Weaponized skepticism:
An analysis of social media deception as applied political epistemology

Regina Rini (rarini@yorku.ca)

Abstract: Since at least 2016, many have worried that social media enables authoritarians to meddle in democratic politics. The concern is that trolls and bots amplify deceptive content. In this chapter I argue that these tactics have a more insidious anti-democratic purpose. Lies implanted in democratic discourse by authoritarians are often *intended* to be caught. Their primary goal is not to successfully deceive, but rather to undermine the democratic value of testimony. In well-functioning democracies, our mutual reliance on testimony also generates a distinctively democratic value: decentralized testimonial networks evade control by the state or powerful actors. The deliberate exposure of deception in democratic testimonial networks undermines this value by implicating citizens in their own epistemic corruption, weakening the resilience of democratic society against authoritarian pressure.   In this chapter I illustrate that danger through a close reading of recent Russian social media interference operations, showing both their epistemic underpinnings and their ongoing political threat.

Keywords: testimony, social media, fake news, weaponized skepticism, political interference

There's more than one way to skin a democracy. Back during the Cold War, the apparent threat was sneaky: a dagger within a cloak, a hypnotized Manchurian candidacy. Those who meant to undermine democratic societies would first have to slip our defenses. We spent millions on intelligence agencies investigating (often spurious) allegations of clandestine propaganda. We thought: so long as we catch them red-handed, they cannot be effective.

Since at least 2014, anti-democratic actors have used social media to distribute fake news and disingenuous provocations with the aim of affecting the course of democratic elections. Their

most famous interventions came in the 2016 UK Brexit vote and US presidential election. It is hard to establish whether this interference played a decisive role in either vote, but few dispute that it had some effect. According to the United States Justice Department, the goal was to "spread distrust toward the candidates and the political system in general".[1]

If we follow the old Cold War mentality, the solution to this problem is obvious: stop the perpetrators. Now that we know they are doing it, we must simply be alert to the next attempt. Shut down the fake news vectors. Educate our citizens about the epistemic risks. Knowing that the problem exists is the first step toward solving it.

But this mentality underestimates the destructive power of social media deception. As I will show, those who turned fake news against democracy may have done so *expecting and wanting to get caught*. The subversive effects of their interference may be most powerful when they are not clandestine; it is our *awareness* of the pervasiveness of fake news that does maximal damage to democratic discourse.

The basic idea is simple: democratic practice requires that citizens trust one another, and awareness of pervasive fake news erodes this trust. This means that democratic political culture is *distinctively* vulnerable to the exposure of fake news transmitted through social media, much more so than politics in non-democratic societies. By participating in defective testimonial chains, citizens become complicit in their own epistemic victimization. And once citizens come to realize this fact, they reasonably begin to distrust one another's competence as co-participants in the collective epistemic projects that make democratic culture possible.

---

[1] *United States of America v. Internet Research Agency* federal indictment (2018, p. 6).

In this essay I aim to illuminate the politico-epistemic damage wrought by Russian social media operations, primarily in the years 2014-2016. (I wrote this chapter in late 2018 and revised it in early 2020; by the time it is published, the 2020 US presidential election will surely have turned up new complications that it cannot anticipate.) In the first half I present an analytically focused narrative of these events, drawing out the malefactors' apparent indifference to disguising their work. In the second half I offer a philosophical analysis of the social epistemic environment that made this possible – which is now being dangerously warped by its effects. The point is both to appreciate the troubling specificity of Russia's attack on democratic culture, and also to draw general lessons about the new epistemic vulnerabilities created by social media.

As we will see, the ultimate goal of anti-democratic epistemic interference is not simply to trick people. The point is not to implant false beliefs. Rather, the point is to *induce skepticism*. By flooding the channels of public discourse with falsehood, then allowing citizens to know that this has happened, anti-democrats make it reasonable for us to trust no one, least of all our co-citizens. That is the story of how weaponized skepticism slices through democratic culture – not down some dark alley, but out where everyone can watch it happen.

Project Lakhta

In late May 2016, an American citizen posed for a photo in front of the White House with a large sign reading "Happy 55th Birthday Dear Boss!" This person was doing a favor for an online political group, who explained only that the sign was meant for "our boss… our funder". Three days later, in St. Petersburg, Yegveniy Viktorovich Prigozhin turned 55.

Prigozhin was the director of Concord Management and Consulting LLC, a Russian-government-linked services provider. The company allegedly shifted millions of dollars to support Project Lakhta, a political interference operation working domestically and in several foreign countries. One of its beneficiaries was the St. Petersburg Internet Research Agency (IRA), which created the fake online group responsible for dispatching that White House sign-bearer.[2]

Of course that was not all it did. In February 2018, Prigozhin and 12 other Russian nationals were indicted by Special Counsel Robert Mueller for conspiracy to defraud the United States. According to the indictment, since at least 2014 the IRA had been using Social Security and date-of-birth information to steal the identities of unwitting American citizens. With these identities, the IRA created hundreds of email accounts, web sites, and social media personae. The fakes were then were used to disseminate false and provocative news stories, to organize divisive political rallies in the physical world, and to coordinate with mostly unwitting American political operators, including members of the Trump campaign. The fake entities ranged across the political spectrum, including an Instagram account called "Woke Blacks", a Twitter account purporting to be the Tennessee Republican Party (@TEN_GOP), and an improbable Facebook community called "Stop A.I." (which, despite its name, was concerned with alleged Clinton

---

[2] *US v. Internet Research Agency*, (2018, p 7-8). Prigozhin is also widely believed to direct the Wagner Group, a Russian mercenary force that is allegedly responsible for stirring conflicts in Ukraine, Syria, and Libya and allegedly murdered three journalists in the Central African Republic. See Lister and Shukla (2019).

voter fraud rather than the coming robot apocalypse).[3] Together these accounts had hundreds of

thousands of social media followers.[4]

All of this happened during the then-little-appreciated rise of fake news on social media.[5] Not all

fake news came from Russian intelligence operatives, of course; some was produced by

disingenuous domestic political operators. Others were just in it for the money, like some now-

infamous Macedonian teenagers, or the click-bait fake headlines that 20th Century Fox used to

promote a new film.[6]

