



## **Ideological Asymmetry in the Reach of Pro-Russian Digital Disinformation to United States Audiences**

Hjorth, Frederik Georg; Adler-Nissen, Rebecca

*Published in:*  
Journal of Communication

*Publication date:*  
2019

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY-ND](#)

*Citation for published version (APA):*  
Hjorth, F. G., & Adler-Nissen, R. (2019). Ideological Asymmetry in the Reach of Pro-Russian Digital Disinformation to United States Audiences. *Journal of Communication*, 69(2), 168-192.

## ORIGINAL ARTICLE

# Ideological Asymmetry in the Reach of Pro-Russian Digital Disinformation to United States Audiences

Frederik Hjorth & Rebecca Adler-Nissen

Department of Political Science, University of Copenhagen, 1353, Denmark

*Despite concerns about the effects of pro-Russian disinformation on Western public opinion, evidence of its reach remains scarce. We hypothesize that conservative individuals will be more likely than liberals to be potentially exposed to pro-Russian disinformation in digital networks. We evaluate the hypothesis using a large data set of U.S.-based Twitter users, testing how ideology is associated with disinformation about the 2014 crash of the MH17 aircraft over eastern Ukraine. We find that potential exposure to disinformation is concentrated among the most conservative individuals. Moving from the most liberal to the most conservative individuals in the sample is associated with a change in the conditional probability of potential exposure to disinformation from 6.5% to 45.2%. We corroborate the finding using a second, validated data set on individual party registration. The results indicate that the reach of online, pro-Russian disinformation into U.S. audiences is distinctly ideologically asymmetric.*

**Keywords:** Disinformation, Russia, Ideological Asymmetry, Social Media, Twitter.

doi:10.1093/joc/jqz006

The suspected hacking by Russian authorities or proxy actors in the 2016 U.S. presidential election has highlighted ongoing efforts by foreign actors to affect domestic politics in the United States and other Western societies using online tools. Long before the 2016 hacks, one such tool, frequently applied by Russian authorities, has been state-sponsored *dezinformatsiya* (disinformation), targeted at publics in Western societies (Boghardt, 2009). While misinformation is typically defined only in terms of incorrectness, disinformation is understood here as false information created for the strategic purpose of deceit. Disinformation thus shares an important property with a related term, “fake news,” in that it is “intentionally and verifiably false” (Allcott & Gentzkow, 2017, p. 213).

Disinformation can be presented in a standard news media format, as fake news is; for example, pro-Russian disinformation is disseminated through news channels

Corresponding author: Frederik Hjorth; e-mail: [fh@ifs.ku.dk](mailto:fh@ifs.ku.dk)

such as *RT* (formerly known as *Russia Today*) and *Sputnik*, but it can also be disseminated through social media accounts (Kragh & Åsberg, 2017). In Western societies, these efforts have been met with concern that the spread of pro-Russian digital disinformation can erode support for national governments, affect electoral outcomes, and impact Western audiences' views on foreign policy and international security (Pomerantsev, 2015).

In spite of these concerns, there is, as of yet, only a modest amount of research on how (if at all) pro-Russian online disinformation spreads to publics in Western societies. Scholarly and media attention has tended to focus on the sources of pro-Russian disinformation (Mejias & Vokuev, 2017) and the role of computational propaganda, meaning automated "bots" and algorithms that can make a message go viral (Woolley & Howard, 2016). In contrast, limited attention has been paid to the reach of pro-Russian disinformation: that is, the nature and extent of its audience online. Instead, discussion of pro-Russian disinformation tends to implicitly assume that its reach is diffuse and pervasive. This lack of knowledge is paradoxical, given that an estimate of the reach would be essential in accounting for its political and societal consequences.

In this paper, we provide one of the first systematic investigations into who is most exposed to pro-Russian digital disinformation, focusing on U.S. audiences. We characterized two large samples of U.S.-based Twitter users who are following those Twitter accounts producing pro-Russian disinformation, dividing them in terms of their political ideology and party registration, as compared to non-followers. The paper uses the term "pro-Russian" to denote a pro-Kremlin stance: that is, supportive of the current Putin regime and its political interests. We focused on Twitter, as it remains one of the most important sites for global struggles over truths in international conflicts. Moreover, fake news spread mainly through those social media platforms that are also the main gateway to news in the United States and many Western countries (for a critical evaluation of computational propaganda research's tendency to rely on Twitter data because of its availability, see Bolsover & Howard, 2017).

We took a case-based approach, studying the Twitter communication flow surrounding the crash of the Malaysia Airlines MH17 aircraft over eastern Ukraine on 17 July 2014. By linking U.S. Twitter accounts following communication about the MH17 crash with data about user characteristics, we could characterize the ideological and demographic profiles of the users most likely to follow accounts disseminating disinformation. Thus, unlike most studies that investigate disinformation based on whether accounts have previously disseminated disinformation, we used a more precise, content-based measure, operationalizing disinformation as information explicitly rejecting Russian responsibility for the MH17 crash (or stating that Ukraine was responsible). Our research question is as follows:

RQ: How does ideology condition potential exposure to online, pro-Russian disinformation?

We found that ideologically conservative users are significantly more likely to follow disinformation accounts, compared to liberal users. Corroborating this result, we found evidence that followership of accounts spreading disinformation is associated with being a registered Republican.

This paper contributes to the existing literature by characterizing the ideological distribution of the audience of a verifiable case of pro-Russian disinformation. To our knowledge, we are the first to provide evidence of ideological asymmetry in exposure to online, pro-Russian disinformation. [Badawy, Ferrara, and Lerman \(2018\)](#) have come closest, with a study of retweets of Russian trolls, which found that conservatives retweeted Russian troll accounts 31 times more often than liberals in the 2016 U.S. election campaign. In doing so, we connect the literature on digital misinformation (e.g., [Berinsky, 2017](#); [Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012](#); [Nyhan & Reifler, 2010, 2015](#)) with the political psychology literature on ideological asymmetry ([Jost, 2017](#)), thus beginning to fill one of the “key research gaps” in studies of social media, polarization, and disinformation ([Tucker et al., 2018](#), p. 62). We also meet calls to examine how social media platforms disrupt traditional flows of political communication from mainstream media to mass publics ([Entman & Usher, 2018](#)). Our findings provide evidence of a political communication environment where high-influence accounts play a crucial role in disseminating disinformation, in what resembles a two-step communication flow ([Katz & Lazarsfeld, 1955](#)). Our case also illustrates an important paradox: in this environment, information searches by politically engaged citizens, usually deemed democratically desirable (e.g., [Bennett & Iyengar, 2008](#)), may, in some cases, diminish rather than increase factual knowledge. We revisit this implication in the concluding section.

Our study also contributes to the literature on partisan selectivity in (online) media use: that is, the tendency of media choice to reflect partisan or ideological considerations ([Iyengar & Hahn, 2009](#)). Like previous studies, we found clear ideological divisions in the types of information consumed online, although our research design cannot disentangle self-selected from inadvertent exposure. Lastly, the focus on pro-Russian disinformation connects this study to literature on authoritarian regimes’ use of social media, including direct censorship of posts with mobilizing potential in China ([King, Pan, & Roberts, 2013](#)) and more subtle control through “networked authoritarianism” in former Soviet regimes ([Pearce & Kendzior, 2012](#)). In contrast to these studies, we highlight the downstream distribution of disinformation within audiences.

In the next section, we present our theoretical framework and develop a hypothesis of the relationship between ideology and exposure to pro-Russian disinformation. We then describe the empirical case—the crash of the MH17 passenger aircraft in 2014—followed by a presentation of our data and statistical approach. We proceed to present and discuss the results of the study. Lastly, we suggest directions for future research and consider the broader implications of our findings.

## Ideological asymmetries in the market for disinformation

In theorizing how ideology conditions an individual to potential exposure to pro-Russian disinformation, we draw on a substantial literature in psychology, communication studies, and political science on ideologically driven selectivity in (online) media exposure. While our design does not allow us to observe cognitive processes, several findings within this literature provide theoretical grounds for expecting the reach of pro-Russian disinformation to be asymmetrical, with respect to ideology.

