[8]

On Knowing Too Much: Technologists' Discourses Around Online Anonymity

Paula Bialski

This chapter focuses on the way technologists approach the data they collect, manage, and analyze; at times feeling they can know too much and see too much about individual users, at times feeling that they know too little, leaving them hungry for gathering more data. Based on preliminary research in San Francisco among data brokers, hackers, activists, privacy teams at large corporations, app developers, bloggers, and cryptographers, I create a typology of characters that handle data. Using the metaphor of weaving, I imagine data as threads that make up a fabric. Using this metaphor, I ask: Who collects these threads? Who gathers them, weaves them, and who cuts them? How are data gathered and treated?

Introduction

There are moments in life when we overhear conversations we do not particularly want to hear. I was sitting on the late train coming home from Lüneburg to Hamburg—with nobody in the train car other than myself, my partner, who was asleep, and two Polish thugs in their thirties. Speaking in Polish, thinking nobody would overhear them, they started discussing, at normal volume, a drug heist they were planning in which they wanted to transport five kilograms of a drug to Sweden by ship using a smuggler. Using my keen understanding of Polish, I started collecting items of information: five kilograms, a boat to Sweden, thousands of euros, endless questions about how to find a smuggler that looked right, that police would not expect, how to not get caught. She should be a small chick. Or a fag. Or a couple. Who would do it? Who could they take advantage of? Even before their sexist and homophobic remarks, I thought to myself, "This has gone too far. I know too much." The train was nearing Hamburg, and I froze, thinking, "What to do now with all this knowledge?" A huge part of me wanted to track them with my iPhone—snap a few photos, record their conversation, and email the information to the Hamburg police, citizen's arrest style. Another part of me didn't want to track and trace them. Why should I be the one with the power to reveal who they were, just because I had this information? Their lack of knowledge of my surveillance of them deemed my tracking practices unjust. Should I strip these two of their intentions and freedoms to disassociate from this drug deal? My partner woke up, and after I told him what was happening, he started getting angry. These guys were being sexist? His chest puffed up, he turned around and started glaring at them. They barely noticed. The train stopped at Hamburg's central station and he stepped out of the train behind them. They still didn't notice. While the story ended with the two thugs leaving the station unaware of our existence, I still couldn't help thinking—what do people do when they really know too much, and what are the affective dimensions among people who know too much?

Each and every person has a particular form or pattern of life. As Gregoire Chamayou explained in *Drone Theory*, our daily actions are repetitive, and our behavior has certain regularities. "For example, you rise at roughly the same hour and regularly make the same journey to work or elsewhere. You frequently meet up with the same friends in the same places. If you are placed under surveillance, it is possible to record all your movements and establish a spatiotemporal map of all your usual doings. Furthermore, by intercepting your telephone calls, observers can superimpose your social network upon this map, determine which are your personal links, and calculate the importance of each one in your life" (Chamayou 2015, 75). As an American army manual explains: "While the enemy moves from point to point, surveillance tracks and notes every location and person visited. Connections between those sites and persons to the target are built, and nodes in the enemy's network emerge" (Chamayou 2015, 76).

These practices, behaviors, daily patterns of doing things are all identifying markers of who we are. Today's digital infrastructures of collection, transmission, analysis, and presentation have made continuous data-mining possible (Couldry and Powell 2014)—continuous mining of what makes up "us." As one of the technologists I met during my fieldwork in San Francisco explained to me, "You would be surprised how unique you really are. All this stuff about us being the same is all wrong when it comes to a data perspective." It is very easy to find that one particular 30-year-old man, born on April 16, who is exactly six meters tall and goes to work at eight in the morning.

Many everyday activities now produce data without requiring human meaning or construction (or even basic consent). Along with the innovation of sensor networks, individuals started producing not "'content' composed of messages containing intrinsic or constructed meaning, but mere data—temperature readings, status updates, location coordinates, tracks, traces and check-ins" (Couldry and Powell 2014, 3). Not one of these individual data

types is necessarily meaningful in itself—but taken together, either through aggregation, correlation, or calculation, such data provide large amounts of information. "We are living through a transformation of governance—both its mechanisms and reference-points—which is likely to have profound implications for practical processes of government and everyday understandings of the social world" (Couldry and Powell 2014, 1).