But the IRA was very much in the business of political manipulation. According to the Prigozhin

et al. indictment, its agents were given the instruction, "use any opportunity to criticize Hillary

and the rest (except Sanders and Trump – we support them)." Six weeks before the election, an

operative running the fake "Secured Borders" Facebook page was scolded for their "low number

of posts dedicated to criticizing Hillary Clinton". On November 3, three days before the election,

the fake Instagram account "Blacktivist" declared, "Choose peace and vote for [third party

candidate] Jill Stein. Trust me, it's not a wasted vote." [7]

Though humans created this deceptive content, it was virally amplified by 'bots', social media

accounts operated by software and programmed to automatically disseminate messages. A study

by Chengcheng Shao and colleagues used mathematical network modeling to conclude that

Twitter disinformation typically propagates outward from a core of non-human accounts that

---

[3] "Stop A.I." stood for "Stop All Invaders". Accounts linked to the group continued to operate into 2017, when they reacted to media reports of Russian interference by posting: "Instead this stupid witch hunt on Trump, media should investigate this traitor [Obama] and his plane [sic] to Islamize our country." *US v. Khusyaynova* criminal complaint, (2018, p. 28).
[4] *US v. Internet Research Agency*, (2018, p. 15).
[5] Silverman and Alexander (2016); Vosoughi, Roy and Aral (2018).
[6] Silverman and Alexander (2016); Rainey (2017).
[7] *US v. Internet Research Agency*, (2018, p. 17-18).

continuously re-tweet one another.[8] Though Twitter purged tens of millions of bot accounts in

2018, many remain active.[9] The Alliance for Securing Democracy, a NATO-funded body,

operates a web dashboard called Hamilton 68, which tracks the evolving focus of Russian-linked

Twitter accounts.[10]

It's difficult to say how much these activities matter to election results. After all, individual

voters are themselves often unsure about exactly when or how they made their decision; they

may not remember the source of the information that pushed them over the edge. But it seems

hard to deny that, in the 2016 US president contest, there was *some* effect. Kathleen Hall

Jamieson, a media scholar with a non-partisan reputation who has studied every presidential

election since 1976, told the *New Yorker* that she thinks Russian digital disinformation

"influenced who voted, or didn't vote, and *that* could have changed the outcome."[11]

But my point does not turn on whether Russia succeeded in throwing the election to Donald

Trump. My point is that Russia's interference aimed and succeeded at weakening American

democracy merely by playing a role in US debates - and indeed by being *caught* playing that

role.

What's remarkable about Project Lakhta is that, for the intelligence operation of a government

headed by a former KGB agent, it didn't do much to cover its tracks. Hiring a person to stand in

front of the White House with a frivolous message to the conspiracy's leader does not look like

good tradecraft. In fact it looks more like trolling, the sort that is meant to be noticed.

---

[8] Shao et al. (2018).

[9] Dellinger (2018).

[10] You can find Hamilton 68 at https://dashboard.securingdemocracy.org/

[11] Mayer (2019). See also Jamieson's (2018) book, *Cyberwar: How Russian Hackers and Trolls Helped Elect a President – What We Don't, Can't, and Do Know*.

Several pieces of evidence suggest the IRA made little effort to go undetected. The pseudo-American accounts' activity patterns correlated with St. Petersburg office hours. If subterfuge were a priority, presumably they could have tasked espionage professionals to operate overnight. Instead they hired ordinary people who expected to work ordinary schedules. Indeed, much of the public knowledge of the IRA's operations comes from the testimony of Lyudmila Savchuk, a former employee who sued the agency in 2015 for abusive labor practices such as successive 12 hour shifts.[12] That Savchuk was hired in the first place suggests minimal operational security; she has a history as a political activist and says she took the IRA job intending to write an exposé.[13]

After Savchuk's lawsuit, office hours seem to have been standardized. In October 2017 one conspirator allegedly gave this advice to colleagues:

> Posting can be problematic due to time difference, but if you make your re-posts in the morning St. Petersburg time, it works well with liberals – LGBT groups are often active at night. Also, the conservative can view your re-post when they wake up in the morning if you post if before you leave in the evening St. Petersburg time.[14]

Perhaps most tellingly, the IRA operations continued, mostly unchanged, even after the operatives learned they were being monitored by American intelligence. Media reports about the investigation began circulating in autumn 2017. On September 13, an IRA operative emailed her family: "We had a slight crisis here at work: the FBI busted our activity (not a joke). So, I got preoccupied with covering tracks together with the colleagues."[15] Yet they continued to use the

---

[12] Chen (2015).
[13] Jones (2018).
[14] *United States v Khusyaynova* (2018, P. 14).
[15] *United States v. Internet Research Agency*, (2018, 24).

same tactics. The advice to time postings within St Petersburg office hours was sent a month *after* the IRA learned the FBI was on to them.

Though Facebook and Twitter shut down the fake accounts used during the 2016 election, the IRA created many more. A September 2018 criminal complaint against Elena Khusyaynova, the IRA's accountant, provides evidence of new fake profiles as recently as May 2018, months after the original indictment was issued.[16] There is every reason to believe these activities continued into the 2018 US midterm campaign, though the end of the Mueller probe made it more difficult to get public information. Presumably similar plans are in place for the 2020 US presidential campaign.

I cannot definitively establish that those behind Project Lakhta intended to get caught. But the informality of their operations and their persistence after being detected are consistent with this theory. Further, as I'll now show, being seen to easily manipulate American voters would accomplish one of the key ideological goals of the Putin regime.

