Normalizing Digital Trace Data

Andreas Jungherr
Assistant professor for *Social Science Data Collection and Analysis*
Konstanz University
andreas.jungherr@gmail.com

In *Digital Discussions: How Big Data Informs Political Communication*, eds. Natalie Jomini Stroud and Shannon McGregor. New York, NY: Routledge. (Forthcoming).

February 5, 2017

Abstract
Over the last ten years, social scientists have found themselves confronting a massive increase in available data sources. In the debates on how to use these new data, the research potential of "digital trace data" has featured prominently. While various commentators expect digital trace data to create a "measurement revolution", empirical work has fallen somewhat short of these grand expectations. In fact, empirical research based on digital trace data is largely limited by the prevalence of two central fallacies: First, the n=all fallacy; second, the mirror fallacy. As I will argue, these fallacies can be addressed by developing a measurement theory for the use of digital trace data. For this, researchers will have to test the consequences of variations in research designs, account for sample problems arising from digital trace data, and explicitly link signals identified in digital trace data to sophisticated conceptualizations of social phenomena. Below, I will outline the two fallacies in greater detail. Then, I will discuss their consequences with regard to three general areas in the work with digital trace data in the social sciences: digital ethnography, proxies, and hybrids. In these sections, I will present selected prominent studies predominantly from political communication research. I will close by a short assessment of the road ahead and how these fallacies might be constructively addressed by the systematic development of a measurement theory for the work with digital trace data in the social sciences.

Keywords: Computation Social Science, Digital Trace Data, Political Communication, Methodology

Acknowledgements: The arguments developed here result from a continuous conversation with colleagues from the social sciences and computer science. Especially, I want to thank Daniel Gayo-Avello, Pascal Jürgens, David Lazer, Helen Margetts, Takis Metaxas, Oliver Posegga, Harald Schoen, Ralph Schroeder, Sebastian Stier, Markus Strohmaier, and Yannis Theocharis.

Author Note: Andreas Jungherr is Assistant Professor for *Social Science Data Collection and Analysis* at the University of Konstanz, Germany. His research focuses on the impact of digital technology on political communication and the use of digital trace data in the social sciences. His research has been published in *Journal of Communication, Journal of Computer-Mediated Communication, The International Journal of Press/Politics*, and *Social Science Computer Review*. (E-mail: andreas.jungherr@gmail.com)

1 Digital Trace Data in the Social Sciences: A Promise Yet to Be Realized

Gradually, over the last ten years, social scientists have found themselves confronting a massive increase in available data sources. The digitalization has, for example, opened up vast textual corpora (Grimmer & Stewart, 2013), provided researchers with cheap and fast alternatives to telephone or face-to-face surveys (Callegaro, Manfreda, & Vehovar, 2015), and the increasing use of digital services in everyday life provides social scientists with an ever increasing reservoir of digital data traces documenting slices of users' everyday interactions with various digital devices or services (Howison, Wiggins, & Crowston, 2011). This increase in variety and size of data available to researchers has been heralded by some as a measurement revolution for the social sciences (Golder & Macy, 2014; Lazer, Pentland, Adamic, Aral, Barabási, Brewer, Christakis, Contractor, Fowler, Gutmann, Jebara, King, Macy, Roy, & Van Alstyne, 2009; Schroeder, 2016; Watts, 2011). Especially, the research potential of digital trace data (Howison et al., 2011) has featured prominently in these accounts.

Digital trace data are data documenting the interactions of users with digital devices or services (Howison et al., 2011). They potentially include a full set of these interactions. This has led some prominent commentators to declare an age of "big data", in their view characterized by the ability of researchers to measure everything and everyone of interest to them. This view is exemplified by the catchy term "n=all" (Mayer-Schönberger & Cukier, 2013). In fact, researchers are usually very far from being able to realize this potential as they depend on data access policies set by service providers. So, in fact, “n=all” becomes “n=sample with unknown properties from unknown populations determined by third parties”. In practice, these data access policies vary between open—e.g. Wikipedia—largely restricted—e.g. Facebook—or completely closed off to outsiders—e.g. Google or mobile phone providers (Resnick, Adar, & Lampe, 2015; Schroeder, 2016).

Another characteristic of “digital trace data” is that they are found data. They are produced and provided not based on research designs. Instead, they document user interactions through the lens of a service’s specific data retention and access policies (Howison et al., 2011). On one hand, this makes these data immensely valuable as they provide researchers a view of actual behavior in the field. On the other hand, researchers have not the ability to fine tune their data collection to specific research questions. Instead, more often than not, they have to adjust their research question to the data available. Through this, the field runs the risk of producing largely data-driven studies instead of testing carefully developed hypotheses based on current theoretical debates. This raises the importance of careful conceptualization of digital trace data in the context of specific research questions, behavior, or populations. This becomes all the more important once we move from an optimistic early reading of digital trace data as true mirrors of human behavior, attitudes, or social phenomena to a more realistic reading. Digital trace data are the result of very specific data generating processes that are framed by, among other factors, the affordances of the service providing the data, its code, cultural usage practices associated with the service, users’ usage motives, and the level of data access granted to researchers. This makes digital trace data the result of highly specific mediation processes that filter social or political phenomena of interests (Jungherr, 2015; Jungherr, Schoen, & Jürgens, 2016a). It is important for researchers to account for the specific mediating steps producing digital data traces if they want to present findings going beyond a digital ethnography of the device or service producing their data source or an isolated proof of concept illustrating the workings of their algorithm or model of choice.

Surprisingly, these two challenges at the center of the work with digital trace data have only received limited attention in the debate on how to use these data sources in the social sciences. This is surprising as the literature has become very sophisticated with regard to other characteristics of these data, such as problems associated with data access, the dissemination of necessary analytical skill sets, or ethics (boyd & Crawford, 2012; Freelon, 2015; King, 2011). Only recently, questions explicitly addressing the need for theory-driven research, research designs, careful conceptual work, or indicator validation are being raised (González-Bailón & Paltoglou, 2015; Howison et al., 2011; Jungherr et al., 2016a; Jungherr, Schoen, Posegga, & Jügens, 2016b; Lazer, Kennedy, King, & Vespignani, 2014; Salganik, 2017).

This has led to a use of digital trace data largely based on an implicitly held naive measurement theory. Implicitly practitioners seem to assume signals found in digital trace data to directly inform on phenomena of interests. This runs counter to measurement theories in the social sciences and statistics. These fields have developed sophisticated understandings of how phenomena of interests have to be translated into theoretical concepts, these concepts in turn have to be translated into measures, these measures have to be tested on their validity, and finally statistical procedures have to be developed as how to draw inferences on the respective phenomena based on available data. The social sciences and statistics look back on a long and fruitful debate on this process (Donoho, 2015; Efron & Hastie, 2016; Gerring, 2012; Goertz, 2005; Hand, 2004).

So far, this debate has no equivalent in the use of digital trace data. This indifference has given rise to two central fallacies in the work with digital trace data in the pursuit of social or political research questions:

1. The n=all fallacy
2. The mirror fallacy

These fallacies permeate much of the most prominent work with digital trace data and often limit its access to the social science mainstream. As I will argue below, these fallacies can be addressed by developing a measurement theory for the use of digital trace data. For this, researchers will have to test the consequences of variations in research designs, account for sample problems arising from digital trace data, and explicitly link signals identified in digital trace data to sophisticated conceptualizations of social phenomena. Continuing to ignore these challenges will limit the reach of work based on digital trace data. The development of a sophisticated measurement theory is a precondition for digital trace data being meaningfully integrated in the social sciences. The alternative is the relegation of this work to a subfield of applied computer research. Instead of meaningfully examining social or political phenomena, this research runs the risk of only speaking to the applicability of a specific algorithm or model to a specific data set. The social or political phenomena ostensibly in the focus of this research become thus nothing more than vehicles to illustrate computational and quantitative prowess.

Below, I will outline the two fallacies in greater detail. Then, I will discuss their consequences with regard to three general areas in the work with digital trace data in the social sciences: digital ethnography, proxies for other data sources, and hybrids. In these sections, I will present selected prominent studies, predominantly from political communication research. Still, the points raised are general and, thus, should hold for work in other fields. I will close with a short assessment of the road ahead and how these fallacies might be constructively addressed by the systematic development of a measurement theory for the work with digital trace data in the social sciences.