First, consider the so-called “hostile media effect” whereby partisans tend to perceive media coverage as biased against their own viewpoint (Vallone, Ross, & Lepper, 1985). Scholars have explained the effect in terms of group psychology: because information provided in media coverage may challenge the in-group’s perceived ideological or moral superiority, individuals are motivated to denigrate coverage as biased against their in-group (Ariyanto, Hornsey, & Gallois, 2007). Supporting this intergroup explanation, Hartmann & Tanis (2013) show that the hostile media effect is stronger among individuals with higher group identification and a perception of their own group as socially subordinate.

The hostile media effect provides a partial explanation for the widespread public mistrust in mass media (Ladd, 2012). However, empirical studies of the phenomenon find that mistrust in the media is not evenly distributed across the ideological spectrum. Instead, in the United States, conservatives consistently exhibit more mistrust in mass media (Eveland & Shah, 2003; Lee, 2010). As noted by Shin & Thorson (2017), this asymmetry may reflect that statements made by conservatives tend to fare worse in media fact checking.

One potential consequence of this ideologically asymmetric mistrust in mass media could be that conservatives would tune out of news media consumption altogether. However, the same mechanism driving denigration of traditional media should also spur a need for an alternative platform that validates the in-group identity. Lending support to this hypothesis, existing research finds that mistrust in media is indeed sharply diminished when the media source represents a salient political in-group (Bolsen, Druckman, & Cook, 2014; Reid, 2012). This leads us to expect conservatives, who are relatively more distrustful of traditional media sources, to be relatively more likely to gravitate away from traditional media sources and toward alternative sources.

Second, recent research has found that this greater motivation for ideologically conformant information among conservatives is matched by available information sources in the online sphere. Studying networks of politically engaged users on Twitter, Boutyline & Willer (2017) found that conservatives are more likely than liberals to select into ideologically homogeneous networks. Studying Twitter-based news media that propagate conspiracy theories about mass shooting events, Starbird (2017) documented an “alternative media ecosystem” with a striking ideological slant: out of 44 media domains coded as politically slanted, Starbird coded half as belonging to the “alt-right,” seven as “anti-globalist,” and just four as “alt-

left.” Further, Starbird (2017, p. 8) found that two of the news sources identified in her sample of domains were “clearly Russian Propaganda”; several of the other sources produced strongly pro-Russian content, a theme that was most widespread among the alt-right domains. This suggested a non-negligible overlap between pro-Russian news sources and sources of dis- and misinformation on Twitter.

Based on these theoretical assumptions and empirical results, we expect conservatives to be more likely than liberals to follow online sources of disinformation. We refer to this as the ideological asymmetry hypothesis: compared to liberal accounts, conservative accounts are more likely to follow accounts disseminating pro-Russian disinformation.

As outlined above, the ideological asymmetry hypothesis follows from “demand-side” factors (i.e., individual, social-cognitive motives for seeking out alternative, ideologically conformant news sources), as well as “supply-side” factors (i.e., the availability of online news sources catering to this demand). Here, we follow Theocharis, Barberá, Fazekas, Popa, and Parnet (2016) in theorizing the observable patterns as an interplay between supply- and demand-side factors. As a consequence, we remain agnostic in this study as to which side of the two takes causal precedence (for a discussion of whether polarization causes partisan selective exposure or vice versa, see Stroud, 2010).

### Case selection

A fundamental problem in any observational study of disinformation is measurement. The most common method to measure online disinformation is by attributing it to a particular source (i.e., accounts or outlets), the output of which is then assumed to be disinformation (e.g., Fletcher et al., 2018; Starbird, 2017). However, not all information spread by these sources is—strictly speaking—disinformation. Measuring disinformation by content raises another problem, intrinsic to the nature of disinformation: the difficulty of assessing the veracity of any given, minimally plausible statement *prima facie* is precisely what makes disinformation potentially effective.

We took a case-based approach to the study of disinformation. Instead of characterizing disinformation by sources or in the abstract, we selected a concrete case where identifying disinformation is feasible in retrospect. Specifically, we studied the information flow after the crash of the Malaysia Airlines MH17 aircraft over eastern Ukraine in July 2014.

Two major concerns guided this case selection. First, the MH17 crash was politically significant. The downing of MH17 was not merely a tragedy attracting global attention, it also helped turn the Ukrainian crisis into a veritable international conflict and prompted the European Union and North Atlantic Treaty Organization to establish tougher sanctions against Russia (Golovchenko et al., 2018). For Ukraine and Russia, embroiled in a war over control of Crimea, assuming responsibility for the downing of the plane would have been severely damaging to either country’s

political standing. Second, studying disinformation in the context of the MH17 crash is analytically tractable. Because of the political significance of responsibility for the crash, discussion about the crash hinged on who shot down the plane. Hence, most communication in both traditional media and on online platforms discussing the causes of the crash took a clear stand on the question of political responsibility. In the context of pro-Russian disinformation writ large, we consider the MH17 crash a pertinent, empirical case in which online pro-Russian disinformation demonstrably occurred.

Much attention has been devoted to how Russia fuels already-polarized debates within the West by, for instance, prior to elections in Western countries, buying Facebook ads discussing race or crime issues. In contrast, the MH17 case was embedded in a broader geopolitical context of competing pro-Russian and pro-Western narratives. The MH17 case shares this geopolitical aspect with the annexation of Crimea in January 2014 (involving pro-Russian disinformation about the presence of Russian troops in Crimea) and the missile strike against a Syrian air base in April 2017 (involving pro-Russian claims that the strike was a U.S.-staged hoax). What these cases of pro-Russian disinformation have in common is that they discussed responsibility for particular international events or atrocities, demonstrating in each case that Russia was innocent and the West was culpable. However, the MH17 crash is distinct, in that it became an instant global tragedy, killing all 298 people on board (including 193 Dutch, 43 Malaysian, 27 Australian, 12 Indonesian, and 10 British passengers, as well as 13 citizens with other nationalities). As such, it was likely to generate more attention in the West than other issues involving Russia and its geopolitical interests.

### The MH17 crash

At around 1:20 PM local time on 17 July 2014, the Malaysia Airlines aircraft MH17 from Amsterdam to Kuala Lumpur crashed near the city of Torez in Donetsk Oblast, eastern Ukraine. The plane carried 283 passengers and 15 crew members, all of whom died in the crash ([Dutch Safety Board, 2015](#)). Discussion of the crash on social media began shortly after the crash. The earliest relevant tweet in our data is from 3:13:58 PM, less than two hours after the crash, which read “#BREAKINGNEWS MALAYSIA AIRLINES FLIGHT #MH17 CONFIRMED SHOT DOWN OVER #DONETSK OBLAST, SHORTLY BEFORE REACHING RUSSIAN AIR SPACE.” From then, discussion spread rapidly; over the following hour, more than 10,000 tweets in our sample alone discussed the crash.

In assessing the key facts of the case, we relied on the report of the Dutch Safety Board (DSB; the Dutch government agency charged with investigating the crash), published in October 2015. The report concluded that the MH17 was shot down with a Buk 9M38 missile, fired from an area southeast of Torez, a town in Crimea that was controlled by pro-Russian separatists at the time ([Dutch Safety Board, 2015](#), p. 144). In 2018, the Dutch-led Joint Investigation Force confirmed that the aircraft was shot down by a launcher belonging to Russia’s 53<sup>rd</sup> anti-aircraft missile

brigade (Joint Investigation Team, 2018). Russian authorities and Russian state-sponsored and mainstream media continue to claim that the MH17 was shot down by a Ukrainian military aircraft in an air-to-air strike, a scenario the DSB explicitly rules out (Joint Investigation Team, 2018, p. 128). Based on these facts, we considered as a ground truth that pro-Russian separatists, with the help of Russia, were responsible for the downing of the MH17, and we considered attempts to argue otherwise as disinformation. In the next section, we explain how we built on this case selection strategy to measure observable instances of disinformation.