Ivan Ilyin and the humiliation of democracy

Ivan Ilyin was a Russian nationalist philosopher, born in 1883 and exiled in 1922 after opposing the Bolshevik Revolution. Living in Berlin through the 1920s and 30s, Ilyin came to praise authoritarianism. But his most central views were negative. He saw democracy as a fundamentally corrupting form of government, destined to be brought down by its own hypocrisies. Ilyin blamed ineffectual Russian democratization in the late czarist era for allowing the Bolsheviks to come to power and he saw strongman leadership as the best guard against the

---

[16] *United States. Khusyaynova*, (2018, 38).

social corruption of democracy. Ilyin told his readers: "The principle of democracy is the irresponsible human atom."[17]

Ilyin died in 1954 and was nearly forgotten. However, as historian Timothy Snyder has documented, from the 1990s Vladimir Putin reanimated Ilyin's ideas to guide his ideological reshaping of Russian political society. Putin regularly quotes Ilyin in national addresses. In 2006 he had Ilyin's archived papers returned to Russia; in 2005 he brought Ilyin's bodily remains home from Switzerland. By 2014, Ilyin's collected works were being distributed to Russia's top civil servants.[18]

Snyder interprets Putin-following-Ilyin as a practitioner of what he calls the "politics of eternity". This "places one nation at the center of a cyclical story of victimhood. Time is no longer a line into the future, but a circle that endlessly returns the same threats from the past."[19] Unlike in democratic narratives of progress, the politics of eternity does not promise that things will get better. Instead, the relentless focus is on holding off external threats of corruption and decay. In its Ilyin-infused Russian form, according to Snyder, the politics of eternity acquires a mystical element, as a heroic authoritarian leader arises to ward off the "seduction" of corrupt forms of government.[20]

Most important for my purposes is Ilyin's analysis of the failure of democracy. Ilyin interpreted democratic traditions of free speech and individual rights as vehicles for cynical or deluded self-interest:

---

[17] Ivan Ilyin, *Kosciol i wladza*, as quoted in Synder (2018, p. 27).
[18] Snyder (2018, p. 17-18).
[19] Snyder (2018, p. 8).
[20] Snyder (2018, p. 22-29).

A citizen is given a limitless right to secret self-seduction, corruption of others, and an imperceptible self-commodification. He enjoys the freedom of insincere, deceitful, treacherous, insinuating speech and an ambiguous but calculated suppression of truth. He is free to "trust" liars and scoundrels or at least pretend to do so, selfishly feigning this or that political persuasion. To freely express all this spiritual temptation, he is provided with an "election ballot".

Not only are democratic institutions vicious and corrupting, according to Ilyin, they are also *vulnerable*. Since democratic governments fail to regulate their citizen's speech, they can do nothing to stop the rot from spreading:

The freedom of opinion is to be absolute, state functionaries dare not subvert and diminish it. The most idiotic, harmful, disastrous, and disgusting of "opinions" are "untouchable" only due to the fact that a certain mischievous idiot or traitor took the time to express them, hiding behind their "untouchable" nature.

Finally, this vulnerability leads to the self-destructive failure of democracy:

This order will endure until the seduction undermines the very idea of voting and the very eagerness to submit to a majority… then the ballot will be supplanted by an uprising, and an organized totalitarian minority will seize power.[21]

It's important to stress that, for Ilyin, this is not simply a *prediction* of democracy's uselessness. It is also an *agenda* for wise authoritarians who find themselves trapped amid democratic corruption. Democracy will inevitably fall – so the advice is to give it a push in the right direction, to make it fall your way, on your terms, at your time.

---

[21] Ivan Ilyin, 'On Formal Democracy'. As printed in his *Our Tasks (Nashi Zadachi)* (1948). There is no public English translation of this book, and even the Russian original is hard to find outside Russia. The excerpts here are from a bespoke translation prepared for me by a Russian academic who prefers to be identified as Fluctuarius Argenteus. The digital source text (in Russian) appears at: http://apocalypse.orthodoxy.ru/problems/094.htm

Putin followed this advice throughout his consolidation of power in the early 2000s. Post-Soviet Russia, riddled with oligarchic factionalism, was never a fully mature democracy. But under Boris Yeltsin it held genuine elections and permitted a relatively free media. After Putin came to power in 1999, these democratic trappings were rapidly subverted. Through selective prosecution for corruption, Putin exposed his political opponents as self-interested pretenders – the liars elected by democracy. He supported witchhunts against independent media and NGOs, which were accused of spreading a western homosexual agenda. After 2010, the state media began to tell Russians that protests against growing authoritarianism were organized by pedophiles and foreign spies.[22]

Fake news was a key element in Putin's extirpation of Russian democracy. As Snyder puts it, the purpose was "both to spread confusion about a particular event and to discredit journalism as such. Eternity politicians first spread fake news themselves, then claim that all news is fake, and finally that only their spectacle is real."[23]

The social media bomb

By 2014, the Putin regime had begun exporting these tactics. After a popular revolution overthrew Ukraine's pro-Kremlin president, Ukrainian and Russian Facebook users began reading false stories of anti-Russian atrocities in Ukraine. Many were later traced to Russian

---

[22] See Gessen (2017), especially chapter 16.
[23] Snyder (2018, p. 11).

operations.[24] The IRA's first major project was a propaganda leadup to the Russian invasion of Crimea and eastern Ukraine in February 2014.[25]

On June 23, 2014, Russian-linked forces in occupied eastern Ukraine shot down a civilian jet, Malaysia Airlines flight 17, killing all 298 people aboard. Over the following days, as Timothy Snyder has since reconstructed, Russian television and social media practiced a sophisticated form of disinformation: instead of flatly denying Russian responsibility, they spun multiple conflicting alternative stories. Russian media claimed the surface-to-air missile came from Ukrainian forces and had in fact been aimed at Vladimir Putin's own plane (which was nowhere near the attack). Then they claimed that Ukrainian fighter jets accidentally shot down MH17 due to a deceptive conspiracy among Ukrainian air traffic controllers. Then another official admitted that Russia *had* shot down the plane – but said there was no atrocity, since the plane contained only already-dead corpses planted by the CIA. Obviously these stories could not all be true at once. Yet, as Snyder puts it, "even if all of these lies could not make a coherent story, they could at least break a story – one that happened to be true."[26]

This would become a regular pattern of Russian interference in other nations' politics. The point was not to gain belief in any one falsehood, but to saturate the epistemic environment with conflicting accounts so that the truth appears to be only one of many bickering narratives. The goal was confusion and disarray, generalized skepticism rather than focused false belief. As Michael Lynch and Zeynep Tufekci have noted, this distortion of the epistemic environment is a common authoritarian tactic.[27] The authoritarian leader establishes control over information by

---

[24] Priest, Jacoby, and Bourg (2018).
[25] Snyder (2018, p. 139).
[26] Snyder (2018, p. 182).
[27] Lynch (2016); Tufekci (2018).

spreading so much disinformation that citizens come to distrust *all* reporting. The authoritarian may not be able to guarantee positive belief in his lies, but he can hinder belief in the truth.

