



The Joseph and Alma Gildenhorn
Institute for Israel Studies
University of Maryland

Poising the Political Discourse: The Role of Online Fake Proxies

Gabriel Weimann

Dept. of Communication, University of Haifa, Israel

and

Gildenhorn Institute for Israel Studies, University of Maryland, USA

May 5, 2019

The ideal subject of totalitarian rule is not the convinced Nazi or the convinced communist, but people for whom the distinction between fact and fiction (i.e. the reality of experience) and the distinction between true and false (i.e. the standards of thought) no longer exist.

Hannah Arendt, *The Origins of Totalitarianism*

Research Paper 18

May 2019



In January 2019, Amy Spiro, an Israeli journalist, received a direct message on her Twitter account linking to a sensational news story. The sender, using the Jewish-sounding name "Bina Melamed", directed her to a fake story falsely alleging former Israeli defense minister Avigdor Lieberman was a Russian spy. Spiro did not fall victim to the ruse, but four Israeli journalists -- hoodwinked by the article appearing on a rogue but convincing duplicate of Harvard University's website -- spread the story, before it was exposed. And who is Bina Melamed? She turned out to be a fake account operating from Turkey, one of many attempting to propagate fake news in Israel through bots. In some cases, like Bina Melamed, disinformation and fake news make their way from social to professional media. This flow is made possible by online "proxies", namely avatars, bots, and trolls, representing various forms of false actors, fake identities and virtual persons. In this paper, we will examine the threat of these fake online proxies. After a brief explanation of the ways these virtual agents are used online, we will explore their alarming impact on elections campaigns, political discourse, and the stability of the democratic system. Finally, several counter-measures will be suggested.

The Agents: Avatars, Bots, and Trolls

How is it possible to interfere with political discourse? Today, through social networks, it is possible to spread false rumors, promote "fake news", incite and radicalize discourse, cause harm to candidates and parties, widen social rifts, and plunge election campaigns into an abyss of extremism, distrust, sectarianism, and violence. There is also growing concern that the business models and algorithms that drive social media companies are fueling extreme partisanship and widening political cleavages, as well as ethnic and religious splits.

To accomplish this, there are several digital tools that are well-known from the online commercial world: "avatars", "bots" and "trolls". An avatar is a fictional digital character that appears on the Net and pretends to be real. For instance, fictitious users can be found on social networks equipped with a name, profile, and pictures that seem real but, in reality, are imaginary figures whose purpose is to promote certain messages.

A bot is a software application designed to perform actions online by mimicking a normal user; that is, a kind of robot that poses as a human user. It is possible to set up a bot network (botnet), meaning a network of dozens or even thousands of fake and automatic profiles. Such bots can be programmed to create greater exposure to certain posts (or more likes, shares, and comments), thereby affecting the algorithm that promotes them on the social network and increases exposure. As was recently reported, the defense establishment in Israel estimated that approximately 30% (!) of all discourse on social networks is the product of bots.¹

A troll is a user whose entire purpose is to provoke and inflame the discourse. A troll will write controversial, false, or slanderous posts in order to promote interest. A troll can be a real user, but also an avatar or a bot.

Already in 2017, studies revealed that 50 million accounts on Twitter are actually automatically run by bot software. On Facebook, social bots—accounts run by automated software that mimic real users or work to communicate particular information streams—can be used to automate group pages and spread political advertisements. A public disclosure from Facebook revealed that a Russian "troll farm" with close ties to the Kremlin had spent around \$100,000 on ads ahead of the 2016 US election and produced thousands of organic posts that spread across Facebook and Instagram. The same firm, the Internet Research Agency, has been known to make widespread use of bots in its attempts to manipulate public opinion through social media.²

