Speaking on Behalf of: Representation, Delegation, and Authority in Computational Text Analysis

Eric P. S. Baumer
Computer Science and Engineering
Lehigh University
Bethlehem, PA 18015

Micki McGee
Sociology
Fordham University
Bronx, NY 10458

Abstract

Computational tools can often facilitate human work by rapidly summarizing large amounts of data, especially text. Doing so delegates to such models some measure of authority to speak on behalf of those people whose data are being analyzed. This paper considers the consequences of such delegation. It draws on sociological accounts of representation and translation to examine one particular case: the application of topic modeling to blogs written by parents of children on the autism spectrum. In doing so, the paper illustrates the kinds of statements that topic models, and other computational techniques, can make on behalf of people. It also articulates some of the potential consequences of such statements. The paper concludes by offering several suggestions about how to address potential harms that can occur when computational models speak on behalf of someone.

Computational techniques such as topic modeling (Blei, Ng, and Jordan, 2003; Griffiths and Steyvers, 2004) are designed to identify latent themes in a corpus. Such themes, or "topics," provide an overview both of the corpus as a whole and of individual documents.

Such techniques offer a double-edged sword. On the one hand, they can provide a comprehensive overview of vast corpora in a fraction of the time it would take a human to read equivalent content.

On the other hand, the model can also be used to make authorial/authoritative claims about the corpus. Appeals to algorithmic authority suggest that the model can tell us what the corpus is *really* about (Eslami et al., 2015). The topics observable as latent patterns can be informative even, and perhaps especially, when the author(s) themselves are unaware of them (Youyou, Kosinski, and Stillwell, 2015; Murnane and Counts, 2014). Thus, techniques such as topic modeling can be used to gain insights not only about a corpus of documents but also about the author(s) themselves.

In short, this paper highlights how computational techniques can be used to speak on behalf of someone (else) (cf. Butler, 2001). It does so with a case study that applies topic modeling (Blei, Ng, and Jordan, 2003) to analyze blogs written by parents of children on the autism spectrum. Autism spectrum disorder (ASD) is a complex and often contentious condition. Claims made in this domain bear significance beyond those made in the kinds of corpora usually used to test topic modeling (academic papers, newspaper articles, etc.)

(Blei, Ng, and Jordan, 2003; Griffiths and Steyvers, 2004).

This case study demonstrates two points about speaking-on-behalf-of. First, it offers examples of the kinds of things that techniques such as topic modeling can be made to say on behalf of someone. Second, it leverages the domain of ASD parenting to illustrate the complex and contentious consequences of an algorithmic system making claims about the topics that a person discusses. After considering some ethical dimensions of this specific case, the paper concludes by describing how the issues raised here apply beyond topic modeling and should be considered more broadly in computational analysis of social data.

Background

Sociologies of Translation

Prior work in a number of fields has explored sociological relations among humans and non-human artifacts (Callon, 1986; Latour, 1987, 1992, 1993; Law, 1984). So-called processes of "translation" deal with how technological artifacts speak and/or act on behalf of humans. As an example, consider how municipal authorities might encourage drivers to slow down where construction crews are working (Latour, 1993). They might: station flaggers near the construction zone; install a cardboard cut out of a person holding a flag; install a sign with a flag connected to it; or even install a speed bump in the road. These are instances of a form of translation referred to as *delegation*, which bestows onto a non-human artifact the agency to perform a task otherwise performed by humans.

More relevant here are cases where one group speaks on behalf of another. Callon (1986) details a case of complex discussions over whether or not to restock scallops in a bay off the northwest coast of France. Throughout the process, certain subsets come to be seen as representative of larger groups. Most notably, three researchers studying the bay scallop population end up speaking on behalf of: the larger community of specialists, the fishers who catch and sell the scallops, and even the scallops themselves. Callon refers to this process as *mobilisation*, where spokespersons are selected as representative of a larger group. Similarly, Latour (1987) develops a concept of *inscription*. Codified laboratory procedures are used to create representations (visual or otherwise) of data about observations of a phenomenon.

These representations then speak as authoritative accounts of (i.e., in place of) the phenomenon, rather than needing others to observe the phenomenon itself.

