distinction. While Google was not responsible for the bias, it has the power and thus the responsibility to correct it. Ethan Zuckerman put forward the idea of professional ethics that includes collective identity, values, and social sanctions against wrongdoing, as one possible way of codifying norms.

The effect of personalization on common and public discourse

Participants raised concerns about how algorithmic media can cause divides, or fragment, distribution of information to various publics. Some of these concerns centered on the value of synchronicity and time within the news cycle. As we move further away from broadcast media in which news reaches similar publics at similar times, there are concerns about how such asynchronous news consumption will affect public discourse and debate. There were also concerns raised about how this may create “monopolies of knowledge” and divides between various publics who are exposed to different types of information at different times. Tying into the concept of “values-in-design,” Zeynep Tufekci voiced a concern over how we think about the way values operate in this context, particularly the role played by feedback cycles. She noted that because these algorithms optimize for engagement, there is an incentive for social movements to game them to their political advantage. The wider public then reacts to these influences and a feedback loop is created. danah boyd added that this process is further complicated by the global nature of systems, and the segmentation of users and user behaviors through personalized algorithms.

Participants expressed concerns that algorithmic personalization undermines media diversity, both in content and in terms of exposure to different viewpoints on cultural/political events. These concerns are related to the notion that limiting exposure to different viewpoints can severely hinder political discourse and debate. It also highlights the underlying tension between the role media plays as a

gatekeeper responsible for balancing what individuals may want to consume with broader public interest values such as the need to inform the public. Participants also brought up the issue that the “network effect” makes it difficult to exercise choice in a meaningful way—if Facebook users are unhappy with Facebook’s algorithm, there’s a lack of other equivalent platforms to choose from because Facebook is where their friends and family choose to connect.

Power, Monopolization, and Scale

Related to the notion of algorithmic gatekeeping, participants expressed concerns about how concentrated ownership and monopolization is affecting algorithmic control of information flows. Increasingly, a smaller number of platforms, such as Google, Facebook, and Twitter, are hosting more and more of the world’s news and information, as well as public and private communications. Participants expressed that this has already led to significant information asymmetries between information producers (users/individuals) and information holders (platform owners and service providers, as well as other parties who have access to this data), and can also affect what news and information is favored over platforms, and for whom. Participants also raised the concern that data collection increases the power of the collector relative to the data subject, highlighting the importance for participants of implied values like privacy, transparency, fairness, and non-discrimination.

The discussion also honed in on issues of scale and the spheres within which power dynamics operate. Zeynep Tufekci observed there is a range of contentious issues for which common ground is not easy to find. Frameworks that are normative in the U.S. setting may be unreasonable in a European context, such as the American emphasis on First Amendment rights. Jonathan Zittrain took an informal poll of the room to gauge how many were in support of the European efforts to institute a Right to Be Forgotten. With about half of participants in favor, Zittrain underscored the ambiguities in identifying to common norms and a set of actors to which algorithms can be accountable.

Session 2: Complicating Solutions

For the first breakout session, participants were asked to articulate the concerns they have related to the role of algorithms in the public sphere, to identify the values that underpin these concerns, and to list other related or competing issues that come up in these discussions. What follows is an overview of major themes that emerged over the course of the session.

Group 1

The first group focused on three main topics: (1) treating platforms like Facebook and Google as “information fiduciaries,” and require audit trails; (2) algorithmic literacy, including knowledge of how to read and write (or design) algorithms; and (3) direct rulemaking.

The concept of **information fiduciaries**, which has been mentioned by workshop participant and legal theorist Jonathan Zittrain, along with other scholars like Jack Balkin, would require that online service providers and other online companies that have implications for the First Amendment would have an additional set of regulations and responsibilities in terms of how the company treats information flows. The concept recognizes that platforms that ‘hold’ user information have significant power and should wield it ethically, not unlike other professionals who have increased access to personal information, such as doctors or lawyers. Advocates for this view argued the remedy would hold institutions that have power over information flows more accountable to users. A limitation of this remedy is that it does not take into account market forces or broader public interest concerns beyond privacy and use of personal information. Several examples were mentioned to show that privately owned platforms could act in the public interest following advocacy campaigns, including Google’s decision not to show payday loan ads, and limiting ads for guns or illegal drugs. Participants acknowledged the remedy suffered from problems, in particular, placing even greater amounts of power and trust in the hands of privately owned corporations.