## Methods and data

In November 2016, we collected a sample of tweets based on a set of MH17-relevant keywords dating from 10 July 2014 (i.e., a week before the crash) until the time of data collection. The keywords and hashtags included in the search query were МН17, малазийский Боинг (meaning “Malaysian Boeing,” and often used in Russian media to refer to the aircraft), #MH17, #Pray4MH17, and #PrayforMH17. The hashtags and keywords were selected so they clearly related to the crash of MH17 (and not the Ukraine crisis more generally), they were neither pro-Ukrainian or pro-Russian by nature, and they covered the major languages in which the debate over MH17 took place. The aim was to get a broad representation of tweets from both the pro-Ukrainian and the pro-Russian sides. We collected the tweets from the Twitter Gardenhose, a 10% random sample of the full Twitter stream, ensuring a representative sample of tweets matching the search constraints. In the following, we refer to this data set as the tweet sample. The tweet sample contains a total of 481,567 tweets. Coming from the Gardenhose, this sample of tweets is not exhaustive of the full universe of MH17-relevant tweets. Furthermore, the sample reflects the keywords included in the search query, the optimal selection of which is not, *ex ante*, well-defined (King, Lam, & Roberts, 2017). These caveats notwithstanding, we consider the tweet sample to be reasonably comprehensive and representative of the population of interest: that is, the full set of MH17-relevant tweets.

## Linking tweets and audiences

In this analysis, we refer to followers of accounts spreading pro-Russian disinformation as having potential exposure to disinformation. While we cannot know whether actual users in the follower set saw the MH17 coverage contained in the tweet subsample, they were potentially exposed to it, insofar as the tweet would have appeared on their Twitter home timeline. As a passive measure of media exposure, the content of a Twitter user’s timeline eschews the problems of recall errors and social desirability bias that typically plague measures based on self-reports (Prior, 2009). On the other hand, passive measures do not directly measure exposure (Vreese & Neijens, 2016). This measurement problem is compounded on social media, where a given user’s actual exposure to an account in her network depends on how many other accounts she follows, as well as the frequency and timing of

activity from each account, relative to her own usage. To keep this distinction between following and actual exposure salient, we refer to users' exposure to disinformation as potential exposure throughout.

Another important caveat in this context is that the queries in November 2016 retrieved each accounts' set of followers at that time, not at the time of the original tweet. In order to evaluate the consequences of this discrepancy, we present an analysis in the Supplementary Appendix C, exploiting information about follower sequences, available from Twitter's application programming interface. We found that our results cannot be explained by changes in the ideological composition of followers over time. We revisit the issue of this time gap in measurement in the concluding section.

We now faced two additional tasks: identifying disinformation within the tweet sample, and learning how potential exposure to disinformation is related to user characteristics.

### Identifying disinformation

We identified disinformation by first drawing a random sample of 10,000 English-language tweets from the tweet sample. In the following, we refer to this as the tweet subsample. The tweet subsample is, thus, representative of English-language tweets in the full tweet sample. To the best of our knowledge, there exists no study investigating differences between English-, Spanish-, or French-speaking, U.S.-based Twitter users. In general, scholars exploring the demographics of U.S. users do not pay attention to languages, which poses problems of external validation. Given that the dominant non-English languages in our tweet sample are Dutch, Russian, German, and Indonesian, with very few in Spanish, we deemed the issue of non-English, U.S.-based tweets to be minimal in this particular case.

We manually coded the entire tweet subsample for whether each tweet contained disinformation. Although ideally we might have used the hand-coded subsample to train a statistical learning model to classify all of the 471,567 tweets in the remaining tweet sample, we found that, in this case, standard classifiers had insufficient precision and recall to be useful. For this reason, we focused on the subsample, where the content of each tweet was known. As a consequence, our inferences in the following relate to English-language, online, pro-Russian disinformation.

Table 1 presents the categories used in the manual coding, the proportions of tweets assigned to each category, and examples of tweets from each category.

The manual coding involved four coders, assigning tweets to one of three categories. Pro-Russia tweets (Category 1) explicitly stated that Russia was not responsible for the MH17 crash (or that Ukraine was responsible). Pro-Ukraine tweets (Category 2) explicitly stated that Ukraine was not responsible (or that Russia was). Lastly, "Other" tweets (Category 3) involved neither of these statements. We assigned each of the four coders a quarter of the subsample, plus a shared subset of 100 tweets, used for assessing intercoder reliability. Intercoder reliability in this

**Table 1** Coding Categories and Examples for Sample of MH17 Tweets

Category	Proportion	Examples	Recorded
1: Pro-Russia	5.5%	<ul style="list-style-type: none"> <li>· Detailed analysis: MH17 shot down by Ukrainian SU-25 cannon fire and air-to-air missile</li> <li>· Why the USA and Ukraine, NOT Russia, were probably behind the shooting down of flight #MH17</li> <li>· Moral Terror: How Critics of Western MH17 Coverage Are Bullied Into Silence</li> </ul>	Disinformation (1)
2: Pro-Ukraine	10.3%	<ul style="list-style-type: none"> <li>· RT @KiritRadia: Ukraine’s PM says they’ll get the “bastards” who shot down #MH17</li> <li>· Video—Missile that downed MH17 ‘was brought in from Russia’ @peterlane5news</li> <li>· RT @mashable: Ukraine: Audio recordings show pro-Russian rebels tried to hide #MH17 black boxes.</li> </ul>	Non-disinformation (2+3)
3: Other	84.1%	<ul style="list-style-type: none"> <li>· #PrayForMH17 :(</li> <li>· AIDS conference honours lost colleagues #MH17</li> <li>· RT @deserto_fox: Russian terrorist stole wedding ring from dead passenger #MH17</li> </ul>	

Proportions sum to less than 100 pct. due to rounding.

shared subset was high (Cohen’s  $\kappa = .8$ , Krippendorff’s  $\alpha_{\psi} = \psi .78$ ). See Supplementary Appendix A for the codebook.

By relying primarily on the DSB report, we allowed for the fact that some of the messaging we labeled as disinformation occurred prior to the publication of the report. While U.S. officials confirmed, only days after the crash, that Russia was implicated (Ackerman & Walker, 2014), 86% of the tweets in the tweet sample were posted prior to the publication of the final report on 13 October 2015. Our classification of tweets is thus, in some cases, post hoc, as the truth value of the tweets may not have been verifiable at the time. Since our definition of pro-Russian disinformation relates to the facts of the case, not to what was publicly known at the time, this does not affect our classification. Still, this is an important contextual feature.

As shown in the “Proportion” column in Table 1, the vast majority of tweets (84%) were assigned to the Other category. This reflects two important features of the coding. First, most tweets simply tended to be expressions of grief or

compassion over those killed in the crash (like the first example) or matter-of-fact news headlines (like the second example). In other words, the vast majority of tweets contained no statement as to who bore responsibility for the crash. Second, the proportion of tweets in the Other category also reflects a deliberately narrow coding scheme. Tweets were assigned to the Other category if they did not clearly imply culpability for the MH17 crash.

The third example shown in Table 1 for the other category exemplifies this coding strategy. The tweet, referring to Russians stealing jewelry from MH17 passengers' corpses, clearly portrayed Russians negatively. But since it did not imply responsibility for the crash, it was assigned to Other. Thus, the coding strategy prioritized precision over recall and our estimated share of pro-Russian disinformation in the hand-coded subsample should be considered a lower-bound estimate.

To simplify the analysis below, we relied on a recoding of these categories, shown in the "Recoded" column in Table 1. For reasons described above, we classified the Pro-Russian tweets implying no Russian responsibility as "Disinformation." Tweets from the other two categories were coded as "Non-disinformation." The labeling of the second category is deliberately residual and, admittedly, somewhat awkward. To be sure, the recoded Non-disinformation category contained not only true statements about the causes of the MH17 crash, but also emotional expressions and, crucially, in all likelihood, disinformation about other phenomena. However, the category of interest is the Disinformation category, containing only the small proportion of tweets we know with high certainty contained disinformation. With the narrow definition of disinformation, we thus opted to tolerate some false negatives in the residual category, in exchange for minimizing the number of false positives in the category of interest.

