Troll Factories: Manufacturing Specialized Disinformation on Twitter

Darren L. Linvill

Clemson University Department of Communication

Patrick L. Warren

Clemson University Department of Economics

Abstract

We document methods employed by Russia’s Internet Research Agency to influence the political agenda of the United States from September 9, 2009 to June 21, 2018. We qualitatively and quantitatively analyze Twitter accounts with known IRA affiliation to better understand the form and function of Russian efforts. We identified five handle categories: *Right Troll, Left Troll, News Feed, Hashtag Gamer,* and *Fearmonger*. Within each type, accounts were used consistently, but the behavior across types was different, both in terms of “normal” daily behavior and in how they responded to external events. In this sense, the Internet Research Agency’s agenda-building effort was “industrial”-- mass produced from a system of interchangeable parts, where each class of part fulfilled a specialized function.

In February 2018, the U.S. Justice Department indicted 13 Russian nationals for interference with the 2016 U.S. Presidential election (United States of America v. Internet Research Agency LLC). The indictment named the Internet Research Agency (IRA), based in St. Petersburg, as central to a Russian effort to sow discord in the U.S. political system, largely through social media. The IRA intervened in the 2016 election (Mueller , 2019), with some even suggesting they may have tipped the balance of the election in favor of candidate Donald Trump (Jamieson, 2018). Regardless of the effect on the election, however, it is undeniable that many real people were taken in by IRA messaging, including mainstream journalists (Lukito & Wells, 2018) and political activists (Birnbaum, 2019).

Researchers have moved to try to understand the strategy and impact of what is, perhaps, the most important foreign influence operation of the social-media age. Concentrating on the discussions on Twitter around the Black Lives Matter movement, Arif, Stewart, and Starbird (2018) showed that the IRA fostered antagonism and undermined trust in authorities. Looking at accounts discussing vaccines, Broniatowski et al (2018) showed that IRA trolls amplified both sides of the contentious debate. In the context of the Twitter discussion of the Malayasian Airline flight (MH17) downed in eastern Ukraine, Golovchenko, Hartmann, and Adler-Nissen (2018) showed that the IRA appeared in the conversation but had no substantial effects on its progress. Finally, Badawy, Lerman, and Ferrara (2018) investigated what sort of accounts shared the content the IRA produced, in the context of the 2016 U.S. presidential election, finding that more conservative and more “bot like” accounts, with fewer followers but more status updates, were more likely to share IRA content.

But, given the topical nature of research to date, each of these investigations sampled a small subset of the IRA activity, often pulling overwhelming from one style of account. The actual

IRA operation was quite broad, multi-faceted, and interlinked. IRA activity has been identified on Facebook, Twitter, Instagram, Stitcher, Youtube, and stand-alone websites. Furthermore, these accounts masqueraded as American citizens and organizations from a wide variety of political orientations or from no obvious political orientation at all.

As we will show, there is enormous heterogeneity in theme and approach across IRA accounts, even just on Twitter. A piecemeal investigation of this or that account risks misleading conclusions about the overall strategy, if any, employed by the IRA. Like the blind monks in the parable of the elephant, the researcher would draw different conclusions depending on which part of the operation they grabbed. To avoid this fate, we take a holistic view of the IRA disinformation enterprise, at the risk of eliding some of the details that would come from a microscopic investigation of some specific account or tweet. Our goal is to understand the overarching strategy that the IRA is pursuing, if any, to affect the political conversation in the United States.

Scholars have examined the role that media, including social media, play in driving an agenda (Lariscy, Avery, Sweester, & Howes, 2009, Parmelee, 2014). The work of the IRA to build a public agenda (Denham, 2010) differs from previous social media cases, however, as the content is not genuine and was created by a single, state-affiliated entity rather than organic to the public discourse. National efforts to use media to influence foreign citizens are not new; Japan broadcasted to U.S. troops throughout World War II, and Voice of America has for decades been a global mouthpiece of the U.S. government. However, Russia’s work on social media has taken agenda-building efforts by nations into a new context. The purpose of this study is not to look at the effect the IRA may or may not have had on genuine online conversations or even the specific agendas behind IRA efforts. Rather, the purpose of this study is to better understand the structure of the IRA’s campaign. Given the covert nature of this campaign, such understanding is essential.

This study asks three questions about the IRA's behavior on the social media platform Twitter:

RQ 1: Can IRA Twitter handles employed between June 19, 2015 and December 31, 2017 be categorized by their content into multiple discrete types, and if so, what characterizes those types?

RQ 2: If distinct content types exist, do individual accounts tweet in a manner consistent with a given type throughout their existence?

RQ 3: If distinct content types exist, are those types employed in ways that are different from one another? Specifically,

RQ 3.1 Are accounts of the different types employed in different circumstances?

RQ 3.2 Do accounts of different types use different mixes of communication actions?

RQ 3.3 Do accounts of different types locate themselves differently in the social network?

We will show that the content of the tweets alone suffices for us to reliably identify a handful of thematic types that capture the behavior of 83% of the English-language IRA accounts, which are responsible for 97% of the English-language content. Within these types, the accounts are quite consistent, not only in content but also in the three other behavioral characteristics that we investigate: activity over time, network location, and communication strategy. We liken these IRA accounts to industrial machines in a modern propaganda factory, both interchangeable (within type) and extremely specialized (across types), and which are best understood as a coherent unit.[1]

[1] This metaphor is best understood as an as-if model of organizational behavior. Just as there is, in reality, no unitary actor called the "firm" that maximizes profits or "party" that seeks political power, the "IRA" does not actually make decisions. Instead, we observe the behavior of a

Method

We employed an exploratory, sequential mixed methods design (Creswell, 2014), first applying qualitative analysis to infer the types of accounts, from the perspective of the content produced, and then using quantitative analysis to explore how other aspects of behavior varied over time and between the types identified in the qualitative analysis.

Sample

Our research employed a data set of 9.04 million tweets released by Twitter on October 17, 2018 (Gadde & Roth, 2018), and updated in January, 2019. These tweets came from 3,613 accounts, which are a subset of the 3,841 accounts given by Twitter to Congress. A list of these account handles was released on June 18, 2018 by the U.S. House Intelligence Committee (Permanent Select Committee on Intelligence, 2018). Exactly how these accounts are identified has not been fully disclosed by Twitter. They have stated that they "employ a range of open-source and proprietary signals and tools to identify when attempted coordinated manipulation may be taking place, as well as the actors responsible for it" (Roth, 2019). The Twitter release included hashed/de-identified versions of account handles for accounts with fewer than 5000 followers. We used an alternate version of the Tweets we collected for an earlier draft of this project to re-identify most of the accounts (see appendix).

collection of individual agents. Whether our model of this information operation as a coherent, unitary actor is fruitful depends, ultimately, on whether it sheds useful light on the behavior we document.

We identified 18 handles with tweets not associated with IRA agenda building. Eight handles engaged in commercial activity (four marketed exercise and diet related activities, and one each that marketed payday loans, essay writing services, exotic dancing, and travel services). It is possible these accounts served some function related to IRA goals, but that function was not apparent in the content. Ten accounts appeared to engage in normal human activity and likely unassociated to the IRA.[2] We removed 163,317 tweets associated with these 18 handles.

5,657,204 tweets from 1,607 separate handles tweeted predominantly in a language other than English. The majority of these were Russian language handles, but handles also tweeted in German, Italian, Arabic, French, and Spanish. To keep the focus of the current study on the IRA's U.S. operations, these handles were removed.

2,962,903 tweets associated with 1,858 IRA handles remained for analysis. Figure 1 presents the overall daily output of the IRA, divided into English and Non-English accounts.

