A Paucity of Data

The Digital Platforms’ Responses to Pillar 5 of the Code of Practice on Disinformation

Rebekah Tromble
Associate Professor, School of Media and Public Affairs
Associate Director, Institute for Data Democracy and Politics
The George Washington University

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Executive Summary

In October 2018 the major online platforms signed the European Code of Practice on Disinformation, a self-regulatory framework designed to "address the spread of online disinformation and fake news" across the European Union (European Commission, 2019). Under Pillar 5 of the Code of Practice, Facebook, Google, and Twitter committed to supporting "good faith independent efforts to track Disinformation and understand its impact…" (European Commission, 2018, Chapter II.E., para. 12).

This report provides an in-depth assessment of the platforms' responses to Pillar 5, focusing on efforts made to support independent academic research in particular. It examines recent scholarly research on two issues at the heart of the Code of Practice—online political ad micro-targeting and disinformation—and seeks to assess the extent to which this research has been enabled and supported by Google, Facebook, and Twitter. It examines the gaps remaining in our knowledge and evaluates the barriers to advancing our understanding in these areas.

Based on detailed analyses of the data, analytical tools, and resources made available by the online platforms, I argue that Facebook, Twitter, and Google have fallen short of their commitments under Pillar 5. Though the public ad archives the platforms have developed do provide advances in data access and transparency, scientific research remains extremely difficult to conduct. Scholars are unable to systematically access data, including data provided via the ad archives. What data are made available cannot be independently verified as either complete or representative. And the platforms are not sharing crucial data points, including data on ad targeting and user engagement with disinformation. As a result, the most important questions about the extent and impact of micro-targeting and disinformation remain unanswered. Unless better mechanisms for systematic access to verifiable data are secured, additional progress will be limited.

The report concludes with a series of recommendations for the platforms and policymakers. These recommendations focus on the types of information that should be included in the public ad archives, as well as regulatory mechanisms needed to ensure that responsible and ethical scientific inquiry can be pursued with the necessary platform data. In the latter case, I recommend that regulatory authorities (a) begin to require that platforms share certain types of data with scholars, (b) help to establish research "safe harbors," and (c) introduce independent audits of data shared by the platforms.

I. Introduction

In October 2018 the major online platforms signed the European Code of Practice on Disinformation, a self-regulatory framework designed to "address the spread of online disinformation and fake news" across the European Union (European Commission, 2019). Under Pillar 5 of the Code, Facebook, Google, and Twitter committed to supporting "good faith independent efforts to track Disinformation and understand its impact…" (European Commission, 2018, Chapter II.E., para. 12).

This report provides an in-depth assessment of the platforms' responses to Pillar 5, focusing on efforts made to support independent academic research in particular. It examines recent scholarly research on two issues at the heart of the Code of Practice—online political ad micro-targeting and disinformation—and seeks to assess the extent to which this research has been enabled and supported by Google, Facebook, and Twitter. It examines the gaps remaining in our knowledge and evaluates the barriers to advancing our understanding in these areas.

Based on detailed analyses of the data, analytical tools, and resources made available by the online platforms, I argue that Facebook, Twitter, and Google have fallen short of their commitments under Pillar 5 of the Code of Practice. Though the public ad archives the platforms have developed do provide advances in data access and transparency, scientific research remains extremely difficult to conduct. Scholars are unable to systematically access data, including data provided via the ad archives. What data are made available cannot be independently verified as either complete or representative. And the platforms are not sharing crucial data points, including data on ad targeting and user engagement with disinformation. As a result, the most important questions about the extent and impact of micro-targeting and disinformation remain unanswered. Unless better mechanisms for systematic access to verifiable data are secured, additional progress will be limited.

The report proceeds as follows: Section II offers an overview of recent scholarly literature, first on online political micro-targeting, then on disinformation. It takes a closer look at research conducted to date using the platforms' ad archives. It also lays out a set of critical questions that require additional scientific investigation. Section III turns to the specific datasets, tools, and resources that the platforms have provided thus far. Some of these are explicitly intended for academic research, but most are not. The latter have been adopted (and sometimes adapted) by scholars to try to address scientific research needs. Section IV provides insights into the types of data needed to answer some of the questions laid out in Section II but which are not covered in the current platform offerings discussed in Section III. Section IV also offers an analysis of some of the ways in which these data, which may involve sensitive information, could be ethically and responsibly shared and analyzed. Section V concludes with a series of recommendations for the platforms and policymakers.

