The Ethics of Countering Digital Propaganda

Corneliu Bjola

More than 150 million Americans were exposed to the Russian disinformation campaign prior to the 2016 presidential election, which was almost eight times more than the total number of people who watched the evening news broadcasts of ABC, CBS, NBC, and Fox stations on a given night in 2016.[1] As of January 2018, some 3,500 examples of pro-Kremlin disinformation ("fake news") that contradicted publicly available facts in a systematic fashion had been identified by the European Commission as key elements of an "orchestrated" propaganda campaign by the Russian government against the European Union.[2] Even more far-reaching, some two hundred unique targets—including politicians, diplomats, UN officials, military personnel from thirty-nine countries, as well as members of twenty-eight governments—were found by a research center affiliated with the University of Toronto to have been part of an extensive Russia-linked phishing and disinformation campaign.[3] The message is hard to miss: Western countries face an unprecedented, systematic, and unrelenting disinformation assault on their cyberspace, mainly from Russia,[4] which has been enabled by the very digital technologies that they have created. And, even more troublingly, they have no clear understanding of how to protect themselves against it.

Attempts to influence "collective attitudes by the manipulation of significant symbols," as communications theorist Harold Lasswell described propaganda efforts,[5] have been around for centuries, but with the rise of social media they have simply exploded. The growing popularity of social media networks makes it possible for digital messages to be disseminated widely, deeply, and quickly whether they are accurate or not, hence the rise of "fake news" and "post-truth" politics.[6] Furthermore, the sheer magnitude of the global production and

Ethics & International Affairs, 32, no. 3 (2018), pp. 305–315.

doi:10.1017/S0892679418000436

consumption of digital data places serious constraints on governments with respect to monitoring, let alone effectively confronting, the sources of propaganda leveraged against them. This deliberate attempt to disseminate information on digital platforms with the purpose to deceive and mislead may be termed "digital propaganda." By fundamentally disrupting the way in which information is generated, aggregated, disseminated, and ultimately interpreted, the new technological landscape ushered in by the arrival of social media and big data ensures that digital propaganda is here to stay.[7]

For many governments, however, combating digital propaganda comes with a serious ethical dilemma: how to react to and confront acts of disinformation directed by other states without losing the moral high ground. This essay argues that the ethically sound solution to combating propaganda lies with the concept of "moral authority," which ensures that an actor can have its arguments treated with priority by others and thus build support for and deflect challenges to certain objectives that it favors. More specifically, in the case of digital propaganda, an actor can maintain moral authority by making the case that it has been harmed, that it has normative standing to engage in counter-interventions, and that it does so in an appropriate manner. Failure to maintain moral authority could make an actor vulnerable to accusations of serving to amplify rather than contain disinformation, and thus help to legitimize the claims of those intentionally promoting disinformation.

Moral Authority as a Power Resource

The idea of moral authority serving as a power resource is, of course, not new. For instance, Rodney B. Hall showed how moral authority was used as a power resource in feudal Europe by both ecclesial and political-military authorities to effect outcomes in all matters of disputes.[8] Kent J. Kille investigated the role of religious and moral leadership as a power resource in the context of the United Nations, and found that the Secretary-General's ability to influence global affairs would often depend on how the international community regarded his moral authority.[9] William C. Wohlforth and others examined the concept in the context of status politics, and argued that small and middle powers employed moral authority as a strategy for gaining status as a "good power" in international relations.[10] As these studies suggest, the concept of moral authority is informed by two interrelated considerations, one normative and the other strategic. The first refers to the set of attributes and requirements that grant moral standing to an actor so that its arguments are treated with priority by others. The second consideration suggests that moral authority constitutes a source of power by which the holder can build support for and deflect challenges to certain objectives that one favors. As such, the concept helps delineate a set of conceptual considerations to guide and justify possible responses to digital propaganda.