Data Analysis

We both worked to qualitatively analyze the handles, as recommended by Corbin and Strauss (2015), and placed handles into emergent categories. First, we engaged in a process of unrestricted open coding, examining, comparing, and conceptualizing the content. We considered elements of tweets including the hashtags employed by a handle, cultural references within tweets, as well as issues and candidates for which a handle advocated. Many tweets included external links, some of which were usable, and external pages were considered. Finally, the name of the

[2] Twitter's misidentification of IRA accounts has been documented. A previous list published by the U.S. House Intelligence Committee in November, 2017 contained four handles belonging to non-IRA affiliated individuals we worked with journalists to identify and speak to (Calderwood, Riglin, & Vaidyanathan, 2018). These individuals, and others, were removed from the updated June, 2018 Congressional list. With this experience in mind, we felt it was reasonable to remove accounts from the dataset.

handle itself often contained information that helped us better understand its nature (e.g. @BLMSoldier). We conducted axial coding to identify patterns and interpret emergent themes. Through axial coding we identified links and relationships between codes and, through both inductive and deductive reasoning, built a frame to better understand the data. To verify the validity of results, near the end of axial coding, peer debriefing was conducted (Creswell & Miller, 2000). This involved bringing in an external individual familiar with the phenomenon to play devil's advocate.

313 handles with 101,089 total tweets could not be categorized due to either insufficient activity or a lack of specificity in content. Many of these appeared in the early days of the IRA's English-language operations and consisted of "junk" content such as song lyrics or quotations (often the same content used across several accounts). Many of the categorized accounts also started in this way, but were eventually put to more specific use, so some of these uncategorized accounts may have simply never been "activated". In later periods, many of the uncategorized handles simply tweeted very few times, often in single digits (the 25th percentile account of this type tweeted only 8 times). We do not know if handles stopped tweeting voluntarily or if Twitter suspended the accounts.

We each independently coded a sub-sample of the same 50 handles and found a Krippendorf's alpha inter-coder reliability of .92 (note: error occurred only in accounts with extremely low tweet counts). After reaching this level of reliability we coded the remaining IRA-associated handles in our data, placing them into one of five categories. We coded a total of 1,726 accounts.

To address RQ3.1, the data were collapsed to account-by-day and account-by-hour units of observation, with total tweets tallied, and each account matched with the account-type codes derived in.[3] We then analyzed the behavior by account type, both over the full period and in specific event windows.

To address RQ 3.2, we further subdivide the tweets into original tweets, retweets, quote tweets, and replies. We then document whether and how the distribution of activity among these actions varies across account types. We also investigate the clients that the accounts use to generate their output and demonstrate how that distribution differs across account types.

To address RQ 3.3, we need to define what constitutes a link in the social network. Following/follower links are not available in our data, so we instead use mention, reply, and retweet connections to define links, where two IRA accounts are defined as linked if one connects to the other in one of those ways, and that link is “directed” in the sense that we distinguish between the account that is the origin of the link and the account which is the target. For tweets with multiple mentions, we use the first mentioned account, only.[4] Using this definition of a link, we will answer RQ 3.3 by documenting the extent to which accounts of each type identified in RQ1 link to IRA accounts of their own and other types. We will also use this definition of a link to investigate the degree to which the same accounts, even those outside the IRA, are linked to by IRA accounts of different types.

Results

RQ1 and RQ2. We identified five categories of IRA-associated Twitter handles, each with unique patterns of behaviors: *Right Troll*, *Left Troll*, *News Feed*, *Hashtag Gamer*, and *Fearmonger*. With

[3] The quantitative data analysis were conducted in pandas, the Python Data Analysis Library.

[4] We use first mention, only, in order to not overcount tweets that are simply strings of mentions. Choosing a random mention from each tweet delivers nearly identical results.

the exception of the *Fearmonger* category, handles were consistent and did not switch between categories.

Right Troll (454 handles, 705,064 tweets, mean = 1,553, s.d. = 8,063).

These handles broadcast nativist and right-leaning populist messages. They employ common hashtags used by similar real Twitter users. Among the ten hashtags most often employed by these accounts were #MAGA (i.e., "Make America Great Again," n = 23,449), #tcot (i.e. "Top Conservative on Twitter," n = 10,322), #AmericaFirst (n = 7,039), and #IslamKills (n = 3,927). Following the nomination of Donald Trump, they uniformly supported his candidacy and his Presidency, e.g. @AmelieBaldwin retweeted on December 13, 2016, "No, Russia didn't elect Donald Trump, the voters did https://t.co/ce70G9gv4h Repeat over and over disbelievers. PRESIDENT DONALD TRUMP!!" As noted, these handles regularly employed #MAGA, Donald Trump's campaign slogan. They routinely denigrated the Democratic Party, e.g. @LeroyLovesUSA, January 20, 2017, "#ThanksObama We're FINALLY evicting Obama. Now Donald Trump will bring back jobs for the lazy ass Obamacare recipients."

These handles' themes were distinct from mainstream Republicanism. They rarely broadcast traditionally important Republican themes, such as taxes, abortion, and regulation, but often sent divisive messages about mainstream and moderate Republicans. During the Republican Party primaries, #GOPStop appears frequently in Right Troll tweets, e.g., @amalia_petty, December 16, 2015, "#VegasGOPDebate Asking who is gonna win #GOPDebate is like asking what sort of crap is your favourite?" Similarly, on October 6, 2016, @hyddrox retweeted "The House voted to impeach Koskinen but that JERK McConnell said he didn't have time to take it up on the senate Time to EXIT THE D.C." in reference to Republican Senate Majority Leader Mitch McConnell.

This category also includes some themed accounts, including @itstimetoseced, which advocated for the secession of Texas, and @Jihadist2ndWife, a parody handle, which presented itself as the wife of an Islamic State fighter. The overwhelming majority of handles, however, had limited identifying information, with profile pictures typically of attractive, young women.

Left Troll (228 handles, 560,744 tweets, mean = 2,459, s.d. = 3,550).

These handles sent socially liberal messages, with an overwhelming focus on cultural identity. They discussed gender and sexual identity (e.g., #LGBTQ) and religious identity (e.g., #MuslimBan), but primarily focused on racial identity. Among the ten hashtags most employed by these accounts were #BlackLivesMatter (n = 13,258), #PoliceBrutality (n = 2,269), and #BlackSkinIsNotACrime (n = 1,981). Many handles, including @Blacktivists and @BlackToLive, tweeted in a way that mimicked the Black Lives Matter movement, with posts such as @Blacktivists, May 17, 2016, "Justice is a matter of skin color in America. #BlackTwitter". Many such tweets seemed intentionally divisive, including @Blacktivists, May 10, 2016, "When you have been handcuffed for no good reason, all you can think about is how not to get shot. Never trust a cop", or @BlackToLive, September 6, 2016, "they treat us today, not like fellow citizens, but as an insurgency which they must suppress...".

`Just as the Right Troll handles attacked mainstream Republican politicians, Left Troll handles attacked mainstream Democratic politicians, particularly Hillary Clinton. Tweets such as @Blacktivists, October 31, 2016, "NO LIVES MATTER TO HILLARY CLINTON. ONLY VOTES MATTER TO HILLARY CLINTON" and a retweet from @JerStoner, October 7, 2016, "#ClintonBodyCount if anyone else had her rap sheet - they'd be on death row". Such tweets undermined Clinton's credibility and spread questionable information about her campaign prior to the 2016 election. In contrast, these handles were supportive of Bernie Sanders prior to the election, with posts such as @blacneighbor, June 13, 2016, “I think many folks took @BernieSanders for granted. I've never seen a politician so passionate about the people!”

It is worth noting that this account type also included a substantial portion of messages which had no clear political motivation. For instance, among the ten hashtags most employed by this category were #NowPlaying (n = 5,200) and #sports (n = 1,685). It seems possible that such messages were employed as a means of camouflage to appear more genuine, as a method of gaining followers, or a combination of both.

News Feed (55 handles, 910,384 tweets, mean = 16,552, s.d. = 15,909). These handles overwhelming presented themselves as U.S. local news aggregators and had descriptive names such as @OnlineMemphis and @TodayPittsburgh. Among the ten hashtags most employed by these accounts were #news (n = 232,145), #sports (n = 94,445), and #local (n = 50,632). These accounts linked to legitimate regional news sources and tweeted about issues of local interest, such as @KansasDailyNews, December 9, 2015, “#news Barton County finds new revenue with oil well” and on the same day, “#news SW Kansas sheriff says he’s getting calls about welfare of some horses”.