II. Recent Research

A. Online Political Micro-Targeting

Though research into micro-targeting in online ads (and related concepts such as ad personalization and behavioral advertising) has accelerated in recent years, this field of inquiry is still in its relative youth. There are many more studies into personalization in commercial advertising than there are examining political advertising, per se (for a recent review, see Varnali, 2019). Searches of Google Scholar, Academic Search Premier, and Web of Science for the terms "micro-target*," "microtarget*," "personalized ad*," and "(behavioral OR behavioural) ad*" published between January 1, 2017 and December 3, 2019 revealed just a handful of studies examining online political advertising. Some of these studies provide normative (Papakyriakopoulos, 2018) or legal (Harker, 2019; Witzleb et al, 2019) frameworks, while empirical work typically either offers a sense of how digital micro-targeting is used in political campaigns (Chester & Montgomery, 2017; Dobber et al, 2017; Dommett, 2019; Kruschinski & Haller, 2017) or provides insights into how people perceive personalized political ads (Baum et al, 2019; Dobber et al, 2019).

In theory, the ad archives provided by Facebook, Google, and Twitter could help expand empirical research. However, very little scholarly inquiry has yet made use of the ad archives. An exhaustive search of the same databases listed above returned just four studies. And the conclusions drawn in these studies remain very narrow. Findings to date address broad categories—e.g., geographic, age, and gender—used to target users (Edelson et al, 2019; Gosh, Venkatadri, & Mislove, 2019; Hegelich & Serrano, 2019), the amounts spent on political ads (Edelson et al, 2019; Gosh, Venkatadri, & Mislove, 2019; Jamison et al, 2019), and audience size (Edelson et al, 2019). Two of the four studies focus exclusively on Facebook ads. Hegelich and Serreno (2019) assess both Facebook and Google advertising, while Edelson and colleagues (2019) analyze data from all three platforms. However, Edelson et al noted that their findings for Google and Twitter are particularly limited because these two platforms provide very little useful targeting data. (p. 9).

Note, too, that three of the four studies focus on the US; Hegelich and Serrano (2019) provide the only analysis from Europe, assessing Facebook and Google ad spends in Germany. Given that the platforms only began implementing their ad archives in Europe in early 2019, the lack of European research is not particularly surprising. However, as discussed in Section III below, unless Facebook, Google, and Twitter make significant changes, these archives are unlikely to generate much new scholarly research from any country or region.

Yet new research is needed to answer fundamental questions about online political micro-targeting. Open and pressing questions include:

- How much political micro-targeting occurs in different political jurisdictions? What variables (e.g., political and electoral system, campaign finance or other election laws) account for observed variation across jurisdictions?
- Do people perceive ad personalization/targeting? What explains different user perceptions?

- If users perceive ad personalization, how do they respond? Do they feel uncomfortable in any way (Boerman et al, 2017; Dobber et al, 2019)? And do users' behaviors match their perceptions? For example, if a user is aware of ad targeting and responds negatively, do they in turn restrict their interactions with the group, party, or candidate associated with the ad? Do they act in privacy-protecting ways (Baum et al, 2019; Boerman et al, 2018)?
- What are the impacts of features and programs designed to inform users about political ad micro-targeting? Do they alter perceptions or change behavior based on these interventions (Dogruel, 2019)?
- Does political micro-targeting change users' attitudes or behaviors? Does personalization, for instance, increase ad click-through or share rates? Does it lead to merchandise purchases? Does it change people's attitudes toward political actors or groups? Does it change their beliefs about political issues? And does it alter voting behaviors?
- What are the impacts of different levels and types of personalization? For example, does direct targeting based on browsing or search history have different effects than indirect targeting achieved through the platforms' "custom audience" and "lookalike" features?

B. Disinformation

In recent years scholarly research on disinformation has burgeoned and is much more plentiful than work on political micro-targeting. Tucker and colleagues (2018) identify five categories of empirical disinformation research:

- Disinformation's producers,
- The strategies and tactics used to spread disinformation,
- The effects of exposure to online disinformation,
- Disinformation and its relationship to political polarization, and
- Disinformation's impacts on political systems, including democratic norms and institutions.

To this list I would add research that investigates how best to correct misperceptions resulting from disinformation (e.g., Porter, Wood, & Bahador, 2019; Vraga et al, 2019).

So far, the ad archives have led to very little scholarly research on disinformation, per se. Only Edelson et al (2019) and Jamison et al (2019) touch on this. Both studies describe disinformation's producers and their tactics. Edelson and colleagues (2019) identify "dishonest advertisers that are not correctly disclosing or are obfuscating the real ad sponsor," on Facebook, categorizing them "into quasi for-profit media companies and corporate astroturfers" (p. 1). Jamison et al (2019) examine pro- and anti-vaccine content on Facebook and find that while many groups have placed pro-vaccine ads, a small number of advertisers account for the vast majority of anti-vaccine advertisements.