While such prior work provides important conceptual devices, the case at hand differs in an important way from those previously analyzed. Namely, this paper does not deal with data or traces from sociological translations that have already occurred (cf. Latour, 1992; Geiger and Ribes, 2011). Rather, this paper points out the ways that computational techniques *could* be used as an instrument of translation. Put differently, we show how topic modeling, and similar techniques, can be used to speak on behalf of those whose data are being analyzed. We demonstrate both the kinds of things that such techniques can be made to say on someone's behalf, as well as potential consequences of such translations.

Doing Justice in Textual Analysis

Another relevant line of inquiry comes from work in gender studies. Butler (2001) describes the case of an individual (pseudonym John/Joan) who underwent multiple gender reassignments, not all of which occurred at their own volition[1]. In critically considering what textual materials are available for analysis – a mix of research papers in psychological and medical journals, along with popular journalism and books – Butler notes that she has no way of directly knowing the person John/Joan themself.

> "I do not know this person and have no access to this person, I am left to be a reader of a selected number of words, words that I did not fully select, *[words] that were selected for me*, recorded from interviews and then chosen by those who decided to write their articles on this person for journals [....] *I have been given fragments of the person, linguistic fragments* of something called a person." (Butler, 2001, p. 630, emphasis added)

Butler asks, under these circumstances, how one might do justice to this person, who is present only in the form of "linguistic fragments."

This paper asks an analogous question. As with Butler, this paper considers how we might treat responsibly the people discussed in texts we analyze. Similar to her case, the linguistic fragments are not selected by us as authors. However, the linguistic fragments in this case are not chosen by medical researchers or by popular journalists. Instead, they are chosen by an algorithm, specifically, topic modeling. This paper considers the unique consequences of this difference for socially just computational research.

Relationship with Algorithmic Bias

Algorithmic bias deals with situations where the results of a computational system disproportionately impact members of certain groups (Barocas and Selbst, 2016; Sweeney, 2013). Often, the algorithms in question involve AI and/or machine learning models. Disparate impacts can occur along numerous lines, such as race (Sweeney, 2013; Hern, 2015;

Angwin et al., 2016), gender (Lenssen, 2007b,a), sexual orientation (Ananny, 2011), or other attributes.

These issues emerge in part because algorithmic systems can carry an air of objectivity. Prior work has shown how people interacting with such systems can perceive a computer as being a more credible than other people or even than themselves (e.g., Ananny, 2011; Eslami et al., 2015). This prior work documents how people at times perceive such judgments as more objective because they are made by a machine rather than directly by a human. This perception occurs despite significant research into the assumptions, commitments, politics, and other human judgments embedded in these systems (Winner, 1980; Barocas and Selbst, 2016; Sweeney, 2013; Ananny, 2011; Gillespie, 2013).

This paper raises a related but distinct issue. Work on algorithmic bias deals with, for instance, racial biases in the classification of individuals into different risk categories (Angwin et al., 2016). In contrast, this paper focuses on how computational systems can, akin to the processes of translation and mobilisation described above (Callon, 1986; Latour, 1987), speak on behalf of people whose data are being analyzed. This paper considers both the kinds of statements that computational tools can make on behalf of others, and the consequences of such sociological translations.

Case Study: Blogging ASD Parenting

Data

The second author drafted a list of 68 candidate blogs based on their familiarity with the autism blogging community. From this initial list, we removed blogs with technical errors (e.g., the owner had stopped paying hosting fees), blogs that did not focus on autism and/or parenting, and blogs with fewer than ten posts.

This process resulted in a list of 46 blogs with publicly accessible content that focused on issues related to parenting ASD children. These blogs were scraped using a custom scraper based on scrapy (2018) and hand-tuned to each blog's HTML structure. Doing so generated 31,976 documents with a total of 17,273,079 words (words per document M=540.2, Mdn=430), of which 8,136,071 were non-stopword tokens.

Methods

A topic model was fitted to these data using LDA (Blei, Ng, and Jordan, 2003) as implemented in MALLET (Mc-Callum, 2002) via the r-mallet package (Mimno, 2013). Different numbers of topics were tested between 20 and 75. A model with 50 topics was selected based on coherence (Lau, Newman, and Baldwin, 2014; Lau and Baldwin, 2016) and on manual review.

