



## Workshop Discussion Notes: Inequalities & Asymmetries

The Social, Cultural & Ethical Dimensions of “Big Data”

March 17, 2014 - New York, NY

<http://www.datasociety.net/initiatives/2014-0317/>

*This document was produced based on notes that were taken during the Inequalities & Asymmetries workshop as part of “The Social, Cultural, and Ethical Dimensions of ‘Big Data’”. Not all attendees were involved in every part of the conversation, nor does this document necessarily reflect the views and beliefs of individual attendees. All workshop participants were given workshop materials prior to the event to spark discussion. This primer can be found at: <http://www.datasociety.net/pubs/2014-0317/InequalitiesAsymmetriesPrimer.pdf>*

### Overview

The discussion in this workshop focused on inserting power and equity into conversations about the big data phenomenon. Some key issues that were raised include technological due process and the ethical dimensions of data analytics when it is applied to various sector; the necessity of addressing power and inequality instead of just privacy; the issue of informed consent or coercive data collection when it comes to uneducated or marginalized groups; the problem of bad actors or those in power abusing it; and the problem of data retention and persistence, and the possibility of marginalized groups being made invisible by data analytics. What if the big data phenomenon leads not only to the entrenchment of existing inequalities, but also to the emergence of new forms of discrimination? For instance, what if genetic information is used as a barrier to entry when it comes to particular jobs or educational tracks? What about the values embedded in new technological forms? Where do these values come from? For instance, if marginalized groups are underrepresented in the engineer class, could this widen the gulf between marginal communities and more privileged ones?

Structural inequality was raised as a crucial issue, both because of its ability to render certain groups invisible and because of its ability to make certain groups subject to more surveillance than others. Individuals have very few choices when it comes to dealing with private and government power, but marginalized groups have even less control. In any event, as some participants mentioned, the matter of rational choice for individual actors is a red herring because even if underprivileged people understand these issues, they do not necessarily have the luxury of refusal. For example, EBT cards

are now used instead of food stamps, but this means that people's purchases are tracked. However, it is also possible that invisibility will deepen existing structural inequalities. StreetBump is an example of what happens when assumptions about technology, i.e. the prevalence of smart phones, mean that certain groups, like the impoverished or the elderly, are excluded from civic projects. Data invisibility can be as much of a problem as too much surveillance.

## Themes and Discussion Topics

### *Power, not just privacy*

Some participants argued that we should address discrimination directly - focusing on discrimination as more than just a privacy problem. How do we go about addressing these issues without relying on discourses about individual privacy rights? Some members noted that equity, due process, and economic security all need to be considered in addition to privacy.

To help form a cohesive way of addressing inequalities and asymmetries in relation to the big data phenomenon, participants spoke of many different conceptual issues. One idea that emerged was the notion of technological due process. New technologies have led to complications in the legal realm, making it difficult to discern what practices are legal and which are not. We do not yet have a means of addressing due process and information or privacy rights with respect to corporations or government. In the face of these forms of power, individuals have very little choice. Individuals have an asymmetrical power relationship with various institutions, i.e. when it comes to users against Facebook or journalists against government. In the case of already marginal groups, this disparity is even more pronounced and individuals are left with even fewer choices.

Participants discussed the ways that new technologies may exacerbate power differentials. Unexpected use of data has shifted power in unforeseen ways. For example, now major telecommunications companies are also operating as financial institutions. Power is also a problem when it comes to relationships between data collectors and data subjects. While data collected about students may benefit parents and students, how can these large datasets be used when parents don't have the same access to technology? Can researchers develop tools that are effective across multiple platforms and in different kinds of spaces, i.e. the researcher's office or classroom versus the domicile? Technologies can reinforce social power, too. For example, "influencers" on Twitter may receive better customer service in the real world or even free gifts. These

practices may then reinforce their social power and attract even more followers to Twitter accounts.

Power discussions also emerged within the context of service agreements. Even tech-savvy individuals may not read the fine print or fully understand what they are agreeing to. For one, reading privacy terms of service takes a lot of time and effort if platforms do not make this information immediately visible. If individuals feel like they are disempowered, they may not bother to read the terms of service because they don't feel that it will make a difference if they read it or not. Rational ignorance was discussed as a factor as well, especially when it comes to privacy experts and other tech-savvy individuals who neglect to read terms of service. Perhaps privacy experts and others don't read policies because they don't view them as efficient mechanisms for protection. Ignorance is sometimes bliss in the context of power's relationship to privacy.