### Measuring user characteristics

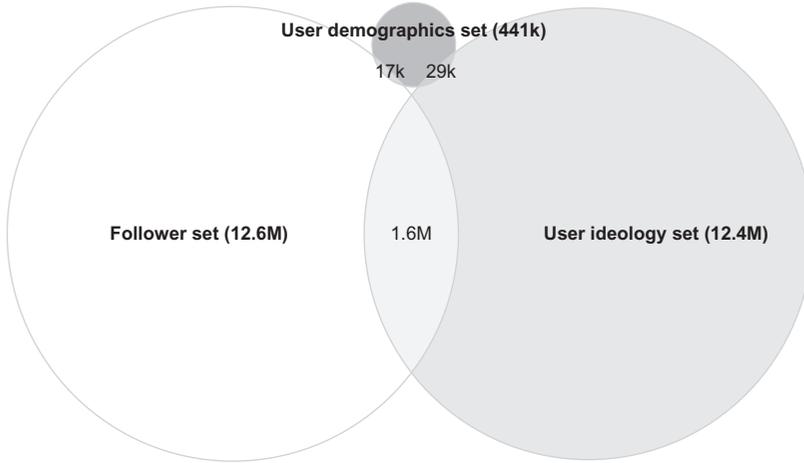
The population of interest was the set of users consuming disinformation. In order to arrive at a sample of this population consisting of individuals who potentially were exposed to coverage or discussion of the MH17 crash on Twitter, we retrieved every Twitter follower of one or more of the accounts included in the tweet subsample for whom data was available. We did so by querying Twitter's REST API, which collects data on Twitter users' networks. The 10,000 tweets in the tweet subsample came from 8,575 unique users, 1,345 of whom had protected or expired accounts. We retrieved a full list of followers for the remaining 7,230 accounts. These accounts had a total of 12,552,843 unique followers. We refer to this set of users as the follower set.

The follower set contains the user identifications of the roughly 12.6 million accounts following one or more of the accounts in the tweet subsample. We matched these user identifications with information about other characteristics of Twitter users from two other data sources. Figure 1 visualizes the relationships between the tweet subsample, the follower set, and the two other data sources.

First, we matched the follower set with data on the estimated ideology of 12.4 million Twitter users, presented in further detail in [Barberá, Jost, Nagler, Tucker, and Bonneau \(2015\)](#). The data set, collected from 2012–2014, includes U.S.-based Twitter users who engaged in discussions of salient contemporary issues, and is restricted to users following five or more political accounts. Based on these followed accounts, user ideology was then estimated using a variation of the procedure presented in [Barberá \(2015\)](#). This procedure, relying on a correspondence analysis ([Greenacre, 1993](#)), estimated users as liberal or conservative, based on whether they followed liberal or conservative accounts. For a recent validation of this ideology measure, see [Rivero \(2017\)](#), though see [Bauer, Barberá, Ackermann, and Venetz \(2017\)](#) for a critique of unidimensional measures of ideology. We refer to this data set as the user ideology set.

An important limitation of the user ideology set is that it may contain automated accounts built to mimic real users: that is, bots ([Howard, Wooley, & Calo, 2018](#)). Since bots will often, by design, follow several political accounts, they would be assigned an ideology score, potentially confounding the results. To address this issue, we also matched the follower set with a data set estimating the characteristics of 441,000 validated accounts of U.S. citizens. This data set, presented in further detail in [Barberá \(2016\)](#), included Twitter users matched by name, home state, and county (inferred from geolocated tweets) to U.S. citizens in public voter files. Since the accounts were matched with real individuals, results using this data set could not be confounded by anonymous or pseudonymous bots. The validity of the data could, theoretically, be challenged by the large-scale impersonation of real U.S. voters using names from voter files, combined with fabricated geolocation data matched to voters' home counties: a scenario we consider unlikely. The matching procedure produced estimates of each individual's gender, age, and, most importantly, party registration. We refer to this data set as the user demographics set. Because some demographic information is imputed, all variables were rescaled to range from 0 to 1, where 0 or 1 indicates that category membership (i.e., male, white, low-income, etc.) is known, and intervening values indicate an estimated probability of membership. Our key measure of interest, party registration, was drawn from the voter file and, thus, has low measurement error. Though analytically distinct from ideology, party registration is strongly correlated with ideology in the U.S. context ([Levendusky, 2009](#)), and we used it to corroborate the finding for ideology. In the interest of completeness, we present the results for age and gender alongside party registration, but note that for these variables, a measurement error is likely to attenuate the observed relationships. See Supplementary Appendix D for summary statistics for the variables in each data set used in the analysis.

As illustrated in [Figure 1](#), there is limited overlap between the follower set and the user ideology and demographics sets. This is to be expected, since the data sets came from entirely different sampling procedures, and we used data only on followers of the tweet subsample. Furthermore, while the follower set contained a diverse set of accounts following English-language tweets about MH17, the user



**Figure 1** Overview of types of data and sizes of their intersections.

ideology and demographics sets were restricted to U.S.-based Twitter-users. Nevertheless, there was sufficient overlap such that an analysis of how ideology and demographics are associated with exposure to disinformation was feasible. Since the tweet subsample was randomly drawn from the full tweet sample, the sampling step resulted in less overlap between the sets (and, in turn, less statistical power), but did not bias our estimates.

For users in the intersections of these sets, we thus had information about ideology/demographics, as well as potential exposure to (dis)information. However, it is not realistic to assume that inclusion in the intersection was random. Individuals who followed Twitter discussions of the MH17 crash were likely to differ systematically from those who did not. In order to assess how observable characteristics are associated with exposure to information about the MH17 crash, we therefore modelled inclusion in the non-overlapping parts of the known ideology and demographics sets (i.e., the outermost parts of the dark circles in Figure 1) as a separate outcome. We can think of these users as individuals who had no observable potential exposure to information about MH17, of any type. Each user in the user ideology and demographics sets is then associated with one of the three outcomes in Table 1 above: no information, that is, not overlapping with the follower set; disinformation; and non-disinformation. While the analyses thus cover the entire ideology and demographics sets, we stress that the observed relationships may not hold for exposure to other cases among users in the non-overlapping subsets.

### Statistical approach

For the 12.4 million users in the user ideology set, we estimated the following equation:

$$Info_i = f(Ideology_i + Ideology_i^2) \quad (1)$$

Where  $Info_i$  is a categorical measure of the outcomes described above for user  $i$ , and  $Ideology_i$  is the estimated ideology for user  $i$ . We included a squared term for ideology to allow for a nonlinear functional form, such as similar outcomes at the ideological extremes. If the true relationship between ideology and potential exposure were U-shaped, a linear specification would misleadingly indicate no association. Since the dependent variable,  $Info_i$ , can take three categorical values—No information, Disinformation, or Non-disinformation—we estimated  $f(\cdot)$  using multinomial logit.

For the 441,000 users in the known demographics set, we estimated the following equation:

$$Info_i = f(Republican_i + Age40plus_i + Male_i) \quad (2)$$

Where the right-hand side variables are indicators for being a registered Republican (relative to Independent or Democrat), 40 years old or more (relative to being younger), and male (relative to female). As above, we estimated  $f(\cdot)$  using multinomial logit.

To render the results of the multinomial logits more easily interpretable, the results section below presents predicted probabilities for the Disinformation and Non-disinformation outcomes across the ranges of the independent variables. We also present an additional quantity of interest, defined as follows:

$$Pr(Disinfo|Any\ info) = \frac{\widehat{Pr(Disinfo)}}{\widehat{Pr(Disinfo)} + \widehat{Pr(Non - Disinfo)}} \quad (3)$$

We calculated the quantity as the predicted probability of exposure to disinformation (conditional on covariate values), as a share of the predicted probability of exposure to any type of information (conditional on covariate values). The quantity captures the intuition: out of all the information about the MH17 crash available to the individual, what proportion is likely to be disinformation? Presenting this conditional probability highlights that, while the estimated probability of exposure to either type of information was relatively rare for most individuals, the composition of types of information could and did vary considerably.