A small number of these handles, including @SpecialAffair and @WarfareWW, tweeted about global issues, often with a pro-Russia perspective. The handle @todayinsyria tweeted on October 11, 2015, “2 civilians killed by terrorists’ gunfire in Sweida countryside http://t.co/lHbleruLq3” and on the next day “Russian Air Force destroys 53 targets for ISIS in several areas in Syria http://t.co/aSBbcfwQkT”. These link directly to the Syrian Arab News Agency, a Syrian state agency allied with the Russian government.

Hashtag Gamer (110 handles, 392,285 tweets, mean = 3,566, s.d. = 4,208). These handles are dedicated almost entirely to playing hashtag games, a popular word game played on Twitter. Users add a hashtag to a tweet (e.g., #ThingsILearnedFromCartoons) and then answer the implied question (Haskell, 2015). Among the ten hashtags most employed by these accounts were #ToDoListBeforeChristmas (n = 3,780), #ThingsYouCantIgnore (n = 3,358), #MustBeBanned (n = 3,129), and #2016In4Words (n = 2,875).

These handles also posted tweets that seemed organizational regarding these games, e.g. @AmandaVGreen's quote tweet, August 31, 2016, "15 minutes till we play @TheHashtagGame with @HashtagRoundup & @HashtagZoo! Who's ready to #hashtag!". Many of these tweets were mundane, including @DonnieLMiller, April 12, 2017, "#OffendEveryoneIn4Words fart in your face." Like some tweets from Left Trolls, it is possible such tweets were employed as a form of camouflage, as a means of accruing followers, or both.

Other tweets, however, often using the same hashtag as mundane tweets, were socially divisive, including @DonnieLMiller, April 12, 2017: "#OffendEveryoneIn4Words undocumented immigrants are ILLEGALS." Many tweets from Hashtag Gamers were overtly political, e.g. @LoraGreeen, July 11, 2015, "#WasteAMillionIn3Words Donate to #Hillary". While many tweets shared themes seen in the Right Troll category, Left Troll themes also appeared, e.g., @LoraGreeen, January 25, 2016, "#ItsSoWhiteOutsideThat Donald Trump thought it was a meeting of his followers."

Fearmonger (698 handles, 293,337 tweets, mean = 420, s.d. = 455). These accounts spread disinformation regarding fabricated crisis events, both in the U.S. and abroad. Such events included non-existent outbreaks of Ebola in Atlanta and Salmonella in New York, an explosion at the Columbian Chemicals plan in Louisiana, a phosphorus leak in Idaho, as well as nuclear plant accidents and war crimes perpetrated in Ukraine. Among the ten hashtags most employed by these accounts were #Fukushima2015 (n = 10,830) and #ColumbianChemicals (n = 5276). These accounts typically tweeted a great deal of innocent, often frivolous content (i.e. song lyrics or lines of poetry) which were potentially automated. With this content these accounts often added popular hashtags such as #love (n = 9,852) and #rap (n = 4,277). These accounts changed behavior sporadically to tweet disinformation, and that output was produced using a different Twitter client than the one used to produce the frivolous content.

The final story fabricated by these accounts was typical of their activity. This story was that salmonella-contaminated turkeys were produced by Koch Foods, a U.S. poultry producer, near the 2015 Thanksgiving holiday. The tweets described the poisoning of individuals who purchased these turkeys from Walmart. These included @RitterTra, November 26, 2015, "OMG Obama and Koch bros. are trying to steal our holidays! nice. #USDA" and also @Peter_Downs_Up, November 27, 2015, "wooow Whut? Poisoned #turkey on Thanksgiving?! #KochFarms #foodpoisoning #USDA". Koch Foods has no connection to the Koch brothers, and the story was an IRA fabrication (Washington, 2018).

The Fearmonger category was the only category where we observed some inconsistency in account activity. A small number of handles tweeted briefly in a manner consistent with the Right Troll category but switched to tweeting as a Fearmonger or vice-versa. We coded accounts in a way consistent with how they tweeted most recently. We observed no such inconsistency after mid-2015.

RQ3. Analysis of account types found that account types were employed differently at various times, often seemingly in response to real world events; account types functioned in largely different networks from one another; and account types differed in their communication actions. The details of these differences are outlined below.

RQ3.1 Are accounts of the different types employed in different circumstances?

Figure 2 displays the daily number of tweets by account type. Panel (a) presents Left Troll and Right Troll accounts, while panel (b) displays News Feed, Fearmonger, and Hashtag Gamer accounts. These figures illustrate many differences in how the IRA employed account types. First, the timing. Fearmongers were operated most intensely in a much different period than the other account types, very early in the campaign, in late 2014 and early 2015. The Left Troll, Right Troll, and Hashtag Gamer accounts, by contrast, were most active in late 2016 and early 2017. News Feeds operated consistently from early 2015 to mid-2017. This suggests that Fearmongers may represent a single, early, abandoned method of the IRA to engage with English-speaking users rather than an account type which was part of a possible larger strategy.

A second marked difference is in variance of output. Left Troll, Right Troll, and Hashtag Gamer accounts had much more variable output than the News Feeds did. As an example, in 2016, the Left Troll, Right Troll, and Hashtag Gamers' daily outputs had coefficients of variation (ratio of standard deviation to mean) between one and two (1.3 for Left Trolls, 1.5 for Right Trolls, and 1.8 for Hashtag Gamers), while News Feeds' daily output has a coefficient of variation of only 0.48. A Kruskal–Wallis test rejects the null that the daily tweet totals for these five accounts types were pulled from the same distribution ($p < .001$). In a 12-month period when they were most active (August 1, 2014 – July 31, 2015), Fearmongers also had highly variable output, with a coefficient of variation of 2.2. There were also differences in the tails of these distributions. Right Troll, Left Trolls, and Fearmongers, have very heavy tails, with maximums close to 20 times their means, while the max/mean ratio of Hashtag Gamers is around ten, and that for News Feeds is around two.

In contrast to these differences in variance and timing, there is a surprising consistency in mean output. In 2016, the mean daily output of Left Trolls, Right Trolls, Hashtag Gamers, and News Feeds was 708, 549, 606, and 686, and, with the large standard deviations, we cannot reject the null that they are all equal. In the year they were most active, the mean daily output of the Fearmongers was 723, which is also not statistically different from the other four.

The underlying differences in variance and skew may result from the way account types differentially reacted to political circumstance. Figure 3 zooms in on two short periods of interest to demonstrate how the account types act and react hour-to-hour. Panel (a) zooms in on a five-day period beginning on Sept. 11, 2016, while Panel (b) displays over a month's worth of activity (from September 20 to October 25, 2016). In addition to presenting hourly output by account type, each timeline also labels several important events with vertical bars and displays the most prominent hashtags appearing in several of the spikes of activity, where the colors of the hashtags coincide with the account type producing them (green for Hashtag Gamers and red for Right Trolls).

Three important things to notice in Panel (a) are the presence of working and "off" periods, the sudden switching between account types, and the way different account types respond to political circumstances. IRA employees tasked with making social media posts are reported to have been organized into twelve-hour shifts, a day shift and night shift, and instructed to make posts at times appropriate to U.S. time zones (United States of America v. Internet Research Agency LLC). We can see these shifts quite distinctly on the 12th-14th.

Even within work periods, the troll operators seem to be mostly specializing on one account type at a time, beginning with Hashtag Gamers and following up with Right Trolls on Sept. 12th, but reversing the order on Sept. 14. Part of this scheduling might be driven by the planned nature of the Hashtag Gamer events. News Feeds are not subject to these shifts, producing at low and consistent volumes throughout the period.