These descriptive findings are useful, but ultimately limited. Indeed, despite the burgeoning of recent scholarship on disinformation, some of the most pressing questions about disinformation and its impacts remain underexplored. In fact, we know particularly little about who is exposed to online political disinformation, on what platforms, and to what effect (Weeks & Gil de

Zuniga, 2019). Lacking much robust, systematic data from the digital platforms themselves, it is virtually impossible to understand:

- Who views disinformation,
- Whether and how people interact with disinformation (e.g., liking or sharing posts),
- How disinformation spreads—especially *across* platforms, and
- How exposure to and interactions with disinformation changes people's beliefs, attitudes, and behavior, including voting behavior.

In their discussion of disinformation research gaps, Tucker et al (2018) emphasize the need for "data that are not currently accessible for open scientific research due to proprietary and/or privacy concerns" (p. 64). In Section III, I analyze the types of data currently made available by Twitter, Google, and Facebook, and in Section IV, I discuss potential avenues for improving scholarly data access that emphasize the principles of ethical and responsible data sharing and use.

III. Platform Data, Tools, and Resources

A. Ad Archives

As the previous section noted, thus far very little scholarly research has been based on data found in Facebook's, Google's, and Twitter's respective ad archives. This is in part because the archives are relatively new. However, it is also clear that the archives are currently insufficient to support scientific inquiry.

There are a number of reasons for this. To begin, the archives are difficult to systematically query. Twitter's Ad Transparency Center can only be queried based on the account names of advertisers (e.g., @EU_Commission). While it is not especially difficult to generate a list of accounts for well-known political organizations and actors, it is effectively impossible to do the same for accounts that actively seek to mask their identities and influence. In other words, researchers cannot surface false and malicious content using Twitter's Ad Transparency Center unless the researchers already know what accounts are suspect. What is more, only ads purchased by certified campaign accounts can be retrieved from the Ad Transparency Center indefinitely. All other ads can be viewed for just seven days. This means that when researchers identify previously unknown suspicious accounts, many of the account's ads are likely to be unavailable in the Ad Transparency Center.

Note that ahead of the European Parliamentary elections, many official political party accounts had yet to be registered as certified campaign accounts. As a result, retrospective research on even well-known political actors' advertising practices is rendered impracticable (Tromble, Jacobs, & Louwerse, 2019).

Facebook, in contrast, does appear to archive all relevant advertisements for (at least) seven years (Facebook, 2019). However, both Facebook's Ad Library and its Ad Library API (application programing interface) must be queried using keywords identified by the end user. It is not possible to retrieve all ads in the archive for a specific geographical region and investigate those as a set. Again, for known entities, such as political parties, this may not be problematic, but searching for previously unknown actors is much more difficult, if not impossible.

Google, on the other hand, does provide downloadable data (in csv format) on all political advertisements. Data can be filtered by country or region, and the advertiser's name and unique ID are provided. In theory, among the three platforms, Google's approach is most useful for uncovering and analyzing the activities of the purveyors of disinformation. However, the data Google provides for download only contain links to the requisite advertisements, not the actual content of the ads themselves. And recent evaluations reveal that significant amounts of the ad content is already missing. That is, many of the ad links are already dead and cannot be studied (Edelson et al, 2019; Tromble, Jacobs, & Louwerse, 2019).

To be sure, all three platforms remove from their archives advertisements deemed to be in violation of their policies, even if they ran for extended periods of time. This means that the content of most interest to researchers studying disinformation is likely to become inaccessible. Furthermore, as the platforms change their policies over time, ads from even legitimate actors may disappear. For example, at the time of writing (early December 2019) all three Dutch political parties that had official Twitter campaign accounts ahead of the European Parliamentary elections—@D66 (Democrats 66), @groenlinks (Green-Left), and @VVD (People's Party for Freedom and Democracy)—have had their campaign account status suspended, and all or nearly all of their ads have been stripped from the Ad Transparency Center.

However, even when content is retrievable, the metadata provided for ads is often unhelpful. Google, for example, reports ad spend and impressions data in such broad ranges that they are of little utility for research purposes. Its ranges for euros spent per ad are < €50, €50-500, and €500-30,000, and ad impressions are reported in ranges of ≤ 10k, 10k-100k, and 100k-1M.

And none of the platforms provide adequate (micro-) targeting data. In fact, only Twitter provides actual targeting data. Facebook and Google merely provide audience reach data. In all instances, however—that is, whether reporting targets or audience reach—information is limited to broad, generic categories such as age, gender, and/or geographical region (city or province). More precise targeting categories, such as specific views or interests (typically inferred from things like browsing and search activities), are not included in any of the archives. Yet these categories are especially important for research into extreme audience segmentation and disinformation.