The stopword list began with a small set of high frequency articles, determiners, conjunctions, prepositions, etc., as well as related contractions. The stopword list intentionally omitted gendered pronouns (she, him, etc.). Contrary to Schofield, Magnusson, and Mimno (2017), we included several proper names in the stopword list. Initial training indicated that many topics were composed of proper names

---

[1]Applying queer theory to autism and neurodiversity (Yergeau, 2017), while generative, exceeds this paper's scope

that were unique to one blog. Because we sought to identify linguistic patterns that transcended multiple blogs, the stopword thus included proper names (and their possessive versions) that appeared in more than 1% of the documents.

Topics and Their Interpretation

This section shows several topics resulting from the above analysis. For each, it lists the top 20 highest probability terms for the topic and describes the content of documents with high proportions of the topic. It does *not* offer extended quotes from any document, as performing an internet search for such quotes would easily identify the bloggers – and identifying individuals with topics is the exact ethical issue this paper raises. Where possible, very short phrases are quoted directly, but only when we determined that they would not enable easy identification of the blogger. See the discussion for further ethical considerations.

Topics were chosen for this section in two ways. First, we describe topics with the highest coherence (Lau, Newman, and Baldwin, 2014; Lau and Baldwin, 2016). Since coherence aligns with human judgments about topic quality, researchers would likely use coherence to guide examination of topic modeling results. Second, the authors hand-picked a set of topics around a particularly contentious issue for ASD parents: early childhood vaccination. All topics presented in this section are numbered using a ranking based on coherence (Lau, Newman, and Baldwin, 2014; Lau and Baldwin, 2016). T_1 is the most coherent topic, T_{50} is the least.

Each of these topics is then analyzed using an inductive, sociological approach (cf. Glaser and Strauss, 1967; Charmaz, 2006), iteratively reading through the representative documents for each topic. Doing so both aligns with recent social scientific and humanist use of topic modeling (e.g., Jockers, 2013; Rhody, 2013; Baumer et al., 2017). It also provides a sense of what these topics *mean* and what the ramifications might be for an individual blogger to be associated with one of these topics.

High Coherence Topics

This subsection lists the five topics with highest coherence (Lau, Newman, and Baldwin, 2014; Lau and Baldwin, 2016). Each topic is listed with the top 20 highest probability words for the topic, as well as a brief description the topic's focus, based on a reading of the documents with the highest proportion of that topic.

T_1 – *spectrum, disorder, autism, diagnosis, disorders, asd, social, developmental, intellectual, syndrome, criteria, dsm, asperger,'s, mental, disability, diagnostic, new, test, functioning, symptoms*

Most of the top documents for this topic deal with the status of autism as a diagnosable condition. These documents discuss in depth the particular criteria used to diagnose autism, especially changes from DSM-IV to DSM-V. Many documents quote the DSM extensively, and many are also critical of it.

T_2 – *food, eat, eating, dinner, ice, cream, store, ate, lunch, chocolate, milk, restaurant, diet, chicken, table, cheese, drink, pizza, breakfast, foods*

Many documents in this topic provide recipes, often with commentary. Some discuss adapting to dietary restrictions, e.g., finding adequate milk and chocolate substitutes to make dairy-free, gluten-free "hot chocolate." Others deal with normative issues around what foods kids *should* or should *not* be eating, e.g., whether it is acceptable for a school to serve ice cream as a snack.

T_3 – *school, education, special, needs, schools, children, students, inclusion, child, classroom, district, learning, teachers, public, educational, mainstream, parents, system, environment, local*

The documents with highest proportions of this topic focus on special education in schools. There is a significant focus on whether (and how) current special education programs do or do not meet the needs of autistic children.

T_4 – *autism, brain, study, genetic, environmental, research, mercury, children, studies, disorders, may, factors, development, disease, cells, exposure, risk, mitochondrial, researchers, immune*

This topic features documents that engage in intricacies of scientific literature around autism. Many of these documents focus specifically on studies about the relationship between autism and vaccines, though rarely discussing such issues as vaccination schedules. Examples include discussing studies about the neurotoxicity of aluminum, or about chemical pathways by which mercury (specifically, mercuric chloride, or $HgCl_2$) may trigger brain inflammation. Others examine issues such as links between folic acid supplements and gene expression, or the relationship between ASD and mitochondrial dysfunction. Many documents include extensive quotes from the scientific literature they cover.