2014 was also the first year Russia employed disinformation tactics beyond its traditional sphere of influence. According to British cyber security expert Ben Nimmo, early IRA troll farms (including some of the same accounts spreading lies about Ukraine) attempted to interfere in the independence referendum in Scotland. When the referendum failed, these accounts claimed that the result was rigged and posted misleading videos of ballot stuffing (the footage was actually recorded during the 2012 Russian elections).[28]

By 2016, the IRA had built a large system of Twitter 'trolls' (humans posting maliciously under false identities), 'bots' (programmed accounts set to automatically redistribute content), and 'cyborgs' (accounts sometimes operated by human trolls and sometimes running as bots). Researchers at the University of Edinburgh have identified 419 IRA Twitter accounts that tweeted about the June 2016 Brexit referendum. Their posting rate spiked on the afternoon of the vote, but ceased at 4pm British time even though voting continued for another 6 hours; it was 7pm in St. Petersburg. Of the 398 Brexit-related tweets sent by the network on that date, just 9 consisted of original content. The rest were only amplification, mostly retweeting other members of the network.[29]

The Edinburgh researchers point out that the Brexit bot/troll network did not seem to be pushing any particular outcome; a content analysis showed that its tweets trended only slightly toward the Leave campaign. This lack of direction suggests, again, that the goal was incoherence rather than influence. Further, the researchers suggest that one explanation for the quick repurposing of

---

[28] Carrell (2017).
[29] Llewellyn et al. (2019, p. 1159).

these accounts (most had cover identities outside the UK and seemed to be designed for election interference in America and Germany) was opportunistic chaos: "Brexit as a controversial issue… provided a suitable topic for generalized disruptive tweeting."[30]

The largest known Russian digital interference operation was, of course, the 2016 US federal election. Sean Edgett, Acting General Counsel for Twitter, later told a Senate committee that Twitter had detected 36,746 automated accounts, totalling 1.4 million tweets in the two months before the election – slightly under 1% of *all* election-related tweets on the network.[31] Of these accounts, Twitter identified 2,752 as run directly from the St. Petersburg IRA. These seem to have been "cyborg" accounts, sometimes operated by a human agent, but around 47% of the time displaying automated posting behavior.[32] Twitter banned and deleted all these accounts after the election.

This seems to be a common pattern in social media interference operations. Researchers at City University of London examined 13,493 Twitter accounts that had opined on Brexit and then disappeared once the vote was complete. (The researchers do not speculate on how many of these accounts were operated from Russia, though they note that few seem to be UK-based.) Analyzing relations among automated accounts, the researchers concluded they were grouped into two distinct bot networks: one responsible for amplifying messages from human-operated accounts, and the other dedicated to simply retweeting other bots.[33]

---

[30] Llewellyn et al. (2019, p 1160-1).

[31] Testimony of Sean J. Edgett before the US Senate Judiciary Committee, Subcomittee on Crime and Terrorism. October 31, 2017. https://www.judiciary.senate.gov/imo/media/doc/10-31-17%20Edgett%20Testimony.pdf . p 9-10.

[32] Testimony of Sean J. Edgett, 12.

[33] Bastos and Mercea (2019, p 38).

Looking at the content of these tweets, the researchers found that more than 50% linked to

"perishable news", website addresses that would be deadlinks only six months later. Checking

the links that were still active, the researchers found content "that blurs the line between

traditional tabloid journalism and user-generated content, which is often anonymous, fact-free,

and places a strong emphasis on simplification and spectacularization".[34] Though the Brexit bots

did prefer the Leave side of the referendum, 17% of their tweets included the term "Remain",

higher than among human Brexit tweeters. In other words, the botnet gave support to both sides,

seeming to be more interested in sharpening partisanship and spreading shocking tabloid

headlines, rather than any particular outcome.[35]

A further sign that divisiveness was the primary goal of the US operation is that interference

continued even *after* the election. On November 12, 2016, four days after voting was finalized,

two political rallies were staged in Manhattan. One, organized by social media group

BlackMatters, was called the "Donald Trump is NOT my president" rally. The other, titled

"show your support for president-elect Donald Trump", was organized by the Being Patriotic

social media group. In fact, both BlackMatters and Being Patriotic were fake accounts operated

by the IRA.[36] The thousands of Americans who attended had no idea that they were being

orchestrated by the same foreign actors. With cinematic irony, the anti-Trump protesters

marched from Union Square up 5th Avenue, until outside Trump Tower they met police barriers

at the site of the pro-Trump rally.[37] It is hard to imagine a more symbolic manifestation of

---

[34] Bastos and Mercea (2019, 45).
[35] Bastos and Mercea (2019, 47).
[36] *United States v. IRA*, (2018, 23).
[37] Haskins (2017).

Russia's divisive achievement than this: the election already over, thousands of American citizens unwittingly steered into confrontation with one another on the streets of the largest city.

But this was not the end of it. The next twist of the knife came the following year, as American citizens began to understand what had been done to them. Reports of Russian cyber interference in the US had circulated as early as 2014, following publication of IRA documents stolen by a group of Russian hackers.[38] More than a year before the election, journalist Adrian Chen published an extensive investigation of the IRA's activities in the *New York Times Magazine*.[39] Still, most people did not pay much attention until autumn 2017, when Facebook and Twitter executives testifying before Congress acknowledged the scope of the problem. It was then that Facebook admitted as many as 126 *million* Americans may have seen content posted by the IRA.[40] Facebook created a tool allowing users to check whether they read IRA-linked material during the campaign.[41] Finally, in February 2018 the Project Lakhta indictments led to top headlines around the world.