Nor do any of the platforms report on the types of indirect targeting that occur as a result of their custom audience and lookalike features. Using these features, advertisers may submit a list of specific, identifiable individuals—from, for example, a list of party members and supporters—whom they would like to view a particular advertisement (custom audience), and the platforms may then algorithmically identify more users who "resemble" the people on that list to target (lookalike). Though custom audience and lookalike advertisers do not specify particular characteristics on which to target users, previous research has highlighted the fact that these features can have discriminatory impacts (Speicher et al, 2018; Venkatadri et al, 2018). If, for instance, an advertiser wishes to target users based on race or religion, they may submit a custom audience list reflecting those parameters without specifically revealing that everyone on their list shares these traits. The lookalike feature will in turn identify more users to target based on those characteristics.

Finally, and most fundamentally, robust, systematic scientific inquiry is hampered by the fact that scholars have no way of independently verifying the provenance, authenticity, or completeness of ad archive data. To ensure that scientific inferences are sound and unbiased, conclusions should be based on either the full population of relevant data (e.g., all ads viewed in a given country ahead of the European Parliamentary elections) or a representative sample of those data. None of the platforms guarantee that they provide either full populations or representative samples. In fact, thanks to ad removals, we can be fairly certain that the data are both incomplete and unrepresentative. Moreover, various evaluations of the ad archives have conclusively demonstrated that relevant ads are missing. In the Netherlands, Dutch political parties reported that they could not find their own ads in the archives (Tromble, Jacobs, & Louwerse, 2019), and in the Czech Republic, where parties are legally required to report all digital ads, researchers were able to locate just 25% of those ads in the respective platform archives (Havlíček, 2019).

B. "Malicious Accounts" Data Releases

Though the ad archives are the primary mechanisms through which the platforms responded to their commitment to support "good faith independent efforts to track Disinformation and understand its impact…" (Code of Practice, Chapter II.E., para. 12), there are a number of other ways that data can be and sometimes is made available by the platforms for scientific research.

Facebook and Twitter have both provided datasets containing information about accounts they have identified as malicious actors, especially Russian IRA and Iranian-backed accounts (Acker & Donovan, 2019; Bail et al, 2019). Unfortunately, however, these datasets tend to be limited in their utility. First, they often need to be transformed—sometimes at great time and expense—by scholars and other volunteers before they are useable by the research community (Summers, 2018). Second, as Acker & Donovan (2019) note, these datasets are stripped of vital metadata, including information on engagement and social connections. These types of metadata are vital to understanding how disinformation spreads both within and across platforms. Third, such datasets typically carry little information about how the datasets were curated. As with ad archive data, it is therefore difficult for researchers to understand what the data in front of them represent and is particularly difficult to determine what scientific inferences, if any, can be drawn from the data. In fact, "close examination" of the malicious account datasets "reveals that these have been curated, in some cases heavily edited and appraised for reputation management" (Acker & Donovan, 2019, p. 1597).

C. Application Programming Interfaces (APIs)

Twitter offers two application programming interfaces (APIs) that scholars regularly use to access public Twitter data (and metadata) without a fee. For historical data, researchers may query the REST API by tweet ID, account ID, or keyword. However, in the latter two cases—queries based on accounts and keywords—the REST API poses rather strict limitations on the volume of tweets that can be returned. For any given account, only the last ~3,000 tweets are retrieved, truncating data from especially prolific (bot) accounts. Keyword searches only return tweets from approximately the last seven days, and the retrieved tweets comprise a non-random sample, not the full population of relevant tweets (Tromble, Storz, and Stockmann, 2017). What is more, tweets that have been deleted or that are associated with suspended accounts cannot be retrieved. Again, this is a significant problem for disinformation research.

For those tracking disinformation in real-time, the Twitter Streaming API provides a much more reliable source of data. The Streaming API can also be queried based on account ID or keywords and returns tweets as they are posted. However, the Streaming API also imposes certain rate limits. If the query parameters a researcher sets match more than 1% of the global volume of tweets at any given time, the returned data will be truncated (i.e., all tweets beyond the 1% threshold will be missing from the dataset). In the past, if researchers thought it likely that their queries would encounter rate limits, it was possible to break them up into a large number of smaller queries. Recently, however, Twitter began limiting the number of access keys that researchers can use to run separate queries of its APIs. For projects seeking to track and collect very large amounts of data—as many disinformation projects do—this has presented a significant constraint.

Still, Twitter's APIs are considered by far the best available tools for researchers. Very little useful information can be obtained from YouTube's API, for example. And in mid-2018 Facebook drastically restricted access to its Graph and Pages APIs, two significant sources of data for disinformation research. Indeed, without access to these APIs, some of the most path-breaking disinformation studies (e.g., Bail et al, 2019; Barfar, 2019; Guess et al, 2019) would not have been possible.