T_5 – *hair, wear, blue, wearing, red, white, color, pink, dance, isaac, green, black, shirt, dress, shoes, art, beautiful, favorite, cut, image*

This topic could likely be best summarized as "fashion." Top documents feature discussion of clothing, haircuts, dancing (with an emphasis on costumes), colors, etc. The few exceptions still feature this language prominently, e.g., a knock-knock joke about colors.

Few of these topics would likely be seen as controversial. That is, it seems unlikely that a blogger in our data set would find it problematic to be associated with any of the above topics. That ambivalence might not hold for the topics described below.

Vaccine Topics

The relationship between ASD and early childhood vaccination has received, and continues to receive, significant attention (Kirby, 2005; Wakefield et al., 1998; The Editors of The Lancet, 2010). This analysis identified three separate topics that include the term "vaccine" within the top 20 highest probability terms. These topics are listed here in order of decreasing coherence (Lau, Newman, and Baldwin, 2014). To reiterate, the topic number indicates each topic's ranking among the 50 topics in terms of coherence.

T_7 – vaccine, vaccines, children, health, thimerosal, mercury, safety, public, cdc, vaccination, parents, flu, medical, vaccinated, disease, information, government, research, year, years

Most of the documents for this topic discuss debates over the relationship between ASD and vaccination. The large majority suggest that early childhood vaccination causes autism. In one example, a blogger describes how their state's official health website includes a claim that no evidence suggests a causal link between autism and vaccines. The blogger argues that this claim is false, and offers a 9000 word post chronicling their attempts to have the claim removed from the state's website, attempts which ultimately failed. Other examples include coverage of vaccine-related deaths, lists of all vaccines that contain thimerosal (a mercury-based preservative), and commentary on changes in vaccine schedules.

The large majority of these documents make an implicit or explicit link between autism and vaccines. Others take the position that current vaccinations are vital for public health, for example, criticizing the delayed vaccination schedules advocated by some. That caveat made, an accurate label for this topic would be "anti-vax," i.e., focused on documents that oppose childhood vaccination.

Contrast this first topic with a second topic also pertaining to vaccines.

T_{23} – autism, vaccines, science, vaccine, case, court, evidence, david, cause, kirby, his, offit, cases, scientific, jenny, link, mccarthy, believe, generation, story

This topic revolves more around coverage of a specific events. Examples include particular court cases (e.g., the US Federal "Omnibus Autism" case (McNeil, 2009)), the retraction by Salon of a story by Robert F. Kennedy Jr. that mercury-based thimerosal in vaccines caused autism (Lauerman, 2011), or the former head of the US National Institutes of Health (NIH) calling for more investigation into autism and vaccines.

In contrast to T_7 above, this topic includes more direct confrontations, sometimes even personal attacks. For example, specific journalists or media personalities (David Kirby, Brian Deer, Jenny McCarthy, etc.) are criticized for propagating rumors and false information. In other cases, bloggers specifically name, link to, and argue with one another. Anecdotally, much of this argumentation occurs in blogs' comment sections, which were not scraped here. Thus, some of the discourses invoked in these arguments may not be entirely captured here. Overall, though, T_{23} deals less with the details of vaccination schedules or thimerosal content and more with contestation and argumentation.

The third topic dealing with vaccines is much more focused than the above two.

T_{45} – wakefield, mmr, his, andrew, research, paper, measles, wakefield's, study, children, medical, deer, brian, vaccine, article, lancet, published, cdc, virus, evidence

In 1998, The Lancet published a paper by Wakefield et al. (1998) with evidence that vaccines containing thimerosal caused autism. In 2010, that paper was retracted, due concerns both over tampering with the study and over the ethics of participant recruitment (The Editors of The Lancet, 2010).

The majority of documents with high proportions of this topic deal with these events. Examples include listing the specific charges against Wakefield and his colleagues, details of how the case was decided, and commentary and criticism about media coverage surrounding the initial paper and its retraction. Most of these posts are not sympathetic toward Wakefield.

Thus, these three topics capture three distinct positions. Documents representative of the first topic, T_7, argue that vaccines *do* cause autism. Documents representative of the third topic, T_{45}, argue that vaccines do *not* cause autism. Documents representative of the second topic, T_{23}, do not take a clear cut position but rather emphasize the conflict and controversy surrounding the issue.