At the normative level, one needs to understand the objectives, tactics, and consequences of disinformation to identify the necessary attributes for achieving moral standing when engaging in counter-interventions. At the strategic level, one needs to reflect upon the way in which one's actions to combat disinformation fruitfully draw on—but do not erode—one's sources of moral authority. The two levels should reinforce each other: normative attributes inform and define the moral standing of the actor, but they also provide the arena in which strategic action takes place. When the gap between the two levels becomes unsustainable, the moral standing of the actor collapses, and by extension, its capacity to effect outcomes is likely to decline as well.

In the context of digital propaganda, three considerations are relevant to normative inquiry into potential sources of moral authority: (1) whether the actor has been harmed as a result of disinformation, (2) whether the actor has standing to engage in counter-intervention, and (3) whether the actor's reaction is appropriate in light of contextual circumstances. These considerations address key elements of moral reasoning and, hence, the evaluative outputs they generate count as important sources of moral authority. Notably, the three considerations parallel the principles of just cause, legitimate authority, and proportionality that one finds in just war theory, but the comparison is rather limited in scope given the differences that exist between disinformation and conventional war, especially with respect to the nature of the harm. By pursuing an analysis of the relevant facts pertaining to disinformation from a moral authority perspective, this inquiry can help us to understand the conditions of validity of certain reasons to act and, by extension, help identify the normative attributes that can best enhance the moral value of these actions.

The Nature of the Harm

The first consideration to guide our inquiry concerns the reason for combating digital propaganda. Why should disinformation be confronted in the first place and, relatedly, under what conditions should it be done? In the battle between the freedom of expression and disinformation, democratic societies have traditionally sided with the former in times of peace, and for good reasons. What seems to have changed the balance of considerations is the arrival of the new concept of hybrid warfare, which advocates a prominent role for disinformation, whereby nonkinetic means are supposed to be deployed in all stages of conflict development alongside military capabilities.[11] By blurring the distinction between war and peace, and between combatants and noncombatants, disinformation thus becomes a form of sharp power to "pierce, penetrate, or perforate the political and information environments in the targeted countries."[12] Simply put, whether by fomenting political discontent, influencing electoral results, or weakening state authority, digital propaganda has now reached a point at which its long-term impact on state functioning can arguably be as great as a conventional military attack.

One might then argue that countering digital propaganda should be seen as the least harmful means for correcting and/or punishing a severe harm perpetrated against the sovereignty of the country. Further, it should do so by setting the response threshold to match the potential harm inflicted by the disinformation. This argument is theoretically appealing, but it comes with two limitations. First, what counts as harm in the case of disinformation is a matter of debate, as the full impact of digital propaganda can sometimes only be accurately assessed with hindsight. The 2016 U.S. presidential election offers a good illustration of this point, as few observers would have agreed before the election that the U.S. political system was acutely vulnerable to digital propaganda. However, with the election results now in the rearview mirror, the U.S. case serves as a critical reference point concerning the potential implications of propaganda for other states. Second, the issue of attribution must be carefully considered, especially since the identity of social media account holders is relatively easy to mask.

Ongoing efforts by certain states to "weaponize information" have thus changed the balance of considerations about whether disinformation should be confronted or not. Disinformation can be harmful, but the nature of the harm and the identity of the perpetrators cannot be easily determined. This makes the issue of normative attributes even more relevant, as victims of disinformation need to demonstrate extra care when making moral claims about their case. Truthfulness and prudence are two normative attributes that could enhance the moral authority of the victims and, by extension, the moral value of the actions to be taken under these circumstances.

Truthfulness requires actors to carefully evidence the nature of the harm as accurately and compellingly as possible given the available information, without deliberate omissions or overstatements. While a certain degree of politicization might be inevitable, a comprehensive disclosure of the raw data of online disinformation operations could strengthen the case about the harmful implications of propaganda. Unlike the case of cyber hacking, data pertaining to disinformation campaigns can be easily aggregated and made available by digital platforms, thus reducing the need for governments to reveal sensitive information or techniques.[13] Prudence, on the other hand, implies that questions regarding attribution should be treated with utmost seriousness. For example, when direct attribution is technically unavailable, both intelligence and open source data could be used to establish a clear pattern of prior actions that point to the likely source of disinformation. Inviting the alleged perpetrator to credibly address the accusations could also demonstrate the disposition to engage in prudential conduct.

