BIG DATA AND ANTHROPOLOGY: CONCERNS FOR DATA COLLECTION IN A NEW RESEARCH CONTEXT

JUSTIN LANE[1]

Introduction

Traditionally, anthropologists have worked within relatively small groups of individuals (at least relative to the scope of modern big-data analytics). Traditionally, we have known our informants and participants and likely have had some personal relationship or connection with them at some level. Such research has carried with it a practice of protection; anthropologists are keenly aware that we often work in fragile parts of human societies and ask personal questions; therefore we have strived to protect the identities of our informants.

The modern digital environment is one where researchers have access to individuals' data—sometimes deeply personal data—at the touch of a button. Participant anonymity becomes a thorny problem. Given relatively easy access to massive amounts of unique individual data, one can reverse-engineer the data in order to obtain the specific identity of the person, even if their name is changed or erased from that data. In addition, it is often the case that, when a researcher obtains social network data—even when assuming complete consent and legal transfer of the information—information concerning real individuals who have not consented to participate in the research is also transmitted.

This paper argues that we have not given enough thought to such problems as online data becomes of increasing interest to anthropology. I outline some of key issues around data security and big data, and highlight the dilemmas that are likely to confront anthropologists in the near future. My conclusion argues that anthropologists must keep in mind a combination of "traditional" research values as well as the fact that we are in a new frontier of information as we enter the world of "big-data". I finish with some suggestions for participant protection.

[1] *Justin Lane*, Doctoral Candidate, University of Oxford; Postdoctoral Fellow in Modeling Simulation and the Scientific Study of Religion, Institute for the Bio-Cultural Study of Religion, Boston University; Research Associate, Laboratory for Experimental Research of Religion, LEVYNA. Contact: justin.lane@spc.ox.ac.uk, Institute for Cognitive and Evolutionary Anthropology, 64 Banbury Road, Oxford. The author would like to thank Darryl Stellmach and Isabel Bashar for their suggestions on earlier drafts of this paper.

Introduction to data security

The digitization of data represents a new horizon for anthropologists, though one that comes with new challenges and ethical questions. Here, I hope to start what I believe is an overdue conversation on the nature of data security in anthropology. It focuses on two aspects of data security: digital data security, and 'big data'. These two aspects of the modern digital world present different problems for anthropology. I begin by outlining the problems generally before discussing some of the initial steps that we can take in order to safeguard the security of our participants and informants, as well as ensure that our research conforms to the most rigorous ethical standards.

Data security can be defined as protection against unwanted or unauthorized access or use of data or of the systems that store and manipulate data. *Digital* data security is the extension of this concern to include digital forms of data such as those stored on a computer or hard drive, or even the data we transmit by phone or email. Securing this data can involve extremely simple physical methods, like locking our hard drives in a drawer, or electronic methods, such as using complex passwords or encrypting our hard drives. Securing this data is important not only because our digital data include intimate details about our own lives, but also because as anthropologists our data include intimate details about the lives of our informants as well. I discuss some basic precautions in greater detail in later sections and offer simple suggestions as to their use and where one might go to learn more information.

Like so many terms in contemporary media, 'big data' is used so frequently that its meaning is becoming lost on many. 'Big data' refers to massive amounts of electronic data that are indexable and searchable by means of computational systems. Generally, such data are stored on servers and analysed by algorithms, since the amount of information to be analysed is too large to be interpreted initially by human coders. 'Big data' is not only a way of describing large electronic datasets, it is also an industry. Massive dot-com companies like Google, Facebook and Twitter, as well as telecommunications companies, are able to study, measure and even buy and sell our data. This has given rise to companies such as Palantir[2] and products such as IBM's 'Watson'[3] that specialize in making sense of big data.

[2] www.palantir.com
[3] www.ibm.com/smarterplanet/us/en/ibmwatson

Ultimately, big data is human data: it is generated by humans and—key for the discussion at hand—it can be reconstructed to identify those who originally produced the data. Although this topic has been of great importance to contemporary media and political debates, to discuss the use and collection of big data by modern government agencies would go beyond the scope of the current article, even though it is conceivable that this too may impact on the anthropologist's research. Instead, I focus on the ability to identify individuals participating in studies conducted in anthropology departments based only on their data or 'meta-data'. This is a concern not only for anthropologists who might conduct research among vulnerable populations or in repressive regimes, but for anyone who, for example, makes use of social media or cloud services when dealing with participants or participant data.