Before we start, let me emphasis that this is a work in progress. The studies listed here, are meant to be illustrative of approaches typical for the work with digital trace data in the social sciences. Also, some aspects of the studies below are discussed critically. This is not meant as a critique of the authors who are pioneers in this field and have proven themselves to be highly original and innovative. Still, as we now look back at nearly ten years of working with digital trace data, specific practices have become limiting to the development of the field but at the time of their first application were highly original and promising. So, the arguments presented below are meant to move the debate further and not to judge the inspiring work of past pioneers whose work enables us to have this debate in the first place.

2 Two Fallacies

There are two central fallacies permeating current research with digital trace data: The n=all fallacy and the mirror fallacy. Both form implicit or explicit foundations of much of current research with digital trace data. As I will argue below, they are rooted in misconceptions about the nature of digital data traces and a widespread indifference among researchers toward sophisticated conceptual debates in the social sciences. They, therefore, severely limit the reach of research based on digital trace data in the established social science core discourses. Yet, as I will argue below, once identified, both fallacies can be addressed by the use of well-established scientific practices. If applied, these practices promise a clearer road forward for the work with digital trace data into the core of social science discourses.

2.1 The n=all Fallacy

In its design, much of current research with digital trace data is largely indifferent toward questions of sampling. To more classically oriented quantitative social scientists this might seem surprising. In fact, this indifference is deeply rooted in an understanding that all relevant interactions with devices and services by all users leave digital data traces and are, therefore, complete. Sampling would be, therefore somewhat beyond the point. Why limit your potential information by actively discarding data when instead you can just analyze everything? In fact, the apparent possibility of working with datasets documenting all interactions with a device or service by a complete population is presented by some as one of the defining characteristics of digital trace data or, as they call it, “big data”:

> Sampling is an outgrowth of an era of information-processing constraints, when people were measuring the world but lacked the tools to analyze what they collected. As a result, it is a vestige of that era too. [...]
>
> The concept of sampling no longer makes as much sense when we can harness large amounts of data. The technical tools for handling data have already changed dramatically, but our methods and mindsets have been slower to adapt.
>
> Yet sampling comes with a cost that has long been acknowledged but shunted aside. It loses detail. In some cases there is no other way but to sample. In many areas, however, a shift is taking place from collecting some data to gathering as much as possible, and if feasible, getting everything: n=all. (Mayer-Schönberger & Cukier, 2013, p. 26).

While some of the early enthusiasm of Mayer-Schönberger and Cukier (2013) begins to ring hollow by now, they clearly captured the mood in the field. n=all is a catchy formula and makes explicit central assumptions underlying much of the work with digital trace data. We find studies concerned with collecting all relevant topically content (Ausserhofer & Maireder, 2013; Bruns & Highfield, 2013) or all content posted by selected populations (Jungherr, 2015; Lin, Keegan, Margolin, & Lazer, 2014). Why limit yourself if you can get it all? Yet in the case of digital trace data, it turns out, the term “all” just means “all that was collected and made available to researchers by commercial third-party entities”. A meaning quite distinct from the word’s more common usage. So, what are the limits of digital trace data’s “all”?

First and foremost, digital trace data come with a series of technical limitations. Central here is the question of data access. While it is certainly true that in theory all interactions by all users of specific devices or services might be collected by said device or service, it does not necessarily follow that this is in fact the case. Even in an age when data storage is cheap, service providers do not save every data trace available to them. Instead, they make a choice as to which data seem valuable to their business case or which data might conflict with their company's policies, such as privacy concerns. But even to this subset of all interactions, researchers do not have full access. Instead, they can only access these data directly from the service—for example through application programming interfaces (API)—or indirectly—for example by web scraping. Both approaches only provide researchers with subsets of all potentially relevant data. The data quality of these subsets and their relationship to all potentially relevant data can only be guessed at (Morstatter, Pfeffer, Liu, & Carley, 2013; Ruths & Pfeffer, 2014). Here, it is always important to keep in mind that the interests of service providers of what to make public and what to keep confidential are not necessarily identical to the interests of researchers. For example, increasingly services like Facebook or Twitter show users algorithmically ordered feeds and allow the pushing of posts' prominence through ad buys. These algorithmic and commercial interventions thus influence the use of these tools considerably but do not necessarily appear in digital trace data available to researchers. What researchers might take for social processes giving rise to specific patterns in these data might in fact be results of algorithmic interventions (Strohmaier & Wagner, 2014). Recently, various scholars have tried to avoid these challenges by partnering directly with the providers of digital tools (Bakshy, Messing, & Adamic, 2015; Bond, Fariss, Jones, Kramer, Marlow, Settle, & Fowler, 2012). While these studies have been highly original and imaginative, this approach raises serious questions with regard to scientific practices. Obvious issues are replicability, potential conflicts of interests between researchers and partners leading to the focus on or avoidance of specific research questions, and ultimately a divide between options open to researchers at prestigious institutions with connections to digital service providers and others.

Varying access policies across digital services lead to another problem for researchers. The use of digital devices and services inherently encompasses the use of many different devices, services, and channels for overlapping or divergent patterns (Rainie & Wellman, 2012). A user might call a family member from her smart-phone and start exchanging views on current events while receiving an email with a link to an article covering politics by a work-colleague. In following, she might post a link to the article on her Facebook page for her contacts to see, exchange acerbic comment on the article's subject with a friend on her favorite chat program, and post an especially funny quip on her Twitter feed. This whole interaction chain comprises political talk online but only slices are available to researchers working with digital trace data. A researcher focusing on only one of the services in this interaction chain will invariably come to different assessments of the nature of political talk online from a researcher focusing on another. This interaction between various communication channels is a central challenge and yet only few studies using digital trace data try to explicitly account for it (Leskovec, Backstrom, & Kleinberg, 2009).

A final limitation to the "all" provided by digital trace data to be discussed here stems from digital services' limited and skewed user base. While it is certainly true that many people use digital devices and services in their daily lives, this user base is far from including all members of a society. In fact, digital devices and services are predominantly used by a skewed subset of the population, tending to be younger, wealthier, and better educated. This skewness becomes stronger when we focus on the use of specific services and usage practices (Blank, 2016; Hargittai, 2015). Also, the composition of this self-selected unrepresentative population sample seems to drift unpredictably and non-randomly over time making the adjustment for the underlying skewness with regard to overall populations infeasible (Diaz, Gamon, Hofman, Kiciman, & Rothschild, 2016). So again, n=some seems to be a more accurate description of what we find in the work with digital trace data than n=all.

All of this does not invalidate the work with digital trace data. Instead, this eerily resembles challenges known from other approaches to data collection in the social sciences. Here, we find sophisticated methodological debates addressing the potential benefits and limits of traditional data collection approaches (Groves, Fowler Jr., Couper, Lepkowski, Singer, & Tourangeau, 2009; Tourangeau, Rips, & Rasinski, 2000). Yet, in the work with digital trace data the n=all fallacy misleads researchers to ignore systematic limitations of their datasets. Limitations that could be addressed head-on by the open testing and discussion of different sampling procedures allowing the systematic assessment and maybe even controlling for potential biases inherent in digital trace data. Only recently, these questions are beginning to be raised in the field (Salganik, 2017). The early exceptionalism, a mood so evocatively captured by Mayer-Schönberger and Cukier (2013), has stood in the way for social scientists working with digital trace data to profit from the deep and long-ongoing methodological debates at the intersections between social science and statistics. Joining this debate holds very high promise in addressing the limitations of digital trace data listed above (Donoho, 2015; Efron & Hastie, 2016).

2.2 The Mirror Fallacy

The second fallacy is a subtler one. Other than the n=all fallacy, the mirror fallacy is seldom stated explicitly but lies implicitly at the core of many works based on digital trace data. The perceived potential of digital trace data is based on the assumption that digital trace data reflect political or social phenomena in users' data traces (Golder & Macy, 2014; Lazer et al., 2009). This position emerged in reaction to the increasing growth and variety of digital services and the subsequent growing availability of data. Some authors have framed it as a shift from using data collected on the internet for the examination of online-phenomena or behavior—cyber-ethnography—to using these artifacts and data traces to draw inferences on larger social or political phenomena (Rogers, 2013). This was a decisive shift in the focus of researchers interested in the use of digital trace data and significantly increased the perceived potential of these data even for researchers who before had not shown interest in the internet or digital services as objects of study. Yet, this originally immensely liberating hypothesis has become a trap for researchers. A lazy reading tempts them to ignore the inherent mediated nature of digital trace data and the skewness this might introduce to the reflection of social and political phenomena.