Taken together, truthfulness and prudence enhance the moral authority of the actor and may serve as a protective "moral shield" against counterclaims that the potential reaction to disinformation has no moral merit. This is illustrated by the case of Sergei Skripal, the former Russian military officer who was allegedly poisoned, along with his daughter, by Russia while living in England. Here, the British government managed to retain moral authority and to win over public opinion and the support of its allies by carefully stating the case against Russia while prudently offering Russia the opportunity to provide a credible explanation for the attack on Skripal and his daughter.[14]

The Standing of the Actor

The second consideration concerns the normative standing of the actor to engage in counter-intervention. Who has the right authority to address digital disinformation and why? A quick look at the current landscape reveals that a wide variety of state and nonstate actors have taken up this challenge. The European Union, for example, coordinates its capacity to forecast, address, and respond to disinformation activities by foreign actors through a small strategic communications group called the East StratCom Task Force, created in March 2015.[15] Not dissimilarly, in the United States, the government-run Global Engagement Center was established by then Secretary of State John Kerry in April 2016. Initially, its stated goal was to defeat terrorist organizations and disrupt their ability to recruit new followers, but its mission was expanded in 2017 to include countering the adverse effects of state-sponsored propaganda and disinformation.[16] In Ukraine, however, the institution that established itself as leading the fight against disinformation is a nongovernmental organization called StopFake.[17] Established in 2014, its goal is "to verify information, raise media literacy in Ukraine, and establish a clear red line between journalism and propaganda."[18] Similarly, in Lithuania a five-thousand-strong volunteer army, which calls itself the Lithuanian Elves, has been patrolling social media since 2014 to find and expose fake accounts and pro-Russian trolls.[19]

Having a governmental institution in charge of combating disinformation can be beneficial in terms of access to human and financial resources, technical expertise, and coordination mechanisms, but sometimes less so in terms of public credibility, as governmental actions in the area of strategic communication are generally viewed with suspicion, both by domestic and (especially) international audiences. Nongovernmental organizations may suffer less from this problem so long as their political independence is well established, but this advantage may come at the expense of reduced effectiveness, as resources for an NGO can be scarce. Both types of actors need to be careful that public assessment of their activity is not undercut by the use of tactics that are misleading in ways similar to the very disinformation they seek to counter. This will be discussed further in the next section.

Still, the fact that an entity (state or nonstate) has the ability to initiate actions to combat digital propaganda does not mean that they automatically enjoy normative standing in this regard. Three normative attributes are particularly important to consider when making this assessment. The first one is *accountability*, which requires the group to make itself subject to public scrutiny for its actions, since mistakes made in the fight against disinformation could have severe consequences.[20] There are often no real checks on how these institutions make decisions about what counts as good reasons for reacting to disinformation and what are the most suitable means for conducting such interventions. Thus, public accountability is key. The second attribute is *integrity*, that is, the need for the actor to demonstrate consistency between its stated objectives and its actions so that potential accusations of hypocrisy, incompetence, or malice can be firmly preempted. The third attribute is overall *effectiveness*, since the capacity to contain or eliminate disinformation has an intrinsic moral value as it helps protect and enhance the "right to communicate" online.[21] These three normative attributes combine to demonstrate that a particular entity (whether state or nonstate) deserves normative standing to take action against disinformation because it does so without abusing its power, in a trustworthy fashion, and with a moral purpose.