More recent evidence is sketchy, but it is clear that Russian interference operations continued into America's 2018 midterm elections.[42] In July 2018 Facebook banned 32 pages and accounts for "coordinated inauthentic behavior" around hot-button social issues.[43] A month later it deleted 652 fake accounts linked to Russia or Iran.[44] In October 2018 Donald Trump announced he would manufacture a "Great Midterm issue for Republicans!" out of a caravan of Honduran

---

[38] Seddon (2014).
[39] Chen (2015).
[40] White (2017).
[41] Romm and Wagner (2017)..
[42] D'Souza (2018).
[43] Fandos and Roose (2018).
[44] Solon (2018).

refugees fleeing north through Mexico.[45] The following week social media filled with fake news about the caravan. According to one estimate, up to 60% of *all* caravan-related tweets were emitted by bots.[46]

But there is a difference from 2016. American social media users now *know* that they are being targeted by Russian interference operations. When they read a political post, they might wonder whether it contains fake news originating from the IRA, or whether the provocative stranger trying to engage them in debate is really a paid troll or even just a bot. This awareness has profound consequences for how citizens share information and conduct public debate.

Testimonial sabotage

To fully understand the political consequences of IRA interference, we need to examine the epistemology of social media. People use social media for many things, such as admiring baby pictures and cyber bullying. But a central use is information-gathering; in 2019, 55% of Americans said they often or sometimes got news from Facebook, Twitter, etc.[47] Even those who don't seek social media news will stumble across newsy stories while scrolling.

People who get news from social media are acquiring beliefs through what epistemologists call *testimony*. If I believe X because someone else said X, then my justification for believing X depends upon the reliability of this person. Much of our knowledge of the world comes about through testimony. For instance, I've never been to Singapore, but I know it is humid because

---

[45] @realDonaldTrump tweet, Oct 17 2018. https://twitter.com/realDonaldTrump/status/1052556222459727872 Accessed Oct 31 2018.
[46] Lapowsky (2018).
[47] Shearer and Grieco (2019).

others have testified to that. Though I haven't done the experiment myself, I know that bronze

melts at 950°C because others have testified to this happening.

Further, much of our testimonial knowledge relies upon being able to trust strangers or

acquaintances who haven't established an epistemic track record. In day-to-day life, we don't

interrogate the personal history of each testimony-giver; instead, we rely upon the belief that

most people are reliable on most topics most of the time. Of course, this trust is defeasible and

context matters. If a stranger in a coastal Japanese village tells me a tsunami is coming, I

probably should believe them. If the same thing happens in downtown Omaha, that's different.

Obvious implausibility blocks default reliance on testimony.

Our reliance on strangers isn't groundless, because we have social *norms* surrounding testimony.

People are expected to be truthful when giving testimony, and to only report on things they are

competent to speak about (or explicitly qualify their uncertainty). People who frequently or

egregiously violate these norms acquire reputations as liars and fools, and fear of such a

reputation motivates most people to testify responsibly. That fact justifies our default reliance on

stranger testimony.[48]

Unfortunately, several crucial features of testimonial practice are weakened on social media (as I

have discussed in greater detail elsewhere).[49] Social media users engage in what I call "bent"

testimony, where neither the speaker nor the audience fully expect the speaker to know what they

are talking about. People share stories thoughtlessly and retweet-without-endorsing. Social media

testimony is not regulated by strong norms of sincerity and competence, perhaps because it is so

hard to remember who shared what and thereby hold unreliable testifiers accountable. Most

---

[48] For varying views on the epistemology of testimony, see Coady (1992), Lackey (2008), and Goldberg (2010).
[49] Rini (2017).

oddly, even though we now *know* that all of these things happen, we often still treat social media posts like ordinary testimonial acts. We act as if social media testimony conveys epistemic justification, even though social media practices have much weaker norms around sincerity and competence.

This situation, combined with partisanship, provides much of the explanation for the rampant spread of fake news on social media. People pass along outrageous stories on the say-so of their "friends", without investigating and without much reason to believe these friends investigated either.[50] Testimonial practices that are sensible in other environments misfire badly on social media, yet the behavior of most users seems insensitive to this difference.

I first wrote about fake news and social media in January 2017, immediately after the 2016 US presidential election. At that time, most people were only just beginning to learn of the phenomenon. Donald Trump hadn't yet co-opted the phrase 'fake news' to refer to any information unfavorable to his interests. In a sense, that was a more innocent time. In the years since, things have changed. Most people are now explicitly aware that social media is awash with deceptive content; a mid-2018 survey found that 57% of American adults think the news they find on social media is "largely inaccurate".[51] Since the IRA indictments in February 2018, many are aware that it is not just social media *stories* that sometimes turn out fake, but also the people telling them. That friendly local Facebook group leader organizing a rally for your party might be just a stolen profile picture covering a St. Petersburg operator – who is simultaneously organizing a rally *against* your party. Or maybe it's just a bot.

---

[50] See above, as well as Rini (2018).
[51] Matsa and Shearer (2018).

Weaponized skepticism

Widespread awareness of coordinated social media fakery has deep consequences for public epistemology. The most important is this: testimonial skepticism is now a much more reasonable position.

In philosophical parlance, a skeptic is a person who challenges the reliability of basic sources of knowledge. A complete testimonial skeptic *denies* that we gain any justification for beliefs from the say-so of others. According to the testimonial skeptic, I can't justifiedly believe that bronze melts at 950 degrees just because others have told me. When the seaside stranger shouts a warning to me, I have no greater justification for anticipating a tsunami than I did a moment before.

Like most types of philosophical skepticism, pure testimonial skepticism is utterly impractical. You cannot go around living a human life in the modern world without according at least some evidential weight to the words of others. But *limited* testimonial skepticism makes more sense. In certain places, at certain times, via certain media, we ought to suspend our usual reliance on testimony. You shouldn't believe things on the say-so of people staggering out of bars at 3:00 am, nor should you give much weight to information conveyed via toner-streaked pamphlets thrust into your hands outside subway exits. These represent sensible local exceptions to a general reliance on testimony.

It is increasingly plausible to say that social media itself has become such an exception. Given how widespread we now know fakery and misinformation to be, perhaps we should accord little or no evidential weight to the testimony of social media strangers. On this view, I should not accept what I am told by social media users unless I can personally trace a testimonial chain back

to a person or organization with a good epistemic track record. In practice, I often won't have the time or resources to do this, so I should typically distrust what I learn on social media.