To tackle this issue, Couldry and Powell explained that emerging cultures of data collection deserve to be examined in a way that foregrounds the agency and reflexivity of individual actors as well as the variable ways in which power and participation are constructed and enacted. While I agree with this statement in that it calls to re-evaluate tensions between structure and agency, plus control and resistance of the actor within our data-driven environment, the "actor" or "data subject" often points inquiry more towards the "user" and less at what is happening behind the screen, within the bodies and minds of the technologists who gather, operate and analyze our data. When Beer (2009, 999) noted that sociology must also "focus … on those who engage with the software in their everyday lives," I would add that sociology must also focus on the way in which software engineers, system admins, and data analysts also envision the everyday lives of users—thus creating a more open inquiry into what types of decision are made, what types of battle are played out, and what obstacles exist in implementing technology that influences our everyday lives. As I will explore in this paper, technologists think about the data they collect, manage, and analyze—at times feeling they can know too much and see too much, at times feeling that they know too little, leaving them hungry for more.

These technologists operating drones or the analysts in San Francisco are the ones who see our patterns of life. Understanding their bird's eye view of us helps us think about their agency, which is in itself "fundamental to thinking about the distribution of data power" (Kennedy et al. 2015, 2). In order to think

through these two dimensions—agency and data power—my research focuses on one key problem today: anonymity.

Bachmann et al. (2014), drawing from Strathern's "Cutting the Network" (1996), have suggested that if you want to understand anonymity, you have to start conceptualizing it as the act of making cuts in identifying markers. To engage in a form of anonymity—such as facelessness, namelessness, or pseudonymity—means that one "cuts" these potentially identifying markers of individuality and difference from a person. "Genuine gains and losses of anonymity occur when a second party links, or fails to link, personal information with the person to whom it belongs" (Ponnesse 2013, 344).

This process of linking and de-linking is, according to Ponnesse, "the result of a specific exercise of control" (2013, 344). Because contemporary societies are increasingly based on networked information and infrastructures, we are facing new questions of how networks of information, properties, and people can be linked or de-linked in order to produce, maintain, abandon or modify anonymity—and who holds that control.

These cuts are today assisted or fully brokered by specific technologies, or specific persons. When these cuts happen—preventing one piece of information from reaching another party (be it a person or server)—anonymity is being played out. A cut could be made by side A of the anonymous interaction, side B of the anonymous interaction, or both, but it is also increasingly other actors who are influencing this cutting moment: for example, system admins, privacy teams, and data analysts. So rather than focusing on the way in which the "user" makes cuts in potentially identifying markers of their own individuality and difference, and rather than focusing on how the "user" creates situational, relational, and partial forms of un-knowability, invisibility, and un-trackability—I wish to focus on people like the drone operator in Chamayou's story, or the technologist I interviewed in San Francisco. For a number of

complex reasons relating to both the material structure and the socio-economic system within which the technologist operates, they are at times a powerful, and at times a powerless, mediating agent in how forms of anonymity become transformed. Moreover, the "technologist" is not just one person—each has their own different agenda. My interest in understanding them—and not the user—also stems from understanding and unpacking the "black box" (Star 1992) of how they often gather and know our "patterns of life," unbeknownst to us.

In order to explore these characters, and how they come to "know too much," I will do a few things in this paper:

Firstly, I will introduce the method in my work, which creates a typology of "characters who know too much." These are the data scientists, technologists, system admins, cryptographers, and app developers who come from various fields and dimensions of the tech industry. Some work for large corporations, some are creating their own start-ups. The reason I create these Weberian ideal types is not only to synthesize and explain the various characters and ideologies of the people who "know too much," but also to camouflage the identity of the subjects I interviewed—focusing less on the person and their identifying markers, and more on their affective dimension of handling data. I realize the methods of anonymizing data while doing a project on anonymity calls for much more explanation, but I will reserve that for another paper, and for the sake of time not take it up here.