Finally, the prominent hashtags reveal how the accounts respond to shifts in political circumstances, in this case the well-known faint/stumble by Hillary Clinton leaving a 9/11 commemoration event, followed by her pausing the campaign with an announcement of pneumonia. By the time of their next work period, on Sept 12th, the trolls are already commenting on this event, but the approach varies by troll type. The Right Trolls directly question Clinton's health, claiming that this event is evidence of a serious underlying condition, including supporting a conspiracy theory about Clinton using an actress to cover up a condition, with #HillarysHealth and #HillaryBodyDouble being the dominant hashtags on the next 2-3 days. The Hashtag Gamers are more subtle, starting with conducting two health-related hashtag games on the day after the stumble, before they take a more direct, but still veiled, opportunity to foster commentary on the body double conspiracy by running a game on the hashtag #IfIHadABodyDouble. Unlike Right Trolls, Hashtag Gamers maintain a veneer of equal-opportunity jokesters, so they take a less direct path. But even the relatively innocuous hashtags provide an opportunity to hide within them the more provocative ones. For example, one of the most common actions from the Hashtag Gamers on Sep. 12 was to retweet "@ChrixMorgan: #ToFeelBetterI get a doppelganger who goes to work instead of me but doesn't ask for any money #HillarysBodyDouble". Over this same period, the Left Trolls are not particularly active, and they are hardly discussing Clinton's health at all.

Panel (b) illustrates two additional facts about the trolls' behavior: how they operate around pre-scheduled political events and what happened on Oct 6, 2016, the most dramatic shift in troll behavior of the entire operation. Three vertical lines in the graph indicate the hours at which each of the three general election U.S. presidential debates began in 2016. Hashtag Gamers were the only troll types that had significant spikes in output immediately around these times, and the prominent hashtags in each of those spikes were directly relevant to the debates (#ThingsMoreTrustedThanHillary, #BetterAlternativesToDebates, and #RejectedDebateTopics), in contrast to apparently non-political prominent hashtag in the week before the first debate (#ReasonsToGetDivorced). But the exact timing of how the Hashtag Gamers tweet, relative to the debate time, varied among the three debates. For the first debate, they were most active in the hour of the debate. For the second, they were most active in the hours after the debate, while for the third they were most active in the hours before the debate. These differences arise despite all the debates taking place at the same time of day. We can only speculate about the motive for the timing, but perhaps they were experimenting to determine the best time to affect the political conversation.

The second striking feature of the troll output in this period is their behavior on and after October 6, 2016. This day was the maximum day for English-language IRA output since very early in 2015, and it was, by far, the maximum day for the Left Trolls. The Left Trolls greatly increase production of tweets for at least 14 hours in a row, beginning at noon UTC (3pm in St. Petersburg, 8am in NYC) on the 6th, and continuing with a second spike around 8:00 a.m. UTC on the 7th. In additional to being a spike, it also demarcated a significant change in behavior by both Left and Right Troll accounts, as they begin to tweet at much more consistent daily levels (See Figure 2) and to engage in much higher rates of retweeting (over 90% for Left Trolls through, at least, the end of 2016). Right Trolls showed a similar, but less marked shift over this period, as well. There were no substantial changes in behavior for the other troll types on or around this day.

This sudden change in behavior illustrates Left Trolls being used in a very different way than they had before or than other account types were being used, and the reason for the shift is unclear. But the change is so dramatic, that we believe some speculation about an underlying strategy is warranted. One strong possibility is that the IRA was using its Left Troll accounts to attempt to encourage the participation in the political debate by factions of the Democratic coalition that were suspicious of Clinton's candidacy in the month leading up to the election, in general, and, more specifically, in anticipation of the release of John Podesta's emails by Wikileaks in the coming day.

At 4:32pm EDT (8:32 UTC) on Oct 7th, Wikileaks released the first trove of emails that the Russian military intelligence had illegally obtained by hacking the email account of John Podesta, Hillary Clinton's campaign chairman. This release is indicated by a vertical line in the timeline. It included several emails that might have undermined left-wing support for Hillary Clinton, especially among Bernie Sanders supporters, include quotations from her paid speeches to Wall Street indicating that she had a "public position and a private position" and one suggesting that she may have had early access to a primary debate question for the March 23 Democratic Town Hall (Goldman, 2016).

There were several other important political events at about the same time as the Podesta release, including a joint statement from the Department of Homeland Security and Office of the Director of National Intelligence, making the first public claim that the Russian government was behind the election hacking that had targeted the Clinton campaign and the DNC, the release of the *Access Hollywood* tape, and the second Presidential debate, but there are good reasons to suspect that the Wikileaks release was the key. First, we know from the Special Counsel's investigation that the emails originated with the GRU, and they had expressed preferences over the timing of their release (Mueller , 2019). Second, the flood of retweets on October 6th, continuing up to and through the election were dominated by Left Trolls. The accounts retweeted by the Left Trolls would be likely targets for the information contained in the hacked Podesta emails, wavering Sanders-Clinton voters who may have already been weakly supportive of her candidacy. Linvill et al. (2019) shows that the messages contained in Left Troll tweets in the month leading up to the election were quite ambivalent about Clinton's candidacy. These sorts of messages would be complementary to the information included in the hacked emails.

The other alternative triggers seem less plausible. There is no evidence that the IRA had any advance knowledge of the *Access Hollywood* tape. A preemptive amplification strategy to shape the response to the release of the intelligence assessment seems hard to understand. Such a strategy would likely have focused on the Right Trolls, as it is the strong pro-Trump partisans who would have been most likely to react with suspicion to such a claim from the Obama Whitehouse. Activating accounts associated with the left of the Democratic Party and with Black Lives Matters makes little sense in that context. Finally, it is hard to understand why this shift would have been made in anticipation of the second debate. We see no similar strategic shifts around either of the other two presidential debates, and much of the activity immediately surrounding pre-scheduled political events such as the debates is carried out by the Hashtag Gamers, rather than the ideological types. Finally, according to Clint Watts, former FBI agent and expert on Russian troll behavior, this activity is consistent with previous observations that activity tends to "ramp up when they know something's coming" (Timberg & Harris, 2018, p. A12).

These periods make clear that the IRA allocated their efforts amongst the account types differently when faced with varying political circumstances or shifting goals. In both periods, a Kruskal–Wallis test rejects the null that the daily tweet totals for these accounts types were pulled from the same distribution (p < .001). Without better information about their goals, it is difficult to speculate about exactly what underlying strategy is driving these shifts, or even whether the shifts are not strategic at all are, but are rather the byproduct of some strategic change along an unobserved dimension, but it is clear from the differential changes that accounts of different types are not simply substitutes for each other--- each plays a differentiated role in this campaign.

RQ 3.2 Do accounts of different types use different mixes of communication actions?

Table 1 presents the results of our investigation of the communication actions taken by the five major English-language account types. Each column reports the share of tweets from the indicated account type that were of the action type indicated by the row. In panel (a), we report the shares of tweets that were retweets, quote tweets, replies, and “original” tweets, which could include copy/pasted content that is not explicitly retweeted. In panel (b), we report the share of tweets that originated from each of the top-15 Twitter clients used by the IRA. The remaining share were lumped into an aggregate “other” category. In both cases, we can reject a null hypothesis of equal distributions with very high confidence (p<0.001), both overall and for every pairwise comparison of account types.

The results in panel (a) point to large differences in the mix of tweet type across account types. Left Trolls are, by far, the most likely to retweet, with over three-quarters of their output being simple retweets of other accounts, leaving about 10% for original stand-alone tweets. News Feeds were at the opposite extreme, with 99.6% of tweets being original stand-alone tweets, although these were overwhelmingly scraped from legitimate news producers. Fearmongers also had high rates of original stand-alone tweets (89.5%), but many, if not most, of these tweets are recycled between multiple accounts. Hashtag Gamers retweeted frequently, about 57% of the time, but make little use of quote-tweets or replies. Finally, Right Trolls were the relatively likely to reply (5%) and quote tweet (9.6%), and they had the second smallest share of original tweets (24.8%).

It is likely some of these differences in communication activity reflect the activity of accounts the IRA are mimicking. The retweeting of preferred tweets seems to be a fundamental element of the hashtag game and so high rates of retweeting by these accounts is natural. A need to engage in community specific behavior may similarly explain differences between other account types.

The Twitter client results in panel (b) of Table 1 also point to significant differences across types. Left Trolls overwhelmingly used the Twitter Web Client (91%) and Tweetdeck (5.3%) to post their content, as did Hashtag Gamers (83.5% and 15.5%, respectively). In dramatic contrast, the Newsfeeds overwhelmingly used Twitterfeed (75.4%) and Twibble.io (21%), and no other account type substantially used either of these clients. Right Trolls and Fearmongers also used the Twitter Web Client as their modal platform (67.1% and 60.9%, respectively), but Right Trolls also made substantial use of Twitter for Android (15.7%), which had no substantial use by other account types, while the Fearmongers very commonly, and uniquely, used vavilonX (34.9%).