## Results

Here, we present the results of estimating the models described above. After presenting how followers' ideology and demographics are associated with exposure to disinformation, we present a qualitative look at the most influential accounts in the data, providing additional face validity to the proposed interpretation of the findings.

### Followers' ideology

Table 2 presents results from estimating Equation 1 on the known ideology set. We present the results with No information as the reference category. Relative to No information, ideology is negatively related with exposure to Non-disinformation and positively related to Disinformation. In other words, relative to not being exposed to any MH17 discussion, more conservative individuals were less likely to be exposed to non-disinformation and more likely to be exposed to disinformation. Because of the inclusion of the squared term in the specification, the coefficients are not easily interpretable in isolation. Figure 2 visualizes the results in a more intuitive form, by plotting predicted probabilities across the range of the ideology variable.

As indicated by the solid blue line in Figure 2, exposure to non-disinformation has a curvilinear relationship with ideology, such that individuals at the extremes were more likely to be exposed to non-disinformation (compared to no information) about the MH17 crash. This result dovetails with existing studies that found that more ideologically extreme individuals are more likely to consume political news (Arceneaux & Johnson, 2013). However, the predicted probability of exposure to disinformation had a quite different trajectory, hovering around zero at the left end and turning sharply upwards at the right end of the ideological spectrum. In other words, while exposure to non-disinformation was symmetrically concentrated at the ideological extremes, exposure to disinformation is ideologically asymmetric, concentrated among the most conservative individuals. This pattern corresponds with Guess, Nyhan, and Reifler (2018), who found exposure to misinformation to be concentrated in the most conservative decile of users.

The dashed black line in Figure 2 shows the conditional probability of disinformation exposure, as described in Equation 3. Because disinformation exposure is concentrated at the right end of the spectrum, the conditional probability slopes sharply upwards. For the most liberal individuals in the data, the conditional probability of exposure to disinformation was 6.5%. For the most conservative individuals in the data, the probability was 45.2%. The results indicate a clear ideological asymmetry in the reach of online pro-Russian disinformation.

One notable feature of the measure of followers' ideology is that it is strongly right-skewed, with an outlier cluster of followers on the far-right end of the scale. We present a density plot of the distribution in the Supplementary Appendix B. We also show that our results are robust to estimating Equation 1 with this cluster omitted from the data.

### Followers' demographic characteristics

We now turn to followers' demographic characteristics: estimating Equation 2 on the user demographics set. Our results are shown in Table 3.

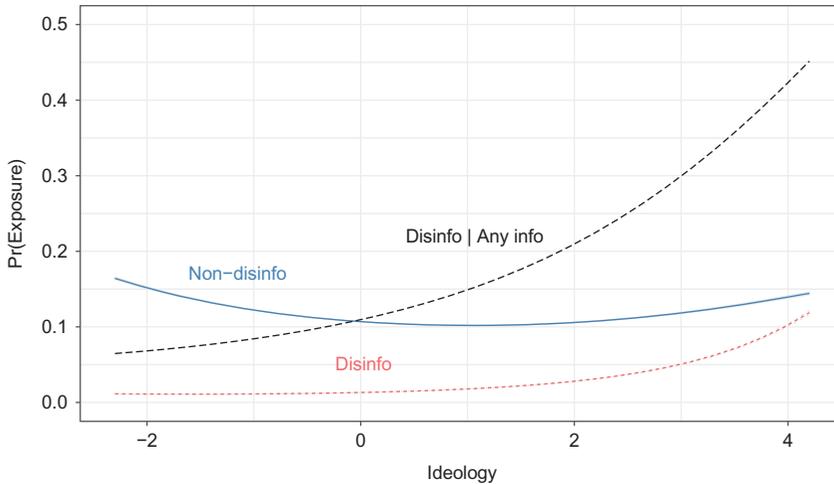
Because registering as a Republican was post-treatment to the other variables, the coefficients on these variables in the full specification were subject to a post-treatment adjustment bias. To show the consequences of various specification

**Table 2** Multinomial Logit Models of Estimated Ideology and Potential Exposure to Types of Information

	Non- disinformation	Disinformation	Non- disinformation	Disinformation
Intercept	-2.11* (.00)	-4.20* (.00)	-2.92* (.00)	-5.15* (.00)
Estimated ideology	-.10* (.00)	.22* (.00)	-.18* (.00)	.06* (.00)
Estimated ideology <sup>2</sup>	.05* (.00)	.08* (.00)	.06* (.00)	.13* (.00)
Political figures followed	-	-	.08* (.00)	.09* (.00)
AIC	10,533,507.22	10,533,507.22	9,974,905.91	9,974,905.91
BIC	10,533,593.20	10,533,593.20	9,975,020.55	9,975,020.55
Log likelihood	-5,266,747.61	-5,266,747.61	-4,987,444.96	-4,987,444.96
Deviance	10,533,495.22	10,533,495.22	9,974,889.91	9,974,889.91
Number observed	12,361,427	12,361,427	12,361,427	12,361,427

Note: \**p* < .001.

AIC: Akaike information criterion; BIC: Bayesian information criterion.



**Figure 2** Predicted probabilities of exposure to types of information across the observed range of estimated ideology. The ideology measure is standardized, so values on the x-axis represent standard deviations from the mean of estimated ideology. The solid blue line plots the probabilities for non-disinformation. The dotted red line plots the probabilities for disinformation. The dashed black line plots the probability of disinformation exposure, conditional on exposure to any information. The gray bands represent 95% confidence intervals.

**Table 3** Demographic Predictors of Potential Exposure

	Non-disinformation	Disinformation	Non-disinformation	Disinformation	Non-disinformation	Disinformation
Intercept	-3.15*	-6.50*	-5.19*	-9.05*	-4.94*	-9.14*
	(.02)	(.05)	(.03)	(.10)	(.04)	(.10)
Republican	-.69*	3.49*	-	-	-1.32*	.96*
	(.08)	(.16)	-	-	(.07)	(.15)
Male	-	-	.68*	1.35*	.72*	1.29*
	-	-	(.04)	(.10)	(.03)	(.10)
Age 40+	-	-	3.35*	5.64*	3.46*	5.40*
	-	-	(.05)	(.13)	(.05)	(.14)
AIC	154,614.64	154,614.64	147,239.10	147,239.10	146,808.47	146,808.47
BIC	154,658.63	154,658.63	147,305.06	147,305.06	146,896.42	146,896.42
Log likelihood	-77,303.32	-77,303.32	-73,613.55	-73,613.55	-73,396.24	-73,396.24
Deviance	154,606.64	154,606.64	147,227.10	147,227.10	146,792.47	146,792.47
Number observed	441,261	441,261	439,739	439,739	439,734	439,734

Note: \* $p < .001$ .

choices, Table 3 also presents a bivariate model with only party registration and a model with party registration excluded.

The first two columns in Table 3 present a result that aligns closely with the findings for ideology: individuals registered as Republicans were less likely to be exposed to non-disinformation and more likely to be exposed to disinformation. This pattern of opposite-sign coefficients is, in fact, unique for party registration.

In the full model, presented in columns 5 and 6, the patterns for the demographic variables were effectively unchanged (all remain significant). However, compared to the bivariate specification, the coefficient on Republican party registration with respect to disinformation exposure was sharply diminished, suggesting that the higher probability of disinformation exposure among Republicans in part reflects the background characteristics associated with registering as Republican.

Figure 3 presents predicted and conditional probabilities for each of the demographic variables. As shown, the predicted probabilities of potential exposure to any kind of information were low and relatively invariant. The exception from this pattern is age, where there was a noticeable uptick in the predicted probability for individuals predicted to be age 40 or older. This finding aligns with [Guess et al. \(2018\)](#), who found that older U.S. Americans were much more likely to visit fake news websites during the 2016 U.S. presidential election.