An online attack using these means against a foreign state or in one's own system, election campaigns, or referendums (e.g., doubts whether the campaign for Brexit referendum has been swamped by bots) has many advantages for the attackers. First, because it allows remote interference at relatively low cost, with complete anonymity. A computer and keyboard are sufficient. Second, because it also allows small groups, organizations, and individuals to conduct effective attacks. Third, because such attacks undermine the social order, corrupt the political discourse, and undermine public confidence in the democratic system and the political structure as

¹ Inai, Itay, "Invasion of the Political Bots", *Yedioth Ahronoth*, weekend supplement, January 18, 2019, pp. 22-30

² "The Bots That Are Changing Politics", *Motherboard*, at: https://motherboard.vice.com/en_us/article/mb37k4/twitter-facebook-google-bots-misinformation-changing-politics

a whole. Hostile elements, including terrorist organizations and their state sponsors, are likely to use avatars, bots, and trolls to harm the election campaign by deliberately defacing public discourse, as revealed by a recent report on digital terrorism and the interference in the Israeli election campaign of 2019 by terrorist organizations and state sponsors of terrorism.³

Interference in Election Campaigns

The possibility of online intrusion into the politics of foreign nations is highlighted, as one would expect, during election campaigns. About two years ago, following clear evidence of Russian interference in the US presidential election campaign, the US government informed dozens of countries around the world that their election campaigns were a target for the "Russian government's cyber activists". There is no doubt that the Russians have used online measures to interfere in the election campaigns of most Western countries, including the US, Germany, Britain, France, Italy, Montenegro, and more. Evidence of this led even Vladimir Putin to admit half-heartedly the possibility that, according to him, patriotic hackers from Russia had carried out cyber-attacks "against countries that had strained relations with Moscow and did so on their own initiative". However, US intelligence authorities thought differently: the US Justice Department filed indictments against 12 Russian intelligence officers, all of whom were employees of the Russian government, on suspicion of acting under its guidance to skew the US elections.

A report prepared for the US Senate Select Committee on Intelligence exposed the Russian disinformation campaign, which included millions of posts on social networks designed to influence the 2016 presidential election. Russian agents exploited all possible social networks, including Twitter, Facebook, and YouTube, to influence the online discourse surrounding Donald Trump's candidacy. The goal of the Russians, according to the report, was to confuse, distract, and influence voters. The Kremlin's efforts were spearheaded by the Internet Research Agency, a Russian governmental body that has published posts on issues such as race, immigration, and

³ Weimann, Gabriel, 2019. "Digital Terrorism in the Israeli Election Campaign?", *International Institute for Counter-Terrorism (ICT), Special Report*, at: [file:///C:/Users/gweimann/Downloads/Digital%20Terrorism%20Israeli%20Election%20Campaign%20\(1\).pdf](file:///C:/Users/gweimann/Downloads/Digital%20Terrorism%20Israeli%20Election%20Campaign%20(1).pdf)

weaponry, in order to sow disputes and divisions among American voters. The Russian agents, some of whom have already had indictments filed against them in the US for criminal interference in the elections, divided the American electorate into key groups and conveyed specific messages to each group.⁴

In Israel, too, fear of Russian interference in the 2019 election campaign increased, as expressed in a statement by the head of the Shin Bet, Nadav Argaman, about a "foreign power" interfering in the Israeli political system via the Internet.⁵ Attention, of course, was focused on Russia as, through a number of organized attacks, the Russians have demonstrated impressive capabilities of infiltration, hacking, interference and disruption using online means. Russian hackers, on behalf of Putin, have already attacked Estonia (2007) and Georgia (2009), hacked into the election headquarters of the Democratic Party in the US, planted advanced viruses within the e-mail systems of government organizations in Germany, as well as in the German parliament and in the offices of Chancellor Angela Merkel's party, waged a cyber-attack against election campaigns in many countries, including the Ukraine (2014), the US (2016), France (2017), Germany (2017) and Holland (2017), and interfered in referendums in Britain, Holland, Italy, and Spain (2017). In these attacks, they combined "hard" measures (such as disrupting the computer systems of election administration, hacking campaign managers' computers, leaking party data, etc.) and "soft" measures, mainly including penetrating and influencing the public discourse by invading social networks.⁶