Meta-data is the data we have about data. For example, rather than recording a conversation (the data), meta-data is the record of how long the conversation lasted and who participated in the conversation. Anthropologists often record both data and meta-data in their research. This is sometimes done directly with our notes or audio recordings, or passively by means of the timestamps generated automatically by our devices and online communication tools. Furthermore, and more to the point, the social sciences are currently moving in a direction of increased digitization and utilizing online social networks either passively or directly in research. Therefore it is important we understand what can happen with the data and meta-data records because this affects the ability of anthropologists to maintain the privacy and protection of their informants and research participants.

How big is big data?

One question that often arises in discussions of 'big data' is how big is 'big'. Largely, this is a semantic issue. Generally, 'big data' refers to datasets that are too large to be manipulated or stored on a single computer. The quantity of such data generally goes far beyond the ability of any one individual or even group of individuals to analyse. For example, one may take weeks to read through the entirety of the New International Version of the Bible (which is roughly 6,000 kb). However, this file is could be one of millions of equally large files stored on a consumer external hard drive available at almost any computer store (a 6TB drive could take 1,000,000 copies of the Bible), representing an amount of text that could not be read within the lifetime of any one individual. To put this in perspective, the ARCUS-b system is the new 'supercomputing'

facility for the University of Oxford and is open for use to researchers in any department; the Institute for Cognitive and Evolutionary Anthropology has been using the system for advanced data analysis and simulation since 2013. This system, while impressive and useful, is not competitive with many modern cloud computation platforms, presently having approximately 1500TB of space.[4]

Currently, there are a number of 'big data' projects in anthropology that really are 'large data', projects such as the SESHAT data archive (Turchinet al. 2012) and to some extent the eHRAF database (Human Relations Area Files, 2015). These databases are archives for works produced by small numbers of individuals, but although they are impressive in their size and scope, they would not be considered 'big data' by most data analysts. Furthermore, the type of data in these aggregation projects rarely if ever records individual-level data points. As such, they represent great archival resources but do not necessarily involve the ethical dilemmas that collecting individuals' personal data would.

Some researchers, however, utilize corporate–academic partnerships or have found ways of obtaining data from websites such as Facebook, Twitter and other online social networking platforms. Other researchers have utilized data produced passively (i.e. without user intervention) by electronic devices such as smart phones and GPS tracking devices for their research (e.g. Backstromet al. 2012; Eagle and Pentland 2005; Gonçalveset al. 2011; Lermanet al. 2010; Leskovec and Horvitz 2008; Pentland 2014; Ritteret al. 2013). This can be done by gaining access to their data servers but can also be done by 'web-scraping' or downloading and restructuring the information (such as usernames, timestamps, posts, replies, 'likes', etc.).

What is obtained?

In principle big data can be almost any type of data; in so far as anthropologists are concerned, it is data about individuals and their beliefs and behaviours. Currently big data ranges from our credit card records, internet usage, social network contacts, phone records to even dating habits (Rudder 2014). However, when it comes to data for human communication, of a sort that would interest anthropologists, big data can provide information about an individual, who they communicate with and what was said. This does more than provide a framework for data analysis – it also provides an opportunity for data reconstruction. By this I mean the use of large

[4] Figure based on the current allocation of 5TB per user and an average of 300 active monthly users.

datasets to interpolate relationships between parts of the data in order to recreate the underlying social networks from which the data were obtained.

For example, when working with Facebook data, the actual social networks of an individual can be downloaded (assuming that the appropriate agreements and consent have been provided by all relevant parties) in a machine-readable format. However, one could also utilize a web-spider in order to harvest the list of friends put on a website and the associated links for that person, then have the program go to each of those links and download the list of friends for each individual, and subsequently for each of those individuals in turn, and so on. Such a process allows us to publicly recreate approximations of social networks without the actual consent of any individual.

Accidental data collection

A second issue with data protection now arises. Specifically, when I grant access for an outside party to gather my data, by implication it also allows them to collect information about other individuals (i.e. my friends). This is the case even though there was no informed consent on the part of any other person besides myself. Given how many friends an individual is likely to have on a social network, what results is that informed consent has not been obtained for most of the 'participants' who have now become part of a study.

What can be done with big data?

Now that there is at least a general overview of what big data is and where it comes from, this leads us to a practical question: what can we do with it?

In theory, we can do almost anything with such data. It can be analysed for correlations, mined for patterns of speech or social interactions, measured for descriptive analyses of sociality, or used to better understand how information is transmitted between individuals (among many other things). A recent monograph has shown the full power of big data to predict human behaviour in its title: *Predictive analytics: the power to predict who will click, buy, lie or die* (Siegel 2013); to these ends, predictive analytics of big data is not a matter of looking at population trends, but of targeting individuals for (mostly) marketing purposes.