Digital data traces are always the product of the design of digital services and devices, cultural usage practices, and users' usage motives. The reflection of social and political phenomena arising from these data traces are, therefore, mediated by these factors (Jungherr, 2015; Jungherr et al., 2016a). If we accept this, we cannot simply take signals contained in digital trace data at face value but instead have to interpret them in the context of likely mediating processes leading to their emergence. This renders the process of interpreting signals in digital trace data and linking them explicitly to concepts of interests of paramount importance (Howison et al., 2011). And yet, this step is nearly always neglected. Instead, researchers seem to share a highly simplistic measurement approach. They mainly take for granted that signals extracted from digital trace data indeed measure their given phenomenon of interest. For such a quantitatively advanced field, this is a surprisingly unsophisticated approach to the measurement of social reality (Hand, 2004). Most of the time, researchers appear to be satisfied by statistically linking signals identified in digital trace data to arbitrarily chosen metrics documenting social or political phenomena. Alternatively, they might simply point to patterns identified in digital trace data and take them as direct expression of underlying social or political processes. This and an often superficial reading of conceptual debates in the social sciences lead to the limited impact works based on digital trace data have had on central debates in the social sciences.

In light of this, the tendency to approach signals in digital trace data uncritically as indicators for micro-level attitudes and behavior or macro-level phenomena seems highly troubling. For example, a very influential study showed that Twitter mentions of a selection of German parties seemed to correspond with their vote shares (Tumasjan, Sprenger, Sandner, & Welpe, 2010). Parties' Twitter mentions thus seemed to mirror their level of public support. This was surprising. Given the well-known skewness in Twitter's user base and the prevalence of political snark and public critique on the service, one would have expected parties with young and online-savvy supporters and those at the center of public controversy to be overrepresented. As it turns out, this pattern emerges once one changes only a few parameters of the original study. Simply by including mentions of the Pirate Party, a party focused on digitalization's social impact and civil rights, the surprising link between Twitter mentions and vote shares breaks spectacularly (Jungherr, Jürgens, & Schoen, 2012).

Tumasjan et al. (2010) clearly fell victim to the mirror fallacy. In their analysis, the team found a promising pattern. After all, who would not jump at the possibility of predicting elections by simply counting tweets? Instead of critically investigating their data guided by a realistic reading of Twitter's data generating process, the authors simply took their welcome results at face value and declared aggregates of Twitter messages as valid indicators of public support for parties. Following this practice, digital trace data become a mirror for everything enterprising researchers want to see in them. For example, Twitter messages have been taken, among other phenomena, as expression of political support (Tumasjan et al., 2010), the onset of depression (De Choudhury, Counts, & Horvitz, 2013), or signs of imminent stock market movements (Bollen, Mao, & Zheng, 2011). If true, this would make the microblogging service the most universally applicable concoction since the discovery of snake oil.

A the heart of the mirror fallacy lies the predominant practice of linking metrics statistically without seriously proposing and testing a mechanism plausibly leading to the emergence of this association (Jungherr et al., 2016b). A research environment characterized by big data sets necessarily comes with many correlations between variables. Most of these correlations will be spurious, they do not speak to a systematic or causal link between variables but instead emerge by chance. Especially in scientific contexts, the goal is to differentiate between spurious correlations and those speaking to a systematic link between variables. Social science methodology is rich in debates on how to correctly identify causal links between phenomena and variables (Gerring, 2012). At the very least, researchers should propose a mechanism linking two concepts or variables of choice, explaining the nature of any reported correlation. This provides a series of explicit steps linking two concepts or variables. Each of these steps can then be tested on its plausibility or correspondence with available evidence. In this process, the identification of statistically linked metrics is only the first step. The true test of a correlation's meaning lies in the proposal and testing of potential mechanism leading to its emergence (Gerring, 2008, 2010).

In practice, most contemporary work with digital trace data ignores this crucial second step and instead follows a purely data-driven logic in reporting statistical links between arbitrarily chosen metrics. Pair this with a publication culture showing a clear bias towards the reporting of positive findings, economic pressures on labs and researchers to provide outside funders and the media with spectacular and often counter-intuitive results and you have the perfect storm leading the field to being drowned in reports of positive but irrelevant statistical associations between digital trace data and social or political phenomena. To avoid this, careful conceptual work (Goertz, 2005), the proposal and testing of likely causal mechanisms (Gerring, 2008), and measurement validation (Adcock & Collier, 2001) are of paramount importance in the work with digital trace data.

As with the n=all fallacy, the mirror fallacy does not render the work with digital trace data futile. Instead, it simply means researchers have to account for the mediated nature of their data sources in their interpretation of data traces. As with other data sources, data generating processes determine the meaning we plausibly can assign to identified patterns. Yet, currently questions in the interpretation and conceptualization of digital trace data are predominantly neglected by the field. This might be due to some researchers' hope of having finally found a data source large enough to turn social science into a "true" science, allowing for the identification of a physics of social life. Ignoring the questions raised by the mediated nature of digital trace data is only logical from this point of view. Interpretation and conceptualization "scale up" badly thereby getting in the way of finding the gravitational laws of social systems. Yet, collectively ignoring a problem usually does not make it disappear. Instead, it is much more likely that work done based on a common consensus in a small community of practice travels not well beyond the borders of said community. Accounting for the mediated nature of digital trace data in research designs, conceptualization, and interpretation, therefore, is very likely a necessary step for research using these data to extend beyond the small group of converted.

3 The Fallacies at Work

The two fallacies discussed above permeate the work with digital trace data. Here, I will exemplarily illustrate their presence by discussing selected prominent papers using digital trace data in political communication research. We can group these studies by their research interests. First, studies using digital trace data to analyze politically relevant behavior online— digital ethnography. Second, studies using digital trace data to draw inferences on offline phenomena, traits, or attitudes—proxies. Third, studies focusing on interactions between political behavior online and offline—hybrids. Especially in the first two groups, we find prominent studies explicitly or implicitly falling for the n=all or the mirror fallacy.

3.1 Digital Ethnography: The Dangers of Conceptual Stretching

Using digital trace data to examine political behavior of users online and on specific services is an intuitive first step. Here, phenomena of interests are closely connected to data generating processes giving rise to signals identified in digital trace data. Yet, as I will show, most authors are not content with simply speaking to political phenomena online but instead take them as expression of larger political phenomena. Here, the question arises if patterns found in digital trace data truly speak of political phenomena as conceptualized in the mainstream social science discourse. Here, authors have to consciously engage with contemporary conceptualizations and provide explicit interpretations of how their signals of choice identified in digital trace data connect with them. In neglecting this step, researchers run the risk of concept stretching— extending concepts developed in specific contexts to others without appropriately testing their fit to the new environments (Collier & Mahon, 1993; Sartori, 1970, 1984). This risk is especially relevant with regard to online communication as it is far from obvious which concepts and theories predominantly developed in the context of mass media travel to this new environment (Neuman, 2016).

Let us have a closer look at two prominent studies ostensibly examining a prominent topic in political science—political polarization—through the lens of digital trace data. One of the early studies using digital trace data to examine specific usage patterns of political online use is

Adamic and Glance (2005). The authors analyzed linking patterns between 1,000 political blogs and the content of 22,884 posts on 40 prominent political blogs. They found:

> [...] a divided blogosphere: liberals and conservatives linking primarily within their separate communities, with far fewer cross-links exchanged between them. This division extended into their discussions, with liberal and conservative blogs focusing on different news, topics, and political figures (Adamic & Glance, 2005, p. 43).

In a similar vein, Conover, Ratkiewicz, Francisco, Gonçalves, Flammini, and Menczer (2011) examined 252,300 politically relevant tweets, focusing on retweet interactions between 18,470 users, and @message interactions by 7,175 users. The authors were interested in whether users sharing the same political leaning were more likely to interact with other users sharing their political convictions. They found that:

> The retweet network is highly polarized, while the mention network is not. To explain these observations we highlight the role of hashtags in exposing users to content they would not likely choose in advance. Specifically, users who apply hashtags with neutral or mixed valence are more likely to engage in communication with opposing communities (Conover et al., 2011, p. 95).

