Influence, War, and Ethics

Beba Cibralic*

Introduction

In his 1928 seminal work *Propaganda*, Edward Bernays wrote, "[t]oday the privilege of attempting to sway public opinion is everyone's. It is one of the manifestations of democracy that anyone may try to convince others and to assume leadership on behalf of his own thesis."[1] His words are a reminder that, at least in principle, contemporary democratic political systems grant an individual or group the privilege of trying to influence others.

To an extent, liberal democracies also grant foreigners – those who are not members of the polity – this same privilege. As largely open political communities, liberal democracies allow information to come in and out with relatively few gatekeepers; this is one of their core strengths. It is also a vulnerability, as the epistemic security of a liberal democracy can be undermined by nefarious actors. The absence of clear and enforceable international rules regarding information-based influence campaigns has provoked debate on how best to conceptualize and respond to such campaigns.[2]

In trying to understand the harms, scholars have reached for different frameworks that, if implemented effectively, give states recourse against campaigns by adversaries. This has been important in the wake of the 2016 Russian influence campaign against the United States, which sought to undermine trust in the democratic process, and remains important as the United States prepares for the 2024 presidential election.[3]

---

* Beba Cibralic is an Associate Fellow at the Leverhulme Centre for the Future of Intelligence, University of Cambridge.

1. Edward L. Bernays, Propaganda 147-48 (1928).

2. States vary on the robustness of domestic laws for regulating influence. Many States such as China, Iran, and India have domestic laws that allow the government to regulate influence campaigns more easily.

3. There have been numerous information-based influence campaigns since 2016. For example, in 2023 the Australian Strategic Policy Institute reported on one conducted by China targeting the United States. Influence campaigns that target a State during an election cycle are of particular importance because of their impact on democratic decision-making. For those less familiar with Russia's influence campaign against the United States, here is additional context: in the lead up to the 2016 U.S. presidential election, the Russian government engaged in an influence campaign to advance Russia's strategic interests. According to the January 2017 Intelligence Community Assessment (ICA) declassified report, "Moscow's influence campaign followed a Russian messaging strategy that blends covert intelligence operations—such as cyber activity—with overt efforts by Russian government agencies, state-funded media, third-party intermediaries, and paid social media users or 'trolls.'" *See* Off. of the Dir. Of Nat'l Intel., Background to Assessing Russian Activities and Intentions in Recent US Elections: The Analytic Process and Cyber Incident Attribution 2 (2017) [hereinafter Russia Background Paper]. Examples from the Russian campaign include hacking Democratic National Committee (DNC) servers, releasing the data obtained (for example, from John Podesta's email account) via WikiLeaks and DC Leaks, and pushing disinformation and fake news on

Some scholars have claimed that foreign influence campaigns are an infringement of sovereignty. Others have argued that they undermine the right to self-determination. Stronger still, some have suggested that these campaigns might, under certain circumstances, constitute an armed attack or be considered a kind of warfare.[4] In Part I, I argue that none of these frameworks is adequate for addressing influence.

Many scholars have already addressed the legal reasons in favor of or against each of the dominant frameworks, and my aim is not to refashion their arguments. While I will draw on the legal literature where useful, I will focus on offering conceptual and normative arguments against grouping influence within any of these dominant paradigms.

In Part II, I explore how we might draw the distinction between permissible and impermissible influence. No simple demarcation is readily available, and the considerations introduced in this Part should be taken as suggestive, not forceful. Ultimately, my aim is to de-center a focus on the specific features of influence we take to be normatively troublesome, such as speaker-identity deception. Before proceeding, a brief word on nomenclature. Foreign influence is, simply put, when one state tries to influence another. How we come to conceptualize foreign influence and all the mechanisms involved is more complex. Different kinds of campaigns and activities have been described as examples of foreign influence: influence operations, information campaigns, psychological operations, information operations, reflexive control, social engineering, and foreign influence efforts (FIEs), to name a few. While there is variation, all information-based influence campaigns tend to have the same underpinning structure: an agent (an individual, group, state, etc.) in some context (environment, country, period of time, etc.) targets an audience (an individual, group, state, etc.) using instruments (a technique, tool, medium, information, etc.) for the sake of some desired end (to sway public opinion, change behavior, disrupt an election, etc.). I will generally use the term "information-based influence campaign," but much of what I write can apply to other categories.[5]