It is probable these differences reflect account types' differing activity: the IRA uses tools that best facilitate implementing the actions required of a given account type. Twibble.io and Twitterfeed, for example, allow easy links between RSS feeds and Twitter, allowing the News

Feed accounts to easily mirror the content of legitimate local news sources. Tweetdeck, on the other hand, allows users to simultaneously retweet the same content from multiple accounts, which would be useful in amplifying content in the hashtag game. The use of specialized Twitter clients for more automated behavior is consistent with some of the markers of Russian bot behavior (Stukal et al, 2017). The use of VavilonX by Fearmongers may simply be a remnant of the pivot from the IRA's domestic operations in Russia and Ukraine.

RQ 3.3 Do accounts of different types locate themselves differently in the social network?

Table 2 presents the results of our investigation of the social network links among the five major English-language account types. Each column reports the share of tweets originating from the indicated account type that target accounts of the type indicated in the rows. The last line in each panel reports the number of tweets from the indicated account type that qualify for analysis in this panel by targeting another IRA account. We report results for mentions, retweets, and replies, in separate panels. In all three cases, we can reject a null hypothesis of equal distributions with very high confidence (p<0.001), both overall and for every pairwise comparison of account types.

Across all metrics of social-network linkage, both Left Trolls and Right Trolls link to other accounts of their own type or to News Feeds, almost exclusively. Left Trolls exhibit more homophily (own-type linkage) in mentions and retweets, while Right Trolls exhibit it more in replies. Overall, both types link more to News Feeds than to any other type of account.

Fearmongers link to other Fearmongers and, in the case of mentions and replies, to accounts that we were unable to encode. Overall, they link more to other Fearmongers than any other account type. Hashtag Gamers link to other Hashtag Gamers or, in the case of replies, to

Fearmongers. Overall, more than 90% of their links are to other Hashtag Gamers. News Feeds overwhelmingly link to other News Feeds (43% of retweets were to non-English accounts, but retweets by News Feeds were extremely rare, less than .001% of the news-feed tweets).

Table 3 presents another piece of evidence about the way that the accounts of different types position themselves in the broader social network. The table investigates the degree to which accounts were linked to by IRA accounts of different types and shows that linkages were quite concentrated. In each panel, the unit of observation is an account that is linked to by the IRA at least three times, using the indicated metric of network linkage, whether or not that target account is IRA-affiliated.[5] For each row, we show the number of qualifying accounts (i.e., those that received 3-5 linkages in the case of the first row), and what fraction of those qualifying accounts received a significant number (more than 5%) of links from one account type, 2 account types, and so forth.

In the final column we present the mean Herfindahl-Hirschman Index (HHI) for account origin types. Higher HHI indicates that the account's IRA linkages are more concentrated in a few categories. It is defined by the sum of the square of the shares of links that the account receives from each of the IRA account types. This measure runs from 1/5, if the account received an equal number of links from each of the five origin types, to 1, if all its links came from the same origin type.

The first panel considers retweet linkages. For those accounts receiving 3-5 linkages, about 85% received them all from the same account type (possibly from the same account). Even as the

[5] We restrict attention to accounts linked to at least three times to avoid the trivial concentration of origin account types that occurs when the number of links is very low.

number of retweets grow, they are still quite likely to all come from the same sort of origin account. Of the 2,750 accounts receiving more than 50 retweets by the IRA, for example, over 75 percent of them received all of those retweets from the same type of troll account. From Table 2, we can calculate that overall index of concentration for retweets, what the HHI would be if all targeted accounts received the population share of retweets from each origin type, is 0.32. But for the individual target accounts it is much higher, over 0.90 regardless of the number of links they receive. [6]

The second panel presents results for reply linkages. They are also considerably more concentrated than random, with more than half of accounts receiving a significant number of links from a single source type only. But the link concentration is much less pronounced than it was for retweets, especially for the accounts that received many replies. Nevertheless, over 90 percent of accounts received a significant number of replies from 2 or fewer origin account types, and the mean HHI was over 0.8 for all levels of reply activity, despite being only 0.3 in the population of replies.

The third panel, includes the statistics for all mentions. They also show high levels of concentration, falling between the reply and retweet results.

The overall pattern of linkage concentration is broadly consistent with the homophily we showed in Table 2. Most accounts that are linked to by the IRA receive all their linkages from the same type of account, and even the accounts that receive many linkages that are very heavily concentrated from a single account type. The concentration seems stronger for retweets than

[6] Overall, about 30% of retweets come from Left Trolls, 32% come from Right Trolls, 35% come from Hashtag Gamers, about 2% come from Fearmongers, and an insignificant number come from News Feeds, so the overall HHI is given by 0.3^2+.32^2+.35^2+.02^2=0.315.

replies, and this pattern contrasts a bit with the pattern for internal links in Table 2, where both Left and Right Trolls retweeted News Feeds quite a lot but only Left Trolls replied to them to a significant degree.

The differences in when different accounts types were most active (Figure 2) may account for some differences in communication networks we see. Fearmongers, for instance, are most active before other account types are created and heavily employed and could therefore not engage with other account types to any large degree. Regardless, it seems probable that the IRA networked their accounts by structured sub-groups, either those types outlined above or ones similar. This was, perhaps, to boost specific messages, gain visibility and followers for accounts, or both.

Discussion

The IRA efforts in our sample period can be understood as systematic. Their system was industrial -- mass produced from a collection of interchangeable parts, where each class of part fulfilled a specialized function. Handles were built into one of five groups used as interchangeable parts depending on organizational needs. Effort was reallocated amongst account types in response to shocks, depending on the segment of the U.S. electorate the IRA wished to engage, changing IRA strategic goals, or both. It is clear from our analysis that the IRA focused on divergent, often contrary agenda in their disinformation campaigns, engaging with opposing, ideologically engaged networks. This supports the narrative that one effort the IRA was engaged in was to divide the country along partisan lines by playing multiple sides against each other (Graff, 2018).

Understanding how governments, government affiliated, and politically motivated organizations work to influence nations is vital, and the IRA social media operation is an important example from the digital age. At a February 13, 2018 U.S. Senate Intelligence Committee hearing, Senator Mark Warner stated that social media companies have been “slow to recognize the threat” that Russian influence poses (Nakashima & Harris, 2018, para. 7). At that same hearing, U.S. Director of National Intelligence Daniel Coats said of Russian efforts to disrupt the 2016 election, “There should be no doubt that Russia perceives its past efforts as successful” (para. 9). The Director then warned of the certainty of future Russian interference. Given the industrialized nature of the production of tweets analyzed here, we agree with Coats.

For this reason, future research will need to examine IRA efforts further, as well as the efforts of other producers of state-affiliated disinformation. The data employed for this study can be used to analyze the qualitative nature of individual tweets and to give a more detailed understanding of the effectiveness of this campaign over time. These data can also be used to better understand how the IRA’s tactics adapt over time and, by analyzing non-English tweets, in various contexts.

Data from other social media platforms should also be systematically analyzed to understand how, if at all, platforms were employed differently by the IRA and how the nature of platforms influenced their use. Only such a broad understanding will allow the public to fully guard against future disinformation attacks. Any such study would potentially face some of the same limitations as ours, however. This research was reliant on data made public by Twitter. It is possible, if not probable, that this sample is not the complete population of IRA associated content during the period explored. This sample was dependent on Twitter’s ability to both accurately and fully identify IRA activity on their platform as well as their willingness to disclose identified activity. Given the number of tweets available in the dataset, however, we argue that while our findings may not be representative of all IRA activity, they certainly point to important strategies employed.

None-the-less, future research should endeavor to explore methods of reliably identifying valid sets of disinformation produced on social media platforms. Any approach to doing so would likely have additional limitations, but understanding this important element of our political discourse cannot remain reliant on content which for profit media platforms do or do not choose to share publicly. Future research should also aim to better understand any potential effects of state sponsored disinformation and other forms of public agenda building. Such questions could not begin to be answered with the data analyzed in this study, however.