What we should concentrate on for the purposes of this article is the more nefarious use of such data. By nefarious, I'm not exclusively referring to its use by 'hackers' or identity thieves

(although this possibility is extremely real). I am generally referring to the use of big data to identify individuals for any reason beyond that intended by the primary researcher who collected the data and the participant who consented to its use in a specific manner.

How can this be done?

Given enough data, individuals can be identified quite easily, even if they are anonymous or have been re-coded in the dataset. Think for a moment about all of your friends, and all of your friend's friends. Imagine that one has all the data necessary to recreate the social network of your friends. Now, given a single extra data point beyond the initial network, you have to specify which individual is X. You know X's friends and you know that X also has a unique data marker (say a specific political leaning). One can then take this unique marker and match it against the publicly available data accessible through basic search engines. After interpolating the unique data marker against a foundation of the social network, one has only a very few statistical targets left.

This doesn't have to be done algorithmically; it can be done manually as well. Given a deep understanding of someone's beliefs, likes and predispositions, one could easily acquire a deep qualitative understanding sufficient to target a needle in a haystack. For example, marketing consultant and entrepreneur Brian Swichkow obtained online quasi-celebrity status recently by playing a prank on his room-mate. Knowing basic information about the latter, such as the fact that he was a professional sword swallower, he was able to quickly construct Facebook ad campaigns that targeted only his room-mate (Holiday 2015). He was helped by the fact that the relevant information was mostly demographic, such as his room-mate's employment and location, and was thus able to create intimate ads that targeted only one of Facebook's 1.5 billion monthly users. Given the ability of one person to target another, the possibilities to reconstruct social network information only become greater given the widely available data-stores on the Internet.

Individual identification data can also be hacked. This is sadly a very real possibility that anthropologists should take seriously. As we move from our pen-and-paper field notes to increasingly digital information storage platforms such as Dropbox[5] or NVivo,[6] we open

[5] Dropbox is computer program that allows anyone to freely store documents and files on their computer and automatically back them up externally 'online' and access them from an internet browser if need be (Dropbox.com).

ourselves up to having our data taken by anyone who can gain access to those digital files. This means that we need to take careful consideration of how and where our data are stored.

Two issues to start with

Given the outline presented above, readers may have many questions that they would like addressed. I will take two issues and discuss them a bit further, namely data security and the propensity to find unexpected results, though they are only two among many important issues. Data security refers to the way in which researchers store, transport and utilize the data they have at their disposal; this is inextricably linked to participant protection. 'Unexpected results' refers to the ability of researchers to discern information about their subjects or participants that they did not intend.

1) Data security

Data security is one of the most talked about and least understood issues in our daily lives. Taking even the simplest steps to secure our data can go a long way. This section will briefly present three ways of increasing data security for our subjects or participants. The first is physical security, which means keeping close tabs on the physical location of our data. The second is encryption, or the process by which we make our data unintelligible to unauthorized entities; this can be done physically or digitally. The last, related to 'physical encryption', is anonymization, or taking steps to ensure that the data cannot be reverse engineered to reveal the identity of someone even if it falls into the wrong hands.

Physical security

One of the first ways to keep data safe is to make sure that we keep them securely stored in a way that only we can access them. This goes not only for digital records, but also physical records such as pictures and field notes, which obviously contain very important and often identifiable information. Although digital data can clearly be hacked and reconstructed, physical records collected by anthropologists often include identifying information; after all, the personal life of people is the professional life of the anthropologist. Therefore, knowing at all times where

[6] NVivo is a popular software program used to store field notes, videos, audio files, transcripts, photos and other materials electronically, thus allowing one to organize and analyze the material.

our physical and digital records are and who has access to them is of the utmost importance. For some, such as myself, big data security means not allowing a computer that has access to the data to be misplaced or stolen. The same goes for the physical data of field notes. In some cases, this can be hard to do; conflict zones often include checkpoints or border crossings where searches and seizures of one's belongings are possible. In such a situation, we can rely on two more concepts in order to protect our data: encryption and anonymization.

Encryption

Encryption is a method by which data are rendered unintelligible without a key. In the digital sphere, even if someone were to get hold of an entire encrypted hard drive, it would be useless without the encryption key. In the physical world, using codes that have keys stored separately can serve the same function.