---

popular social media platforms like Facebook and Twitter. Adam Entous, Ellen Nakashima, & Greg Miller, *Secret CIA Assessment Says Russia Was Trying to Help Trump Win White House*, WASH. POST (Dec. 9, 2016, 10:45 PM), https://perma.cc/CM7F-SZK4; Raphael Satter, Jeff Donn, & Chad Day, *Inside story: How Russians Hacked the Democrats' Emails*, AP NEWS (Nov. 4, 2017, 3:34 PM). The 2017 ICA report assesses, with a high degree of confidence, that the aim was to "undermine public faith in the United States' democratic process, denigrate Secretary Clinton, and harm her electability and potential presidency." RUSSIA BACKGROUND PAPER, *supra* note 3, at ii.

4. The phrase "armed attack" is derived from Article 51 of the United Nations Charter and refers to use of force that authorizes a use of force in response. However, not every use of force rises to the level of an armed attack. *See* Military and Paramilitary Activities in and against Nicaragua (Nicar. v. U.S.), Judgment, 1986 I.C.J. Rep. 14, ¶¶ 191, 210 (June 27). "Warfare" is not a legal word. Warfare equates to "armed conflict" in the law, where an armed conflict involves a level of intensity and duration greater than that of a single armed attack. A state may respond to an armed attack with a use of force but has the discretion to determine if this means the initiation of an ongoing armed conflict. *See id.* at ¶¶ 210-11.

5. For a longer explanation, *see* Beba Cibralic, *A Topography of Information-based Influence*, *in* HYBRID THREATS AND GREY ZONE CONFLICT: THE CHALLENGE TO LIBERAL DEMOCRACIES (Mitt Regan & Aurel Sari eds., forthcoming 2024).

I. BEYOND THE DOMINANT PARADIGMS: WAR, SOVEREIGNTY, AND SELF-DETERMINATION

In Part I, I consider and ultimately reject the dominant paradigms in international law for capturing the harms of foreign influence. In Section A, I present my case against relying on the warfare paradigm. While there is some merit in referring to "information warfare" in the context of an armed conflict, using the language of warfare outside of this context hinders us in various ways: it reifies a construct that may not always be useful, it confuses us, and it may lead us to certain conclusions that are not justified. In Section B, I argue against using the sovereignty and self-determination paradigms and show the foreign/domestic distinction is inadequate for framing online influence.

A. Influence and War

Legal scholars examining information-based influence campaigns have persuasively shown that, under the existing laws of armed conflict, these influence campaigns do not constitute warfare as international law defines it. As Jens David Ohlin explains, "international lawyers have generally assumed that an armed attack for purposes of Article 51 [of the United Nations Charter, which allows States to use force in self-defense] requires some physical destruction. In the *Nicaragua* opinion, the International Court of Justice (ICJ) argued that an attack must meet a 'scale and effects' criterion in order to qualify as an armed attack—a standard that has had enormous influence in contemporary international legal and diplomatic practice."[6] This is reflected in Rule 71 of the Tallinn Manual, for example, writes, "A State that is the target of a cyber operation that rises to the level of an armed attack may exercise its inherent right of self-defence. Whether a cyber operation constitutes an armed attack depends on its scale and effects."[7] The rule deliberately limits itself to cases where the cyber operation *rises to the level of an armed attack*; that is, causes physical damage to the target state. Mere influence does not do that.

Ohlin is not alone in his interpretation. Herbert Lin suggests that information operations are hostile non-kinetic activity, and not warfare under the laws of war or in the Clausewitzian sense. In his view, information operations are "better characterized as hostile or adversarial psychological manipulation . . . [they have] connotations of soft power (more properly, a mix of soft power and sharp power): propaganda, persuasion, culture, social forces, confusion, deception."[8]

There are, however, conceptual and normative questions beyond the legal question. Put differently, should we expand our understanding of war so that

6. JENS DAVID OHLIN, ELECTION INTERFERENCE: INTERNATIONAL LAW AND THE FUTURE OF DEMOCRACY 52 (2020); *see also,* Nicar. V. U.S., 1986 I.C.J.