This is a sad outcome for anyone who once evangelized the democratic potential of social media. A decade ago, activists cheered the emergence of information channels evading the controls of government and powerful interests. Executives at Facebook and Twitter were quick to trumpet their products' role in the Arab Spring; knowledge of corruption and protests that would once have been suppressed by the state flowed instead through testimonial chains of smartphone-equipped citizens.[52] The assumption was not only that we *could* rely on social media testimony, but that doing so would enable a flourishing democratic culture even in societies where self-rule had been a thin dream.

That heroic vision now seems deluded at best. Authoritarian governments rocked by the Arab Spring quickly deployed mass disinformation to addle social media democrats. As we've seen, Putin's Russia began doing the same in 2011, and by 2014 had packaged the technique for export.

It is crucial to understand that these regimes' legitimization of testimonial skepticism is not accidental. This is *weaponized skepticism* – a calculated deployment of landmines across the epistemic commons. By forcing democratic citizens to confront the fallibility of their testimonial channels, authoritarians have neutralized a challenge to their hold on information.

This skepticism is not a purely defensive weapon for authoritarians. It also contributes to the proto-fascist arsenal of Ivan Ilyin's followers. To see this, notice the crucial way in which social media misinformation differs from the radio and TV broadcast lies of earlier authoritarians.

---

[52] See Tufekci (2017).

When the state Propaganda Ministry continuously bleats falsehoods over the radio, it may succeed in confusing some citizens. But citizens *themselves* aren't necessarily implicated in the deception. They may be passive and relatively innocent epistemic victims. Among friends, they may even achieve a kind of solidarity, sharing a quiet laugh at the latest absurd broadcast. Traditional misinformation does encourage skepticism, but it is a type of skepticism directed at broadcasters, journalists, and, lurking behind them, the regime itself.

But social media disinformation is different. Because it is spread directly *by* users, its unreliability implicates *them* as well. The gullible citizen who gets caught passing on fake news or unwittingly retweeting a propaganda bot thereby diminishes *their own* credibility and weakens their future ability to serve as a channel for information contrary to autocrats' interests. Recall Ilyin's depiction of democratic discourse: "the freedom of insincere, deceitful, treacherous, insinuating speech… free to "trust" liars and scoundrels or at least pretend to do so…" Every time a citizen is seen passing testimony from a liar or fake, Ilyin's insulting image looks a little more accurate. And this is exactly what the authoritarian wants.

This severely compounds the harm of disinformation. It is not *just* that citizens, increasingly unsure what to believe, are deprived of the benefits of knowledge transmission. Testimonial skepticism goes beyond this, to undermine citizens' trust in one another *as citizens*, not just as epistemic vectors. A person you cannot trust to convey truth is also not a person you can trust to negotiate mutual self-government in good faith, nor a person you can trust to stand beside you in solidarity against autocrats. A citizenry riven by mutual charges of fake news and gullibility is not a citizenry prepared to resist the pressures of a truth-indifferent incipient strongman.

This is the insidious genius of the Putin regime's social media strategy. It is distinctively harmful to democratic political cultures, inverting the epistemic resilience of open public debate. And the

effect is most virulent precisely when democratic citizens know it is happening; local testimonial

skepticism becomes reasonable because we now *know* a malicious actor is planting

disinformation, even fake 'citizens', into our testimonial networks. In the age of social media,

Ilyin's anti-democratic prescription needn't be subtle or covert. It's fine if the targets know that a

hostile autocratic power is at work; the fact that their co-citizens have foolishly allowed

themselves to be co-opted is itself encouragement to see democratic debate as corrupt and venal.

Centrifugal partisanship and the disarm/detonate choice

There is one further complication to this story: the confusing role of partisanship. Partisanship

leads people to apply skepticism selectively. We have all witnessed this on social media. A

person seems happy to rely on testimonial backing for news that confirms their partisan

worldview, but when they encounter opposing evidence they declare "fake news!" or "troll!" or

"bot!".

This is a dangerous form of local testimonial skepticism, made selective not by medium, but by

content. It leads to a winnowing of permissible evidence. Welcome information gets a pass, but

unwelcome evidence is dismissed on the grounds that the medium cannot be trusted.[53] Donald

Trump is a fan of this style of reasoning; as he told a crowd in Erie, Pennsylvania on October 10,

2018: "I believe in polls. Only the ones that have us up, because they're the only honest ones.

Other than that, they're the fake news poll. Fake news. Fake news."[54]

---

[53] See Kahan (2016) on cultural cognition, and Anderson (This Volume). See also Levy (2017) regarding the psychological mechanisms supporting the efficacy of fake news.
[54] Factba.se, transcript: 'Speech: Donald Trump Holds a Political Rally in Erie, PA – October 10, 2018'. https://factba.se/transcript/donald-trump-speech-maga-rally-erie-pa-october-10-2018 [Accessed December 5, 2018]

In one sense this is just confirmation bias, a phenomenon as old as human cognition. But it is important to see how these patterns are deliberately aggravated by social media influence operations. As we saw earlier, IRA bots were particularly active in the days leading up to the Brexit vote and the US presidential election, but they did not campaign only for Leave or Trump. Their aim seems to have been simply to whip up division. And that goal continued after, and beyond, electoral politicking. In May 2016, IRA trolls used false identities to arrange two competing rallies – "Stop Islamification of Texas" and "Save Islamic Knowledge" – at the same time outside a Houston Islamic center.[55] Public health researchers found that IRA troll accounts posted inflammatory content both for and against the use of vaccines; according to Mark Dredze, they "seem to be using vaccination as a wedge issue, promoting discord in American society."[56] Even pop culture can be weaponized; in 2018 Russian social media trolls fanned criticism of the Star Wars film *The Last Jedi,* apparently baiting a cultural fight over inclusiveness in Hollywood.[57]

The point of all of this is to transform social media from a neutral information-vector into a perpetual partisan battleground. Doing so also changes our co-citizens into perpetual antagonists, who are rendered epistemically suspicious the moment they present any information we do not like. The cumulative effect mutates confirmation bias from a pathology of individual information acceptance into a disease of testimonial networks, spreading corruption via the same lines that might otherwise have conveyed democratic cooperation. It is exactly as Ilyin imagined: democratic citizens coming to see one another as playthings of liars and cheats.