The Level of Reaction

The third consideration concerns the ethical implications of the means by which such counter-interventions are pursued. Some argue that potential for the "structural virality" of disinformation—that is, the ability of a disinformation message to propagate itself in a multilayered, cascading fashion[22]—makes it more difficult for the counter-intervention to address the content and source of disinformation in an effective and timely fashion. If this argument holds, should the counter-intervention respond only to ongoing disinformation activities or should it consider preventing them from occurring in the first place, similar to the concept of a preemptive strike? Most importantly, what specific ethical considerations should inform judgments about engaging in reactive versus preemptive strategies of counter-intervention?

Defensive counterstrategies like the one currently undertaken by the European Union's East StratCom Task Force are useful for exposing patterns of digital propaganda, identifying nodes of influence in the disinformation network, and improving media literacy about how propaganda works and how not to play its game.[23] The main goals of such strategies are to raise public awareness about the role of propaganda in amplifying societal vulnerabilities and to build public resilience against disinformation so that its potential effects become less corrosive over time. Despite these important goals, however, taken alone these defensive strategies cannot stop adversaries from launching future disinformation attacks.

Preemptive strategies, by contrast, seek to anticipate potential disinformation operations and inflict some degree of pain on the opponent so that such operations would become more difficult to initiate in the future. The EU, Canada, and U.S. joint operations against media outlets affiliated with the Islamic State terror group[24] offer an ethically plausible template for such interventions, not only against terrorist-inspired networks but also against state-sponsored propaganda networks that seek to undermine key functions of state sovereignty (elections, referendums, foreign policy decisions). At the same time, one should also be mindful of the fact that offensive counter-interventions might lead to online profiling of certain communities and risk amplifying the message they seek to delegitimize.

Drawing on just war theory, we can use the distinction between *narrow* and *wide* proportionality to assess the appropriateness of the counter-intervention. The narrow version of proportionality would limit liability to the circle of direct perpetrators and in a manner that is proportional to the harm inflicted by disinformation. The wide conception of proportionality would extend liability to a broader circle of actors (for example, perpetrators and decision-makers) or even to nonliable persons (for example, the general public) if a "lesser-evil" justification can be offered, such as to prevent a substantially greater amount of harm from being suffered.[25] The problem with applying this logic to harm from disinformation is that the harm is not immediately manifested as with conventional military attacks, and it usually takes some time until it becomes fully visible. Consequently, the concept of proportionality must be adapted to take into account the delayed nature of the harmful effects.

To adapt proportionality when the nature of the harm is not easy to ascertain due to the temporal effect, the *frequency* of a disinformation operation should serve as a qualifying consideration. The moral permissibility of different courses of action against disinformation would then be in part determined by how frequently a party is subjected to disinformation, alongside considerations of harm. This leads to the following rough prescriptions for maintaining moral authority when responding to digital propaganda: defensive measures when the incidence of disinformation is minimal and the nature of the harm is diffuse; narrow actions against direct perpetrators when the frequency is significant and the nature of harm is plausibly serious; and wide offensive counter-interventions against direct perpetrators and decision-makers, even at the risk of affecting nonliable persons, when the frequency of disinformation operations is high and the possibility of a great harm is reasonably certain.

The discussion above leads us to the key normative attribute that an actor must cultivate as a source of moral authority for combating disinformation in a proportionally appropriate manner: *responsibility*. As a source of moral authority, responsibility dictates that the counter-intervention be conducted in a measured manner that takes into account the contextual circumstances and the likely nature of the harm generated. Following Peter Strawson's account of moral responsibility,[26] it can be argued that when the basic demand for goodwill toward one another is not met, the parties are justified in developing reactive attitudes toward each other. Through such reactive attitudes as resentment or anger, the side subjected to disinformation communicates to the alleged perpetrator a reasonable expectation of goodwill. If this expectation is not met, stronger reactive attitudes will be morally justified. The suspension of reactive attitudes is possible if the alleged perpetrator offers plausible excuses or exemptions, thus potentially leading to a return of goodwill. In sum, a party exposed to disinformation can demonstrate responsibility by tailoring its reactive attitudes to the degree of ill will that it faces, while signaling openness for future goodwill. For example, one should give the benefit of the doubt to the other party when the nature of the harm produced by disinformation is yet to manifest itself, but this excuse loses its validity when the frequency of such acts is all too visible.