To the best of my knowledge, at the time of writing, no scholars have successfully navigated the approval process for (re)acquiring access to either of these APIs specifically for academic research purposes. This is in large part because the terms of service for Facebook API access are designed for corporate use. Academics and academic research do not meet the use cases defined within those terms of service. And the same is true for Instagram's API.

Facebook has begun to offer scholars access to its CrowdTangle API, however. Facebook bought CrowdTangle, a social media analytics tool, in 2016. The tool has been available to many journalists and media companies for use in tracking public posts' performance, but to date, only a limited number of scholars have been given access. However, Facebook does appear to intend to provide academic access more broadly. If the company does so, CrowdTangle will undoubtedly become a powerful tool for disinformation research. (Jamison et al, 2019 combined CrowdTangle data with information from Facebook's ad library in the study of anti-vaccine disinformation described above.) Yet even CrowdTangle's utility is limited, as it provides only aggregated data and does not allow researchers to explore the comments and replies associated with public posts.

D. Academic-Platform Partnerships

Academic-platform partnerships represent the final mechanism for supporting scientific inquiry. These partnerships take a variety of forms. On one end of a partnership continuum, academic researchers may become temporary contractors, collaborating directly with a platform from the inside. On the other end of the continuum, platforms may supply data, money, or both for research to be conducted entirely externally. These models each have their drawbacks, and thus far none has proven effective at providing consistent access to key data for significant numbers of scholars or studies.

When academics contract with and conduct research inside the platforms, they are inevitably constrained by non-disclosure agreements and typically cannot share the results of their work publicly. This means that vital findings remain siloed inside the platforms, failing to benefit science, policymaking, or the broader public.

Facebook and Google also regularly provide unrestricted funding to academic researchers for projects the companies deem important. In general, the research priorities set by Facebook and Google focus on important questions of broad societal significance, and the funds go to very worthwhile research. These no-strings-attached grants should certainly be commended and further encouraged. However, without additional data access—accompanied by agreements that scholars may publish the results of their research based on that data—additional money will only go so far.

Aside from the occasional, ad hoc agreement struck with (typically well-connected, elite) scholars, Facebook and Twitter have each recently undertaken one major initiative to provide broader data access. In spring 2018, Twitter issued a request for proposals under its "healthy conversations" initiative. The company called for help from academic researchers to develop metrics that would allow Twitter—and others—to better understand and diagnose healthy and unhealthy dynamics on the platform (Twitter, 2018). The request for proposals made clear that research results generated as part of the initiative would be shared publicly. Both funding and data access would be part of the initiative, as well.

Twitter ultimately selected two projects. One of those projects has since been abandoned (Wagner, 2019). And, almost 18 months later, the other project—which I lead—has yet to begin. Negotiating the terms of an academic-industry partnership has proven extremely time-consuming and difficult, and aligning priorities across institutions with very different incentives has at times seemed Herculean.

So too with Facebook's major data partnership: Social Science One. Social Science One seeks "a new type of partnership between academic researchers and private industry" to help understand and solve "society's greatest challenges." Its main focus is on providing scholars with "privacy-preserving access" to data, while also guaranteeing scholars' "freedom to publish research findings on agreed upon topics without pre-approval" from Facebook (www.socialscience.one).

Unfortunately, the initiative is currently in serious jeopardy. More than a year-and-a-half since its inception, researchers are still without the promised data. The first dataset—comprised of aggregated data for URLs shared on the platform—was intended to be just one of many that would allow scholars to examine the spread and effects of online disinformation, while still protecting the privacy of Facebook users. But because Facebook has not been able to provide even this first dataset, Social Science One's philanthropic funders have withdrawn their support.

Facebook representatives have frequently cited the European Union's General Data Protection Regulation (GDPR) as a barrier to releasing the promised dataset, arguing that GDPR is unclear and they fear liability for sharing data with academic researchers. Though the URLs dataset is aggregated, it is sometimes possible to identify individuals within very large aggregate data, and Facebook cites this as a privacy concern and risk. While the rights and interests of users should always be foregrounded in data transfer and use decisions (see Section IV.B), GDPR also recognizes the value of research conducted in the public interest and provides a great deal of leeway for academic researchers to acquire and analyze personally identifiable information on this basis. In conversations with a number of Data Protection Authorities throughout Europe, members of the European Advisory Committee of Social Science One (myself included) have been advised that GDPR should not be seen as a barrier to the platforms making data—even sensitive, personally identifiable data—available for academic research. Both the companies and scholars must take measures to mitigate risks, evidenced in detailed data management plans and actions, but with such measures in place, data sharing is certainly possible under GDPR. In Section IV.B I discuss some of the options available for ensuring that scholarly research is conducted in an ethical, responsible manner.