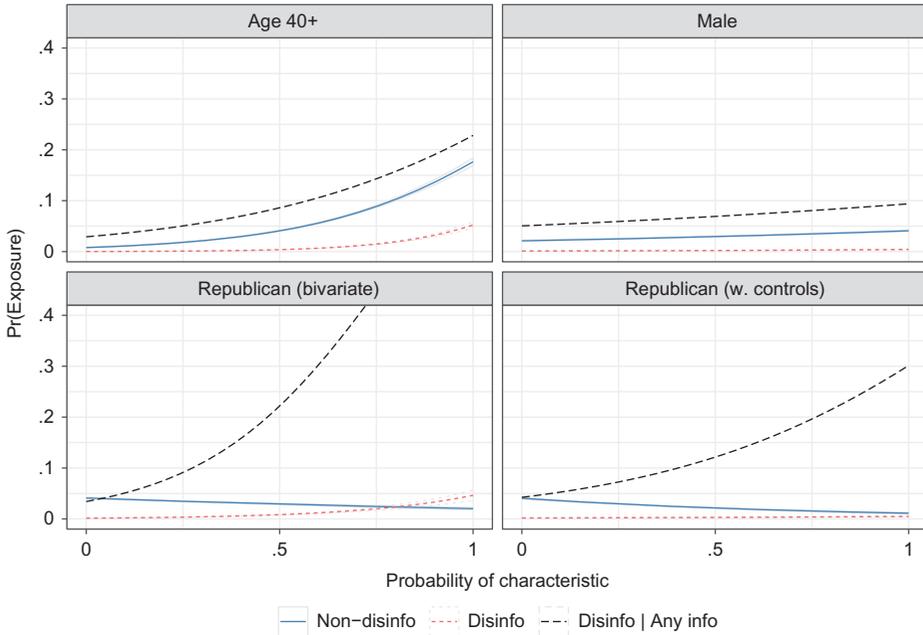
For the bivariate model with Republican party registration (bottom left panel), Republican-registered individuals were less likely to be exposed to non-disinformation and more likely to be exposed to disinformation. Consequently, the conditional probability of exposure to disinformation sloped very steeply upwards. For individuals not registered as Republican, the conditional probability of exposure to disinformation was 3.4%. For individuals registered as Republican, the conditional probability was 69.7%. This mirrors the association we observed for ideology, only stronger.

For Republican party registration in the full model (bottom right panel), the results were similar overall, albeit without the uptick for disinformation for the highest values, the same result captured by the insignificant coefficient in Table 3. In conjunction with the results for the demographic variables, this suggests that the higher conditional probability of exposure to disinformation for conservative and Republican individuals reflected, to a significant extent, age differences in exposure to disinformation, as well as ideology and party registration.

### **Understanding the mechanism: Influential accounts**

While the analysis thus far had established a link between twitter users' ideology and exposure to disinformation, the mechanisms of the association remained unclear. Specifically, who were the original disseminators of the disinformation that reached conservative audiences in particular?

To shed light on this, we turned our attention to the accounts behind the 10,000 annotated tweets. Since some accounts occurred more than once in the tweet subsample, the subsample contained information on 8,575 unique accounts. Within



**Figure 3** Predicted probabilities of exposure to disinformation for demographic variables. The probabilities are all based on the full model in columns 5 and 6 of Table 3, except the panel in the bottom right-hand corner, which is based on the bivariate model in columns 1 and 2. The gray bands represent 95% confidence intervals.

this set, we zeroed in on a set of accounts that were particularly likely to be driving the relationship between user ideology and exposure to disinformation. We refer to these as high-influence accounts.

An account that played a key role in driving the ideological asymmetry in exposure to disinformation would need to have fulfilled three of these criteria in order to disseminate disinformation to a large, ideologically distinct audience. Firstly, we defined high-influence accounts by a large followership, operationalized as a number of followers in the top decile of the distribution. In our data, this cutoff appeared at 1,916 followers. Secondly, we defined high-influence accounts by an ideologically distinct followership, operationalized as an average user ideology in the most extreme decile, either liberal or conservative. Users in our data were in the most liberal decile if they had an average user ideology score of less than .297, and in the most conservative decile if they had a score above .133 (since the ideology score was standardized, these scores are expressed in terms of standard deviations in the full distribution of user ideology). Thirdly, we limited high-influence accounts to those disseminating disinformation, operationalized as having sent out at least one tweet containing disinformation, as annotated in our data.

The accounts captured in this definition were thus highly influential in a dual sense: because of their ideological skew and dissemination of disinformation, they

had a high degree of statistical leverage on the relationship between ideology and exposure to disinformation shown above. Owing to their large numbers of followers, these accounts were also influential in the more intuitive sense of reaching a large online audience. The top panel of Figure 4 illustrates the selection process.

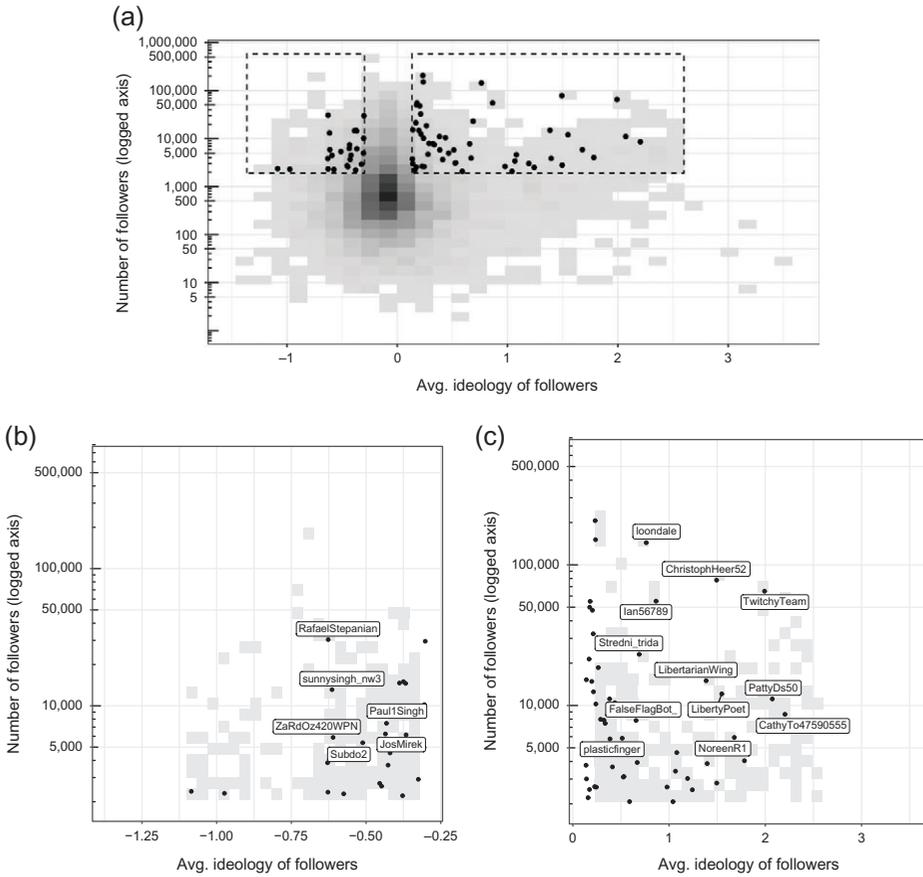
In the bottom two panels of Figure 4, we present close-ups of the areas highlighted in the top panel, showing select liberal (left panel) and conservative (right panel) high-influence accounts. As shown, the sets of high-influence accounts mirrored the finding of ideological asymmetry from above. Whereas the criteria captured 25 accounts on the liberal side of the ideological spectrum, the same criteria captured 52 accounts on the conservative side of the spectrum. In other words, the density of widely followed accounts disseminating disinformation was considerably higher on the conservative end of the spectrum, compared to the liberal end.