Russia's attempts to distract, divide, and demoralize has been called a form of political war (Galeotti, 2018). The U.S. military certainly takes foreign attempts to influence genuine U.S. political discourse seriously, going so far as to remotely disrupt Russian social media troll activity on the day of the 2018 midterm elections (Nakashima, 2019). This analysis has given insight into the methods the IRA used to engage in this war. One former employee of the IRA described the feeling of working there as though "you were in some kind of factory that turned lying, telling untruths, into an industrial assembly line" (Troianovski, Helderman, Nakashima, & Timberg, 2018). The systematic and organized nature of the messaging we have analyzed here suggests this employees' feeling was correct. The IRA engaged in what is not simply political warfare, but industrialized political warfare.

References

Arif, Ahmer, Stewart, L, and Starbird K. (2018) Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, No. CSCW, Article 20.

Badawy, A., Lerman, K., and Ferrara, E. (2018) Who Falls for Online Political Manipulation? The case of the Russian Interference Campaign in the 2016 US Presidential Election. arXiv:1808.03281v1.

Birnbaum, E. (2019, April 18). Mueller identified 'dozens' of US rallies organized by Russian troll farm. *The Hill*. Retrieved from https://thehill.com/policy/technology/439532-mueller-identified-dozens-of-us-rallies-organized-by-russian-troll-farm

Broniatowski, David A., Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze (2018) Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health*, 108, no. 10 (October 1, 2018): pp. 1378-1384.

Calderwood, A., Riglin, E., & Vaidyanathan, S. (2018, July 20). How Americans wound up on Twitter's list of Russian bots. *WIRED*. Retrieved from https://www.wired.com/story/how-americans-wound-up-on-twitters-list-of-russian-bots/

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research.* Thousand Oaks, CA: Sage.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage.

Denham, B. E. (2010). Toward a conceptual consistency in studies of agenda-building processes: A scholarly review. *The Review of Communication*, 10, 306-323. doi:10.1080/15358593.2010.502593

Gadde, V. & Roth, Y. (2018, October 17). Enabling further research on information operations on Twitter. Retrieved from https://blog.twitter.com/official/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html

Galeotti, M. (2018, March 5). I'm sorry for creating the 'Gerasimov Doctrine'. *Foreign Policy*. Retrieved from http://foreignpolicy.com/2018/03/05/im-sorry-for-creating-the-gerasimov-doctrine/

Goldman, J. (2016, October 7) Podesta emails show excerpts of Clinton speeches to Goldman. *CBS News* Retrieved from https://www.cbsnews.com/news/podesta-emails-show-excerpts-of-clinton-speeches-to-goldman/

Golovchnko, Y., Hartmann, M., and Adler-Nissen, R. (2018). State, media and civil society in the information warfare over Ukraine: citizen curators of digital disinformation. *International Affairs*, *94*, 975-994.

Graff, G. M. (2018, October 19). Russian trolls are still playing both sides—even with the Mueller probe. *WIRED.* Retrieved from https://www.wired.com/story/russia-indictment-twitter-facebook-play-both-sides/

Nakashima, E. (2019, February 27). U.S. Cyber Command operation disrupted Internet access of Russian troll factory on day of 2018 midterms. *The Washington Post.* Retrieved from https://www.washingtonpost.com/world/national-security/us-cyber-command-operation-disrupted-internet-access-of-russian-troll-factory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9-af5b-b51b7ff322e9_story.html

Nakashima, E. & Harris, S. (2018, February 13). The nation's top spies said Russia is continuing to target the U.S. political system. *The Washington Post.* Retrieved from https://www.washingtonpost.com/world/national-security/fbi-director-to-face-questions-on-security-clearances-and-agents-independence/2018/02/13/f3e4c706-105f-11e8-9570-29c9830535e5_story.html?utm_term=.9d39f53cf636

Haskell, W. (2015). People explaining their 'personal paradise' is the latest hashtag to explode on Twitter. *Business Insider*. Retrieved from http://www.businessinsider.com/hashtag-games-on-twitter-2015-6

Jamieson, K. H. (2018). Cyber-War: How Russian Hackers and Trolls Helped Elect a President. Oxford University Press: New York.

Lariscy, R. W., Avery, E. J., Sweetser, K. D., Howes, P. (2009). An examination of the role of online social media in journalists' source mix. *Public Relations Review, 35,* 314-316. doi:10.1016/j.pubrev.2009.05.008

Linvill, D.L., Boatwright, B. C., Grant, W.J., & Warren, P.L. (2019) The Russians are hacking my brain!: Investigating Russia's Internet Research Agency Twitter tactics during the 2016 United States presidential campaign. *Computers in Human Behavior, 99,* 292-300. doi.org/10.1016/j.chb.2019.05.027

Lukito, J., & Wells, C. (2018). Most major outlets have used Russian tweets as sources for partisan opinion: Study. *Columbia Journalism Review*. Retrieved from https://www.cjr.org/analysis/tweets-russia-news.php.

Mueller, R. S. III (2019) *Report on the Investigation into Russian Interference in the 2016 Presidential Election, Volume I.* Washington, D.C. U.S. Department of Justice

Parmelee, J. H. (2014). The agenda-building function of political tweets. *New Media & Society, 16*, 434-450. doi:10.1177/1461444813487955

Permanent Select Committee on Intelligence (2018, June 18). Schiff statement on release of Twitter ads, accounts and data. Retrieved from: https://democrats-intelligence.house.gov/news/documentsingle.aspx?DocumentID=396

Roth, Y. (2019, June 13). Information operations on Twitter: Principles, process, and disclosure. Retrieved from https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html

Stukel, D., Sanovich, S., Bonneau, R., and Tucker, J. A. "Detecting Bots on Russian Political Twitter" *Big Data* 5:17, 310-324. DOI: 10.1089/big.2017.0038

Timberg, C., & Harris, S. (2018). Burst of tweets from Russian operatives in October 2016 generates suspicion. *The Washington Post*, p. A12.

Troianovski, A., Helderman, R. S., Nakashima, E., & Timberg, C. (2018, February 17). The 21st-century sleeper agent is a troll with an American accent. *The Washington Post.* Retrieved from https://www.washingtonpost.com/business/technology/the-21st-century-russian-sleeper-agent-is-a-troll-with-an-american-accent/2018/02/17/d024ead2-1404-11e8-8ea1-c1d91fcec3fe_story.html?noredirect=on&utm_term=.d5906ace8983

United States of America v. Internet Research Agency LLC. District of Columbia (2018). Retrieved from https://www.justice.gov/file/1035477/download

Washington, B. (2018, February 22). Inside Russia’s fake news HQ. *The Australian.* p. INQUIRER 11.