IV. Data Needed for Scientific Research

A. Example Question and Data

Scholars seeking a better understanding of disinformation are therefore left without adequate sources of platform data. But what types of data are really needed? In this section, I provide an example of an important, unanswered research question and a variety of data needed to answer that question. However, this research question should be understood as *just one example among many others*.

Research Question: *Is micro-targeting being used to spread disinformation in a way that (a) alters people's beliefs or attitudes and/or (b) alters their behaviors, including voting behavior?*

To answer this question most effectively, both parts (a) and (b) require real-world exposure data at the individual level. Individual-level data are essential for research that seeks causal explanations. If, for example, we want to understand whether micro-targeting alters people's beliefs about the safety of vaccines, we need to be able to assess whether Twitter User A changed her mind after having seen an anti-vaccine advertisement tailored to her interests, psychological profile, etc. Data could include:

- *Categories used to target users, including the categories that are indirectly deployed via custom audiences and lookalike features*. This information will undoubtedly reveal sensitive PII—for example, about people's race, religion, or political views. Yet analyzing such sensitive data is essential if we are to understand how that sensitive information *is itself employed in micro-targeting*, and to what effect. We cannot, for instance, study whether and how purveyors of false information target certain racial groups without ourselves being aware of users' racial identities.

- *Who was exposed to and interacted with both organic and inorganic content.* Targeting data is a first step, but to understand whether micro-targeting is particularly likely to change people's minds and actions, we must also examine the impacts of exposure to the same message but by those who do not share the targeted characteristics. This is most likely to occur with organic content. A great deal of inorganic (paid) content actually starts as organic content. Thus, data on the impacts of exposure both before and after a post was converted to an ad could provide incredibly useful information for answering our research question.
- *Socio-demographic characteristics of users exposed to and interacting with such content.* This again includes sensitive characteristics such as race, religion, and political ideology. If we are to understand whether and how marginalized and vulnerable communities are disproportionately impacted by disinformation, such data are essential.
- *Internal labels for post content.* These are tags or labels generated by the platforms themselves, sometimes through human labeling, other times through automated classification. Labels might include whether a post was fact-checked and the outcome of that fact-check, the topic(s) of a post, and so on.
- *User panel surveys, with links to exposure and interaction data*. In order to understand changes in beliefs and attitudes, as well as off-platform behaviors, exposure and interaction data need to be paired with survey responses about people's beliefs, attitudes, and actions.

This list of data needs requires a host of qualifications. First, as mentioned above, the question I have posed should be considered just one, limited example. There are many other important questions that could be asked. And there are many other important questions still as of yet conceived. These still known questions will naturally emerge as further research is conducted, scandals bring new problems to our attention, etc.

Second, in answering even the single question I have posed, different researchers will want to use different types of data. Indeed, one of the strengths of scientific research is the ability to triangulate inquiries to assess whether different data lead to similar conclusions. In particular, the data I have listed assume large-scale, quantitative analyses will be undertaken. Yet in-depth, fine-grained qualitative analyses are also essential to the study of political micro-targeting and disinformation.

Third, it is particularly important to bear in mind that different platforms generate different types of data. What scholars need to answer a given question on one platform may differ rather significantly on another platform. Moreover, each platform changes over time, meaning that even if a study is focused on a single platform, the data needs to support that research may evolve.

Fourth, scholars' current questions and data requests are shaped by what we know exists on the platforms. Yet the platforms are proprietary black boxes. We ultimately do not know what we could or should request.

B. Ethical, Responsible Approaches to Data Access and Analysis

In the final analysis, then, scientific inquiry requires data access mechanisms that are flexible and adaptable, across platforms and over time. At the same time, these mechanisms must foreground the rights and expectations of the platforms' users. Research must be conducted, and data accessed, in a way that is consistent with both the law and ethical scientific principles. Political micro-targeting and disinformation research should be conducted, first and foremost, in the public interest, and that requires respecting the users who are subjects of the research themselves.

Perhaps the easiest way to achieve this is by obtaining informed consent from the users studied. In the scenario I described just above, users could (and should) be informed about the data to be processed as part of the survey, as well as from their on-platform activities; notified about the risks inherent in participating in the study; and given the opportunity to refuse or withdraw participation at any time. When user consent can be secured for a research project, it should not be difficult for platforms to transfer the requisite data to the scholars in question.

Unfortunately, however, direct user consent cannot always be obtained. In such circumstances—and in keeping with European Union data protection regulations—the risks to research subjects must be weighed against the public interest in conducting said research. And both platforms and researchers must be particularly careful to mitigate any risks posed.