Both studies are highly instructive with regard to the benefits and limitations of work based on digital trace data in political communication research. Both studies use digital services and technology to collect, prepare, and analyze large data sets of political behavior online. They then focus on features of the data sets allowing for easy automated analysis. In this, they jettison some of the potential depth of their analysis. Nevertheless, this approach allows them to use advanced quantitative methods—in both cases network analysis—to identify specific structures of interactions between users on the services under analysis, i.e. blogs and Twitter. In this, both research teams are highly creative and original.

Still, the relevance of the analysis of linking behavior between blogs and the network structures of political talk on Twitter might not be apparent to all. To account for this, both teams situate their work within the larger discourse on political polarization in the USA and the internet’s perceived impact on this. Here, we find both studies to be representative of many others in this field. The connection to the theoretical debate on political polarization is provided either in the most fleeting of ways (Adamic & Glance, 2005) or in a highly selective and suggestive reading of the literature (Conover et al., 2011). Neither team acknowledges the depth and conflicting empirical evidence in the larger debate on political polarization in the USA (DiMaggio, Evans, & Bryson, 1996; Fiorina & Abrams, 2008) or the rich debate on the perceived role of media technology (Prior, 2013; Scheufele & Nisbet, 2013).

What constitutes a graver shortcoming is that both studies do not discuss nor propose convincing mechanisms of why the data traces analyzed—links between political blogs, retweets, or @messages—truly can be interpreted as expressions of political polarization. This depends on the data generating process giving rise to the signals used in the analysis. If we interpret links and retweets as predominantly driven by the intent to point to supporting evidence, additional information, or to point out new content to readers, it is far from surprising to find predominantly ideological homogenous linking patterns in politically focused blogs or Twitter feeds. After all, who while in a political debate points consistently to conflicting evidence or content provided by the opposing side? But is this form of public political expression and the performance of allegiance truly a sign of political polarization?

If we take the influential definition provided by DiMaggio et al. (1996) we can either see polarization as a state or a process:

> "Polarization as a state refers to the extent to which opinions on an issue are opposed in relation to some theoretical maximum. Polarization as a process refers to the increase in such opposition over time." (DiMaggio et al., 1996, p. 693)

To truly speak of political polarization online as a state, studies would thus have to either measure positions on political issues and demonstrate their systematic and extreme divergence between specific populations or political partisans. To detect a process of polarization online, studies would have to identify an increased spread between issue positions between specific populations or sources over time. In contrast, the approach chosen by Adamic and Glance (2005) and Conover et al. (2011) in interpreting clusters in public communicative interactions as evidence of polarization seems a far from ideal measurement strategy.

Both studies illustrate the problems arising from falling for the mirror fallacy by simply assuming chosen signals to be unmediated reflections of phenomena of interest. In doing so, the authors take a specific pattern in the data—homogenous linking or retweet patterns among groups of politically likeminded sources—as evidence of a larger phenomenon—political polarization in the USA—without providing any evidence why the signals chosen by them should actually speak to the underlying phenomenon as expressed in contemporary conceptualizations. Instead of speaking to political polarization online, their studies thus fall victim to conceptual stretching.

Again, my goal is not to single studies out. In fact, the chosen examples are highly cited and very influential for research using digital trace data. This makes identifying their strengths and limitations all the more important. Both studies are highly data-driven. They seem primarily motivated by the analytical potential provided by digital trace data and only secondarily by engaging with the theoretical debate about the phenomena used to situate their findings. This leads to the studies being very sophisticated in their use of data but at the same time very superficial in linking signals identified in digital data traces meaningfully to concepts used in larger debates in the social sciences.

Here lies a central challenge in the work with digital trace data. As long as researchers interested in in the use of digital trace data cannot or will not provide links between their data traces and larger political phenomena their research remains isolated and potentially only relevant to enthusiasts of internet phenomena. Given the growing social, political, and economic importance of digital communication, for some this might constitute no issue. Yet, in connecting their findings with more general concepts and phenomena at the center of social science discourse many researchers neglect the contingency of behavior documented in digital trace data. In this, they fail to provide evidence of whether their data can truly speak to the concept ostensibly at the center of their study or if their operationalization stretches concepts beyond their accepted use. To show the potential of digital trace data in contributing to social science mainstream requires careful conceptualization of the phenomena giving rise to specific data patterns. These conceptualizations have to consciously account for the mediated nature of the data traces under analysis and of social processes beyond the service providing the data (Jungherr et al., 2016a). Here, it is necessary for researchers working with digital trace data to engage more sophisticatedly with the current state of scientific debate in their chosen fields of interest as well as to develop a research culture more aware of the need for interpretation in linking signals to concepts of interest.

3.2 Proxies: Recognizing What's Too Good To Be True

A second group of studies uses digital trace data as proxies for other measures of political phenomena and individuals' traits and attitudes. Here, digital trace data become substitutes for other traditional data sources, such as surveys (Schoen, Gayo-Avello, Metaxas, Mustafaraj, Strohmaier, & Gloor, 2013). For this approach, the central question is how to ensure that signals identified in digital trace data truly measure or predict phenomena whose traditional and tested measurements they are supposed to replace.

Of these, one of the most prominent approaches—tellingly also one of the most controversial—is the attempt to use digital trace data as indicator of political support in election campaigns (DiGrazia, McKelvey, Bollen, & Rojas, 2013; O'Connor, Balasubramanyan, Routledge, & Smith, 2010; Tumasjan et al., 2010). These studies often share a highly sophisticated approach to data collection and analysis. Their authors also share an understanding of digital trace data containing explicit or implicit signals on political attitudes of individuals or public opinion in general. Yet, the quality of digital trace data in replacement of survey results is inconclusive (Murphy, Link, Childs, Tesfaye, Dean, Stern, Pasek, Cohen, Callegaro, & Harwood, 2014; Schober, Pasek, Guggenheim, Lampe, & Conrad, 2016). The problems with this approach become apparent once we examine some of the most prominent studies more closely.

Let us start with a study not directly concerned with political communication but with public health. Nevertheless, the study is highly influential for expectations in the diagnostic and predictive power of digital trace data for offline phenomena. In 2009, Choi and Varian (2012) demonstrated the apparent possibility of predicting—among other things—the spread of influenza based on Google search terms. Google even built a tool, Google Flu Trends[1], with which users could see real-time predictions of influenza trends worldwide. Quickly, this became one of the most highly cited cases for the perceived diagnostic and predictive potential of digital trace data. Yet, in 2013 the model started to fail spectacularly. This eventually led to the website being shut down. A detailed examination of this failure by Lazer et al. (2014) offers a very instructive account of the potential and limitations of digital trace data to infer offline phenomena.

Lazer et al. (2014) point out that the algorithm initially provided by Google fundamentally focused on identifying signals in a large collection of digital trace data that were successful in statistically predicting a much smaller set of data points identifying flu dynamics:

[1] https://www.google.org/flutrends/about/

> Essentially, the methodology was to find the best matches among 50 million search terms to fit 1152 data points. The odds of finding search terms that match the propensity of the flu but are structurally unrelated, and so do not predict the future, were quite high. [...] In short, the initial version of GFT [Google Flu Trends] was part flu detector, part winter detector. (Lazer et al., 2014, p. 1203)

This points to a central problem for theory-free statistical predictions based on big data sets. The larger the data set the higher the probability you find correlations. These might point to systematic relationships between two data points, yet they also might be spurious correlations. In specific contexts, this might not be much of an issue. For example, if the online-vendor Amazon uses theory-free algorithms to detect correlations in the consumption patterns between product A and product B, the nature of these correlations does not matter much. Amazon will use this information to post contextually relevant ads or might even adjust prices somewhat to point consumers of product A to product B. Once it detects a weakening of the correlation between the sales of both products it might simply adjust its recommendations accordingly. In cases like this correlations might in fact be all that is needed to achieve an increase in sales (Mayer-Schönberger & Cukier, 2013). Yet, once we try to use digital trace data for other purposes— be it scientific research or large scale diagnostics of public health—eliminating spurious correlations becomes of central importance. Instead of falling for the mirror fallacy by accepting at face value a signal found in digital trace data as indicator of a phenomenon of interest we have to demand and perform serious tests of indicator validity and reliability (Gerring, 2012).