This group also discussed **algorithmic literacy**, in terms of reading, writing and making algorithms. This solution is a response to a perception of growing “illiteracy” and inequality in access to and control of algorithmic mechanisms. Algorithmic literacy programs are, in general, designed to enable more individuals to impact information flows and perceive when or if they or others are being marginalized. However, participants noted that it is unlikely algorithmic literacy programs will reach large swaths of the population they are intended to reach, and that these programs would most likely have limited effects as knowledge about manipulation doesn’t necessarily “empower” better decisions (decision-making power is still largely out of the hands of individuals). The proponents argued that algorithmic literacy could help people targeted by online censorship to show that they had been targeted (though this would require access to mechanisms, as well as data flows). Potential externalities of this position include a false sense of security in a technological environment that is always changing – how would people update their knowledge of algorithms and technology to reflect dominant ideas affecting design?

Lastly, this first group discussed **direct rulemaking** options with advocates of this position stating that working within existing regulatory bodies would allow the public to have a greater say in how algorithms affect information flows. However, others pointed out that regulatory systems can often be over-inclusive – for instance, the implementation of the Right to Be Forgotten court decision in the EU has caused controversy since defining what content qualifies for removal ideally would be done on a case-by-case basis, preferably by a judge; however, outsourcing this decision process to private companies to implement leaves open the potential for over-censorship if companies are concerned about incurring financial penalties. The other issue is that regulations don’t necessarily benefit the public interest – in many cases, private industry has more time and resources to attend regulatory hearings and make sure their side is well represented. One instance where direct rulemaking could be

seen as beneficial is in the regulation of revenge porn – both governments and corporations have implemented rules and structures to prevent its distribution.

Participants in this group also discussed several other additional ideas for remedies or solutions that could work to decentralize or de-monopolize platforms, including the application of anti-trust laws, data ‘portability’ and other regulatory frameworks to increase interoperability between different social networks or other like-applications. Members of this group also raised questions about other market-based solutions, including public sources of funding, as well as how to increase public-private cooperation for governance models (similar to the FTC’s cooperation with the national advertising bureau to govern false advertising).

Group 2

Participants in the second group discussed which existing solutions they agreed with or thought should be removed from the list, and proposed several new ideas within the broader frameworks of direct rulemaking and values-in-design, including instituting accountability processes, third-party interventions, and public ombudspersons.

Bodó Balázs suggested that both transparency efforts and reverse engineering may prove unrealistic solutions and should be removed from the list. Tarleton Gillespie and Corynne McSherry argued that transparency may not be a full solution but can create some pressures internally; however, publics need to manage their expectations about what can happen from transparency measures. Fenwick McKelvey argued that in many cases policy is becoming less transparent (citing, for example, the Trans Pacific Partnership). Zeynep Tufekci said she is not “anti-transparency” but that it is a misleading remedy, arguing that transparency of algorithmic code can lull people into a false sense of security, or it can take away from the value of the system by enabling users to game algorithms. Representing the side of technologists, Gilad Lotan of Betaworks argued against publishing code as a solution, but thinks that transparency mechanisms, such as Google’s Marketplace, have worked to maintain and advance some systems. Robyn Caplan of Data & Society mentioned that transparency mechanisms frequently benefit the needs of private industry over public interest, arguing that we need to ask who are benefitting from transparency mechanisms, and whether those benefits match the aims. Another proposed transparency solution, the idea of implementing **accountable processes**, entails some transparency but only *some* of the time. In this system, secondary controls would be put in place to flag whether problems (such as bias) had entered into the system. If the threshold is triggered, for example, an audit would take place to independently certify whether the algorithm was running in a biased way.

This group primarily discussed options for *direct rulemaking*. The idea of a **public ombudsperson**, or an individual who can respond to public concerns while having access to decision makers, was presented as a form of “mediated transparency.” It was seen as one potential solution to the “expertise problem,” or the idea that meaningful transparency requires both understanding of algorithmic

systems and the ability to press for accountability. This solution also aims to alleviate one of the problems with transparency, which recognizes that corporations have no incentive to reveal proprietary information (their algorithms) to the broader public, and that doing so may be ineffective in any case. Similarly, participants presented the idea of a **third-party** as a form of transparency mediation. An example of this was given by Fenwick McKelvey, who proposed looking at the Office of the Privacy Commissioner in Canada as an example of a type of independent body that presents opinions on privacy-related issues and represents the public.