During the 2019 elections in Israel, an alarming rise in fake accounts and false identities was revealed by an Israeli watchdog group. The watchdog group, the Big Bots Project, an independent organization that aims to expose the malicious use of social media, found a network

⁴ "Five Takeaways From New Reports on Russia's Social Media Operations", *New York Times*, December 17, 2018, <https://www.nytimes.com/2018/12/17/us/politics/takeaways-russia-social-media-operations.html>

⁵ On July 7, 2018, the head of the Shin Bet, Nadav Argaman, appeared at a conference of the Friends of Tel Aviv University, where he spoke of a "move that could affect the results of the upcoming election" due to be held on April 9, and warned of "interference by a foreign country". See: <https://www.ynet.co.il/articles/0,7340,L-5443404,00.html>

⁶ Weimann, Gabriel. "Will Putin Also Interfere in the Election Campaign in Israel?", *Ha'ayin Ha'shvi'it*, December 18, 2018, <https://www.the7eye.org.il/313117>; "Senate reports show scope of Russian election meddling", *CNN Politics*, December 2018, <https://edition.cnn.com/videos/politics/2018/12/17/russia-2016-report-social-media-ath-vpx.cnn>

of hundreds of social media accounts, many of them fake, used to smear opponents of Prime Minister Benjamin Netanyahu in the April 2019 election and to amplify the messages of his Likud party. According to the report, “The network operates through manipulation, slander, lies and spreading rumors. On its busiest days, the network sends out thousands of tweets a day.”⁷

Spreading Hate

Elections campaigns draw digital attacks and consequently media attention. But less noted and yet much more powerful are the almost routine attacks on political discourse. Challenges to democracy posed by fake proxies in social media surfaced in the European elections in 2017 and in the Baltic states and Ukraine. The NATO Stratcom Center of Excellence in Latvia has tracked Russian operations over the years, and according to its director, Janis Sarts, the use of social media as a military tactic is part of the so-called Gerasimov doctrine, named after the current chief of Russia's General Staff. The goal is permanent unrest and chaos within an enemy state. "This was where you first saw the troll factories running the shifts of people whose task is using social media to micro-target people on specific messaging and spreading fake news. And then in different countries, they tend to look at where the vulnerability is. Is it minority, is it migration, is it corruption, is it social inequality? And then you go and exploit it. And increasingly the shift is towards the robotization of the trolling."⁸

Fake proxies are used, in a growing rate to spread hate among groups. Take, for instance, antisemitism or Islamophobia. Antisemitism and hate crimes have surged in the U.S. over the past couple of years and bots were very active in promoting them. According to a study by the Anti-Defamation League (ADL), almost 30 percent of accounts repeatedly tweeting against Jews on

⁷ Bergman, Ronen, “Twitter Network Uses Fake Accounts to Promote Netanyahu, Israel Watchdog Finds”, *New York Times*, March 31, 2019, <https://www.nytimes.com/2019/03/31/world/middleeast/netanyahu-fake-twitter.html>

⁸ Quoted by [Bob Abeshouse](#), 2018, “Troll factories, bots and fake news: Inside the Wild West of social media,” at: <https://www.aljazeera.com/blogs/americas/2018/02/troll-factories-bots-fake-news-wild-west-social-media-180207061815575.html>

Twitter appear to be bots.⁹ The ADL analyzed 7.5 million Twitter messages between Aug. 31 and Sept. 17, 2018. The study reports that while human users still account for the majority of derogatory Twitter traffic in the lead-up to the midterm elections, "political bots—which explicitly focus on political communication online—are playing a significant role in artificially amplifying derogatory content over Twitter about Jewish people".