Also with regard to the n=all fallacy, Google Flu Trends proves to be instructive. As Lazer et al. (2014) show, changes to a platform's code might break a prediction algorithm based on historical data collected on the platform at an earlier development stage:

> Google reported in June 2011 that it had modified its search results to provide suggested additional search terms and reported again in February 2012 that it was now returning potential diagnoses for searches including physical symptoms like "fever" and "cough". The former recommends searching for treatments of the flu in response to general flu inquiries, and the latter may explain the increase in some searches to distinguish the flu from the common cold. [...] Oddly, GFT bakes in an assumption that relative search volume for certain terms is statically related to external events, but search behavior is not

> just exogenously determined, it is also endogenously cultivated by the service provider. (Lazer et al., 2014, p. 1204)

This discussion shows that researchers need a clear understanding of the data generating process and its changes over time on the platform providing their data. This data generating process essentially guides user behavior and as such filters which phenomena of interests leave imprints in digital trace data and which remain invisible. Inherently, the data generating process determines how thick the slice of social or political phenomena is, that we can expect to find in digital trace data collected on specific services. To simply declare the specific slice available to us as "all" might be at best an attempt to define the problem away but at worst invalidate our studies.

The central issues with Google Flu Trends raised by Lazer et al. (2014) also can be found in other studies using digital trace data to predict or infer attitudes, behavior, or phenomena. Probably the most visible effort of this kind is the attempt to predict election results based on Twitter data. Efforts of this kind range from the use of very simple approaches—for example based on the counting of political actors' mentions (DiGrazia et al., 2013; Tumasjan et al., 2010)—to the use of very sophisticated advanced computational methods— such as attempts at theory-free identification of statistical properties in Twitter corpora statistically linked to positive or negative showings of political actors in opinion polls (Contractor & Faruquie, 2013; Marchetti-Bowick & Chambers, 2012). The literature has been heavily criticized for its methodological inconsistencies (Gayo-Avello, 2012, 2013; Metaxas, Mustafaraj, & Gayo-Avello, 2011) and the arbitrariness of some of its choices (Jungherr et al., 2012). These arguments need no restating here. Instead, I want to focus on what this literature tells us about the persistence of the mirror fallacy.

The claim that Twitter data would allow us to infer present levels of political support or even the prediction of elections rests on a series of papers showing successful statistical associations between some metrics identified in Twitter data and some metrics of political support, such as opinion polls or election results. At first sight, this might seem like a successful collective exercise in indicator validation. But, in fact, these studies turn out to be largely isolated and only seemingly examine the same phenomenon. While studies abound showing positive statistical associations between some signal identified on Twitter and some metric of political support, we do find next to no successful replication of the relationship between specific signals and specific metrics of support (Jungherr et al., 2016b). One of the few studies explicitly testing the stability of the link between Twitter-based metrics and election results over time clearly illustrates the instability of once established statistical relationships over time (Huberty, 2015). This points to the likelihood of positive cases being simply cases of over-fitting, like the one identified by Lazer et al. (2014) in their discussion of Google Flu Trends.

To make matters worse, the practice of exclusively reporting positive statistical associations between Twitter-based metrics and measures of political success without seriously testing mechanisms potentially given rise to these links might lead researchers to mistake the phenomenon indicated by their data. A central element of indicator validation is discriminant validation. A check for an indicator validly measuring a concept of interest and not a related one. In the case of measuring political support through Twitter-based metrics a likely candidate for a concept sometimes—but far from always—related is public attention. So instead of showing positive cases of Twitter identifying political support studies finding positive results might simply identify cases in which Twitter was successful in identifying public attention towards political actors which, in these cases, was positively linked with their success (Jungherr et al., 2016b).

Either way, be it through over-fitting or by mistaking the phenomena measured by digital trace data, the practice of relying on face validity in the demonstration of statistical associations between signals identified in digital trace data and other measures of phenomena of interest has proven to produce unreliable research of inconclusive meaning. Given the obvious problems illustrated above, the ongoing popularity of using digital trace as substitutes for other data sources to directly infer political phenomena or individuals' traits and attitudes surprises somewhat. Apparently exaggerated hope renders myopic. Through this collective myopia, the attempt to use digital trace data as proxies for other data sources has risen to considerable prominence among quantitatively oriented social scientists not interested in internet-based phenomena as such. Here, a central weakness in overly quantitatively oriented social science research becomes apparent. To a regression model every data point seems the same no matter the data generating process. Yet, as the examples above have shown, the research community should be very careful in accepting this. Balanced reviews in the public opinion research community point to the limits of digital trace data as substitutes for other measurement approaches in the social sciences (Murphy et al., 2014; Schober et al., 2016). To ensure we are using digital trace data meaningfully, we have to move from the early enthusiasm-driven stage and its reliance on a proof-of-concept logic.

Instead, we should apply and demand the rigorous application of social scientific methodology and indicator validation. If those are applied, the chances for digital trace data providing substitutes for other data collection approaches to measure individuals' attitudes, traits, or social phenomena seem slim, rendering the endeavor a likely case of misplaced academic industry. Instead, the true potential of digital trace data might lie in their combination with established measurement approaches, not their substitution.

3.3 Hybrids: Connecting Different Data Sources

A third group of studies is characterized by the use of a combination of digital trace data and traditional data sources. Here lies significant research potential. On the one hand, hybrid designs allow validity checks of signals identified in digital trace data by comparing them to a ground truth established by traditional methods, which are much better understood with regard to the benefits and problems arising from their measurement approaches. On the other hand, hybrid designs are tailor-made to better understand the growing interconnection between digital technology and society. Here, I want to focus on studies examining this second question with regard to politics.

The analysis of the interconnection of political behavior on- and offline can take many forms. For example, various authors have looked at the content of Twitter messages and linked it to content in media coverage (Jungherr, 2014; Neuman, Guggenheim, Mo Jang, & Bae, 2014; Vargo, Guo, McCombs, & Shaw, 2014; Wells, Shah, Pevehouse, Yang, Pelled, Boehm, Lukito, Gosh, & Schmidt, 2016), the dynamics of specific media programs and comments on Twitter (Trilling, 2015), or tried to link online interventions to offline behavior (Bond et al., 2012). These studies follow the lead of various pieces theorizing and qualitatively demonstrating that political communication on online channels appears to be highly interconnected with political phenomena—such as political coverage in traditional mass media or political participation—but potentially following distinct patterns (Anstead & O'Loughlin, 2015; Chadwick, 2013; Kreiss, 2016). By explicitly addressing this link, researchers are able to anchor their analyses of behavioral patterns online in central debates of political communication research by demonstrating that online communication has become an integral part of the contemporary political communication environment. They thereby offer a convincing answer to the relevancy question that pure digital ethnography remains vulnerable to. By conceptualizing political behavior online only as one element of political communication in general, albeit one potentially following distinct and channel-specific patterns, these studies avoid overgeneralizing their findings, a problem inherent in the use of digital trace data as substitutes for other measurement approaches.

One promising area for the linking of digital trace data with other data sources is the analysis of political media coverage and public reactions to it. Here, various studies have demonstrated the research potential inherent in Twitter data by comparing the mentions of specific topics or actors on Twitter and their coverage in traditional media either over extended time periods or over the course of one program (Jungherr, 2014; Neuman et al., 2014; Trilling, 2015; Vargo et al., 2014; Wells et al., 2016). These studies allow the analysis of attention dynamics in a political vocal population on Twitter. Thus, digital trace data hold the potential of providing insights into the interconnection between everyday political talk and political coverage in mass media, a topic largely neglected for methodological reasons but of clear interests to political communication research (Dayan & Katz, 1992; Gamson, 1992). These studies also allow a deeper understanding of the potential effects of political media coverage on online activity and vice versa. Thus, the combination of digital trace data and conventional data sources promise insights in the dynamics of the "hybrid media system" (Chadwick, 2013).