Digital encryption uses mathematical transformations of information in a computer to make the data appear essentially random. This is done by using extremely large prime numbers which could not be factored due to current computational limitations. That is to say, if your data are encrypted, they will not be understood unless you want them to be. This is used by banks, governments and journalists to secure the information sent between two people. Many operating systems, webservers and software programs have settings that allow you to encrypt your data. For example, the free operating system Linux allows the user to automatically encrypt all the data on their computer. Email systems, such as the free email client 'Thunderbird', allows the sending and receiving of encrypted emails on all operating systems. Being knowledgeable about what you can and cannot encrypt on your own computer is crucial.

When creating our own field notes, we include a great deal of personal information. We can, for the sake of argument, take this information as similar to the information that is collected in online social networks. This information can allow an individual to pinpoint who it was that provided that information by attempting to resituate the information back to its original context. Because the physical location of our field notes is often either tied our field site, or in the field site itself, it is easy to pinpoint the context from which the information was drawn. Allowing our physical notes to become separated from us under any means therefore represents a security breach that can have detrimental effects on the anonymity of our informants and research participants.

This form of security breach also comes in a quasi-digital form. It is common practice at border crossings for individuals to have their personal belongings searched. This provides the potential for physical notes to be taken, especially in areas where governments claim expanded powers of search and seizure and can legally access your belongings (Schoen et al. 2011). This is also the case in contexts of political instability or when internal leadership exerts further endogenous controls on a population. These are examples of situations in which information carried across borders could potentially harm our informants if they are linked. The loss of direct control of our data represents a similar, if not more serious threat to the security of our informants, and this holds whether the data are stored digitally or physically.

In the modern world, we communicate through online social networks (e.g. Facebook, Twitter), email, or programs like Skype or Google Hangouts that allow us to 'call' or 'video chat' over an internet connection. This information is all trackable and—almost definitely—tracked. This can be recorded either as it happens by tracking information as it goes between internet connections, or by compromising the physical security of electronic devices (Nakashima and Wan 2011; Timmer 2015; see Waksman and Sethumadhavan 2011 for an analytical overview from the computer science perspective). This too is an opportunity for our data to be taken out of our control and therefore represents a potential breach of data security. If someone has access to our laptop and contacts, they know who our informants are and could potentially use this information for nefarious means. Encrypting files, hard drives and email accounts is the least we can do to protect our data in this regard.

Securing data

One potential solution to the threats of physical *and* digital security breaches is a form of 'two-factor authentication'. In the digital security world, two-factor authentication is a system that requires two types of authentication before someone has access to the information. Typically, this is something held by the user and something known by the user. For example, a digital two-factor authentication system could potentially be unlocked by physically inserting a USB (aka 'memory stick' or 'flash drive') into the system and then providing a password or answering a question only the user would know (e.g. where one met their spouse, the name of their first pet, etc.); almost anyone who has ever had to deal with a bank online is familiar with such a system. This principle also applies to physical data (field notes, audio/video recording devices, etc.).

Quite simply, the first key can be physical: a lock on baggage or physical storage. Such physical locks are required for a lot of research and are currently used by researchers in our department to store physical files. The second form of physical data security is to split it up. If you have recorded participants' responses, store their responses separately from their names or consent forms (if such a form is collected physically). Anthropologists often change informants' names, but changing one name to another can be potentially useful for finding the source of the information, such as gender, race, or age. Instead, we should anonymize participants and informants using strings of letters and numbers simultaneously. For example, we could anonymize participants based on the site, year and researcher, combined with a unique identifier. So, if your research group knows that you are researcher 633848, your field site is coded as 145 and the year of research is 2014, you could code the information as 633848781452014, where 78 is a code for a specific participant. The researcher can then store a list of names and simply the number 78. Doing this means that, if you lose that piece of data, the receiver would only have access to a list of names and numbers.[7] If one loses the data itself (interview, survey, transcript, etc.), the receiver would have a lot of data, potentially enough to reverse-engineer if enough contextual details are included. However, they would have that data and the number 633848781452014. If by chance someone finds their data, they would also need to know how to decipher the embedded strings of numbers included in 633848781452014 to discern who the individual was by name. This technique can be strengthened by taking the responses of individuals and breaking them into smaller pieces, all stored separately, thereby increasing the difficulty of reconstructing the dataset without knowing the key to its reconstruction.