7. *See generally* Tallinn Manual 2.0: On the International Law Applicable to Cyber Operations 339 (Michael N. Schmitt et al. eds., 2nd ed. 2017).

8. Herbert Lin, *On the Organization of the U.S. Government for Responding to Adversarial Information Warfare and Influence Operations*, 15 J. L. & POL'Y FOR INFO. SOC'Y 1, 2 (2019).

these kinds of influence activities count as war? Ought we to use the war paradigm to understand these campaigns?

Scholars interested in showing that information-based influence campaigns fit within the paradigm of warfare have asserted that these activities are a kind of information warfare (IW). While the United States and the North Atlantic Treaty Organization (NATO) have not concretely defined information warfare, the term has become more prevalent.[9] Commentators have used the term in different ways, but there is a set of characteristics that most definitions share captured in a definition offered by Catherine Theohary in a defense primer.

According to Theohary, information warfare is a strategy that encompasses a range of activities from disinformation or deception to misinformation to the use of factual information for strategic goals.[10] Information warfare could involve propaganda, defined broadly as "the propagation of an idea or *narrative* that is intended to influence, *similar to psychological or influence operations*. It can be misleading but true . . . A government communicating its intent, policies, and values through speeches, press releases, and other public affairs can be considered propaganda."[11] Notably, this capacious notion of information warfare is not new. In 1993, John Arquilla and David Ronfeldt published a RAND paper called "Cyberwar is Coming!" In the article, Arquilla and Ronfeldt predicted "netwar," a term that refers to:

---

9. Alicia Wanless & James Pamment, *How Do You Define a Problem Like Influence?*, 18 J. INFO. WARFARE 1, 6 (2019).

10. CATHERINE A. THEOHARY, CONG. RSCH. SERV., IF10771, DEFENSE PRIMER: INFORMATION OPERATIONS 1 (2020) [hereinafter DEFENSE PRIMER].

11. *Id*. at 1. Theohary's use of the term "propaganda" is not wrong per se, but it does not reflect all the common ways in which the term is used. Moreover, it is not the most helpful use of the term. Where propaganda is used to mean too much, it means very little. It gets more complex and no clearer in the Defense Primer. Theohary explains:

> In 2018, [the Department of Defense (DOD)] issued a Joint Concept for Operations in the Information Environment [(IE)]. According to this document, *the IE comprises and aggregates numerous social, cultural, cognitive, technical, and physical attributes that act upon and affect knowledge, understanding, beliefs, world views, and, ultimately, actions of an individual, group, system, community, or organization*. The IE also includes technical systems and their use of data. The IE directly affects and transcends all operating environments. New DOD policy would define Operations in the Information Environment (OIE) as actions taken to generate, preserve, and apply informational power against a relevant actor in order to increase or protect competitive advantage or combat power potential within all domains of the operating environment. OIE span the competition continuum (cooperation, competition short of armed conflict, and warfighting). This definition of the continuum would align with the 2018 National Defense Strategy, which emphasizes information warfare [(IW)] as competition short of open warfare. *IW is defined not as a strategy but as a subset of OIE conducted during both competition below armed conflict and during warfighting in order to dominate the IE at a specific place and time*. *Id*. at 2 (emphasis added).

OIEs span all competition continuum (peace, war, and everything in between) and occupy all domains of the social world (from the technical to the cultural). The actions that comprise OIEs are broadly defined as "actions taken to generate, preserve, and apply informational power." *Id*. To say that information warfare is a subset of OIEs does not make information warfare more precise. The concept of OIEs is too broad and vague to be useful.

> information-related conflict at a grand level between nations or societies. It means trying to disrupt, damage, or modify what a target population "knows" or thinks about itself and the world around it. A netwar may focus on public or elite opinions, or both. It may involve public diplomacy measures, propaganda and psychological campaigns, political and cultural subversion, deception of or interference with the local media, infiltration of computer networks and databases, and efforts to promote a dissident or opposition movements across computer networks . . . In other words, netwar represents a new entry on the spectrum of conflict that spans economic, political, and social as well as military forms of "war."[12]

From the quotation above, it is clear that the concept of netwar is similarly capacious to the definition introduced by Theohary.[13] While we should commend Arquilla and Ronfeldt for their prescient contributions, we ought to be wary of using vague and all-encompassing terms in a national and epistemic security context.