Another possibility is the linking of digital trace data with data collected by surveys. For example Jungherr, Schoen, Posegga, and Jürgens (2015) compared political topics identified in Twitter messages with topics mentioned in survey responses to the query for an assessment of the most important problem facing Germany. We showed that there was only very limited overlap between topics identified from tweets referring to politics and the most-important-problem question. On one hand, this shows that Twitter provides a poor proxy for identifying public agendas—another refutation of the mirror hypothesis. On the other hand, the findings also point to the necessity to explicitly conceptualize the nature of political talk on Twitter. Is the frequency of topic mentions on Twitter an indicator of their central importance to political life? Or is the frequency of mentions better understood as an indicator which objects were central to public attention? If so, what drives differences between these lists? Is it that people only publicly post on topics they want to be seen as being associated with, in other words a bias arising from social desirability? Or is it that not everything we pay attention to in everyday interactions is automatically prominent if we are asked to reflect and provide a ranked list of important political problems? In other words, do the data speak to a divide between salience and relevance of topics? Currently, these are unanswered questions potentially providing insights into the nature of political talk and political communication in general. The difference between the measurements of seemingly related concepts in digital trace data and conventional data collection approaches thus might point to promising research puzzles to which we would have remained blind if we had only used established data sources. In these cases, error, deviations between signals identified in different data sources, might carry meaning. Meaning, we would have remained oblivious to if we had focused only on trying to make digital trace data fit patterns identified in established data source.

The prominence of studies combining different types of data sources is only set to increase. Increasingly journalists, campaigners, and politicians take to digital trace data to identify supposed trends in public opinion or test messages (Anstead & O'Loughlin, 2015). This can lead patterns in digital trace data to influence politics or media coverage without truly mirroring the supposed phenomena (Jungherr et al., 2016b). This process can only be understood and evaluated using hybrid designs.

Hybrid designs are of course also vulnerable to the n=all and the mirror fallacies. Yet, by connecting digital trace data to other more conventional data sources they become less exceptional and questions of adequate research designs and conceptual linkages arise naturally. Still, also with studies falling within this category, we have to critically ask ourselves if digital trace data accurately speak to the phenomena of interest or if authors simply followed a data-driven impulse and, after settling on a data set, linked their data only superficially to a seemingly vibrant concept in social science discourse. The link between concepts and measurements is as important here as it was with the other approaches and the social sciences in general.

4 The Road Ahead: A Measurement Theory for Working with Digital Trace Data in the Social Sciences

As discussed above, the use of digital trace data in the social sciences is dominated by two fallacies severely limiting the application of these data. These fallacies are deeply rooted in a naive reading of signals extracted from digital data traces as directly representing social or political phenomena of interest and the indifference of leading practitioners to sampling procedures accounting for inherent limits in coverage and availability of digital trace data. Overcoming these fallacies is central for digital trace data to develop into an accepted data source in the social science mainstream. The alternative is that digital trace data end up as a sandbox for applied computer scientists.

To social scientists, the widespread influence of these fallacies is surprising. The social sciences and statistics look back on a long and fruitful debate on capturing phenomena of interest in concepts, linking these concepts to measurements, and statistical procedures allowing researchers to draw valid inferences based on these data (Donoho, 2015; Efron & Hastie, 2016; Gerring, 2012; Goertz, 2005; Goertz & Mahoney, 2012; Hand, 2004). As shown above, digital trace data share much more characteristics with traditional measurement approaches than early enthusiasts claimed, rendering the former subject to much of the same limitations as the latter. Yet, the current debate in the field focuses much more on the perceived exceptionality of digital trace data than the characteristics they share with other data sources.

Measurement theories are a central feature of quantitative approaches in the social sciences and psychology. Measurement theories conceptualize and test the relationship between signals found in data and their relationship towards phenomena of interest (Hand, 1996, 2004). Central is the acknowledgement that phenomena of interests are not directly connected with signals found in data. Instead, they are linked through a chain of mediating steps. Phenomena of interest have to be translated into meaningful and sophisticated concepts (Goertz, 2005). These concepts then have to be linked to signals found in data. It is crucial that this step accounts for the data generating process underlying specific data sets and their consequences for the applicability of the linked concepts (Gerring, 2012). In a final step, measurement theories have to discuss valid modes of inference based on the underlying data sets and research interests (Efron & Hastie, 2016). Only a sophisticated understanding of this mediating process allows for the robust use of specific data sets and research methods in scientific work.

In the case of digital trace data this mediating process is generally ignored in favor of a simplistic implicit reading of signals found in digital trace data as direct expressions of social or political phenomena of interest. As shown above, this limits the use of digital trace data in the social sciences severely. For their work to be taken seriously, researchers interested in the use of these data for significant contributions to core discourses of the social sciences should therefore push for the development of a measurement theory. This means focusing on the development or adaption of concepts for the research environment of digital devices and services and linking these concepts critically to measurements found in digital trace data. Here, we also need to move away from a computer science publication logic based on proofs of concepts and prototyping of algorithms or models. This approach, while giving rise to a quick succession of innovative research methods, more often than not accepts face validity as test for plausible links between signals and phenomena of interest. This leads papers to have only dubious connections with underlying social or political phenomena ostensibly of interest and speak instead only to the application of a specific method or algorithm in a specific case with ill-understood contextual conditions and effects. Instead, social scientists working with digital trace data should push actively for a culture of serious and sophisticated concept validation (Jungherr et al., 2016). Similarly, researchers should focus not only on the novelty of new approaches but instead on checking very consciously for the potential biasing effects of specific approaches to data collection and the reliability of specific quantitative methods (González-Bailón & Paltoglou, 2015; Ruths & Pfeffer, 2014).

Of similar importance is the development of explicit research designs for the work with digital trace data in the social sciences (Salganik, 2017). By now, it has become obvious that the early expectations of being able to neglect addressing research designs since one was wandering through an n=all world were misplaced. Instead, the inherent limits in reach and coverage of digital trace data have become common knowledge. Now it is time for this common knowledge to find expression in research designs. This includes addressing questions of sampling, the reliability of different approaches to data selection and collection, and the valid combination of digital trace data and other data sources—such as surveys or digital trace data collected on various services following different data access politics. The relevance of this is only likely to increase when data access policies of digital service providers become more restrictive as they increasingly try to monetize this aspect of their services.

These demands might seem overly ambitious for researchers interested in testing the use of digital trace data in service of specific research questions. Yet, it is important to note that the research agenda sketched above does not necessarily have to be realized at once. Instead, it is highly likely that the steps sketched above constitute a collective project over many years. Still, while the establishment of a complete measurement theory of the work with digital trace data in the social sciences might seem prohibitively ambitious, it is necessary for researchers to address the issues raised above. More often than not, this will mean becoming more transparent in linking signals found in digital trace data to concepts of interest and underlying phenomena. This is very much in the reach of single research papers and would be a very big step forward from the state of "unconscious thinking" (Sartori, 1970) we as a field currently seem to be more fond of than appropriate.

This adjustment in the work with digital trace data in the social sciences is necessary for this work to seriously contribute to the core debates in the social sciences. This means taking the “social science” in “computational social science” seriously. The social sciences have a rich tradition in grappling with social and political phenomena. This has led to the emergence of sophisticated pluralistic approaches to conceptualization, measurement, and inference (Brady & Collier, 2010; Gerring, 2012; Goertz & Mahoney, 2012). Currently, these debates are largely ignored in the work with digital trace data. This seems mostly due to an approach to data analysis predominantly driven from a software developer perspective where the inherent divide between signal and concept of interest features less prominently. Therefore, it is no surprise to find that work with digital trace data seemingly addressing questions on social and political life currently can best be characterized as a subfield in applied computer science. Instead of meaningfully contributing to contemporary debates in the social sciences, these studies run algorithms and models on data without validly linking them to relevant concepts. Thus, these studies speak more to the workings of underlying algorithms and methods than the phenomena ostensibly being examined. Instead of following this lead, social scientists interested in working with digital trace data should push actively for the development of a serious measurement theory of digital trace data in the social sciences.

5 Normalizing Digital Trace Data

It is time to dispense with the exceptionalist rhetoric of the early days and to integrate digital trace data in the social science workflow and toolset. This means adapting established research practices to the work with digital trade data instead of proclaiming a new age based on the perceived exceptionalism of this data source. As shown above, obvious tasks for this normalization lie in the conceptualization of digital trace data in ways that reflect their service-specific mediating data generating processes (Jungherr, 2015; Jungherr et al., 2016a), the development of a measurement theory of the work with digital trace data in the social sciences as to identify which phenomena of interest are reflected by the data and which remain hidden from it (Jungherr et al., 2016b), and approaches that allow the linking of digital trace data with traditional social science data sources, such as surveys or content analyses of political media coverage.