Some group members emphasized that this is even more of a problem for undereducated groups. How do you taxonomize and expose rational ignorance? A disadvantaged population may not have means of understanding the terms and conditions. How do you clarify that through a relatively simplified taxonomy of what information you're collecting so that both the educated and uneducated consumer can make a decision? An abbreviated notice of terms and services might serve this purpose, but this may also permit websites to obfuscate what they are actually doing. We can't expect people to act like rational individuals, nor can we attribute rationality to institutions and corporations.

Structural forces compel people to use particular technologies – some jobs, for example, require employees to have Facebook accounts. Even providing information for government services leads to data collection. Terms of service inherently involve an asymmetry of power between users and the companies they're bargaining with. Some participants argued that there should be a mechanism by which we can collectively exercise bargaining power with these companies. Such a tool could possibly evaluate terms and enforce certain data limitation and protections individuals might not think about in the moment.

### *Invisibility versus coercion*

Some participants raised structural inequality as a crucial issue, both because of its ability to render certain groups invisible and because of its ability to make some groups subject to more surveillance than others. Many different aspects of invisibility were discussed. One issue that emerged was inequality and a general lack of diversity in the engineer class, which has disproportionate power. The way that technologies are designed may reinforce structural inequalities. Assumptions are built into software

systems and create implicit policies for consumers, users, and even the institutions deploying them. While many technologists -- whether they be at large companies in Silicon Valley or in a college dorm room-- believe that the platforms they create are neutral and without socially contingent values, this is not really the case. Designers may not consider the social implications of data collection methods. One coder, or a team of them, making a choice about encryption or lack thereof in combination with institutional choices and users' desires can have a far-reaching impact. Inequalities in users' power will be reflected in the technology. For example, Facebook was originally designed for Harvard students, an undeniably privileged group. The discussion then moved a more abstract focus, asking questions like, how can traditionally disadvantaged groups gain entry into elite classes, like engineers? Would this have an impact on narrowing existing inequalities?

Marginalized communities are sometimes unaccounted for because they don't supply as much data as privileged groups. StreetBump was discussed as an example of this problem, when elderly and poor neighborhoods were not represented on a city map of potholes because most community members did not have smart phones. Not enough of a certain class had access to the technology that provided the data used for civic improvement, which meant that wealthier areas of the city received faster help than poorer neighborhoods, perhaps deepening existing structural inequalities.

On the other hand, people in vulnerable positions are often compelled to share data by law enforcement, employers, and institutions and almost never get insight into what happens to that data. Nor are they privy to data on those actually collecting the information, which creates a power divide. The matter of rational choice for individual actors is a red herring because even if underprivileged people understand these issues, they do not necessarily have the choice of refusal. For example, EBT cards are now used instead of food stamps, but this means that people's purchases are tracked. We can't assume that people have the ability to opt-out of data collecting technology they don't want to use.

Within a law enforcement domain, participants raised the example of gang databases that are maintained in most major US cities as evidence that opting out of data collection may not be a choice. Individuals may be put on the list for a variety of reasons - search entry terms, gang symbols, sartorial choices, or social networks. Even if you aren't a gang member or have subsequently reformed, your data is still part of the database. The Rampart list from an LA police program, which was disbanded because of concerns regarding racism, still exists. This type of durable information with racial, gender and class assumptions built in means that some groups lack access to hypermobility. We also don't know if this data was collected at gunpoint or under other forms of duress. The concept of "opting out" is an outdated idea in an era where smart

cities that track your movements exist. The Turnstile system in Toronto, for example, is set up to help traffic become more efficient by tracking individuals' phones but the system also indicates your location, where you buy things, and other personal information. Some group members cautioned that we need to pay attention to these systems soon because once they are concretized and built into cities, they are difficult to roll back. Another issue that was discussed is the matter of access to data. Who can use collected data for beneficial purposes and do the data subjects have access to how those data points were created? Some group members argued that we should address inequality of access to data and force greater transparency from government and other major actors.