After a review of the new solutions, the group agreed that most fell under the category of direct rulemaking or regulatory solutions, involving the introduction of a third party or actor to oversee the use of algorithms in information systems. For this reason, the group focused on discussing “Direct Rulemaking” and “Values-in-Design” for the rest of the session. C.W. Anderson argued that most self-regulation approaches have come about because of the threat of direct rulemaking, and thus can be included within this heading. Corynne McSherry of EFF concurred that we needed to get “better” at direct rulemaking, stating that most of our rules have been created to defend against hackers or piracy, but that this has made it more difficult for others to do good accountability work. C.W. Anderson concurred arguing that not only do we not have direct rulemaking, but it is often actively discouraged by some groups (i.e often favors one group’s interest over others).

Participants mentioned several ways governments are already using or evaluating direct rulemaking options to govern algorithmic media. Fenwick McKelvey mentioned recent efforts by the CRTC (Canadian Radio-television and Telecommunications Commission). With the rise of online media, the Canadian government is interested in understanding how algorithms, optimized for American audiences, are having an effect on the promulgation and production of Canadian content, and the CRTC has engaged in investigations into how to govern algorithms to mitigate this issue. The EU court ruling on the Right to be Forgotten was mentioned by the participants, though they did not agree whether this was an example of “bad” or “good” governance of algorithms. Corynne McSherry brought up the notion of “safe harbor” provisions of the DMCA as an example of good rulemaking of platforms by one participant.

The participants in this session also discussed the potential impact of codes of ethics or professional associations for engineers or designers of algorithms. Several issues were noted however, including precarious labor conditions in the tech industry, and the lack of unionization and/or historical ties to union organizing within these professions. The participants noted that it would be easier for engineers/designers to enforce a code of ethics if their labor was organized in such a way, though more research is needed in this area. Additionally, the lack of diversity in engineering/computer science professions means that engineers and tech workers would still be severely limited by their own point of view. Lastly, this group decided that placing too great an emphasis on the role of engineers is a consequence of the ‘fetishization’ of these professions, which should be stopped. This group conceded that engineers have no larger role in the outcome of algorithms than other participants or factors in

the decision-making process. Though we may be able to make engineers aware of their role, they are still mostly executioners of plans designed by teams.

Group 3

Participants in the third group put forward many solutions outside of the list already proposed. Ethan Zuckerman was a proponent of professional ethics and associations, arguing that engineers have a great deal of power in shaping the public sphere. Fred Turner concurred but said that “professional association support” is what matters. Jason Ng from CitizenLab said that associations could serve to review the code and act as a translator between audiences. This act of individuals who could work to ‘translate’ between groups was mentioned before, for instance, in the form of a ‘public ombudsman’ at a newspaper. It was also mentioned by this group in the form of “research and translation between stakeholders.” Cynthia Wong recognized a need for more research into harms, and Solon Borocas said where research is lacking is “tech expertise with domain understanding of what kinds of questions to ask,” which could potentially improve conversations between technical experts and journalists. This form of research would invite third-parties, such as academics and researchers, to play this transparency and translation function. Latanya Sweeney’s work on racial bias in Google’s search engine algorithm was brought up as one example of this type of research. Policy research undertaken by EFF and Human Rights Watch, which use rankings and indexes to improve understanding of issues like ‘transparency’ was also mentioned as a way to translate algorithmic public sphere issues to multiple stakeholders. danah boyd brought up the point that corporations already have some oversight mechanisms, such as boards, which can serve to increase fiduciary responsibility. However, boyd also acknowledged that this would mean implementation would vary from organization to organization.