Bots and fake actors appear to be used as well for anti-Muslim online campaigns. As a recent study revealed, online anonymous bots are helping to spread and amplify Islamophobia. Researchers from the non-profit organization *Hope not Hate* monitored several anti-Muslim figures and blogs to see how they used bots, image manipulation and fake news to increase the size of their audiences.¹⁰ One person who benefited from fake accounts was Pamela Geller, whose views were found to be automatically propagated on Twitter by 102 bots. Her blog, the Geller Report, reportedly doubled its monthly audience to two million viewers.

One common tactic is to manipulate or digitally alter images to express Islamophobic views. After the Westminster attack in March 2017, far-right activists shared an image of a woman in a hijab appearing to walk away from a victim across Westminster Bridge. The Twitter user @Southlonestar who first used the image to express Islamophobic sentiment was revealed to be a fake account created in Russia, designed to influence UK politics. The *Hope not Hate* research reportedly said that bots tend to be recognizable as anonymous accounts, tweeting and sharing the same content at the same time. While simple bots follow and retweet other users, helping those they follow to appear more legitimate, more sophisticated bots can be difficult to detect.

Several studies revealed how a global network of anti-Muslim activists is using Twitter bots, fake news, and the manipulation of images to influence political discourse. An article entitled "Hate in a Tweet: Exploring Internet-Based Islamophobic Discourses" describes the spread of Internet-based racism that often contributes to diffusing an image of Islam as incompatible with

⁹ "Computational Propaganda, Jewish-Americans and the 2018 Midterms: The Amplification of Anti-Semitic Harassment Online", ADL Center for Technology and Society, at: <https://www.adl.org/resources/reports/computational-propaganda-jewish-americans-and-the-2018-midterms-the-amplification>

¹⁰ "Islamophobia on the internet 'amplified by army of bots'", *The Independent*, November 26, 2017. At: <https://www.independent.co.uk/news/uk/home-news/bots-online-islamophobia-pamela-geller-tommy-robinson-twitter-terror-attacks-hope-not-hate-a8076691.html>

Western values. Bots and fake actors promoted the spread of Islamophobic discourses on social networks such as Twitter.

The data analyzed in this article suggest that specific characteristics of online Islamophobia are rather linked to three other elements: the global connections allowed by the Internet, the proliferation of certain actors (“trolls” and “bots”), and the circulation of so-called “fake news.” The conclusion is rather clear: “The presence of trolls and bots perpetuating Islamophobic narratives is a characteristic of social networks and risks decisively contributing to the visibility of anti-Islam discourses. These actors might be partially responsible for the high number of Islamophobic post-Brexit tweets, and their messages can influence the perception of Islam of all Twitter users that do not recognize them for trolls and bots”.¹¹

The spread of hate may result in real violence: the case of Myanmar is alarming evidence revealing the power of disinformation online and its brutal consequences. The unrest in Myanmar, very often violent, between minority Muslims and Buddhists in the majority-Buddhist country of around 51 million is fueled by online rumors and fake news. Waves of violence broke out in the country, targeting mainly Rohingya Muslims. Hate speech and fake user pages were pervasive in Myanmar, causing the deadly attacks on the Muslim minority. The UN criticized Facebook’s conduct in the Myanmar crisis, which the US says, “bears the hallmarks of genocide,” by serving as a platform for hate speech and disinformation, and that Facebook had “turned into a beast.”¹²

Spreading Chaos and Extremism

While much attention has been paid to the 2016 Russian campaign to boost Donald Trump’s presidential run, much less attention has been devoted to the ongoing disinformation campaigns online. “What we see is that the Russians are still involved in trying to sway the political narrative

¹¹ “Hate in a Tweet: Exploring Internet-Based Islamophobic Discourses”, *Religions* 2018, at: <https://www.mdpi.com/2077-1444/9/10/307/htm>

¹² Quoted in “How Facebook’s Rise Fueled Chaos and Confusion in Myanmar”, *Wired*, July 6, 2018, at: <https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/>

in the United States,” argues Priscilla Moriuchi, a former National Security Agency official and the head of strategic threat development at Recorded Future, a cybersecurity firm. “It works to their advantage that we don’t know as much about it this time around.”¹³ It appears that this time around, Russian trolls, bots, and propagandists appear to have switched their tactics.