Secondly, in order to unpack the actions of these figures who "know too much," I will work with this metaphor of "cutting" and liken data collection to textile production. This approach is inspired by the likes of Donna Haraway with her metaphors of yarn and culture, and more specifically, Janis Jefferies. Jefferies is a British artist and theorist who uses the metaphor of textiles to produce new knowledge around computing and digital technology. She suggests we focus on a material knowledge afforded by textiles, and pattern specifically, where surfaces of patterning

make visible what was once invisible—the conceptual, emotional, textured (Jefferies 2012). In that vein, I imagine data as threads that make up a fabric. Using this metaphor, I ask: Who collects these threads? Who gathers them, weaves them, and who cuts them? How are data gathered and treated? What types of scissors make these cuts? Are they sharp, do they make clean, indiscernable cuts, or are they dull, leaving behind scars and shreds when cutting? Who is the seamstress or tailor that holds the scissors in this cut? Do some hold the scissors, but not make any cuts at all? Why is a cut made in the first place? These seamstresses and tailors have different agendas, and in this paper I will only begin my analysis of the techniques of cutting, showing you who the people are who know too much, and how they deal with what they know.

Introduction to Methods

The fieldwork for this study was conducted for a larger project titled "Reconfiguring Anonymity—Contemporary Forms of Reciprocity, Identifiability, and Accountability in Transformation." This three-year project, which began in August 2015, is a transdisciplinary endeavor bringing together social anthropologists, sociologists, media scientists, and artists to produce new insights into regimes of maintaining, modifying, or abandoning anonymity in contemporary, hybrid online-offline worlds.

I spent nearly two months in San Francisco in August 2015, and during this time I interviewed hackers, activists, privacy teams at large corporations, app developers, bloggers, and cryptographers. In total, I conducted 20 in-depth interviews that lasted from half an hour to a number of days. I also conducted one focus group with the privacy team of a browser provider, attended tech privacy meetups, and gave a public lecture (at the Wikimedia Foundation). This preliminary research then led me to participate in conferences and workshops for technologists, such as the "European Workshop for Trust and Identity" in Vienna in

 December 2016, which brought together technologists working on various topics of transorganizational trust and identity matters.

My interviews were unstructured, and I found my contacts mainly through "hanging out" and asking my interviewees who to talk to next. Our discussions would be mainly around the way in which these actors treat data and the user's personhood, and the tools being developed to help anonymize the user, as well as to help store and encrypt data. We also discussed the future for anonymity or pseudonymity on the net.

Based on this fieldwork, I began seeing conflicts and congruencies in the way in which these technologists or data brokers handled, exchanged, and ethically approached personal data. In this paper, I will limit my ideal types to three "Information Tailors": aggregators, allocators, and analysts. While this paper marks merely the beginning of my analysis, I think these first three "ideal types" can help us think through the distribution of data power and the agency and reflexivity of the technologist in knowing and un-knowing information linked to individual persons while handing data. Again, to help visualize this process, I will liken data collection to textile production.

The Information Tailors

The Aggregator

These agents collect, log, and store data from users. They are a human-machine hybrid. They can be a technical mechanism, like a data packet storage system, which, crudely speaking, collects data packets from any information transferred from one IP address and stores it on a server. Data aggregation is a central structure of the net. Data aggregators can be found all over the net, from Google and online dating websites to small apps. When it comes to knowing too much, data aggregators are the ones who gather and prepare the data—or to use the fabric metaphor—gather tens, thousands, millions, billions of threads

to make yarn or string. A "thread" here is an Item of Information (IOI), and they are combed, separated, and directed towards one server, or data store, or another. While aggregators do not necessarily "know" too much, they collect and log a multitude of data in order to create more knowledge for the users and their platform and product developers.

As one of the data aggregators who was building his own app mentioned, "Humans are giving up their privacy in order to engage in all sorts of beneficial practices (e.g., quantified-self apps)," and as an app designer, he decides which exact data needs to be aggregated, based on the premise of the app (e.g., a running app would aggregate the user's running speed and frequency, their running route, etc.).

This app designer felt that the more data we aggregate, the better—explaining data as a helpful, global brain. He stated: "With any system, once you start recording it, it exists somewhere. So the question is rather, do I trust the overall system to look out for my own interests? And if I don't, how hard am I willing to work to make sure it does? Humans who engage in various practices that they hope is kept private or anonymous should not think about *disengaging* from sharing this information, but must help optimize a central system that can act as a reputation system, but also must collect and protect its user data." Returning to our tailoring metaphor, this app designer was excited to see more data, do more with personal data, while at the same time expressing his general feeling that those giving up their data should trust people like him who thread their data and store it—promising users that he can be trusted to encrypt this data and store it in the right, secure place.