2) Unexpected results

One other issue that often arises from some research is finding unexpected results. Typically, scientific studies are approved for a specific purpose, and confidential or identifying data are kept secure. Therefore, the usage and results obtained by studying such data are restricted for specific pre-specified use. However, in large datasets we can often find unintentional patterns in

[7] Although speaking with an anthropologist or outsider could be potentially threatening to a participant. It is the researcher's responsibility to understand the risks associated with the research and to make all risks explicitly clear to all participants *prior* to initiating any data collection.

the data. While naturally many such correlations or 'significant results' are spurious at best,[8] some may have vast repercussions. If results are both significant in a statistical sense and imply broader consequences for the subjects of or participants in a study, they could be either beneficial or detrimental in their effects. For example, the data could reveal a pattern that may compromise participants if the information got into the wrong hands, as if they were to reveal the specific importance of an individual as the crux of a social movement, such that opponents of the social movement could target that individual.

Big data answers big questions

The above information may come off as very bleak and negative, as if, by utilizing electronic data, anthropologists are compliant in an Orwellian dystopia. I assure you this does not need to be the case. On the contrary, I am personally very optimistic about the research prospects of big data. This is primarily because anthropologists often ask very big questions concerning human sociality and what sets humanity apart from much of the biological world in very interesting ways. To answer these big questions, we can use big data to acquire better understanding through statistical inference and data analysis. We can also use this information and data to generate further questions about human sociality and how individuals in different cultures act similarly or uniquely.

To an extent, big data overcomes issues of sampling and generalization known to the more empirical schools within anthropology. However, we should not think that the issues it raises are unique to anthropological approaches reliant on large sample sizes. As seen above, the in-depth qualitative data that are the hallmark of more qualitative approaches within anthropology can also be abused in the world of big data.

Conclusion: (towards) a framework for consent and the responsible storage of data

Data security issues are an undeniable aspect of contemporary research in anthropology. As studies relying on large samples (i.e. big data) become increasingly common, a host of ethical issues are raised that are both familiar and new to anthropologists. As researchers, we have the responsibility to be informed about the ways in which we can protect our informants and their data. We also have a responsibility as members of the academic community to push our review

[8] This is so common in datasets with large variables that statistical procedures such as Bonferonni Correction have been devised in order to account for studies that test for many relationships simultaneously (see Abdi 2007).

boards to address the problems while understanding the potential for research that utilizes digital data.

So far this article has largely posed questions and only offered brief answers. However, if we are to tackle the problems noted above systematically in such a way that a single ruler can be used to measure the merits and ethics of a research proposal for use in review committees, we must create some systematic way of approaching these questions.

One proposition has been offered by MIT Media Lab's Sandy Pentland (Pentland 2014), who is a world expert on big data gathering and analytics both online and through 'reality mining' (see Donget al. 2011; Eagle et al. 2009; Eagle and Pentland 2005; Waber et al. 2007). Pentland argues that informed consent and the ability of participants to delete their data at will is the key to protecting the data of individuals. On the whole, Pentland's framework starts a great discussion. However, it is not always enough. As noted earlier, so much big data results in information about non-participants that is passively collected. They must be protected as well, and Pentland's framework (presented at the end of Pentland 2014) is insufficient in this regard.

Ultimately, passive data aggregation is the result of an individual's lack of knowledge about what information is presented publicly about them. As such, I suggest, it is the responsibility of the researcher to protect all data, whether or not they are tied in any way to a direct participant in the research. For example, many in cognitive anthropology use psychometrically validated scales. If these are deployed on a social network platform, both social network data and psychometric data are collected in the research. Currently, the ethics review boards of most institutions feel that encrypting and storing the data is sufficient for protection. I argue that this is not the case because a single lapse in the security of that data results in a breach of both social and psychological data, easily allowing participants to be identified. Therefore, different aspects of a project (i.e. the psychometric data, the social network data, the ethnographic or qualitative data, etc.) should be stored in different physical locations using different storage systems and different encryption methods; this still does not make the data impervious to being 'hacked', but nonetheless one can argue that reasonable and necessary precautions have been taken to protect the identities of those who have entrusted their personal information to us.

Clearly, this article is in no way an attempt to finalize a proposal or even nail down what is likely to be the best course of action; it surely fails in this regard. It is only intended to initiate a conversation among anthropologists about what can happen to our data and therefore to our

informants and participants. This will hopefully result in a more rigorous conversation at the institutional level whereby minimum standards can be implemented that all researchers must adhere to in order to best protect the data of those with whom we work. As always, the onus is on the researcher to take the necessary and sufficient action to ensure the security and safety of themselves and their informants or participants. Traditionally, anthropologists have attempted to prioritize the anonymity and welfare of their informants. This priority is well suited for the age of big data. Our intimate knowledge of communities—and what can happen if anonymity is not maintained—makes anthropologists particularly well-suited for this discussion, not only amongst themselves, but within the greater debates that are currently happening in the academic and corporate worlds.