In September 2018, New Knowledge, a company tracking disinformation campaigns, began tracking a network of social media accounts that appeared to be coordinating messages to disseminate Russian propaganda and advance Russian interests. The network included about 100 Facebook pages and 1,400 Twitter accounts that posted between 50,000 and 60,000 times a day! To the casual observer, they look similar to any other online news outlet, providing a way to push pro-Russian misinformation into the mainstream. However, New Knowledge discovered the network by first identifying a set of accounts that appeared to be engaging in propaganda activity and then looking for similarities between those accounts and others.

The use of trolls and bots is done mainly for political motives mostly negative. Foreign governments and groups can promote certain messages and narratives (generally fake) in order to sow chaos and discord in the targeted state. For example, after the April 2018 joint missile strikes on Syria by the U.S., the U.K. and France, the Pentagon reported that it saw a 2,000 percent increase in activity by Russian “trolls.”¹⁴ For years, the Russian online propaganda machine used a network of bots and fake accounts to pump out a steady stream of digital disinformation aimed at destabilizing enemies abroad. Now, since Trump’s election, studies report, the Kremlin has doubled down on its dissemination of fake news.¹⁵ Sometimes the stories are completely made up. More often, they are simply misleading or biased, tidbits of real reporting repackaged to serve Russian goals.

Russia may have started the trend but it has been joined by dozens of states. A 2017 study by the human rights group, Freedom House, concluded that 30 foreign governments are now

¹³ Quoted in “Battling the Bots”, *Foreign Policy*, November 12, 2018, at:

<https://foreignpolicy.com/2018/11/12/battling-the-bots-ai-russia-disinformation-fake-news/>

¹⁴ “Russian Trolls Increased ‘2,000 Percent’ After Syria Attack, Pentagon Says”, *Newsweek*, April 14, 2018. At: <https://www.newsweek.com/russian-trolls-increased-2000-percent-after-syria-attack-pentagon-says-886248>

¹⁵ “Russia Has Weaponized Fake News to Sow Chaos”, *The New Republic*, May 12, 2017, at: <https://newrepublic.com/article/142344/russia-weaponized-fake-news-sow-chaos>

using online propaganda proxies such as bots, to manipulate and distort information online.¹⁶ The study revealed that trolls and bots were used to spread false news and propaganda, concluding that “The use of paid commentators and political bots to spread government propaganda was pioneered by China and Russia but has now gone global”.

What can be done?

The spread of fake online proxies, namely bots, avatars and fake identities, is a serious risk to the integrity of the democratic system. Since they currently represent two-thirds of the activity on Twitter, the bots may well be the ones who get to decide: see this, not that or believe this, not that. The terms “truth decay” or “death of truth” have joined the post-truth lexicon that includes such now familiar phrases as “fake news” and “alternative facts”. These are promoted by fake actors and fake followers on social media generated by bots. This is the worst challenge to free speech, caused by the vicious displacement of truth and reason by lies, rumors, and emotions.

How can we challenge the threat of virtual proxies? The answer combines several measures because an easy, one-measure approach is not feasible. Since Israel is one of the most attacked societies in the world by virtual actors, some of the potential solutions proposed here have been used and tested in Israel. The first is the monitoring and identification of this malicious activity. Detection of phony accounts is largely based on algorithms using artificial intelligence to assess irregular behavioral patterns. In January 2019, it was reported that Iranians have been using hundreds of fake accounts on Israeli social media pages in an effort to sow social division and influence the then upcoming Israeli general elections.