Yet not all data aggregators have the same vision, that "having and collecting more is better." In a lecture given by an operating system developer and system admin, trying to motivate his fellow technologists, he suggested they should "aggregate less" by "logging less." As a background for those who are not familiar

with logging—an essential part of data aggregation—this technologist explained: "Logs are produced by networked services," e.g., a system administrator must log for debugging and have an audit trail and usability studies (how a website gets used), which is useful for analytics. The data that's being logged cover many areas, but in particular, he said, "there are some details which are more identifiable that produce these patterns of information that can be used about someone, but maybe that won't be used by that person. So IP addresses, *who* logged into a machine, there are mail headers that get logged, there are cryptographic parameters that get logged, there is a whole bunch of different stuff that creates finger-printable trails in these data sets."

"Logging less" is part of the practice among system admins and information scientists called "data minimization." It is a theoretical approach that originated in the 1980s along with networked infrastructure and information sciences, and is now seemingly only promoted among "identity management" activists who make it their business to think through personal identity protection and data management. Information scientists Pfitzmann and Hansen explained that this approach "means that first of all, the possibility to collect personal data about others should be minimized. Next within the remaining possibilities, collecting personal data should be minimized. Finally, the length of time collected personal data is stored should be minimized" (Pfitzmann and Hansen 2010, 6).

One aggregator explained: "By default, not even intentionally, we collect data, if we do nothing the data gets stored. But who is allowed to store the data? Deleting is also a conscious decision. And there is also a responsibility issue—who is deciding to delete what? There is an awareness problem."

Speaking passionately, he said,

> I think it's worth thinking about this—people often don't make this simple realization: if somebody is trying to get data about somebody else, from you, there are lots of different

> ways you can resist them getting that data from you. But the simplest way to resist is to not have that data. It's a super stupid thing to come to, but that is the easiest way to resist giving data away to someone else. Just don't have it.

While I do not have time here to explain all the variations of data aggregators, their ideology and agendas, the two I have mentioned show that both sparsely knitted and thickly woven threads of data are in play. The technologists I mentioned favored sparse threads of data out of fear that these threads will fall into the wrong hands. The app developer believed that thickly woven threads would be more useful in making better-quality garments and that trusting the tailor and his technologies will help users share more data, in turn allowing the technologists to know more about the user's patterns of life.

The Allocator

In the game of "knowing too much" about the subject, data allocators are the actors who allocate which threads go into which fabrics. Allocators are usually the privacy teams in companies—the intermediary between the data aggregators, collecting the threads, and the data analysts who weave the various threads together to make a given cloth. The agenda of the allocators is to protect users from "knowing too much" about what data the company collects. These allocators think about their company's user, the image of the company, how much can be "known," and how much should be "left unknown" to the public. Allocators not only make decisions about what to do with the threads being gathered, but about which threads, or items of information, to gather in the first place. During my fieldwork, I learned that various large companies have entire privacy teams that protect the data of users and that these privacy teams act as gatekeepers. Smaller apps, where money is still scarce and the teams are composed of three to five people rather than a few hundred, might not feature a very thorough information allocator. One person can act as an allocator, a designer, and a

manager—having many other jobs—and the amount of effort invested in protecting these data is perhaps not as great as in a privacy team, where the team's sole responsibility is guarding data.

One data analyst I spoke to explained their data aggregation and data allocation team: "There is room for new tools but at the moment a lot of data is just aggregated and not used." The reason it is not being used? The privacy team doesn't allow them to use it. He explained that users must be led to think: "We trust the companies that are aggregating this data, that they won't do anything with it that's too sensitive or gives away our privacy." In this case, the privacy team has to make sure this trust is not breached—allocating only a small ration of data to use, not allowing the users to know too much about other users. We can imagine allocators as gatekeepers in the game of knowing and forgetting.

The Analyst

Much as the name suggests, the analyst analyzes information about a user. They do so for various reasons—in order to gauge the user's engagement in their product or in order to create a new product for their company. The analyst collects various pieces of thread, or items of information—made accessible to them by the allocator—creates the fabric, and assembles the garment. Analysts are at times overwhelmed with the amount of data they have and the amount of knowledge they have about a user. One analyst at a large social networking platform said, "I have more information than you can ever imagine. The amount of things I know about the users is insane. I might think 'Hey, I don't know if I should be tracking this,' but I see that we have to do it. This is something that I have problems with sometimes."