---

[55] Allbright (2017).
[56] George Washington University (2018).
[57] VanDerWerff (2018).

I think this process has come in three steps. First, insertion of demonstrable falsehoods and fake 'citizens' into democratic testimonial networks. This step began in Russia in 2011 and in western democracies in 2014. Second, exposure of many of these fakes, leading to rapid-onset testimonial skepticism. This began as early as 2015, but in the United States reached fever pitch in recriminations over fake news and IRA interference in 2017. Third, channeling of testimonial skepticism along lines of cultural and political partisanship. This last step has been building throughout, and we may not yet have seen its peak.

We may be approaching an inflection point, a period where the mechanisms of democratic testimonial culture begin to seize up and shake themselves apart. People now openly talk about the untrustworthiness of social media users, yet at the same time continue to (sometimes) get their information from them. We are witnessing the emergence of a particularly dangerous sort of testimonial skepticism - one that is selective, partisan, and blatantly hypocritical. We now watch our fellow citizens condemn the gullibility of the other side while falling for their own tribe's misinformation. Our awareness of social media deception has *itself* become yet another partisan battlefield, another breeding ground for mutual suspicion and animosity.

Is there anything we can do to stop this process? I think that we have two options for handling weaponized skepticism: disarm or detonate. In an ideal world, we would disarm the threat. We would focus on educating social media users to make thoughtful testimonial decisions, or on pressuring the platforms to provide tools for better testimonial hygiene.[58] But I am increasingly doubtful that this approach will work. Individual users seem to have little motivation toward epistemic scrupulousness. Perhaps this is not surprising, given that most people use social media

---

[58] See Rini (2017) for the latter.

for entertainment or boredom alleviation, not to do epistemic labor. The problem may be motivation rather than education. And there is so far little evidence that the social media platforms will institute serious reforms.

Instead, the best option may be detonation: blow it all up. Rather than try to inculcate responsible use of social media testimony, we might do best to *embrace* testimonial skepticism, but in a stronger form that does not permit partisan selectivity. If we can convince people to become *universally* skeptical of social media content – to disbelieve everything conveyed through the medium, not just what their partisan affiliation makes disagreeable – then we can at least blunt the effects of induced turmoil, and slow the Ilyinic souring of democratic discourse. On this strategy, social media needs to be brought down to the epistemic level of children's cartoons; a source of diversion, but not a place any adult would think to go for information.

How do we do this? How do we change implicit expectations and norms across vast informal networks of millions of information-consuming citizens? I am afraid that, as a philosopher, this sort of question is beyond my paygrade. If this approach is to work, it will require the coordinated attention of social scientists, media scholars, and policymakers. For the moment, the question is a normative one: is this indeed the best outcome we can realistically seek?

Personally, I am not happy to advocate the strategy of detonation. I would like to believe in the bright social media vision of a decade ago: a revolutionary medium for direct engagement between citizens, a decentralized engine of organization in nascent democracies. It is sad to

admit that Ilyin's descendants have corrupted the tool. Yet it does seem, for now at least, that this is our best remaining option. The casualties of weaponized skepticism can be mitigated at best.[59]

References

Allbright, C. (2017). 'A Russian Facebook page organized a protest in Texas. A different Russian page launched the counterprotest'. *Texas Tribune* November 1 2017 https://www.texastribune.org/2017/11/01/russian-facebook-page-organized-protest-texas-different-russian-page-l/ [Accessed Dec 5 2018].

Anderson, E. (this volume). 'Epistemic Bubbles and Authoritarian Politics'.

Bastos, M. T., & Mercea, D. (2019). 'The Brexit botnet and user-generated hyperpartisan news'. *Social Science Computer Review*, 37(1), 38-54.

Carrell, S. (2017, December 12). 'Russian cyber-activists "tried to discredit Scottish independence vote."' *The Guardian* December 12 2017. https://www.theguardian.com/politics/2017/dec/13/russian-cyber-activists-tried-to-discredit-scottish-independence-vote-says-analyst [Accessed Oct 29 2018].

Chen, A. (2015, June 21). 'The Agency'. *New York Times Magazine* June 21, 2015. https://www.nytimes.com/2015/06/07/magazine/the-agency.html [Accessed Oct 30, 2018]

Coady, C. A. J. (1992). *Testimony: A Philosophical Study*. Oxford University Press.

D'Souza, S. (2018). 'Bots, trolls and fake news: Social media is a minefield for U.S. midterms'. *CBC* Oct 21 2018. https://www.cbc.ca/news/world/national-us-midterm-elections-bots-trolls-fake-news-1.4863258 [Accessed Oct 31 2018].

Dellinger, A. (2018). 'Twitter is Suspending More Than One Million Accounts Per Day in Latest Purge'. *Gizmodo* July 7 2018. https://gizmodo.com/twitter-is-suspending-more-than-one-million-accounts-pe-1827409235

Fandos, N., & Roose, K. (2018. 'Facebook Identifies an Active Political Influence Campaign Using Fake Accounts'. *New York Times* August 1 2018. https://www.nytimes.com/2018/07/31/us/politics/facebook-political-campaign-midterms.html [Accessed Oct 31 2018].

George Washington University. (2018). 'Bots and Russian trolls influenced vaccine discussion on Twitter, research finds: New study discovers tactics similar to those used during 2016 US election'. *ScienceDaily* August 23 2018. www.sciencedaily.com/releases/2018/08/180823171035.htm [Accessed Dec 5 2018].

Gessen, M. (2017). *The future is history: How totalitarianism reclaimed Russia*. Granta Books.

Goldberg, S. (2010). *Relying on Others: An essay in epistemology*. Oxford: Oxford University Press.

Haskins, C. (2017). 'I Unknowingly Went to a Trump Protest Organized by Russian Agents'. *Motherboard* November 22 2017. https://www.vice.com/en_us/article/ywb9kx/nyc-trump-election-protest-hack-russian-agents-trolls-government . [Accessed Oct 30 2018].