If individual-level, or even potentially re-identifiable aggregated, data are involved—and especially if those data contain sensitive information—one particularly attractive option is to establish so-called research "safe harbors." These physical or virtual spaces would allow scholars to directly access and analyze platform data, but they also place clear and substantial limitations on the data researchers can access, as well as the forms of analysis researchers could employ. To be approved for access, researchers might agree to monitoring and could face significant liability for abuse or misuse of the data. Models for this approach exist in medical and health research, as well as in work with sensitive government data (e.g., census data).

V. Recommendations

The Code of Practice on Disinformation calls for the platforms to support "independent efforts to track Disinformation and understand its impact…" (European Commission, 2018, Chapter II.E., para. 12). Yet scholarly research on disinformation and its impacts has been severely hampered by a continued lack of platform transparency, by failed (or failing) platform-academic partnerships, and by continued—and sometimes even increasing—barriers to data access. The proceeding analysis of these barriers, as well as the discussion of scientific research data needs, lead me to offer the following recommendations:

- As part of their public ad archives, the platforms should provide more precise data on ad spending and impressions.
- The platforms should also provide more precise targeting data in the ad archives. This should include direct targeting data, as well as information about categories targeted indirectly through custom audience and lookalike features.
- For sensitive categories (e.g., race or political ideology), audience reach data might be substituted for targeting data. Alternatively, sensitive targeting data could be reported to regulatory authorities, with researchers given the opportunity to access the data under controlled conditions.
- The platforms should preserve deleted ad content, including content removed for violation of ad policies, for analysis by researchers.
- The platforms should provide formal analyses identifying their specific concerns regarding data sharing for independent academic research under GDPR. Such analyses will provide a starting point for resolving areas of ambiguity and uncertainty.
- In turn, Data Protection Authorities should offer formal guidance on permissible data sharing practices under GDPR.
- Regulatory authorities should begin to require that the platforms share data for research purposes. The types and amounts of data should remain flexible, with priorities set based on public interest as defined by the regulatory authorities, in consultation with both the platforms and scholars. The platforms' proprietary interests should not be neglected, but these should be balanced against the public's interest in platform transparency.
- In particular, I recommend the establishment of "safe harbors" designed for independent scholarly research using platform data. Models from the health and medical sectors, as well as the government statistics offices, should be consulted.
- Finally, in order to promote data authenticity and completeness, regulatory authorities should establish mechanisms for independent audits of data shared by the platforms with researchers, the public, or both.

References

Acker, A., & Donovan, J. (2019). Data craft: A theory/methods package for critical internet studies. *Information, Communication & Society*, *22*(11), 1590–1609. https://doi.org/10.1080/1369118X.2019.1645194

Bail, C. A., Guay, B., Maloney, E., Combs, A., Hillygus, D. S., Merhout, F., … Volfovsky, A. (2019). Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1906420116

Barfar, A. (2019). Cognitive and affective responses to political disinformation in Facebook. *Computers in Human Behavior*, *101*, 173–179. https://doi.org/10.1016/j.chb.2019.07.026

Baum, K., Meißner, S., Abramova, O., & Krasnova, H. (2019). DO THEY REALLY CARE ABOUT TARGETED POLITICAL ADS? INVESTIGATION OF USER PRIVACY CONCERNS AND PREFERENCES. *Research Papers*. Retrieved from https://aisel.aisnet.org/ecis2019_rp/77

Boerman, S. C., Kruikemeier, S., & Borgesius, F. J. Z. (2017). Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising*, *46*(3), 363–376. https://doi.org/10.1080/00913367.2017.1339368

Chester, J., & Montgomery, K. C. (2017). The role of digital marketing in political campaigns. *Internet Policy Review*, *6*(4). Retrieved from https://policyreview.info/articles/analysis/role-digital-marketing-political-campaigns

Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. (2019). Spiraling downward: The reciprocal relation between attitude toward political behavioral targeting and privacy concerns. *New Media & Society*, *21*(6), 1212–1231. https://doi.org/10.1177/1461444818813372

Dobber, T., Trilling, D., Helberger, N., & Vreese, C. H. de. (2017). Two crates of beer and 40 pizzas: The adoption of innovative political behavioural targeting techniques. *Internet Policy Review*, *6*(4). Retrieved from https://policyreview.info/articles/analysis/two-crates-beer-and-40-pizzas-adoption-innovative-political-behavioural-targeting

Dogruel, L. (2019). Too much information!? Examining the impact of different levels of transparency on consumers' evaluations of targeted advertising. *Communication Research Reports*, *36*(5), 383–392. https://doi.org/10.1080/08824096.2019.1684253

Dommett, K. (2019). The Rise of Online Political Advertising. *Political Insight*, *10*(4), 12–15. https://doi.org/10.1177/2041905819891366

Edelson, L., Sakhuja, S., Dey, R., & McCoy, D. (2019). An Analysis of United States Online Political Advertising Transparency. *ArXiv:1902.04385 [Cs]*. Retrieved from http://arxiv.org/abs/1902.04385

European Commission (2019). Code of Practice on Disinformation. Retrieved from: https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation (Accessed 1 December 2019).