This also means thinking more consciously about research design. Currently, the n=all fallacy tempts researchers to spend very little time thinking about potential biases to their data collection. Digital trace data are found data, i.e. data produced as byproducts of online interactions by users (Howison et al., 2011). Currently, researchers predominantly have taken these data as found and constructed research based on their features available to them, not necessarily the features providing the best basis for interesting inferences. Various studies have shown that taking a more active approach in their research design, for example through experiments, provides interesting avenues for future research (Bond et al., 2012; Salganik & Watts, 2009). Other studies have shown that taking a more conscious approach to sampling specific populations on Twitter is providing a more robust footing than simply claiming to collect "all" relevant data (Diaz et al., 2016; Lin et al., 2014).

In short, in order to realize the potential of the "measurement revolution" (Watts, 2011) provided by digital trace data in the social sciences, we have to subject this new data source to the same rigorous tests in conceptualization and methodology as we do other data sources (Gerring, 2012; Hand, 2004). Currently, the novelty of digital trace data has given researchers somewhat of a free pass with regard to these questions. This free pass has to be revoked as the work with this data source slowly moves closer to the center of social science discourse. For researchers working with these data this means on one hand a considerable chance to contribute foundational studies at the center of social science core discourses. On the other hand, this also means they have to become more disciplined in addressing the limits of their data sources. In this, the way forward very likely holds fewer exceptional, counter-intuitive, or spectacular findings than we are made to believe. Instead we might witness incremental progress based on highly contextually dependent findings. While this might not be the most promising way to make one's way into airport bookstalls or Buzzfeed's list of "10 research results that will absolutely shock you", there is no reason for disappointment. After all, this is the natural progression of scientific fields and method development.

References

Adamic, L. A., & Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election. In J. Adibi, M. Grobelnik, D. Mladenic, & P. Pantel (Eds.), *LinkKDD 2005: Proceedings of the 3rd international workshop on Link discovery* (pp. 36-43). New York, NY: ACM.

Adcock, R., & Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review, 95*(3), 529-546. doi:10.1017/S0003055401003100

Anstead, N., & O'Loughlin, B. (2015). Social Media Analysis and Public Opinion: The 2010 UK General Election. *Journal of Computer-Mediated Communication, 20*(2), 204-220. doi:10.1111/jcc4.12102

Ausserhofer, J., & Maireder, A. (2013). National Politics on Twitter: Structures and topics of a networked public sphere. *Information, Communication & Society, 16*(3), 291-314. doi:10.1080/1369118x.2012.756050

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science, 348*(6239), 1130-1132. doi:10.1126/science.aaa1160

Blank, G. (2016). The Digital Divide Among Twitter Users and Its Implication for Social Research. *Social Science Computer Review*. doi:10.1177/0894439316671698

Bollen, J., Mao, H., & Zheng, X. J. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science, 2*, 1-8.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature, 489*(7415), 295-298. doi:10.1038/nature11421

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society, 15*(5), 662-679. doi:10.1080/1369118x.2012.678878

Brady, H. E., & Collier, D. (Eds.). (2010). *Rethinking Social Inquiry: Diverese Tools, Shared Standards* (2 ed.). Lanham: Rowman & Littlefield Publishers Inc.

Bruns, A., & Highfield, T. (2013). Political Networks on Twitter: Tweeting the Queensland State Election. *Information, Communication & Society, 16*(5), 667-691. doi:10.1080/1369118x.2013.782328

Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web Survey Methodology*. London, UK: SAGE Publications.

Chadwick, A. (2013). *The Hybrid Media System: Politics and Power*. New York, NY: Oxford University Press.

Choi, H., & Varian, H. R. (2012). Predicting the Present with Google Trends. *Economic Record, 88*, 2-9. doi:10.1111/j.1475-4932.2012.00809.x

Collier, D., & Mahon, J. E. (1993). Conceptual “Stretching” Revisited: Adapting Categories in Comparative Analysis. *American Political Science Review, 87*(4), 845-855. doi:10.2307/2938818

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Political Polarization on Twitter. In N. Nicolov, J. G. Shanahan, L. A. Adamic, R. Baeza-Yates, & S. Counts (Eds.), *ICWSM 2011: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (pp. 89-96). Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI).

Contractor, D., & Faruquie, T. A. (2013). Understanding Election Candidate Approval Ratings Using Social Media Data. In D. Schwabe, V. Almeida, H. Glaser, R. Baeza-Yates, & S. Moon (Eds.), *WWW 2013: Proceedings of the 22nd International Conference on World Wide Web* (pp. 189-190). Geneva, CH: International World Wide Web Conferences Steering Committee.

Dayan, D., & Katz, E. (1992). *Media Events: The Live Broadcasting of History*. Cambridge, MA: Harvard University Press.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social Media as a Measurement Tool of Depression in Populations. In H. Davis, H. Halpin, A. Pentland, M. Bernstein, & L. A. Adamic (Eds.), *WebSci 2013: Proceedings of the 5th Annual ACM Web Science Conference* (pp. 47-56). New York, NY: ACM.

Diaz, F., Gamon, M., Hofman, J. M., Kiciman, E., & Rothschild, D. (2016). Online and Social Media Data As an Imperfect Continuous Panel Survey. *PLoS One, 11*(1), e0145406. doi:10.1371/journal.pone.0145406

DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior. *PLoS One, 8*(11), e79449. doi:10.1371/journal.pone.0079449

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have Americans' Social Attitudes Become More Polarized? *American Journal of Sociology, 102*(3), 690-755. doi:10.1086/230995

Donoho, D. (2015). *50 Years of Data Science*. Paper presented at the John W. Tukey 100th Birthday Celebration at Princeton University, Princeton, NJ. http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf

Efron, B., & Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge, UK: Cambridge University Press.

Fiorina, M. P., & Abrams, S. J. (2008). Political Polarization in the American Public. *Annual Review of Political Science, 11*(1), 563-588. doi:10.1146/annurev.polisci.11.053106.153836

Freelon, D. (2015). On the Cutting Edge of Big Data: Digital Politics Research in the Social Computing Literature. In S. Coleman & D. Freelon (Eds.), *Handbook of Digital Politics* (pp. 451-472). Northampton, MA: Edward Elgar Publishing.

Gamson, W. A. (1992). *Talking Politics*. Cambridge, UK: Cambridge University Press.

Gayo-Avello, D. (2012). No, You Cannot Predict Elections with Twitter. *IEEE Internet Computing, 16*(6), 91-94. doi:10.1109/mic.2012.137

Gayo-Avello, D. (2013). A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data. *Social Science Computer Review, 31*(6), 649-679. doi:10.1177/0894439313493979

Gerring, J. (2008). The Mechanismic Worldview: Thinking Inside the Box. *British Journal of Political Science, 38*(1), 161-179. doi:10.1017/S0007123408000082

Gerring, J. (2010). Causal Mechanisms: Yes, But ... *Comparative Political Studies, 43*(11), 1499-1526. doi:10.1177/0010414010376911

Gerring, J. (2012). *Social Science Methodology: A Unified Framework* (2 ed.). Cambridge, UK: Cambridge University Press.

Goertz, G. (2005). *Social Science Concepts: A User's Guide*. Princeton, NJ: Princeton University Press.

Goertz, G., & Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.

Golder, S. A., & Macy, M. W. (2014). Digital Footprints: Opportunities and Challenges for Online Social Research. *Annual Review of Sociology, 40*(1), 129-152. doi:10.1146/annurev-soc-071913-043145

González-Bailón, S., & Paltoglou, G. (2015). Signals of Public Opinion in Online Communication: A Comparison of Methods and Data Sources. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 95-107. doi:10.1177/0002716215569192

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis, 21*(3), 267-297. doi:10.1093/pan/mps028

Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2 ed.). Hoboken, NJ: Wiley.

Hand, D. J. (1996). Statistics and the Theory of Measurement. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 159*(3), 445. doi:10.2307/2983326

Hand, D. J. (2004). *Measurement Theory and Practice: The World Through Quantification*. London, UK: Wiley.