Jamieson, K. H. (2018). *Cyberwar: how Russian hackers and trolls helped elect a president: what we don't, can't, and do know*. Oxford University Press.

Jones, D. S. (2018). 'Russian Troll Factory Whistleblower Blasts Media'. *Shooting the Messenger* Feb 25, 2018. https://shootingthemessenger.blog/2018/02/25/russian-troll-factory-whistleblower-blasts-media/ [Accessed Oct 31 2018].

Kahan, D. (2016). 'The Politically Motivated Reasoning Paradigm.' *Emerging Trends in Social & Behavioral Sciences*. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118900772.etrds0417.

Lackey, Jennifer (2008). *Learning From Words: Testimony as a Source of Knowledge*. Oxford: Oxford University Press.

Lapowsky, I. (2018). 'Here's How Much Bots Drive Conversation During News Events'. *Wired* October 30 2018. https://www.wired.com/story/new-tool-shows-how-bots-drive-conversation-for-news-events/ [Accessed Oct 41 2018].

Levy, N. (2017). 'The bad news about fake news'. *Social epistemology review and reply collective*, 6(8), 20-36.

Lister, T. and Shukla, S. (2019). 'Murdered journalists were tracked by police with shadowy Russian links, evidence shows'. *CNN* January 10 2019, https://www.cnn.com/2019/01/10/africa/russian-journalists-car-ambush-intl/index.html

Llewellyn, C., Cram, L., Hill, R. L., & Favero, A. (2019). 'For Whom the Bell Trolls: Shifting Troll Behaviour in the Twitter Brexit Debate'. *Journal of Common Market Studies*, 57(5), 1148-1164.

Lynch, M. (2016). 'Fake News and the Internet Shell Game'. *The New York Times* November 28 2016. Retrieved from https://www.nytimes.com/2016/11/28/opinion/fake-news-and-the-internet-shell-game.html

Matsa, K. E., & Shearer, E. (2018). 'News use across social media platforms 2018'. Pew Research Center. Chicago

Mayer, J. (2019). 'How Russia Helped Swing the Election for Trump'. *New Yorker* July 9 2019. https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump

Priest, D., Jacoby, J. and Bourg, A. (2018). 'Russian disinformation on Facebook targeted Ukraine well before the 2016 U.S. election'. *The Washington Post* Oct 28, 2018. https://www.washingtonpost.com/business/economy/russian-disinformation-on-facebook-targeted-ukraine-well-before-the-2016-us-election/2018/10/28/cc38079a-d8aa-11e8-a10f-b51546b10756_story.html [Accessed Oct 29, 2018.]

Rainey, J. (2017). '20th Century Fox Apologizes for 'A Cure for Wellness' Fake News Promos'. *Variety* February 16 2017. https://variety.com/2017/film/news/cure-for-wellness-fake-news-promos-studio-apologizes-1201990634/

Rini, R. (2017). 'Fake news and partisan epistemology'. *Kennedy Institute of Ethics Journal*, 27(2), E-43.

Rini, R. (2018). 'How to Fix Fake News'. *New York Times* October 15 2018. https://www.nytimes.com/2018/10/15/opinion/facebook-fake-news-philosophy.html

Romm, T., & Wagner, K. (2017). 'Here's how to check if you interacted with Russian propaganda on Facebook during the 2016 election'. *Recode* Dec 22, 2017. https://www.recode.net/2017/12/22/16811558/facebook-russia-trolls-how-to-find-propaganda-2016-election-trump-clinton [Accessed Oct 30 2018].

Seddon, M. (2014). 'Documents Show How Russia's Troll Army Hit America'. *Buzzfeed News* June 2 2014. https://www.buzzfeednews.com/article/maxseddon/documents-show-how-russias-troll-army-hit-america [Accessed Oct 30 2018].

Shao C, Hui P-M, Wang L, Jiang X, Flammini A, Menczer F, et al. (2018) 'Anatomy of an online misinformation network'. PLoS ONE 13(4): e0196087. https://doi.org/10.1371/journal.pone.0196087

Shearer, E., & Grieco, E (2019) 'Americas Are Wary of the Role Social Media Sites Play in Delivering the News'. Pew Research Center, October 2019, https://www.journalism.org/2019/10/02/americans-are-wary-of-the-role-social-media-sites-play-in-delivering-the-news/

Silverman, C., & Alexander, L. (2016). 'How Teens In The Balkans Are Duping Trump Supporters With Fake News'. *Buzzfeed News* November 3 2016. https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo

Snyder, T. (2018). *The Road to Unfreedom: Russia, Europe, America*. Tim Duggan Books.

Solon, O. (2018). 'Facebook removes 652 fake accounts and pages meant to influence world politics'. *The Guardian* August 22 2018.

https://www.theguardian.com/technology/2018/aug/21/facebook-pages-accounts-removed-russia-iran [Accessed Oct 31 2018].

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

Tufekci, Z. (2018). 'It's the (Democracy-Poisoning) Golden Age of Free Speech'. *Wired* February 15 2018. Retrieved from https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/

United States v. Internet Research Agency LLC. (2018). 18 U.S.C. §§ 2, 371, 1349, 1028A. Retrieved from https://www.justice.gov/file/1035477/download

United States v. Khusyaynova (2018). 18 U.S.C. §§ 2, 371. Retrieved from https://www.justice.gov/file/1035477/download

VanDerWerff, E. (2018). 'Russian trolls used Star Wars to sow discord online. The fact that it worked is telling'. *Vox* October 4 2018. https://www.vox.com/culture/2018/10/4/17930308/star-wars-russia-trolls-bots-alt-right-gamergate-fandamentalism [Accessed Dec 5 2018].

Vosoughi, S., Roy, D., & Aral, S. (2018). 'The spread of true and false news online'. *Science*, 359(6380), 1146-1151.

White, J.B. (2017). 'Facebook says 126 million Americans may have been exposed to Russia-linked US election posts'. *The Independent* Oct 31 2017. https://www.independent.co.uk/news/world/americas/us-politics/facebook-russia-adverts-americans-exposed-trump-us-election-2016-millions-a8028526.html [Accessed Oct 30 2018].