In this way, the topic model can speak on behalf of the bloggers (Callon, 1986; Latour, 1992, 1993; Law, 1984). Rather than read through thousands of blog posts, an individual may consult only the topic modeling results to understand the position an individual blogger holds on vaccines and ASD. Journalists, the lay public, and even academic reviewers may lack the time, the data, the expertise, or the interest to investigate the original source texts being analyzed. While researchers employing topic modeling sometimes return to these texts (Rhody, 2013; Baumer et al., 2015; Klein, Eisenstein, and Sun, 2015; Mohr and Bogdanov, 2013), they virtually never check their interpretations with the original texts' authors on whose behalf the model speaks.

Discussion

Topic Modeling, Legitimacy, and Authority

"To speak for others is to first silence those in whose name we speak" (Callon, 1986, p. 216). This potential for associating specific topics with individual bloggers/authors has numerous ramifications. First, knowing the authors that most commonly discuss a given topic can alter the interpretation of that topic. For example, T_{45} above comes primarily from neurodiversity advocates, who claim that ASD is part of natural genetic variation. The same topic might mean something different if it occurred most often in the blogs of activists who regularly organize protests to decry conspiracies between government regulators and the pharmaceuticals industry. Just as individual statements are interpreted through the lens of the person who speaks them, so, too, could computationally-identified patterns be interpreted through the lens of the persons in whose writings they occur.

Second, knowing the prominent topics for a document or author can change perceptions thereof. For example, in one post primarily about T_3, a parent discusses their decision to enroll their own children in a separate special-needs school, but makes no prescriptive statements. In contrast, most other posts for that topic advocate strongly that separate learning environments are better for many children than "inclusive" policies that require all children, regardless of abilities or disabilities, to receive instruction in a regular classroom. Not

only might the first blogger wish to avoid association with this view, but seeing the rest of the topic gives that first post a more judgmental connotation than it might have otherwise.

Third, this summarization capacity can co-opt authors' voices. Much like the devices used to measure the attachment of bay scallops (Callon, 1986), there is a potential for topic models to represent the people whose data they analyze. This representational capacity, which lies at the heart of this paper's argument, comes into stark relief in this case.

Parents are, in some contexts, granted a unique kind of authority in understanding their own children. For example, Jenny McCarthy famously described her "mommy instincts" that led her to suspect vaccination had a negative impact on her child (Yochim and Silva, 2013). In Weber's terms, such charismatic authority contrasts with the rational authority used in the scientific discourse around vaccines (Kalberg, 1980). A topic model, then, has the potential to co-opt such discourses, to present its results as a representation of what parents of ASD children *really* say and think. Put differently, using a topic model trained on texts written by parents can channel the legitimacy of parental voice, without needing to check one's interpretation of the topics against those on whose behalf the topic model is speaking.

Fourth, claims about what a corpus *really* says can have significant ramifications, particularly with respect to ASD. As alluded above (e.g., in T_1), the very status of ASD is a point of contention with significant consequences. If ASD results from deliberate collusion between regulators and pharmaceutical companies, then legal action is warranted that could likely result in reparations for families, as well as support from agencies such as Social Security. However, if the argument advanced by neurodiversity advocates were taken to its logical conclusion – that ASD is not a disorder, but a difference in the wide range of human functioning – then its status as a medical and/or psychiatric condition could be destabilized, potentially leading to a reduction in supports and services. Thus, families might not be eligible to receive the kinds of public aid provided for children with neurological disorders.

Our Ethics of Translation

The case presented here – ASD parenting blogs – requires some unique ethical considerations. This paper's contribution is founded on the claim that the kinds of analyses presented here can be highly sensitive. Topic modeling results are sometimes presented along side quotes from highly highly representative documents for each topic (Rhody, 2013; Baumer et al., 2015; Klein, Eisenstein, and Sun, 2015; Mohr and Bogdanov, 2013). However, internet search engines make it easy to link such verbatim quotes to the blogs from which they came, a link we argue can cause harm. Thus, individual quotes or references to specific blogs were carefully obfuscated to protect bloggers' identities from deanonymization.

However, doing so slightly undermines our own authority. Since we have not named our data sources, others cannot replicate our analysis. Similarly, omitting verbatim quotes makes it difficult to validate our interpretations of these topics. Comparisons could be made with Goffman's (2014) in-

tentional obfuscation of sources to protect them, and, potentially, with the mixed reactions to Goffman's work (Singal, 2015; Kotlowitz, 2014). Since this paper does not aim to make definitive claims about autism, parenting, etc., the veracity of our results is less of an issue. Instead, the focus is how the results *could* be used to speak on behalf of these bloggers without their consent.