A number of discussants stressed the need for a more critical media environment around algorithms and algorithmic accountability. Nick Diakopoulos was one proponent of this view, and said that though there has been healthy media criticism in the form of ‘press about press’ in the past, there is little right now in terms of ‘press about algorithms’ or investigative journalism into technological practices. Ethan Zuckerman argued that people who think algorithms may have a strong negative impact should be encouraged to come forward. Karine Nahon argued that alongside more traditional “whistleblowing,” there could be other approaches such as acts of disclosure (though this was not expanded upon), passive or responsive transparency with platforms answering to complaints form acts of disclosure, and creating spaces for stakeholders to convene and share information. Nick Diakopoulos suggested “activist engineered overlays,” an engineering approach where activists create “plug-ins” to undo things, such as reveal biases.

The group then discussed what challenges or barriers to implementation existed for these solutions. Fred Turner brought up the point that algorithms change rapidly over time, requiring modes of reporting that sample routinely over time and that facilitate feedback. Jason Ng pointed out that it’s important to distinguish between companies or actors who aren’t intentionally doing harm but may be

doing so incidentally, and those who are intentionally acting out of malice. Ethan Zuckerman noted that for many cases what matters is not the algorithm itself but the *training data*, and what may be needed is a “right of external algorithm review” that would entail others using different datasets against an algorithm to judge whether it is fair or consistent. Karine Nahon brought up the point that it’s unclear how people are judging what makes an algorithm “good.” For example, when we judge research, we can also check for rigorousness, but what is the equivalent for algorithms? How do we know the algorithm is doing what it’s supposed to do? This entails not just consistency, but what it is *supposed* to represent.

The participants grouped these ideas into several categories: (1) **Reshaping the professional unconscious**, merged together professional ethics and professional association support with increased awareness of the unconscious bias affecting design; (2) **Internal processes and procedures**, such as responsive transparency to complaints, whistleblowing, index or ranking of transparency processes, redress and responsive transparency, board oversight and institutional distrust; (3) **External auditing of algorithms**, in the form of privacy laws and transparency auditing, code review and translation, on-site regulation (which would be similar to what occurs within the fishing industry), auditing human decisions with algorithms, critical evaluation of algorithms; and (4) **Government and multi-stakeholder regulation**, in the form of government management panels, such as the Article 15 Working Group in the EU Data Directive, and other reliability mechanisms such as meetings for stakeholders to discuss common issues.

Session 3: Diving Deeper into Interventions

Participants were again split into three groups, but done so on the basis of self-selection. The conference organizers identified three main topics for discussion that seemed to re-emerge throughout the day: (1) Regulation (government intervention or self-regulation); (2) Professional ethics; and (3) The role of individuals. The groups were given the prompt, “If you had 8 million dollars, how would you use it?” and were asked to do a “deep-dive” interrogation of these proposed interventions.

Regulation/Rule-making

Individuals in this group considered issues related to government intervention and oversight of companies, and were tasked with answering the question: Where would you invest 8 million dollars to make sure that Facebook/Google couldn’t sway or buy an election?

Participants discussed the details of the question and brought up other cases where, for example, Facebook would be able to sway political beliefs or where companies have used the concept of “nudging” for other purposes of social engineering. One participant mentioned how Facebook had partnered with suicide prevention organizations to roll out a feature that would give users more options to express concerns about a friend’s content. Another participant questioned whether

Facebook was engaging in “manipulation” in both the suicide example as well as their “I Voted” experiment, and whether “nudge theory” could be exercised through algorithms without designers or engineers thinking about the potential influence. Zeynep Tufekci argued that not only is “nudging” occurring but it is happening in a “huge information asymmetry environment” wherein users don’t know they are being nudged. This is more powerful than nudge theory as originally proposed by Cass Sunstein (for example, retirement savings policies that “nudge” people toward making better long-term choices—in this example, people are aware of the policies in place).

Pranesh Prakash proposed: (1) Funding research that would lead to a better understanding of the analytical categories of the harms, benefits, or other effects of algorithms; (2) “Forcing” interoperability requirements on other platforms to lower the inter-communication costs with network effects; (3) Getting civil society to push corporations to publish effects of experiments (for example, A/B testing). With regard to regulation, participants also brought up the point that trust is highly localized and cultural—for example, U.S. companies have more power *because* citizens, including the individuals that make up most tech companies in Silicon Valley, are generally more wary of excessive government power than corporate power, which may be the reverse elsewhere. With the example of companies being away to sway or buy an election, Pranesh brought up the point that there are already existing regulations around elections – the question is how to reform election laws and bring them up to date, if they are no longer adequate given the new media environment. Zeynep Tufekci also brought up the point that electoral laws differ from country to country, and, relatedly, that some of the issues may have more to do with features of electoral systems (first-past-the-post vs. other voting systems) rather than algorithms, and may need to be solved outside of technology.