It is sensible, I suggest, to talk about information warfare in the context of an armed conflict. During an armed conflict, a number of tools and techniques, including information, are used to advance a party's strategic aims. Since Clausewitz, if not earlier, information has been viewed as a critical part of warfare.[14] What is more controversial is whether, outside of an armed conflict context, it is appropriate to categorize the kinds of activities Theohary describes as information warfare.[15] In my view, it is wrong to use the label information warfare outside of an armed conflict.

To begin with, conceptualizing information-based influence campaigns as either an armed attack or a war risks escalation.[16] If an information-based influence campaign is an armed attack, this warrants the use of military force in response to a case in which the campaign may not have caused any physical damage. Furthermore, if an ongoing information campaign is itself conceived of as a war between states, then the implicit assumption is that the threshold for the permissible use of military force has *already* been crossed, so a forcible response is always an option. Either way, escalation is possible. More generally, when we claim that certain activities constitute warfare, those activities take on a new

12. ATHENA'S CAMP: PREPARING FOR CONFLICT IN THE INFORMATION AGE 28 (John Arquilla & David Ronfeldt eds., 1997) [hereinafter Arquilla & Ronfeldt]. The 1993 article was republished as Chapter Two therein.

13. In that same paper, Arquilla and Ronfeldt developed the concept of "cyber-war", which is now part of everyday parlance. Netwar never caught on in quite the same way.

14. Susan Carruthers writes that Hitler believed it was Britain's information campaign in World War I (WWI), an "inspired work of genius," that caused German morale to collapse. [1] Hitler was not alone: General Ludendorff claimed that German soldiers had been "hypnotized . . . as a rabbit by a snake" in WWI; similarly, Nazi propagandist Eugen Hadamovsky believed "German people were not beaten on the battlefield but were defeated in the war of words." Susan L. Carruthers, Media at War: Communication and Conflict in the Twentieth Century 66 (2011).

15. Defense Primer, *supra* note 10.

16. I am grateful to Milton C. Regan Jr. for making this point to me.

normative and legal status. We may also embolden those in power to respond to the so-called armed attack in ways that would not otherwise be permitted. Saying "this is war" changes the rules of the game, pushing us onto different terrain.

There is another reason to resist using the term information warfare: once these categories are created, they take on a life of their own and can be difficult to erase. In *The Construction of Human Kinds*, Ron Mallon explores the feedback relationship between our cognitive architecture and the world around us. Mallon draws on the influential work of Ian Hacking on "looping kinds." A looping kind emerges where there is feedback between the pattern of conceptualization and the structures in the world that sustain it, such that these structures become self-generating and self-sustaining.[17] Once a construct is put out into the world, it becomes embedded in various political and social institutions that reinforce the "realness" of that construct.

In the event that we found the construct to be normatively problematic, inadequately matched to reality, or otherwise unhelpful, doing away with the construct would require more than explaining it away because it would have by then become a part of the architecture of our world. To avoid these pitfalls, creating or reifying a category that carries force – such as information warfare – ought to be done with intentionality and purpose.

A third reason to avoid the construct information warfare is that even if the language is supposed to be taken figuratively, not literally, it can do us damage. If we understand the term information warfare as purely metaphorical, it is worth pausing to consider how the metaphor helps or hinders us. It is unsurprising that in trying to describe new kinds of threats and challenges for which we have no (or a very limited) vocabulary, we would reach for the language of war to make sense of the phenomena. In the book *Metaphors We Live By*, cognitive linguists George Lakoff and Mark Johnson explain the significant role that conceptual metaphors play in our understanding of the world.[18] A conceptual metaphor refers to our

17. To illustrate what looping kinds are, Hacking introduces the example of "multiple personality disorder" (today more commonly referred to as "dissociative disorder"). Hacking suggests that in the eighteenth century, there was nobody experiencing multiple personality disorder. There were, of course, people who could hear voices, claimed demonic possession, and so forth. But none of those were *conceptualized* by the people having those experiences or by the cultures in which they were embedded as multiple personality disorders: they were seen as an external force acting on the person, and the person persisted throughout the changes. However, after the first diagnosis of multiple personality disorder in the late nineteenth century, psychiatric professionals began looking at people displaying those behaviors as potentially having multiple personalities. This resulted in an increased number of diagnoses of the disorder, even if nothing had changed the material structure of the world. Once there was an increased number of diagnoses, doctors began telling people about the presence of this kind of state and, in turn, people began to interpret their own mental life through the lens of multiple personality disorder. *See generally* IAN HACKING, REWRITING THE SOUL: MULTIPLE PERSONALITY AND THE SCIENCES OF MEMORY (1995).