### *New forms of inequality*

Some members of the group discussed the possibility that data analytics are fomenting new forms of inequality in addition to emphasizing existing structural inequalities. One issue involves the ability of algorithms to bypass protections for marginalized groups. Different sectors tend to have different regulatory protocols, and it is challenging to change structures across varied domains. This may lead to mismatches when rules are applied, where rules are too segmented. The group discussed how these mismatches may occur sector by sector, based on the type of data (public versus private information), the institutions at play, which are often undifferentiated because there are eroding boundaries between different stakeholders, and public versus private entity categories. In practice, inequalities in the analog world are being transferred into the digital space. For example, there is increasing surveillance technology in poor neighborhoods and programs like stop-and-frisk. What can we do to ensure fair interpretation of data and make sure that data is not used in a discriminatory way?

There is a tension between what technology is making possible and the protections put in place to prevent certain types of discrimination. Some group members raised juvenile records as an example of this. As a society, we have generally decided juvenile records should not be accessible to anyone after a given period of time in the interests of not only the former convicts but of society as a whole because it could reinforce a pattern of bias or crime. Now, technology is making that information available again. The collection of data and the birth of new forms of technology may encroach upon earlier protections. While legally certain types of information should not be used as a part of the hiring process (e.g. race, gender, family plans, religious affiliation) and aren't asked about at interviews, technology allows employers to access this information if people have revealed it elsewhere.

New technologies mean that updated rules and protocols may be needed in order to prevent new forms of discrimination from gaining traction. Algorithms, like laws and social theories, are based on assumptions, but technological revolutions tend to upset the underlying facts upon which those assumptions are based and challenge whether they are still useful heuristics for understanding the complexities of society. Some participants again referenced how telecommunications companies may be more like banks now to exemplify this. In a broader example, the industrial revolution challenged notions of freedom of contract and definitions of labor and markets - we realized that we needed things like workplace safety rules, minimum wage, maximum hours and laws regarding child labor and racial discrimination. Today, we can see that in words like “choice.” What do we mean by that? What does a meaningful choice look like? Is choice relevant to privacy and self-management?

Even asymmetries between tech companies and platforms can inhibit individuals’ choices. Asymmetry between companies with dominant market share and users is tricky because there is an economic improvement at scale and big data reinforces that. For example, Google has a better set of data than DuckDuckGo. This creates a barrier to entry for newcomers and makes it harder for any program more geared to consumers to compete. What is the role of policy in leveling the playing field where there’s an oligopoly and where the barrier for entry is too high for newcomers to disrupt that status quo? Being competitive and honoring privacy are in tension in a number of markets. Some group members argued that we should account for and perhaps look to examples coming from the bottom up. For example, in Kenya, citizens are using mobile phones to exchange money because holding cash may be dangerous. It also provides a way for aid to go directly to individuals and entities and possibly go around corrupt intermediaries.

## Further Exploration

In general, the conversations circled around the fact that we seem to lack a vocabulary for discussing inequality and power when it comes to data analytics. Participants worried about the fact that individuals facing government or private power may be further compromised by their race, gender, and class. The digital divide was also discussed as a contributing factor to inequality, indicating that the concerns about inequalities and asymmetries are about structural deficiencies rather than privacy, specifically. Power differentials mean that certain groups have the ability to collect and then utilize data for their own ends, potentially deepening existing inequalities. Generally speaking, individuals lack meaningful choices in the face of government and private power and this fact is exacerbated when it comes to marginalized groups who

are attempting to negotiate the world of data analytics. In order to think more effectively about the role of power, we should perhaps look to terms like choice, private versus public, and notice.

Along with the entrenchment of existing forms of discrimination, the creation of new forms of invidious discrimination was discussed. Algorithms may have the ability to circumvent existing protections. At the moment, discourses about data rely on traditional economic models with rational actors, free choice, and quantitative metrics. These forms of measurement, however, do not consider the role of power. How can we address the ethical dimensions? Different sectors tend to have different regulatory protocols. This may lead to mismatches. The group discussed how these mismatches may occur sector by sector, based on the type of data (public versus private information), the institutions at play, which are often undifferentiated because there are eroding boundaries between different stakeholders, and public versus private entity categories.