A qualitative examination of the specific accounts in the high-influence sets provided some additional face validation. Particularly high-influence accounts on the conservative side included CristophHeer52 and Ian56789, both high-frequency English-language accounts consistently tweeting pro-Russian messages, a pattern consistent with the behavior of pro-Russian bot or troll accounts (Sanovich, Stukal, & Tucker, 2018; Stukal, Sanovich, Bonneau, & Tucker, 2017). However, the conservative side also included accounts originating in domestic U.S. politics, such as the libertarian account LibertarianWing, and TwitchyTeam, a news aggregator established by the conservative commentator Michelle Malkin. On the liberal side, some of the accounts, such as RafaelStepanian and Paul1Singh, appeared to be troll-like accounts, tweeting left-wing messages with very high frequencies. In contrast to the conservative side, no large news outlets or online communities appeared on the liberal side. This pattern is consistent with previous research finding a distinct, right-wing media ecosystem to be particularly susceptible to disinformation (Benkler, Faris, Roberts, & Zuckerman, 2017).

## Conclusion and discussion

This paper examined the reach of online pro-Russian disinformation using a case-based approach analyzing the discussion of the MH17 crash over eastern Ukraine on Twitter. Consistent with the ideological asymmetry hypothesis, we found that exposure to disinformation was concentrated among the most conservative individuals. In other words, the reach of online pro-Russian disinformation into U.S. audiences was distinctly ideologically asymmetric. In a follow-up analysis of high-influence accounts disseminating disinformation, we found a corresponding asymmetry, in that high-influence accounts in the most conservative decile, in terms of audience ideology, outnumbered liberal accounts by about two to one. Our study thus provides (one of) the first systematic investigations into who is most exposed to pro-Russian digital disinformation.

Our results show, on the one hand, that in absolute terms, the degree of exposure to disinformation for most Twitter users is limited. This is in line with a recent



**Figure 4** Panel a shows the full distribution of the average ideology of followers and the number of followers for accounts in the tweet subsample. To avoid overplotting, the full distribution of the data is binned. The dots in the dashed rectangles are accounts tweeting disinformation, with numbers of followers in the top decile and average user ideology in the most liberal or conservative deciles. Panel b provides close-ups of the top left rectangle in panel a: that is, liberal, high-influence accounts. Panel c provides close-ups of the top right rectangle in panel a: that is, conservative, high-influence accounts. The gray rectangles show the distribution of accounts not tweeting disinformation.

study showing that Russian Twitter accounts contributed relatively little to the Brexit debate (Narayanan, Howard, Kollanyi, & Elswah, 2017). On the other hand, in relative terms, the exposure is concentrated among a particular, ideologically distinct segment, in which the potential exposure to disinformation is considerable for some individuals.

This conclusion comes with some important caveats. For one, our research design did not allow for directly observing exposure, but only potential exposure, as measured by followership links. Moreover, we observed what amounts to an

equilibrium in the market for online disinformation. As a consequence, we cannot disentangle the separate effects of “demand-side” vis-à-vis “supply-side” factors. Here, experimental approaches, like recent work using Twitter bots to distribute treatments (e.g., [Munger, 2017](#)), may provide analytical leverage.

Another significant caveat is the time gap between the MH17 crash in 2014 and our collection of follower data in 2016. Although we found no evidence of changes in the ideological compositions of followers after the crash (see Supplementary Appendix D), a stronger research design could close this time gap by implementing our sampling, coding, and data retrieval procedure in the immediate wake of an event giving rise to disinformation.

Lastly, as in all case-based approaches, case representativity is a relevant concern. We only focused on a population of U.S. Twitter users, representing a limited and more polarized audience when it comes to the global exposure to pro-Russian digital disinformation. While 24% of U.S. adults say they use Twitter ([Smith & Anderson, 2018](#)), studies have shown an overrepresentation of urban, young, male users, whereas representations of race and ideology are more uncertain ([Barberá & Rivero, 2015](#)). Future research may evaluate the representativity of this case by conducting analyses on both similar cases and different cases, including already polarized issues within the West.

These reservations notwithstanding, our study highlights important consequences of the disruptive communication processes often identified as key constituents of the “fourth age of political communication” ([Blumler, 2013](#)). These processes contribute to the breakdown of traditional media systems, but also form a new media system, characterized by “disrupted public spheres” ([Bennett & Pfetsch, 2018](#), p. 245). Our study suggests that one important manifestation of this system is the ideologically uneven spread of online disinformation.

While our study thus, in many respects, complements theories of political communication in this “fourth age,” it challenges them in others. Specifically, our findings question the notion that greater media choice leads to inequalities in factual political knowledge between more and less politically engaged citizens (e.g., [Bennett & Iyengar, 2008](#); [Holbert, Garrett, & Gleason, 2010](#); [Prior, 2005](#)). Paradoxically, the presence of mis- and disinformation in online media environments implies that engaged citizens’ information searches may, in some cases, diminish rather than increase their factual knowledge. This complicates not only theories of political communication, but also widely subscribed recipes for informed democratic citizenship. Understanding the causes and consequences of online flows of mis- and disinformation remains a crucial task for scholars of political communication.

## Supplementary Material

Supplementary material are available at *Journal of Communication* online.

## Acknowledgments

This study was conducted as part of the “Digital Disinformation” project (project no. CF16-0012), funded by the Carlsberg Foundation, directed by Rebecca Adler-Nissen. The authors thank Yevgeniy Golovchenko, JungHwan Yang, Jason Reifler, Zoltán Fazekas, Rune Slothuus, Rasmus Kleis Nielsen, Steven Livingston, Pablo Barberá, Silvia Maja-Vazquez, Sune Lehmann, Anders Søgaard, Federico Botta, Jon Kyst, Silvio Waisbord, and four anonymous reviewers for their feedback on earlier versions of this article. They thank Professor Alan Mislove, Northeastern University, for access to MH17 tweets, based on the Twitter Gardenhose feed.

## References

- Ackerman, S., & Walker, S. (2014, July 22). MH17: US intelligence says Russia ‘created conditions’ for plane disaster. *The Guardian*. Retrieved from <https://www.theguardian.com/world/2014/jul/22/mh17-us-intelligence-russia-separatists-report>. Accessed February 25, 2019.
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. doi:10.1257/jep.31.2.211
- Arceneaux, K., & Johnson, M. (2013). *Changing minds or changing channels? Partisan news in an age of choice*. Chicago, IL: University of Chicago Press.
- Ariyanto, A., Hornsey, M. J., & Gallois, C. (2007). Group allegiances and perceptions of media bias. *Group Processes & Intergroup Relations*, 10(2), 266–279. doi:10.1177/1368430207074733
- Badawy, A., Ferrara, E., & Lerman, K. (2018). Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. *Proceedings of The Web Conference (WWW’18)*. doi:10.1109/ASONAM.2018.8508646
- Barberá, P. (2015). Birds of the same feather Tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23(01), 76–91. doi:10.1093/pan/mpu011
- Barberá, P. (2016). *Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data*. Manuscript in preparation. Retrieved from <https://bit.ly/2KQbwdb>
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right. *Psychological Science*, 26(10), 1531–1542. doi:10.1177/0956797615594620
- Barberá, P., & Rivero, G. (2015). Understanding the political representativeness of Twitter users. *Social Science Computer Review*, 33(6), 712–729. doi:10.1177/0894439314558836
- Bauer, P. C., Barberá, P., Ackermann, K., & Venetz, A. (2017). Is the left-right scale a valid measure of ideology? *Political Behavior*, 39(3), 553–583. doi:10.1007/s11109-016-9368-2
- Benkler, Y., Faris, R., Roberts, H., & Zuckerman, E. (2017). Study: Breitbart-led right-wing media ecosystem altered broader media agenda. *Columbia Journalism Review*, 3, 2017. Retrieved from <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>. Accessed February 25, 2019.
- Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*, 58(4), 707–731. doi:10.1111/j.1460-2466.2008.00410.x