Hargittai, E. (2015). Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 63-76. doi:10.1177/0002716215570866

Howison, J., Wiggins, A., & Crowston, K. (2011). Validity Issues in the Use of Social Network Analysis with Digital Trace Data. *Journal of the Association for Information Systems, 12*(12), 767-797.

Huberty, M. (2015). Can We Vote with Our Tweet? On the Perennial Difficulty of Election Forecasting with Social Media. *International Journal of Forecasting, 31*(3), 992-1007. doi:10.1016/j.ijforecast.2014.08.005

Jungherr, A. (2014). The Logic of Political Coverage on Twitter: Temporal Dynamics and Content. *Journal of Communication, 64*(2), 239-259. doi:10.1111/jcom.12087

Jungherr, A. (2015). *Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research*. Cham, CH: Springer.

Jungherr, A., Jürgens, P., & Schoen, H. (2012). Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment". *Social Science Computer Review, 30*(2), 229-234. doi:10.1177/0894439311404119

Jungherr, A., Schoen, H., & Jürgens, P. (2016a). The Mediation of Politics through Twitter: An Analysis of Messages posted during the Campaign for the German Federal Election 2013. *Journal of Computer-Mediated Communication, 21*(1), 50-68. doi:10.1111/jcc4.12143

Jungherr, A., Schoen, H., Posegga, O., & Jügens, P. (2016b). Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support. *Social Science Computer Review*. doi:10.1177/0894439316631043

Jungherr, A., Schoen, H., Posegga, O., & Jürgens, P. (2015). *Characterizing Political Talk on Twitter: A Comparison Between Public Agenda, Media Agendas, and the Twitter Agenda with Regard to Topics and Dynamics*. Paper presented at the American Political Science Association (APSA) Annual Meeting 2015, San Francisco.

King, G. (2011). Ensuring the Data-Rich Future of the Social Sciences. *Science, 331*(6018), 719-721. doi:10.1126/science.1197872

Kreiss, D. (2016). Seizing the Moment: The Presidential Campaigns Use of Twitter During the 2012 Electoral Cycle. *New Media & Society, 18*(8), 1473-1490. doi:10.1177/1461444814562445

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science, 343*(6176), 1203-1205. doi:10.1126/science.1248506

Lazer, D., Pentland, A., Adamic, L. A., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J. H., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational Social Science. *Science, 323*(5915), 721-723. doi:10.1126/science.1167742

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-Tracking and the Dynamics of the News Cycle. In J. Elder, F. Soulie-Fogelman, P. Flach, & M. Zaki (Eds.), *KDD 2009: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 497–506). New York, NY: ACM.

Lin, Y.-R., Keegan, B., Margolin, D., & Lazer, D. (2014). Rising Tides or Rising Stars? Dynamics of Shared Attention on Twitter During Media Events. *PLoS One, 9*(5), e94093. doi:10.1371/journal.pone.0094093

Marchetti-Bowick, M., & Chambers, N. (2012). Learning for Microblogs with Distant Supervision: Political Forecasting with Twitter. In W. Daelemans (Ed.), *EACL '12: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 603-612). Stroudsburg, PA: Association for Computational Linguistics.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. New York, NY: Houghton Mifflin Harcourt.

Metaxas, P. T., Mustafaraj, E., & Gayo-Avello, D. (2011). How (Not) to Predict Elections. In A. Vinciarelli, M. Pantic, E. Bertino, & J. Zhan (Eds.), *SocialCom 2011: The 3rd IEEE Interna- tional Conference on Social Computing* (pp. 165-171). Washington, DC: IEEE.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In E. Kiciman, N. B. Ellison, B. Hogan, P. Resnick, & I. Soboroff (Eds.), *ICWSM 2013: Proceedings of the 7th International AAAI Conference on Weblogs and Social Media* (pp. 400-408). Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI).

Murphy, J., Link, M. W., Childs, J. H., Tesfaye, C. L., Dean, E., Stern, M., Pasek, J., Cohen, J., Callegaro, M., & Harwood, P. (2014). Social Media in Public Opinion Research: Executive Summary of the Aapor Task Force on Emerging Technologies in Public Opinion Research. *Public Opinion Quarterly, 78*(4), 788-794. doi:10.1093/poq/nfu053

Neuman, W. R. (2016). *The Digital Difference: Media Technology and the Theory of Communication Effects*. Cambridge, MA: Harvard University Press.

Neuman, W. R., Guggenheim, L., Mo Jang, S., & Bae, S. Y. (2014). The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. *Journal of Communication, 64*(2), 193-214. doi:10.1111/jcom.12088

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In M. Hearst, W. Cohen, & S. Gosling (Eds.), *ICWSM 2010: Proceedings of the 4rth International AAAI Conference on Weblogs and Social Media* (pp. 122-129). Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI).

Prior, M. (2013). Media and Political Polarization. *Annual Review of Political Science, 16*(1), 101-127. doi:10.1146/annurev-polisci-100711-135242

Rainie, L., & Wellman, B. (2012). *Networked: The New Social Operating System*. Cambridge, MA: The MIT Press.

Resnick, P., Adar, E., & Lampe, C. (2015). What Social Media Data We Are Missing and How to Get It. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 192-206. doi:10.1177/0002716215570006

Rogers, R. (2013). *Digital Methods*. Cambridge, MA: The MIT Press.

Ruths, D., & Pfeffer, J. (2014). Social Media for Large Studies of Behavior. *Science, 346*(6213), 1063-1064. doi:10.1126/science.346.6213.1063

Salganik, M. J. (2017). *Bit By Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

Salganik, M. J., & Watts, D. J. (2009). Web-Based Experiments for the Study of Collective Social Dynamics in Cultural Markets. *Topics in Cognitive Science, 1*(3), 439-468. doi:10.1111/j.1756-8765.2009.01030.x

Sartori, G. (1970). Concept Misformation in Comparative Politics. *American Political Science Review, 64*(4), 1033-1053. doi:10.2307/1958356

Sartori, G. (1984). Guidelines for Concept Analysis. In G. Sartori (Ed.), *Social Science Concepts: A Systematic Analysis* (pp. 15-85). London, UK: Sage.

Scheufele, D. A., & Nisbet, M. C. (2013). Online News and the Demise of Political Disagreement. In C. T. Salmon (Ed.), *Communication Yearbook 36* (pp. 44-53). New York, NY: Routledge.

Schober, M. F., Pasek, J., Guggenheim, L., Lampe, C., & Conrad, F. G. (2016). Social Media Analyses for Social Measurement. *Public Opinion Quarterly, 80*(1), 180-211. doi:10.1093/poq/nfv048

Schoen, H., Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The Power of Prediction with Social Media. *Internet Research, 23*(5), 528-543. doi:10.1108/IntR-06-2013-0115

Schroeder, R. (2016). Big Data and Communication Research *Oxford Research Encyclopedia of Communication*. Oxford, UK: Oxford University Press.

Strohmaier, M., & Wagner, C. (2014). Computational Social Science for the World Wide Web. *IEEE Intelligent Systems, 29*(5), 84-88. doi:10.1109/mis.2014.80

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

Trilling, D. (2015). Two Different Debates? Investigating the Relationship Between a Political Debate on TV and Simultaneous Comments on Twitter. *Social Science Computer Review, 33*(3), 259-276. doi:10.1177/0894439314537886

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In M. Hearst, W. Cohen, & S. Gosling (Eds.), *ICWSM 2010: Proceedings of the 4rth International AAAI Conference on Weblogs and Social Media* (pp. 178-185). Menlo Park, CA: Association for the Advancement of Artificial Intelligence (AAAI).

Vargo, C. J., Guo, L., McCombs, M., & Shaw, D. L. (2014). Network Issue Agendas on Twitter During the 2012 U.S. Presidential Election. *Journal of Communication, 64*(2), 296-316. doi:10.1111/jcom.12089

Watts, D. J. (2011). *Everything Is Obvious: How Common Sense Fails Us*. New York, NY: Random House.

Wells, C., Shah, D. V., Pevehouse, J. C., Yang, J. H., Pelled, A., Boehm, F., Lukito, J., Gosh, S., & Schmidt, J. L. (2016). How Trump Drove Coverage of the Nomination: Hybrid Media. *Political Communication, 33*(4), 669-676. doi:10.1080/10584609.2016.1224416