European Commission (2018). Code of Practice on Disinformation. Retrieved from: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454 (Accessed 30 October 2019).

Facebook (2019). Facebook Ad Library. Retrieved from: https://www.facebook.com/ads/library/.

Ghosh, A., Venkatadri, G., & Mislove, A. (n.d.). *Analyzing Political Advertisers' Use of Facebook's Targeting Features*. 8.

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

Harker, M. (2019). Political advertising revisited: Digital campaigning and protecting democratic discourse. *Legal Studies*, 1–21. https://doi.org/10.1017/lst.2019.24

Hegelich, S., & Serrano, J. C. M. (n.d.). *IN GERMANY FOR THE 2019 EUROPEAN ELECTIONS*. 17.

Jamison, A. M., Broniatowski, D. A., Dredze, M., Wood-Doughty, Z., Khan, D., & Quinn, S. C. (2019). Vaccine-related advertising in the Facebook Ad Archive. *Vaccine*. https://doi.org/10.1016/j.vaccine.2019.10.066

Kruschinski, S., & Haller, A. (2017). Restrictions on data-driven political micro-targeting in Germany. *Internet Policy Review*, *6*(4). Retrieved from https://policyreview.info/articles/analysis/restrictions-data-driven-political-micro-targeting-germany

Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., & Serrano, J. C. M. (2018). Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data & Society*, *5*(2), 2053951718811844. https://doi.org/10.1177/2053951718811844

Porter, E., Wood, T. J., & Bahador, B. (2019). Can presidential misinformation on climate change be corrected? Evidence from Internet and phone experiments. *Research & Politics*, *6*(3).

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F., Arvanitakis, G., Benevenuto, F., … Mislove, A. (2018). Potential for Discrimination in Online Targeted Advertising. *FAT 2018 - Conference on Fairness, Accountability, and Transparency*, *81*, 1–15. Retrieved from https://hal.archives-ouvertes.fr/hal-01955343

Summers, E. (2018). *Irads: Working with 3,517 Internet Research Agency Facebook ads released by Congress*. Python. Retrieved from https://github.com/edsu/irads.

Tromble, R., Jacobs, K., & Louwerse, T. (2019). *Transparency in digital political advertisements during the 2019 European Parliament elections: country report on the Netherlands*. The Hague: Netherlands Helsinki Committee; Brussels: European Partnership for Democracy.

Tromble, R., Storz, A., & Stockmann, D. (2017). We don't know what we don't know: when and how the use of Twitter's public APIs biases scientific inference. SSRN, http://dx.doi.org/10.2139/ssrn.3079927.

Tucker, J. A., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., … Nyhan, B. (2018). *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature* (SSRN Scholarly Paper No. ID 3144139). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=3144139

Twitter. (2018). Twitter health metrics proposal submission. Retrieved from: https://blog.twitter.com/en_us/topics/company/2018/twitter-health-metrics-proposal-submission.html.

Varnali, K. (2019). Online behavioral advertising: An integrative review. *Journal of Marketing Communications*, *0*(0), 1–22. https://doi.org/10.1080/13527266.2019.1630664

Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K. P., Loiseau, P., & Goga, O. (2018). Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface. *2018 IEEE Symposium on Security and Privacy (SP)*, 89–107. https://doi.org/10.1109/SP.2018.00014

Vraga, E. K., Kim, S. C., & Cook, J. (2019). Testing Logic-based and Humor-based Corrections for Science, Health, and Political Misinformation on Social Media. *Journal of Broadcasting & Electronic Media*, *63*(3), 393–414. https://doi.org/10.1080/08838151.2019.1653102

Wagner, K. (2019). Inside Twitter's ambitious plan to change the way we twee. *Recode*. Retrieved from https://www.vox.com/2019/3/8/18245536/exclusive-twitter-healthy-conversations-dunking-research-product-incentives.

Weeks, B. E., & Gil de Zúñiga, H. (2019). What's Next? Six Observations for the Future of Political Misinformation Research. *American Behavioral Scientist*, 0002764219878236. https://doi.org/10.1177/0002764219878236

Witzleb, N., Paterson, M., Richardson, J., Paterson, M., & Richardson, J. (2019). *Big Data, Political Campaigning and the Law: Democracy and Privacy in the Age of Micro-Targeting*. https://doi.org/10.4324/9780429288654