Finally, an additional layer of delegation (Latour, 1993) or mobilisation (Callon, 1986) occurs in this case. Many of these blogs discuss children, but the blogs rarely include the words or voices of the children themselves. Similar to Butler (2001), we are analyzing linguistic fragments about a person that were chosen neither by us nor by the person themself. While important to note here, full treatment of this issues exceeds the scope of this paper.

Broader Implications

This paper uses the example of topic modeling to illustrate both how computational systems can speak on behalf of people and some of the consequences thereof. However, these arguments apply much more broadly to a wide variety of computational models and tools.

For example, Niculae et al. (2015) analyze and group news sources based on their patterns of quoting excerpts from US presidential addresses. Notably, their analysis suggests that a media outlet's style has more to do with its quoting patterns than its partisan leanings do. For instance, Fox News (known for its conservative slant) is highly similar to the New York Times (known for its liberal slant). These two mainstream media sources are both more similar to each other than either is to smaller, expressly partisan sites, such as The Blaze (conservative) or The Nation (liberal). While analytically interesting, these various sites may not necessarily appreciate being labeled as similar to or different from one another. Put differently, this analysis speaks on behalf of these sites, declaring for them sets of affiliations with which each site's authors or owners may not agree.

As another example, machine learning models are used for a variety of predictive tasks, such as when a smoker who is trying to quit will likely relapse (Murnane and Counts, 2014), or an individual's personality traits (Youyou, Kosinski, and Stillwell, 2015). In these and similar cases, the model speaks on behalf of the person whose data are being analyzed. Such pronouncements may be associated with the claim that this person may not be aware of their own personality, mental health state, likelihood for smoking relapse.

Such issues can arise even with seemingly simple technologies. Sentiment analysis (Pang and Lee, 2008; Hutto and Gilbert, 2014), as an example, is often used to determine the sentiment expressed about a given entity. For instance, reviews may collectively indicate that a given restaurant has good sushi but bad service (Yatani et al., 2011). However, such techniques could be used equally readily to make claims about the authors of documents, e.g., that author X tends on average to be more positive than author Y. Once again, the computational tool assessing sentiment can speak on behalf of authors, expressing both their sentiments about certain entities and their general disposition.

Conclusion and Future Directions

To address the issues raised here, future researchers can work directly with the people whose data they are analyzing. Participatory design (Ehn, 1988; Asaro, 2000) has a history of incorporating multiple stakeholders into design processes. This paper's authors are developing an analogous approach, incorporating the participants whose data we have collected into the processes of analyzing and interpreting those data.

The questions described above can also be incorporated into ethics review processes. However, the practicalities of doing so are tricky at best. It is virtually impossible to know detailed results from computational models before running them. This constraint creates a challenges, not only for researchers, but also for an Institutional Review Boards (IRB) reviewing study proposals.

Researchers can also obfuscate their data, as done here. On the one hand, doing so makes it more difficult for others to verify a paper's results and interpretations. On the other hand, social scientific research has a history of using "composites" when presenting results. In this approach, all the details for each composite come from the data, but individual aspects may be synthesized from different participants' experiences to obfuscate participants' identities. Analogous to techniques for differential privacy (Dwork and Roth, 2014), future research could benefit from exploring means of creating computational composites that resemble the original data in a meaningful way but eschew verbatim quotes and other identifying information.

As a final note, not every domain is equally sensitive. Using corpora such as Associated Press articles or scientific journal proceedings, both of which are common in topic modeling research (Blei, Ng, and Jordan, 2003; Griffiths and Steyvers, 2004), may not carry the same implications as blogs written by parents of ASD children. Thus, this paper offers but one case illustrating potential consequences of computational models that speak on behalf of people. More generally, the paper offers a conceptual vocabulary that researchers across multiple domains can use in considering both the degree and kinds of sensitivity associated with computational analyses of social data.

Acknowledgements

References

Ananny, M. 2011. The Curious Connection between Apps for Gay Men and Sex Offenders. *The Atlantic*.

Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2016. Machine Bias. *Pro Publica*.