18. GEORGE LAKOFF & MARK JOHNSON, METAPHORS WE LIVE BY (1980).

understanding of one domain (source domain) in terms of another (target domain). Conceptual metaphors are embedded in our everyday language.[19]

The conceptual network of battle and war (source domain) is used to help explain other concepts in a wide variety of target domains from sport to relationships to politics. The "war on drugs" is one prominent example. In the context of technology, the "weaponization of social media" and "media warfare" are frequently used.[20] But these metaphors obscure and confound rather than clarify. In using these metaphors, we think the phenomenon in question is a kind of warfare, even though it is outside of the formal warfare category. Labeling it warfare grants permission to respond in a warlike way without the hard moral work of offering reasoned justifications.

This kind of confusion does not occur in all instances of metaphor use: sports are often described as a species of conflict, of battle, and yet no one confuses a game of football for a military battle. The relevant difference between the war/sport metaphor and the war/information-based influence campaign application lies in the closeness between the source domain and the target domain. Because many commentators argue that information-based influence campaigns should be considered within the warfare paradigm, and there persists open debate about where information campaigns fit in gray zone conflict, there is sufficient closeness between the source domain (warfare) and the target domain (information-based influence campaigns) to render the metaphor confusing.

Finally, using the label information warfare attributes coordination where there may be none. In doing so, the use of the label inadvertently amplifies the significance of each of the phenomena because they are now tied to a broader threat. The Russian government's puppetry of Sputnik and RT (formerly Russia Today) – unsurprising given the long history of state media control in authoritarian regimes – seems all the more impressive and concerning when it is paired with other activities, like the spread of mis/disinformation by internet trolls and conspiracy theorists. While we should be concerned about the roles that Sputnik and RT, internet trolls, and conspiracy theorists play in the information economy, and there are often feedback loops between these activities (a story on RT might be picked up by conspiracy theorists, or vice versa), the umbrella term information warfare suggests a coordinated enemy campaign where there may not really be one. In using this term, we balloon the severity of the threat. In some cases, there is in fact coordination – see, for example, Operation Infektion –

19. For example, when we express sentiments such as "My wife and I are at a crossroads," "This relationship is a dead-end street," and "We are going nowhere in this relationship," we use the language of journeys (our source domain) to understand love (our target domain). *Id*. Other common conceptual metaphors include the understanding of time as money ("I spent time on this paper") and quantity in terms of directionality ("The price of inaction is rising"). *Id*.

20. *See generally* UNDP, WOULD I LIE TO YOU? THE WEAPONISATION OF SOCIAL MEDIA (Dec. 5, 2022), https://perma.cc/Y9PV-CMKT; Sarah Oates, *The Easy Weaponization of Social Media: Why Profit Has Trumped Security For U.S. Companies*, 1 Digital War 117 (2020), https://perma.cc/YN9S-GZMD.

coordination should be proven, not assumed.[21] In those cases where there is in fact coordination, the other reasons against using the term "information warfare" apply.

I hold that we should avoid using the term information warfare in outside of contexts where an armed conflict already exists and do not consider information-based influence campaigns under the category of war. Influence might be used in warfare, and it might also help states win a war. But the language of war is a dangerous metaphor outside of an armed conflict context; it is also an intellectual cheat because it licenses the use of force without the necessary moral *ad bellum* justifications.

In Section B, we will evaluate other prominent frameworks in international law that have increasingly been used in the context of information-based influence.