Conclusion

Like many other technologies, social media platforms come with a dual-use challenge, that is, they can be used for peace or war, for offense or defense, for good or evil. By allowing for the decentralization and diffusion of power away from traditional stakeholders (states and governments), digital technologies can serve to empower the powerless, such as happened during the Arab Spring, or they can be deliberately weaponized to undermine the social fabric of modern societies, as in cases of foreign electoral subversion. For many governments, countering digital propaganda under these conditions presents a major ethical dilemma: How can a state react to acts of disinformation without losing the moral ground that it seeks to occupy?

I have argued that the concept of moral authority as a power resource can provide a suitable toolkit to approach this dilemma. More specifically, in order to ensure that its arguments are treated with priority by others, a state or organization needs to make the case that it has been harmed, that it has normative standing to engage in counter-interventions, and that it does so in a proportionate and responsible manner. Otherwise, its moral authority slowly decays to the point that it becomes vulnerable to accusations of serving to amplify rather than contain disinformation and, by extension, its capacity to effect outcomes to counter propaganda is likely to deteriorate as well.

NOTES

[1] Marissa Lang, "Number of Americans Exposed to Russian Propaganda Rises, as Tech Giants Testify," *San Francisco Chronicle*, November 1, 2017, www.sfchronicle.com/business/article/Facebook-Google-Twitter-say-150-million-12323900.php.

[2] Jon Stone, "Russian Disinformation Campaign Has Been 'Extremely Successful' in Europe, Warns EU," *Independent*, January 17, 2018, www.independent.co.uk/news/uk/politics/russian-fake-news-disinformation-europe-putin-trump-eu-european-parliament-commission-a8164526.html.

[3] Adam Hulcoop et al., "Tainted Leaks: Disinformation and Phishing with a Russian Nexus," *Citizen Lab*, May 25, 2017, citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/.

[4] Anti-Western disinformation campaigns originating from China are much less visible, but this may change in the future, as the German intelligence services recently indicated; see Thomas Escritt and Michael Martina, "German Intelligence Unmasks Alleged Covert Chinese Social Media Profiles," *Reuters*, December 10, 2017, www.reuters.com/article/us-germany-security-china/german-intelligence-unmasks-alleged-covert-chinese-social-media-profiles-idUSKBN1E40CA.

[5] Harold D. Lasswell, "The Theory of Political Propaganda," *American Political Science Review* 21, no. 3 (1927), p. 627.

[6] Stephan Lewandowsky, Ullrich K. H. Ecker, and John Cook, "Beyond Misinformation: Understanding and Coping with the 'Post-Truth' Era," *Journal of Applied Research in Memory and Cognition* 6, no. 4 (2017).

[7] The discussion in this article strictly focuses on the use of information for propaganda purposes, not on the methods by which information may be illicitly acquired, which is the field of cyber hacking.

[8] Rodney Bruce Hall, "Moral Authority as a Power Resource," *International Organization* 51, no. 4 (1997).

[9] Kent J. Kille, ed., *The UN Secretary-General and Moral Authority: Ethics and Religion in International Leadership* (Washington, D.C.: Georgetown University Press, 2007).

[10] William Wohlforth et al., "Moral Authority and Status in International Relations: Good States and the Social Dimension of Status Seeking," *Review of International Studies* 44, no. 3 (2018), pp. 526–46.

[11] Flemming Splidsboel Hansen, "Russian Hybrid Warfare: A Study of Disinformation," Danish Institute for International Studies, Copenhagen, Denmark (2017), pure.diis.dk/ws/files/950041/DIIS_RP_2017_6_web.pdf.

[12] Christopher Walker and Jessica Ludwig, "The Meaning of Sharp Power: How Authoritarian States Project Influence," *Foreign Affairs*, November 16, 2017, www.foreignaffairs.com/articles/china/2017-11-16/meaning-sharp-power?cid=int-fls&pgtype=hpg.