Asaro, P. M. 2000. Transforming society by transforming technology: The science and politics of participatory design. *Accounting, Management and Information Technologies* 10(4):257–290.

Barocas, S., and Selbst, A. D. 2016. Big Data's Disparate Impact. *California Law Review* 104(3):671–732.

Baumer, E. P. S.; Guha, S.; Quan, E.; Mimno, D.; and Gay, G. K. 2015. Missing Photos, Suffering Withdrawal, or Finding Freedom? How Experiences of Social Media Non-Use Influence the Likelihood of Reversion. *Social Media + Society* 1(2).

Baumer, E. P. S.; Mimno, D.; Guha, S.; Quan, E.; and Gay, G. K. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology (JASIST)* 68(6):1397–1410.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3(Jan):993–1022.

Butler, J. 2001. Doing Justice to Someone: Sex Reassignment and Allegories of Transsexuality. *GLQ: A Journal of Lesbian and Gay Studies* 7(4):621–636.

Callon, M. 1986. Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of Saint Brieuc Bay. In Law, J., ed., *Power, Action and Belief: A New Sociology of Knowledge?* London: Routledge. 196–223.

Charmaz, K. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: SAGE Publications.

Dwork, C., and Roth, A. 2014. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science* 9(3–4):211–407.

Ehn, P. 1988. *Work-Oriented Design of Computer Artifacts*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Eslami, M.; Rickman, A.; Vaccaro, K.; Aleyasen, A.; Vuong, A.; Karahalios, K.; Hamilton, K.; and Sandvig, C. 2015. "I Always Assumed That I Wasn't Really That Close to [Her]": Reasoning About Invisible Algorithms in News Feeds. In *Proc ACM Conf on Human Factors in Computing Systems (CHI)*, 153–162. Seoul: ACM.

Geiger, R. S., and Ribes, D. 2011. Trace Ethnography: Following Coordination through Documentary Practices. In *Hawaii International Conference on System Sciences*.

Gillespie, T. 2013. The Relevance of Algorithms. In Gillespie, T.; Bockzkowski, P.; and Foot, K., eds., *Media Technologies*. Cambridge, MA: MIT Press.

Glaser, B. G., and Strauss, A. L. 1967. *The Discovery of Grounded Theory*. Oxon and New York: Routledge.

Goffman, A. 2014. *On the Run*. Chicago: U Chicago Press.

Griffiths, T. L., and Steyvers, M. 2004. Finding scientific topics. *Proc Nat'l Acad Sci* 101:5228–5235.

Hern, A. 2015. Flickr faces complaints over 'offensive' auto-tagging for photos. *The Guardian*.

Hutto, C., and Gilbert, E. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proc Int'l Conf on Weblogs and Social Media (ICWSM)*, 216–225. Ann Arbor, MI: AAAI.

Jockers, M. L. 2013. *Macroanalysis: Digital Methods & Literary History*. Chicago: U Illinois Press.

Kalberg, S. 1980. Max Weber's Types of Rationality: Cornerstones for the Analysis of Rationalization Processes in History. *The American J of Sociology* 85(5):1145–1179.

Kirby, D. 2005. *Evidence of Harm: Mercury in Vaccines and the Autism Epidemic*. New York: St. Martin's Press.

Klein, L. F.; Eisenstein, J.; and Sun, I. 2015. Exploratory Thematic Analysis for Digitized Archival Collections. *Digital Scholarship in the Humanities* 30:i130–i141.

Kotlowitz, A. 2014. Deep Cover: Alice Goffman's 'On the Run'. *The New York Times*.

Latour, B. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, Mass: Harvard University Press.

Latour, B. 1992. Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. In Bijker, W. E., and Law, J., eds., *Shaping Technology / Building Society: Studies in Sociotechnical Change*. Cambridge, MA: MIT Press. 225–258.

Latour, B. 1993. Ethnography of a "High-Tech" Case: About Aramis. In Lemonnier, P., ed., *Technological Choices: Transformations in Material Culture since the Neolithic*. London: Routledge. 372–398.

Lau, J. H., and Baldwin, T. 2016. The Sensitivity of Topic Coherence Evaluation to Topic Cardinality. In *Proc Annual Mtg of the North Amer Chapter of the Assoc for Comp Ling (NAACL)*, 483–487. San Diego, CA: ACL.