Cynthia Wong, from Human Rights Watch, brought up the idea of third-party audit systems, such as voluntary transparency reports or ranking/indexing projects, which are separate from government regulations or interventions. Tarleton Gillespie proposed the idea of a public ombudsperson who could play an oversight role, and also noted that platforms may balk at regulatory impositions, such as requiring unfiltered content streams, if they make the platform less user-friendly or less visually appealing – a better option would be to allow users to opt-in to algorithmically curated feeds. Fenwick McKelvey brought up the example of ProPublica’s Message Machine, a project wherein ProPublica collected political campaign emails from hundreds of users to reverse engineer political micro-targeting. Another participant mentioned the idea of creating a Nielsen-type group so that people could run surveys inexpensively, independently, and swiftly, as a way of tracking algorithms’ influence in a more dynamic, nimble way and which would enable people to intervene more quickly.

The Level of Individuals

The second group was tasked with discussing what interventions at the individual level might be necessary or useful in understanding and combating the manipulative power of algorithms. Some of the proposed ideas overlap with other categories of discussion (ethics, regulation, ombudsperson, etc.).

The group first discussed whether there *should* be a role for individuals, and what that might look like. Is it an individual's responsibility to master digital literacy with regard to privacy, for example, otherwise it's their fault? Or is it up to companies to be responsible? The participants pointed out that most people aren't going to solve these problems for themselves, even if they wanted to, because many problems are on system level. On the other hand, the role of the individual is already implicated in a way, since individuals are active in these systems and their actions impact the way algorithms function. This brought up a potential role of individuals in distributed research projects in which users could provide information about what content they are receiving (similar to reverse engineering). Elizabeth Eagen brought up the point that distributed research projects could also potentially have the added value of a constituent-building component – bringing in voices and participation of people impacted by algorithms.

Nick Diakopoulos also suggested funding a series of public opinion surveys or other research projects to determine what the public wants to know about algorithms and on which algorithmic platforms, and then funding technologists or journalists to find the answers. Gilad Lotan questioned whether it's important to think about not just individuals as consumers but also individuals as engineers/designers involved in creating algorithms who can question, for example, why a company is choosing to store vast amounts of user data.

Another suggested model for the role of individuals involved a union-like structure for users of a given platform (for example, a union of Facebook users). Participants noted that there's been an important conceptual shift in public understanding: people see the transactional process behind having access to a product for free and in many cases are willing to give up their data as long as they are getting a product they like in return. Aggregate user data is valuable, which means in theory a big enough group of users should be able to wield some power. Problems with this model include the diffusion of power, limited buy-in, and the fact that users may have very different wants or needs from each other. The participants also discussed how groups of individuals could inform collective rule making for algorithms. The example was brought up of an algorithm for ranking schools nationally based on certain criteria that is applied across states uniformly, and where people in one state might organize to create a local version to prioritize their concerns differently from the nation-wide system.

Other ideas to empower individuals included a LEED-like certification process (in which, if there were multiple platforms to choose from, people could choose ones that have something like a “public interest” certification); algorithmic ombudspersons; participatory design processes; and algorithmic literacy projects.

Professional Ethics and Associations

The third group considered the idea of professional ethics for those involved in creating and implementing algorithms, particularly with regard to online news media and social media platforms.

Participants first discussed who would be the subject of such a project—what professions would apply? Should professional ethics apply to a set of code, a particular field, or a type of actor or institution? Jonathan Zittrain proposed new emerging roles, like chief data officers, that could be the focus of ethics training and accountability. Fred Turner expanded on this idea, noting that algorithms are becoming a part of everyday processes and institutions, and suggested developing a CIO version inside institutions, or developing a CIO training within an institution or a position. Being domain specific is also very helpful because algorithms are becoming important across all domains –focusing first on public institutions may be useful.