Several companies and organizations united in the effort to track down this activity. One of these was the company Vocativ. This monitoring company, founded by Israeli entrepreneur Mati Kochavi, noted an increasing presence of bots operating from Iran that automatically respond to events by posting on social media in an effort to influence the discourse in Israel. The company contacted the social networks and asked them to stop what appears like an organized Iranian move.

¹⁶ “More governments manipulate media with 'bots,' trolls: study”, *Phys.Org. report*, November 14, 2017. at: <https://phys.org/news/2017-11-media-bots-trolls.html>

Vocativ used an intelligence system to scan the web to search for keywords based on the geographic location they originated from, as well as ties between bots, leading them to the fake profiles. Their algorithm determined whether the profile was real or fake by, *inter alia*, examining whether the profile was posting in regular intervals or at an unusual pace. The algorithm could also uncover who was behind the bots, with most of the tweets and posts posted by the fake accounts traced back to an Iranian English-language website called Countdown 2040, which claims Israel will cease to exist by 2040.

The monitoring also revealed that the bots were initially used to enflame the Israeli-Palestinian conflict but, once elections were announced, they were adjusted to focus on influencing their outcome. The analysis noted a 783% increase in bot account activity after elections were announced. Of the 350 Iranian bot accounts identified, only 50 had been active prior to the elections announcement.

The second measure involves removal or blocking the identified malicious actors. This can be done mostly by involving the social media firms. As suggested by Mounk in his 2018 book, *The People vs. Democracy*, “We have to recognize that platforms like Facebook and Twitter can do a lot to stop the spread of fake news or hate speech without going all the way to outright censorship”.¹⁷ Facing the threat of using social media as channels for interfering with the Israeli elections in 2019, Judge Hanan Melcer, the chairman of Israel’s Central Elections Committee and a veteran justice on the Supreme Court, summoned representatives from major U.S. social media and technology companies. He told them that he expected them to play a role in curbing online deception during the country’s election. “You say you’ve learned from 2016,” Melcer told them, according to a government official who was present. “Prove it!”

Indeed, these platforms had already started to scan postings for such proxies’ activities. In the recent Israeli elections, Facebook, for example, noted that it removed fraudulent and automatic accounts (bots) from the network and the pages of the parties’ candidates. Facebook also offered several tools to those dealing with media for reporting networks of fraudulent users. Several other platforms have taken similar measures in the past two years in election campaigns throughout the

¹⁷ Mounk, Y. (2018). *The People vs. Democracy: Why Our Freedom Is in Danger and How to Save it*. Harvard University Press, p. 239.

world, following severe public criticism regarding the 2016 US presidential campaign. Beside public pressure, an additional motive for such counter-activity is that the bots and avatars are also directly costing these companies revenue. In contrast to the purchase of advertising to influence surfers, they earn nothing from the activity conducted through fake accounts. After Israel's election in April 2019, it was estimated that the number of fake messages related to the campaign had fallen by more than fifty per cent when compared to the 3 months before the elections. "Israel went first. Everybody was looking to see how we handled it," a government official told *The New Yorker*, "We set a precedent and it may be a model."¹⁸

The calls for regulating social media are matched by equally forceful calls by those who fear the return to old-style censorship. Yet, as Mounk argues, "Since a wide gulf separates the sides of this debate, it would be tempting to think that we are faced with two equally unappealing alternatives: intrusive regulation or outright censorship on the one side; inaction and fatalism on the other".¹⁹ However, in reality there are some pragmatic and plausible alternatives to these extremes. Cooperation between the public and private sectors in the struggle against fake proxies is a key to the success of the counter-measures suggested above. With such cooperation, the "golden path", minimizing the threat to free speech and yet maximizing the removal of malicious online content, is the ideal model.

¹⁸ "How Israel Limited Online Deception During Its Election", *The New Yorker*, April 11, 2019, at: <https://www.newyorker.com/news/news-desk/how-israel-limited-online-deception-during-its-election>

¹⁹ Mounk, *Ibid.*, p. 238.