REFERENCES

Abdi, H. H. (2007). The Bonferonni and Šidák Corrections for Multiple Comparisons. In *Encyclopedia of Measurement and Statistics.* SAGE. doi:10.4135/9781412952644

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., and Vigna, S. (2012). *Four Degrees of Separation* (No. arXiv:1111.4570v3). Retrieved from http://arxiv.org/abs/1111.4570

Dong, W., Lepri, B., and Pentland, A. (2011). Modeling the Co-evolution of Behaviors and Social Relationships Using Mobile Phone Data. In *MUM '11 Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia* (pp. 134–143). Beijing, China: ACM, New York.

Eagle, N., and Pentland, A. (2005). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, *10*(4), 255–268. doi:10.1007/s00779-005-0046-3

Eagle, N., Pentland, A., and Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(36), 15274–8. doi:10.1073/pnas.0900282106

Gonçalves, B., Perra, N., and Vespignani, A. (2011). Modeling users' activity on twitter networks: validation of Dunbar's number. *PloS One*, *6*(8), e22656. doi:10.1371/journal.pone.0022656

Holiday, R. (2015). EXCLUSIVE: Behind the Facebook Prank That Gamed Reddit And Reached 1M Pageviews. *Observer*. Retrieved June 8, 2015, from http://observer.com/2015/05/exclusive-behind-the-facebook-prank-that-gamed-reddit-and-reached-1m-pageviews/

Human Relations Area Files. (2015). *Human Relations Area Files: Cultural Information for Education and Resources*. Retrieved March 12, 2015, from http://hraf.yale.edu/

Lerman, K., Ghosh, R., and Surachawala, T. (2010). Social Contagion: An Empirical Study of Information Spread on Digg and Twitter Follower Graphs. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*.

Leskovec, J., and Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. *Proceeding of the 17th International Conference on World Wide Web - WWW '08*, 915–924. doi:10.1145/1367497.1367620

Nakashima, E., and Wan, W. (2011, September 26). In China, business travelers take extreme precautions to avoid cyber-espionage. *The Washington Post*. Washington, D.C. Retrieved from https://www.washingtonpost.com/world/national-security/2011/09/20/gIQAM6cR0K_story.html

Pentland, A. (2014). *Social Physics: How Good Ideas Spread – The Lessons from a New Science*. London: Scribe.

Ritter, R. S., Preston, J. L., and Hernandez, I. (2013). Happy Tweets: Christians Are Happier, More Socially Connected, and Less Analytical Than Atheists on Twitter. *Social Psychological and Personality Science*. doi:10.1177/1948550613492345

Rudder, C. (2014). *Dataclysm: Who We Are (When We Think No One's Looking)*. New York: Crown Publishers.

Schoen, S., Hofmann, M., and Reynolds, R. (2011). *Defending Privacy at the U.S. Border: A Guide for Travelers Carrying Digital Devices*. Retrieved from https://www.eff.org/files/eff-border-search_2.pdf

Siegel, E. (2013). *Predictive Analytics: the Power to Predict Who Will Click, Buy, Lie, or Die*. Hoboken, NJ: Wiley and Sons.

Timmer, J. (2015, January). Behind the Great Firewall: using my laptop and phone in China. *Ars Technica*. Retrieved from http://arstechnica.com/staff/2015/01/personal-computing-behind-the-great-firewall/

Turchin, P., Whitehouse, H., Francois, P., Slingerland, E., and Collard, M. (2012). A Historical Database of Sociocultural Evolution. *Cliodynamics: The Journal of Theoretical and Mathematical History*, *3*(2), 271–293. Retrieved from http://www.escholarship.org/uc/item/2v8119hf

Waber, B. N., Olgu, D., Kim, T., Mohan, A., Ara, K., and Pentland, A. S. (2007). *Organizational Engineering using Sociometric Badges*. Cambridge, MA. Retrieved from http://ssrn.com/abstract=1073342 or http://dx.doi.org/10.2139/ssrn.1073342

Waksman, A., and Sethumadhavan, S. (2011). Silencing Hardware Backdoors. In *Proceedings of the IEEE Symposium on Security and Privacy* (pp. 49–63). Washington, D.C.: IEEE Computer Society. doi:10.1109/SP.2011.27