Jason Ng expanded on this point that algorithms are becoming pervasive: the group of people involved in dealing with algorithms is expanding, therefore making it potentially more difficult to regulate as responsibility becomes diffused. Angèle Christin brought up the fact that much of this has to do with traditional organizational theory: as metrics and algorithms become part of the organization, it becomes more of a managerial problem than a tech problem (who has access to the algorithm, how are decisions made, etc.). Mike Ananny argued for a role similar to an ombudsman in newsrooms: a person who is specifically tasked with an auditing role and who has some level of independence from the organization itself.

Several kinds of ethics intervention were discussed, including 1) explicitly developing a set of certifications and institutions to train an identified set of professionals who work with data, 2) working on branding of data scientists in order to try to get the broader public to understand what data scientists do, 3) intervening in the educational system to train people, and 4) mapping existing best practices.

Solon Barocas brought up the point that many of these platforms are confronted with situations where there isn't a clear answer, and it would be useful if professionals had a means or a forum through which to share best practices (like the Global Network Initiative). Krishna Gummadi asked whether ethics change with the scale of the system – for example, one large company serving many people and with greater resources may be subjected to a different level of scrutiny versus smaller companies. danah boyd mentioned how in the past, EFF was an organization that every company would turn to in order to do an ethics check in the early 90s. In comparison to processes like the IRB, this kind of process would be like ethics consulting, and would be nonbinding.

In addition to possible interventions, the group discussed implicit and explicit barriers to the possible implementation of a code of ethics. Part of this discussion revolved around the current culture of technology companies. danah brought up the point that many people in Silicon Valley switch companies every 1.5-2 years, resulting in a very interconnected culture with lots of informal networks. Jason Ng also added that due to competition, there's also a lot of secrecy—compared the environment to an arms race. This raised the question of whether formalized ethics processes would work well in

this kind of environment. There are significant PR risks, as well as legal and policy considerations that could limit tech professionals' ability to speak openly.

Concluding Remarks

The workshop concluded with a short conversation aimed at getting participants to think about what was missing from the day's discussions and what might be needed in order to move the conversation forward to engage with broader constituencies. Several participants mentioned the importance of bringing in more non-Western voices, particularly given that people around the world are working on these issues and some may be working for companies or governments that have specific, non-neutral positions (including those in the West). Additionally, the issue of disparity came up as it relates not only to different levels of internet access, but also differences in access to and ability to lobby private companies.

While the focus on this workshop was primarily on the ways in which algorithms are being used in the online media environment, several participants questioned how to connect with conversations around other civically relevant areas where algorithmic governance comes into play. One participant noted that it might also be helpful to have this conversation with representatives from other industries—fiscal, extractive, etc.—where accountability frameworks are used. In addition, those participants who work more closely with policymakers suggested focusing on finding and highlighting more stories in which clear harms resulted from algorithmic processes, acknowledging that this kind of strategy can make it easier to show the real impact of algorithms when engaging with policymakers who aren't directly involved in the tech sector. At the same time, others made the point that it's important not only to focus on cases where harm has already and attempt to correct it, but also look for examples where the public may want algorithms to do something “deliberatively and constructively,” and seek to encourage those models.

This document was produced as part of the Algorithms and Publics project at the Data & Society Research Institute. This effort was funded by Open Society Foundations and the John D. and Catherine T. MacArthur Foundation. We are very grateful for the contributions and insights made by our funders, as well as the participants of the workshop on Who Controls the Public Sphere in an Era of Algorithms held in February, 2016.

Data & Society is a research institute in New York City that is focused on social and cultural issues arising from data-centric technological development. Data & Society is committed to identifying issues at the intersection of technology and society, providing research that can ground public debates, and building a network of researchers and practitioners who can offer insight and direction. <http://www.datasociety.net>.

Contact

Robyn Caplan
Robyn@datasociety.net
Data & Society Research Institute
36 West 20th Street, 11th Floor New York, NY 10011
Tel. 646-832-2038
datasociety.net

Notes

Workshop Notes, Publics and Algorithms

¹ Workshop summary compiled by Laura Reed, Alexandra Mateescu, and Robyn Caplan.

A special thank you to the participants involved in this workshop, as well as those workshop participants who offered comment and feedback on this draft of the workshop notes, including Tarleton Gillespie, Taina Bucher, and Nick Couldry.