Lau, J. H.; Newman, D.; and Baldwin, T. 2014. Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. In *Proc of the Annual Mtg of the Euro Chapter of the Assoc for Comp Ling (EACL)*, 530–539. Gothenburg, Sweden: ACL.

Lauerman, K. 2011. Correcting our record — Salon.com. https://www.salon.com/2011/01/16/dangerous_immunity/.

Law, J. 1984. On the Methods of Long-Distance Control: Vessels, Navigation and the Portuguese Route to India. *The Sociological Review* 32:234–263.

Lenssen, P. 2007a. Did You Mean: "He Invented"? http://blogoscoped.com/archive/2007-05-07-n56.html.

Lenssen, P. 2007b. Google Stops "Did You Mean: He Invented". http://blogoscoped.com/archive/2007-05-24-n36.html.

McCallum, A. K. 2002. MALLET: A Machine Learning for Language Toolkit.

McNeil, Jr., D. G. 2009. Court Says Vaccine Not to Blame for Autism. *The New York Times*.

Mimno, D. 2013. Mallet: A wrapper around the Java machine learning tool MALLET.

Mohr, J. W., and Bogdanov, P. 2013. Topic models: What they are and why they matter. *Poetics* 41(6):545–569.

Murnane, E. L., and Counts, S. 2014. Unraveling abstinence and relapse: Smoking cessation reflected in social media. In *Proc ACM Conf on Human Factors in Computing Systems (CHI)*, 1345–1354. Toronto, ON: ACM Press.

Niculae, V.; Suen, C.; Zhang, J.; Danescu-Niculescu-Mizil, C.; and Leskovec, J. 2015. QUOTUS: The Structure of Political Media Coverage As Revealed by Quoting Patterns. In *Proc ACM Conf on World Wide Web (WWW)*, 798–808.

Pang, B., and Lee, L. 2008. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* 2(1-2):1–135.

Rhody, L. M. 2013. Topic Modeling and Figurative Language. http://journalofdigitalhumanities.org/2-1/topic-modeling-and-figurative-language-by-lisa-m-rhody/.

Schofield, A.; Magnusson, M.; and Mimno, D. 2017. Pulling Out the Stops: Rethinking Stopword Removal for Topic Models. In *Proc Annual Mtg of the Euro Chapter of the Assoc for Comp Ling (EACL)*, 432–436. Valencia, Spain: ACL.

scrapy. 2018. Scrapy, a fast high-level web crawling & scraping framework for Python. https://github.com/scrapy/scrapy.

Singal, J. 2015. The Internet Accused Alice Goffman of Faking Details in Her Study of a Black Neighborhood. I Went to Philadelphia to Check. https://www.thecut.com/2015/06/i-fact-checked-alice-goffman-with-her-subjects.html.

Sweeney, L. 2013. Discrimination in online ad delivery. *Communications of the ACM* 56(5):44–54.

The Editors of The Lancet. 2010. Retraction—Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet* 375(9713):445.

Wakefield, A. J.; Murch, S. H.; Anthony, A.; Linnell, J.; Casson, D. M.; Malik, M.; Berelowitz, M.; Dhillon, A. P.; Thomson, M. A.; Harvey, P.; Valentine, A.; Davies, S. E.; and Walker-Smith, J. A. 1998. RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet* 351(9103):637–641.

Winner, L. 1980. Do Artifacts Have Politics? *Daedalus* 109(1):121–136.

Yatani, K.; Novati, M.; Trusty, A.; and Truong, K. N. 2011. Review Spotlight: A user interface for summarizing user-generated reviews using adjective-noun word pairs. In *Proc ACM Conf on Human Factors in Computing Systems (CHI)*, 1541–1550. Vancouver, BC: ACM.

Yergeau, M. 2017. *Authoring Autism: On Rhetoric and Neurological Queerness*. Durham, NC: Duke U Press.

Yochim, E. C., and Silva, V. T. 2013. Everyday Expertise, Autism, and "Good" Mothering in the Media Discourse of Jenny McCarthy. *Communication and Critical/Cultural Studies* 10(4):406–426.

Youyou, W.; Kosinski, M.; and Stillwell, D. 2015. Computer-based personality judgments are more accurate than those made by humans. *Proc Nat'l Acad Sci* 112(4):1036–1040.