[13] See, for instance, recent reports on the website of the U.S. House of Representatives Permanent Select Committee on Intelligence, exposing Russia's effort to sow discord online; "Facebook Ads: Social Media Advertisements," democrats-intelligence.house.gov/facebook-ads/social-media-advertisements.htm.

[14] Lizzie Dearden, "Why is the UK Accusing Russia of Launching a Nerve Agent Attack on Sergei Skripal in Salisbury, and What Is the Evidence?" *Independent*, March 16, 2018, www.independent.co.uk/news/uk/crime/uk-russia-nerve-agent-attack-spy-poisoning-sergei-skripal-salisbury-accusations-evidence-explanation-a8258911.html.

[15] European External Action Service, "Questions and Answers About the East Stratcom Task Force," August 11, 2017, eeas.europa.eu/headquarters/headquarters-homepage_sv/2116/%20Questions%20and%20Answers%20about%20the%20East%20StratCom%20Task%20Force.

[16] U.S. Department of State, "Global Engagement Center," www.state.gov/r/gec/.

[17] Vijai Maheshwari, "Ukraine's Fight against Fake News Goes Global," *Politico*, March 15, 2017, www.politico.eu/article/on-the-fake-news-frontline/.

[18] StopFake.org, "About Us," www.stopfake.org/en/about-us/.

[19] Anne Sofie Schrøder, "Lithuania Has a Volunteer Army Fighting a War on the Internet," *Euronews*, September 28, 2017, www.euronews.com/2017/09/28/lithuania-has-a-volunteer-army-fighting-a-war-on-the-internet.

[20] Michael Peel, Mehreen Khan, and Max Seddon, "EU Attack on Pro-Kremlin 'Fake News' Takes a Hit," *Financial Times*, April 2, 2018, www.ft.com/content/5ec2a204-3406-11e8-ae84-494103e73f7f.

[21] Article 19, "Statement on the Right to Communicate," London, U.K. (February 2003), www.article19.org/data/files/pdfs/publications/right-to-communicate.pdf.

[22] Sharad Goel et al., "The Structural Virality of Online Diffusion," *Management Science* 62, no. 1 (2016).

[23] Corneliu Bjola and James Pamment, "Digital Containment: Revisiting Containment Strategy in the Digital Age," *Global Affairs* 2, no. 2 (2016).

[24] Aleksandra Wróbel, "Europol Claims Successful Crackdown on ISIS Propaganda," *Politico*, April 27, 2018, www.politico.eu/article/isis-islamic-state-europe-europol-claims-successful-crackdown-on-propaganda/.

[25] For a discussion of the narrow versus wide version of the proportionality principle in just war theory, see Jeff McMahan, "Proportionality and Necessity in *Jus in Bello*," in Seth Lazar and Helen Frowe, eds., *The Oxford Handbook of Ethics of War* (New York: Oxford University Press, 2018), p. 420.

[26] Peter F. Strawson, "Freedom and Resentment," in Gary Watson, ed., *Proceedings of the British Academy, Volume 48: 1962* (Oxford: Oxford University Press, 1963), pp. 1–25.

Abstract: How can a state react to being a target of disinformation activities by another state without losing the moral ground that it seeks to protect? This essay argues that the concept of moral authority offers an original framework for addressing this dilemma. As a power resource, moral authority enables an actor to have its arguments treated with priority by others and to build support for its actions, but only as long as its behavior does not deviate from certain moral expectations. To develop moral authority, an actor engaged in combating digital propaganda must cultivate six normative attributes: *truthfulness* and *prudence* for demonstrating the nature of the harmful effects of disinformation; *accountability*, *integrity*, and *effectiveness* for establishing the normative standing of the actor to engage in counter-intervention; and *responsibility* for confirming the proportionality of the response.

Keywords: digital propaganda, disinformation, ethics, moral authority, hybrid warfare