An analyst has access to the data allocators and they weave the threads of data in one way or another to create a certain cloth; here meaning a certain function of an app. This same analyst,

who described himself as a hippie, also explained his moral dilemma: "This is the job you have, to help people make decisions. So the more data, the better. But sometimes we also say 'Why are we doing this? The less data, the better.'" This dilemma seemed to me to be a dilemma of data power and his feeling of control: on the one hand, he was hungry to know more about the user and create more features, and on the other, he felt he was invading the user's privacy.

Another analyst, when speaking about the critique of big data and surveillance, lamented that "full anonymity will not give us precise enough data." What he meant by this statement is that data security means often having less data, deleting it, or storing it securely. But in order to make systems faster, provide more features, and make these systems more usable, he has to have more data, and know more about users. This is the usability-data security tension. "How do we prioritize somebody's need for anonymity over the functionality of a system? Those who design and implement products that deal with the user's privacy often want to do their job well, and in order to do so, need to have the most data possible" (data analyst, San Francisco, August 2015).

This moral dilemma is not one that happens on an everyday basis for these technologists. As another analyst said, "Those who design and implement anonymous systems are just technologists, they aren't philosophers or sociologists, their decisions are not completely thought through—they don't consider all possible thoughts going through their heads. The efficiency of developing a product suffers from not having all eyes on everything. In extreme cases, the developer won't think of all of the problems (i.e., privacy or anonymity issues)."

This again creates an instability in the user's sense of anonymity, or what they think they revealed and what they think their receiving parties know about them. One data analyst I spoke to only collects information about a user's transport routes and cell phone provider. He explained that he often came into conflict

with his privacy team because they did not allocate enough data for him to use. This constant linking and cutting of information is at work in the tension between what the analyst is allowed to know, what they are allowed to invent, what they want to know, and what they feel is personally crossing their moral boundary of "knowing too much."

Conclusion

This paper explores the first stages of analysis in an ongoing description of "people who know too much," in which I hope to unravel the stories of the anonymity tailors who make cuts or links in how anonymity is practiced online. I believe that to fully understand how anonymity is done today, and more generally how personal data are handled, qualitative research should investigate the nature of "cutting" data, the tools that are used to cut and link, and the ideologies and agendas for doing so. Further investigation around big data should also take into account the voices of the software engineers, system admins, and data analysts who affect—both directly and indirectly—the everyday lives of users. Doing so will reveal what types of decisions are made, what types of battles are played out, and what obstacles exist when handling personal data. This description of the affective dimensions of cutting and linking can hopefully further reveal how anonymity is being reconfigured, and explain the entangled weave of the technical and the social.

References

Bachmann, Goetz, Michi Knecht, Gertraud Koch, Nils Zurawski, and Ulf Wuggenig. 2014. "Reconfiguring Anonymity: Contemporary Forms of Reciprocity, Identifiability and Accountability in Transformation." In *Grant Application: Volkswagen Stiftung*. Accessed November 2016. http://reconfiguring-anonymity.net/blog/wp-content/uploads/2015/07/Project_description_web.pdf.

Beer, David. 2009. "Power through the Algorithm? Participatory Web Cultures and the Technological Unconscious." *New Media & Society* 11: 985–1002.

Chamayou, Gregoire. 2015. *Drone Theory*. New York: The New Press.

Couldry, Nick, and Allison Powell. 2014. "Big Data from the Bottom Up." *Big Data & Society* 1 (2), July–December: 1-5.

Jefferies, Janis. 2012. "Pattern, Patterning." In *Inventive Methods: The Happening of the Social*, edited by Celia Lury and Nina Wakeford, 125–136. New York: Routledge.

Kennedy, Helen, Thomas Poell, and José van Dijck. 2015. "Data and Agency." *Big Data & Society* 2 (2), July–December: 1-7.

Pfitzmann, Andreas, and Marit Hansen. 2010. "A Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management." Accessed January 30, 2015. http://www.maroki.de/pub/dphistory/2010_Anon_Terminology_v0.34.pdf.

Ponnesse, Julie. 2013. "Navigating the Unknown: Towards a Positive Conception of Anonymity." *The Southern Journal of Philosophy* 51 (3): 320–344.

Star, Susan Leigh. 1992. "The Trojan Door: Organizations, Work, and the 'Open Black Box'". *Systems Practice* 5 (4): 395–410.

Strathern, Marilyn. 1996. "Cutting the Network." *The Journal of the Royal Anthropological Institute* 2 (3): 517–535.