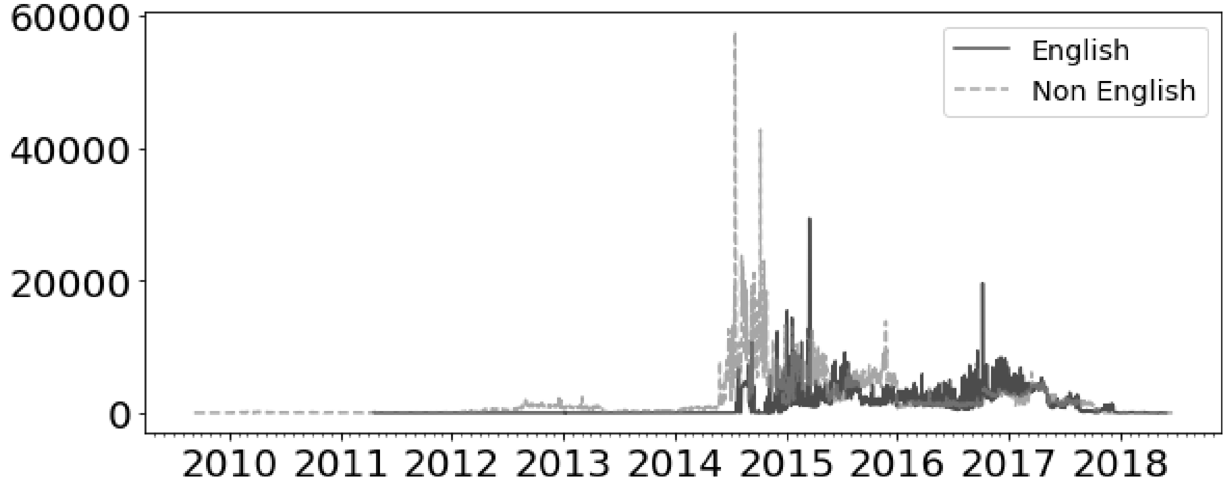
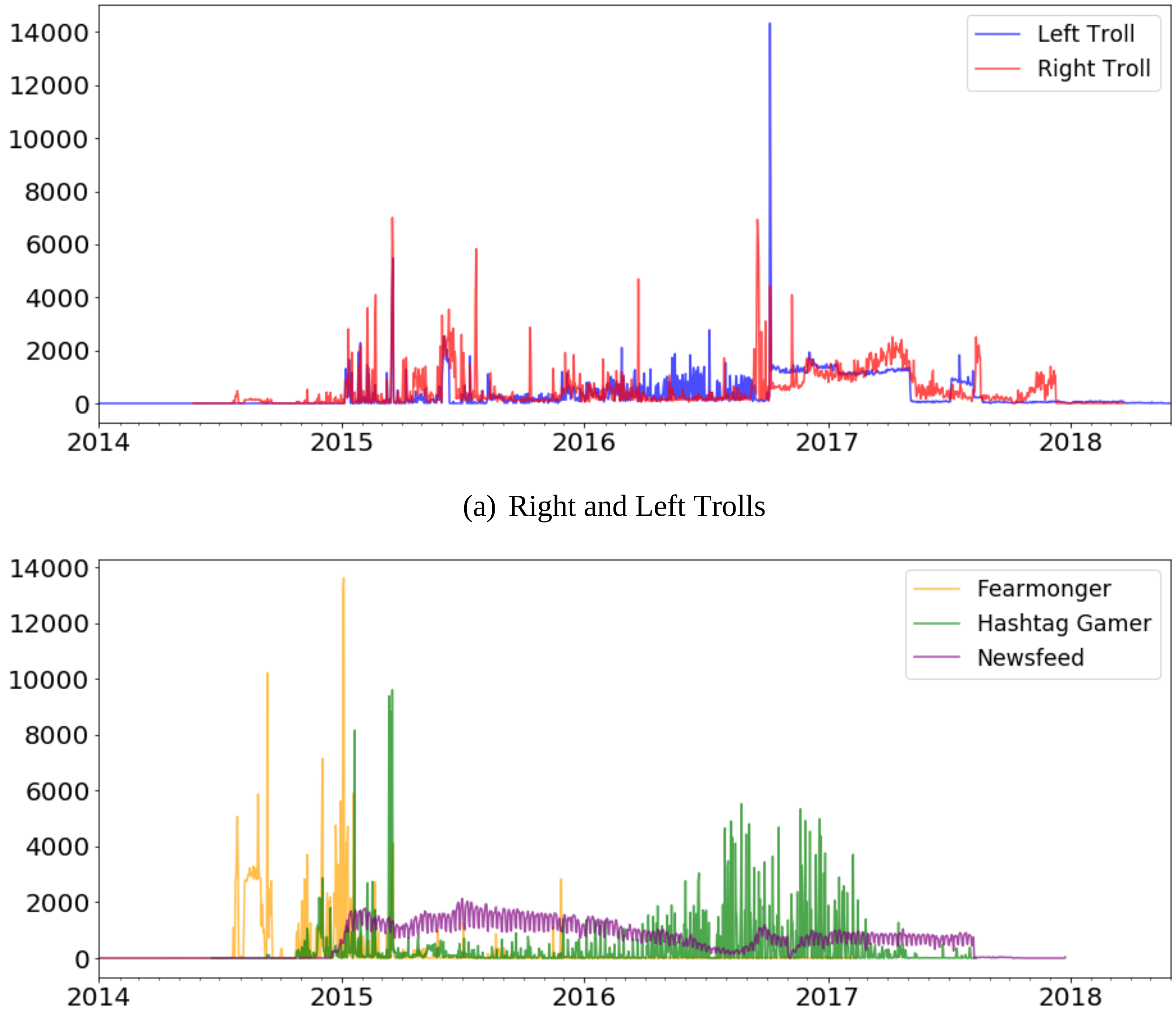


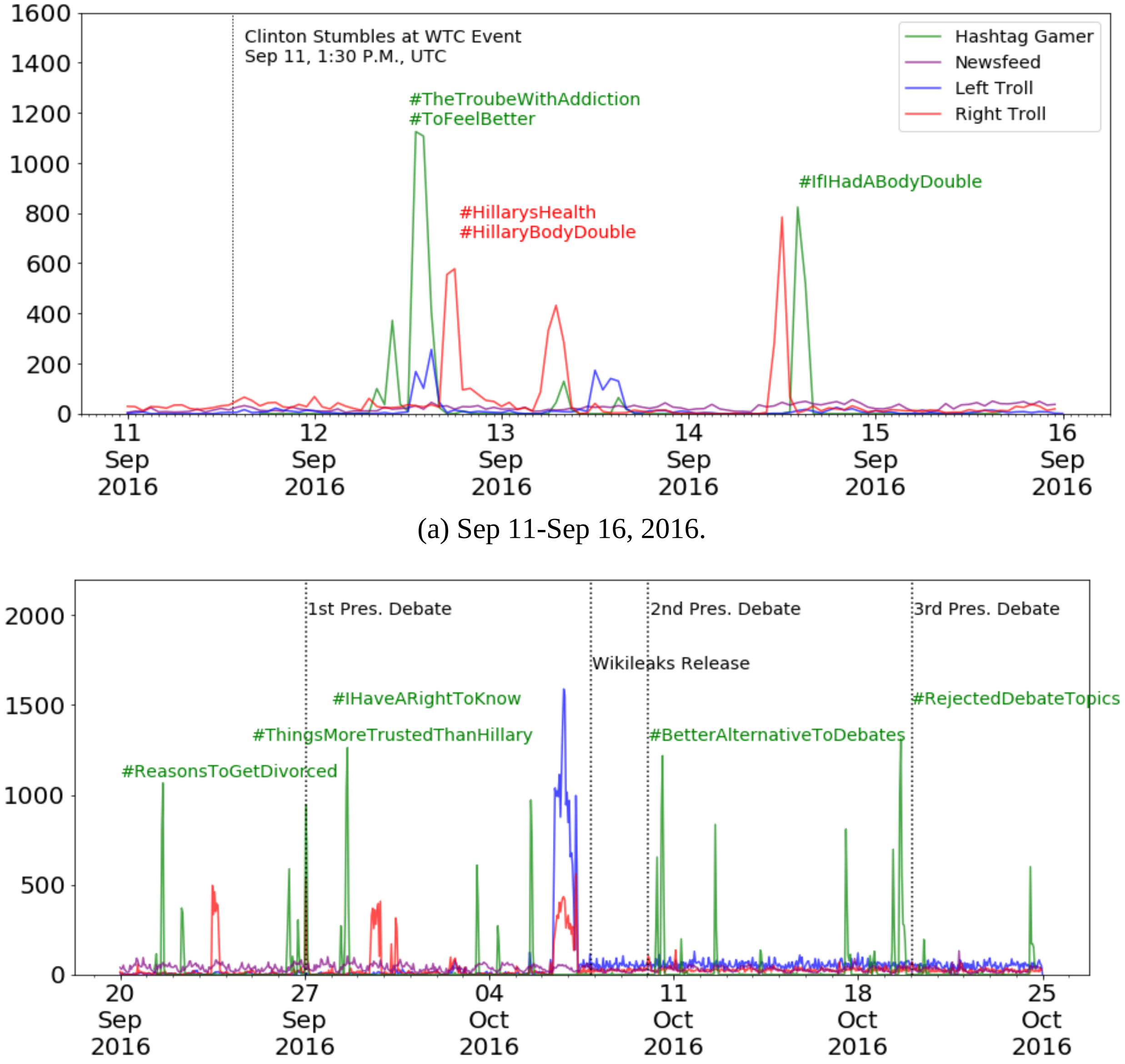
Figure 1. Daily tweets by English and Non-English accounts, Sep 9, 2009 – June 21, 2018.