B. Sovereignty, Self-Determination, and Foreignness

In *Nicaragua v. United States*, we find the canonical source for the definition of an unlawful intervention: "A prohibited intervention must accordingly be one bearing on matters in which each State is permitted, by the principle of State sovereignty, to decide freely. One of these is the choice of a political, economic, social, and cultural system, and the formulation of foreign policy."[22] Some scholars have argued that an influence campaign against a state could be an infringement of the target state's sovereignty. In the wake of Russia's influence campaign against the 2016 U.S. presidential election, this case was made enthusiastically by some.[23]

Most legal scholars, however, have argued that under existing law, information-based influence campaigns do not generally infringe on state sovereignty. Ohlin explains that when the Tallinn Manual 2.0 was being written, experts considered the case of social media election interference. The majority of experts involved concluded "the activity is not coercive in nature and therefore does not constitute prohibited intervention" because "the coercive act must have the potential for compelling the target State to engage in an action that it would otherwise not take (or refrain from taking an action it would otherwise take)."[24]

The exception to the requirement of coercion is usurpation: if an intervening state has usurped some governmental function, it is unnecessary to establish coercion. It would be a radical stretch, however, to suggest that this is what happened during the 2016 Russian campaign.[25] Similarly, Michael Schmitt writes:

---

21. *See* U.S. DEPT. OF STATE, THE KREMLIN'S NEVER-ENDING ATTEMPT TO SPREAD DISINFORMATION ABOUT BIOLOGICAL WEAPONS (Mar. 14, 2023), https://perma.cc/5WB5-7FPZ.

22. Military and Paramilitary Activities in and against Nicaragua (Nicar. v. U.S.), Judgment, 1986 I.C.J. Rep. 14, ¶ 205 (June 27).

23. *See, e.g.*, Steven J. Barela, *Cross-Border Cyber Ops Erode Legitimacy: An Act of Coercion*, JUST SEC. (Jan. 12, 2017), https://perma.cc/LBD6-4HAX (making the case that it did, in fact, constitute an infringement on sovereignty).

24. OHLIN, *supra* note 6, at 83.

25. *Id.* at 85.

> It is equally clear that merely engaging in election propaganda does not amount to interference, at least as a matter of law. This conclusion is supported by the extensive State practice of engaging in both truthful and untruthful propaganda during foreign elections. Of course, such activities may be condemned, as the efforts of RT and Sputnik and the purchase of advertising on social media were in the ODNI Report, but such condemnation is seldom based on assertions of a breach of international law, specifically the obligation to respect sovereignty. This paucity of *opinio juris* and surfeit of contrary practice corroborates the conclusion that election propaganda by cyber-means does not violate a target State's sovereignty.[26]

Schmitt's interpretation is one that many scholars have endorsed, including Ohlin. Ohlin himself believes that the best way to characterize the harm of Russian interference in the 2016 election is to suggest it undermined self-determination.[27] Democracy, Ohlin writes, requires democratic deliberation. For deliberation to be exercised freely, four basic conditions need to be met: legally protected freedom of thought, legally protected freedom of speech, legally protected freedom of the press, and unrestricted access to information.[28] These conditions are not meant to protect the freedom of speech and press of foreigners, but rather of members of the domestic political community.[29]

Ohlin proposes the distinguishing feature of Russian interference in 2016 was that it violated the core "boundary" regulations involving election integrity: the prohibition on outside participation in elections.[30] In so doing, it undermined the United States' right to self-determination. If domestic persons had conducted the kinds of activities that the Russian actors did, the "corruption would have come from within the polity itself and therefore would not have displaced the people's will with the will of an outside actor."[31] Ohlin argues that while Russian social media activity "corrupted the political discourse precisely because it had a *foreign source*," American social media activity expressed, and did not compromise, the will of the American people.[32]

Many scholars believe that in order for liberal democracies to flourish, people need access to quality information. As Michael Lynch argues, democracies have a special interest in institutions, methods, and sources that help citizens reliably pursue truth because, in order to be healthy, democracies need to have an informed populace to enable effective deliberation and promote epistemic justice.[33] What sets Ohlin's account apart from that of Lynch and others concerned

26. Michael N. Schmitt, *Virtual Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law*, 19 CHI. J. INT'L L. 30, 46 (2018).

27. OHLIN, *supra* note 6.

28. *Id*. at 99-100.

29. *Id*.

30. OHLIN, *supra* note 6, at 119.

31. *Id*. at 101.

32. *See id*. (emphasis added).