- Bennett, W. L., & Pfetsch, B. (2018). Rethinking political communication in a time of disrupted public spheres. *Journal of Communication*, 68(2), 243–253. doi:10.1093/joc/jqx017
- Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, 47(02), 241–262. doi:10.1017/S0007123415000186
- Blumler, J. G. (2013, September 12). The fourth age of political communication. Keynote address presented at the Workshop on Political Communication Online, Free University, Berlin. Retrieved from <http://www.fgpk.de/en/2013/gastbeitrag-von-jay-g-blumler-the-fourth-age-of-political-communication-2/>. Accessed February 25, 2019.
- Boghardt, T. (2009). Soviet Bloc intelligence and its AIDS disinformation campaign. *Studies in Intelligence*, 4, 1–20. doi:10.2139/ssrn.2108098
- Bolsen, T., Druckman, J. N., & Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior*, 36(2), 235–262. doi:10.1007/s11109-013-9238-0
- Bolsover, G., & Howard, P. (2017). Computational propaganda and political big data: moving toward a more critical research agenda. *Big Data*, 5(4), 273–276. doi:10.1089/big.2017.29024.cpr
- Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: variation in ideological homophily in online networks. *Political Psychology*, 38(3), 551–569. doi:10.1111/pops.12337
- Dutch Safety Board. (2015). *Crash of Malaysia airlines flight MH17*. The Hague, The Netherlands: Dutch Safety Board.
- Entman, R. M., & Usher, N. (2018). Framing in a fractured democracy: Impacts of digital technology on ideology, power and cascading network activation. *Journal of Communication*, 68(2), 298–308. doi:10.1093/joc/jqx019
- Eveland, W. P., & Shah, D. V. (2003). The impact of individual and interpersonal factors on perceived news media bias. *Political Psychology*, 24(1), 101–117. doi:10.1111/0162-895x.00318
- Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). Measuring the reach of “fake news” and online disinformation in Europe. *Reuters Institute, University of Oxford*. Retrieved from <https://bit.ly/2GC16wT>
- Golovchenko, Y., Hartmann, M., & Adler-Nissen, R. (2018). State, media and civil society in the information warfare over Ukraine: citizen curators of digital disinformation. *International Affairs*, 94(5), 975–994.
- Greenacre, M. J. (1993). *Correspondence analysis in practice*. San Diego, CA: Academic Press.
- Guess, A. M., Nyhan, B., & Reifler, J. (2018). *Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign*. Manuscript in preparation.
- Hartmann, T., & Tanis, M. (2013). Examining the hostile media effect as an intergroup phenomenon: The role of ingroup identification and status. *Journal of Communication*, 63(3), 535–555. doi:10.1111/jcom.12031
- Holbert, R. L., Garrett, R. K., & Gleason, L. S. (2010). A new era of minimal effects? A response to Bennett and Iyengar. *Journal of Communication*, 60(1), 15–34. doi:10.1111/j.1460-2466.2009.01470.x

- Howard, P. N., Wooley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2), 81–93. doi:10.1080/19331681.2018.1448735
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1), 19–39. doi:10.1111/j.1460-2466.2008.01402.x
- Joint Investigation Team (2018). *Update in criminal investigation MH17 disaster*. Retrieved from <https://bit.ly/2shCPV>
- Jost, J. T. (2017). Ideological asymmetries and the essence of political psychology. *Political Psychology*, 38(2), 167–208. doi:10.1111/pops.12407
- Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. New York, NY: Free Press.
- King, G., Lam, P., & Roberts, M. E. (2017). Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science*, 61(4), 971–988. doi:10.1111/ajps.12291
- King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, 107(02), 326–343. doi:10.1017/S0003055413000014
- Kragh, M., & Åsberg, S. (2017). Russia's strategy for influence through public diplomacy and active measures: the Swedish case. *Journal of Strategic Studies*, 40(6), 773–816. doi:10.1080/01402390.2016.1273830
- Ladd, J. M. (2012). *Why Americans hate the media and how it matters*. Princeton, NJ: Princeton University Press.
- Lee, T. T. (2010). Why they don't trust the media: An examination of factors predicting trust. *American Behavioral Scientist*, 54(1), 8–21. doi:10.1177/0002764210376308
- Levendusky, M. (2009). *The partisan sort: How liberals became Democrats and conservatives became Republicans*. Chicago, IL: University of Chicago Press.
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction. *Psychological Science in the Public Interest*, 13(3), 106–131. doi:10.1177/1529100612451018
- Mejias, U. A., & Vokuev, N. E. (2017). Disinformation and the media: the case of Russia and Ukraine. *Media, Culture & Society*, 39(7), 1027–1042. doi:10.1177/0163443716686672
- Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, 39(3), 629–649. doi:10.1007/s11109-016-9373-5
- Narayanan, V., Howard, P., Kollanyi, B., Elswah, M. (2017). *Russian involvement and junk news during Brexit* [Data memo 2017.10]. Oxford, UK: Project on Computational Propaganda. Retrieved from <https://bit.ly/2s6K1nX>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. doi:10.1007/s11109-010-9112-2
- Nyhan, B., & Reifler, J. (2015). Displacing misinformation about events: An experimental test of causal corrections. *Journal of Experimental Political Science*, 2(01), 81–93. doi:10.1017/XPS.2014.22
- Pearce, K. E., & Kendzior, S. (2012). Networked authoritarianism and social media in Azerbaijan. *Journal of Communication*, 62(2), 283–298. doi:10.1111/j.1460-2466.2012.01633.x

- Pomerantsev, P. (2015). The Kremlin's information war. *Journal of Democracy*, 26(4), 40–50. doi:10.1353/jod.2015.0074
- Prior, M. (2005). News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science*, 49(3), 577–592. doi:10.1111/j.1540-5907.2005.00143.x
- Prior, M. (2009). The immensely inflated news audience: Assessing bias in self-reported news Exposure. *Public Opinion Quarterly*, 73(1), 130–143. doi:10.1093/poq/nfp002
- Reid, S. A. (2012). A self-categorization explanation for the hostile media effect. *Journal of Communication*, 62(3), 381–399. doi:10.1111/j.1460-2466.2012.01647.x
- Rivero, G. (2019). Preaching to the choir: Ideology and following behaviour in social media. *Contemporary Social Science*, 14(1), 54–70. Advance online publication. doi:10.1080/21582041.2017.1325924
- Sanovich, S., Stukal, D., & Tucker, J. A. (2018). Turning the virtual tables: Government strategies for addressing online opposition with an application to Russia. *Comparative Politics*, 50(3), 435–482. doi:10.5129/001041518822704890
- Shin, J., & Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication*, 67(2), 233–255. doi:10.1111/jcom.12284
- Smith, A., & Anderson, M. (2018). Social media use in 2018. *Pew Research Center*. Retrieved from <https://pewrsr.ch/2FDfiFd>
- Starbird, K. (2017). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)*, May, Montreal, Canada.
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(2010), 556–576. doi:10.1111/j.1460-2466.2010.01497.x
- Stukal, D., Sanovich, S., Bonneau, R. & Tucker, J. A. (2017). Detecting bots on Russian political Twitter. *Big Data*, 5(4), 310–324. doi:10.1089/big.2017.0038
- Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & Parnet, O. (2016). A bad workman blames his Tweets: The consequences of citizens' uncivil Twitter use when interacting with party candidates. *Journal of Communication*, 66(6), 1007–1031. doi:10.1111/jcom.12259
- Tucker, J., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., ... Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. doi:10.2139/ssrn.3144139. Retrieved from <https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf>. Accessed February 25, 2019.
- Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, 49(3), 577–585. doi:10.1037//0022-3514.49.3.577
- Vreese, C. H., & Neijens, P. (2016). Measuring media exposure in a changing communications environment. *Communication Methods and Measures*, 10(2–3), 69–80. doi:10.1080/19312458.2016.1150441
- Woolley, S. C., & Howard, P. N. (2016). Automation, algorithms, and political political communication, computational propaganda, and autonomous agents—Introduction. *International Journal of Communication*, 10(2016), 4882–4890. Retrieved from <https://ijoc.org/index.php/ijoc/article/view/6298>