(a) Right and Left Trolls

(b) Fearmonger, Hashtag Gamer, and News Feeds

Figure 2. Daily tweets by English Language accounts, by account type, Jan 1, 2014 – – June 21, 2018.



(a) Sep 11-Sep 16, 2016.

(b) Sep 20-Oct 25, 2016

Figure 3. Hourly tweets by English-language accounts, by account type.

Table 1. Post Type and Client Usage Shares by Account Type.

	Left	**Right**	**Fearmonge**	**Hashtag**	**Newsfee**
	Troll	**Troll**	**r**	**Gamer**	**d**
			(a) Post Type Shares		
Retweet	75.9%	60.1%	1.5%	56.8%	0.0%
Quote Tweet	11.2%	9.6%	0.0%	1.2%	0.1%
Reply	2.6%	5.5%	9.0%	2.5%	0.3%
“Original” Tweet	10.3%	24.8%	89.5%	39.5%	99.6%
			(b) Client Usage Shares		
Twitter Web Client	90.9%	67.1%	60.9%	83.5%	1.0%
twitterfeed	0.0%	0.0%	0.1%	0.0%	75.4%
Twibble.io	0.0%	0.0%	0.0%	0.0%	21.0%
TweetDeck	5.3%	7.0%	0.5%	15.5%	2.6%
Twitter for Android	0.0%	15.7%	0.0%	0.1%	0.0%
vavilonX	0.0%	0.8%	34.9%	0.0%	0.0%
IFTTT	0.1%	2.0%	0.0%	0.0%	0.0%
POTUSADJT Bot	0.0%	1.4%	0.0%	0.0%	0.0%
Jerusalem	0.0%	1.3%	0.0%	0.0%	0.0%
Tweefilter	0.0%	1.0%	0.0%	0.0%	0.0%
masss post4	0.0%	0.0%	2.0%	0.0%	0.0%
Crowdfire - Go Big	0.0%	0.6%	0.0%	0.0%	0.0%
Twitter for Android Tablets	0.0%	0.0%	0.1%	0.8%	0.0%
Uptwitter	0.2%	0.3%	0.0%	0.0%	0.0%
masss post5	0.0%	0.0%	1.1%	0.0%	0.0%
Other	3.5%	2.5%	0.4%	0.1%	0.0%
Obs.	560,744	705,064	293,337	397\2,285	910,384

Note: Each entry reports the share of overall tweets by accounts of the type indicated in the column that have the characteristic indicated in the row.

Table 2. Target Account Type Shares by Origin Account Type: Mentions, Retweets, and Replies

	Origin Account Type				
	Left Troll	Right Troll	Fearmonger	Hashtag Gamer	Newsfeed
Target Account Type					
			(a) Mentions (n=122,633)		
Left Troll	44.6%	0.5%	0.3%	0.5%	0.0%
Right Troll	1.5%	34.0%	4.3%	2.9%	0.0%
Fearmonger	0.1%	1.4%	64.5%	2.0%	0.0%
Hashtag Gamer	1.4%	2.5%	2.0%	93.6%	0.0%
Newsfeed	52.4%	60.8%	0.3%	0.8%	97.7%
Non-English	0.0%	0.1%	3.0%	0.0%	2.2%
Unknown	0.1%	0.7%	25.6%	0.2%	0.0%
Obs.	35,782	35,004	15,117	34,626	2,104
			(b) Retweets (n=94,615)		
Left Troll	50.4%	0.5%	2.4%	0.5%	0.0%
Right Troll	1.6%	27.6%	3.6%	3.0%	0.0%
Fearmonger	0.1%	0.3%	81.3%	0.1%	0.0%
Hashtag Gamer	1.9%	2.9%	3.7%	95.4%	0.0%
Newsfeed	45.9%	68.5%	1.4%	0.9%	56.9%
Non-English	0.0%	0.0%	6.9%	0.0%	43.1%
Unknown	0.1%	0.0%	0.8%	0.1%	0.0%
Obs.	28,939	30,941	2,053	32,573	109
			(c) Reply (n=27,356)		
Left Troll	13.3%	0.5%	0.0%	0.1%	0.0%
Right Troll	1.1%	79.6%	4.4%	0.8%	0.0%
Fearmonger	0.1%	10.4%	61.7%	51.5%	0.0%
Hashtag Gamer	0.0%	0.2%	2.4%	42.8%	0.0%
Newsfeed	85.4%	3.4%	0.1%	0.8%	100.0%
Non-English	0.0%	0.6%	2.5%	0.2%	0.0%
Unknown	0.1%	5.4%	29.0%	3.9%	0.0%
Obs.	6,442	3,949	13,299	1,304	2,362

Note: Each entry reports the share of links from the account types indicated in the column that are targeted at the account types indicated in the rows, where links are defined as indicated in each panel. Links between IRA-affiliated accounts only are included in the analysis.

Table 3. Link Origin Types for Accounts Linked to by IRA.

	# of Accounts	(a) Retweeted by…			Herf.
Number of IRA Retweets		**1 Type**	**2 Types**	**3+ Types**	**0.32**
3-5	10152	86.3%	13.0%	0.7%	0.95
6-10	8430	77.8%	20.1%	2.1%	0.92
11-20	5122	72.1%	24.8%	3.2%	0.91
21-50	3500	74.6%	21.9%	3.6%	0.91
50+	2750	76.5%	20.1%	3.5%	0.91
Number of IRA Replies		**(b) Replied to by…**			**0.32**
3-5	861	79.6%	18.7%	1.7%	0.92
6-10	671	73.9%	24.3%	1.8%	0.91
11-20	385	47.5%	43.1%	9.4%	0.86
21-50	251	53.7%	37.8%	8.4%	0.83
50+	223	43.9%	46.2%	9.9%	0.81
Number of IRA Mentions		**(c) Mentioned by…**			**0.26**
3-5	10898	82.3%	16.5%	1.2%	0.93
6-10	9332	73.4%	23.3%	3.2%	0.90
11-20	5636	67.0%	27.4%	5.7%	0.89
21-50	3862	69.4%	24.1%	6.6%	0.88
50+	3056	70.5%	24.5%	5.0%	0.89

Each row presents statistics for accounts that were linked to by the IRA the number of times indicated in the first column, the count of which is provided in the second column. The middle three columns present the share of these accounts that receive a significant number (more than 5%) from the indicated number of origin categories. The final column presents the mean HHI (the sum of squared origin shares) for the targeted accounts.

Appendix—Matching Twitter Data to Social Studio Data

We used Salesforce's Social Studio social listening platform to download all tweets in the study period from known IRA accounts. To gather the original data from Social Studio, we searched for each account name on the list provided by Twitter to the U.S. House Intelligence Committee. If the account names were reused, we collected the one with the account user ID that matched the ID on the House Intelligence list, if one was provided. If one was not provided, we collected any account with that handle that was suspended at the time of collection. The Social Studio data before 2015 were incomplete, and a 3-year look-back window was imposed in the middle of data collection. Some accounts changed screen names and we were not always able to collect the tweets prior to the name change, so some accounts yielded no data or incomplete data for the early period. This Social Studio dataset included screen name, user id, tweet text, and tweet id, among other features.

The Twitter data (Gadde & Roth, 2018) include the full population of tweets by accounts associated to the IRA by Twitter. Each tweet includes tweet text and tweet id, and any tweet by an account that amassed at least 5000 followers also includes user id, and screen name.

We matched each tweet in the Twitter dataset to its associated tweet in the Social Studio data, if one existed. For 2016-onward this match was nearly perfect, but some earlier tweets were missing in the Social Studio data. For any account with a matched tweet, we imputed the screen name as the last screen name that account had in the last tweet that was matched to social studio data. This aligns with Twitter's practice for the larger accounts, where they identify accounts using the final screen name. For every case in which Twitter released the account's screen name (those with 5000+ followers), the Social Studio screen name matched. There were a handful of accounts that tweeted only very early in the campaign for which we did not have a Social Studio match, in which case we simply assigned them their hashed screen name.