33. *Watch Michael Lynch on the Democratic Value of Truth*, PERITIA (May 4, 2021), https://perma.cc/MK7Z-GPYN.

with the importance of truth for democracy is Ohlin's emphasis on the foreign source of influence. Ohlin maintains it is this deceptive foreignness that renders the influence an impediment to democracy (and therefore to self-determination), not the quality of information.[34]

Ohlin's account has merit to it, but the self-determination framework is not without its flaws. As Ohlin recognizes and other scholars, including Schmitt, have underscored, there are legal reasons that scholars might hesitate to use the self-determination argument. First, arguments based on self-determination are typically used when groups are trying to create a state; second, the application of self-determination does not typically apply where "the people are all citizens of a State rather than a distinct group therein that is denied the right to govern itself, as in cases of colonialism, apartheid, alien subjugation, and perhaps occupation."[35] Legal arguments based on self-determination do not usually apply to a people like the American people.

These legal arguments against the use of the sovereignty and self-determination frameworks need not be the final word on the matter. It is worth exploring whether influence should be considered a *kind* of infringement of sovereignty or impediment to self-determination even if, strictly speaking, it does not fit within the legal category. I argue that the answer is no, but not because of legalism. Rather, it is because accounts that strive to prove otherwise rest on the mistaken assumption that a speaker's "foreignness" is what matters the most.[36] In what follows, I will focus my critique on Ohlin's argument concerning self-determination because it is one of the strongest presented in legal scholarship.

To begin with, it is necessary to separate two prongs in Ohlin's argument. The first is that speaker-identity deception is wrong; the second is that foreign participation in online exchanges during election season undermines self-determination because foreigners are not part of the polity. These two points are often tied in the argument. For example, Ohlin writes:

> [I]t is unclear why one should view the expression of lies by foreigners as any more corrosive to the political system than the expression of pure opinion by

34. OHLIN, *supra* note 6.

35. Schmitt, *supra* note 26, at 55-56.

36. I am not the first person to critically appraise our growing preoccupation with foreignness. As Alicia Wanless and James Pamment articulated in "How Do You Define a Problem Like Influence?," the emphasis on foreignness raises a number of questions:

> How relevant is a country's legal definition of foreign to determining how acceptable a participant is in public debate? Are diasporas in their new home legitimate actors in public debate in their country of origin? Are illegal immigrants legitimately able to influence the politics in the countries in which they hope to stay? How does one account for proxy or sympathetic actors ('useful idiots') who may be persuaded or coerced into supporting the goals of a foreign state? What about public diplomacy or grant programs that help activists in other countries—or even one community in the same country who disagrees with the decisions of another? Wanless & Pamment, *supra* note 9, at 5.

While Wanless and Pamment did not directly respond to the questions they raised, they drew attention to a critical issue.

> foreigners . . . what made the participation of foreign actors so corrosive in the 2016 election was not just the content of the social media postings but the fact that the opinions were made by Russians pretending to be Americans. The most salient characteristic of the deception was not deception in the content of the statements but rather deception regarding the identity of the speaker.[37]

Ohlin is right to be concerned with speaker-identity deception, a point I will return to when I consider normatively troublesome kinds of influence. Ohlin is also right that self-determination is an important principle in international affairs.[38] States are often considered the correct locus for a variety of important social, political, and ethical questions, and scholars have offered justifications in both ideal and non-ideal theory for why using the state as an organizing principle is appropriate or even desirable. For example, Michael Walzer, among others,[39] has argued that self-determination within states generates a sense of collective identity.[40] It is not for nothing that Kosovo wanted to secede from Serbia, or that the Kurdish people want their own state.

What Ohlin gets wrong is positing that democratic deliberation influenced by online foreign speakers undermines self-determination. Since 2019, there has been increased scholarly attention to domestic information dysfunction, with new studies suggesting that many of the information woes plaguing democracies – from disinformation to inflammatory rhetoric to dangerous speech – are domestic in origin and connected to traditional media outlets, such as Fox News.[41] Ohlin's account does not sufficiently accommodate the complex interplay online between domestic and foreign actors, domestic and foreign speech, and domestic and foreign ideas, which I crystallize in three objections.

The first problem with Ohlin's account is that a foreign influence effort cannot succeed without a willing domestic audience to take it up, echo it, and amplify it.

---

37. OHLIN, *supra* note 6, at 153.

38. I am grateful to Milton C. Regan Jr. for emphasizing the importance of this point.

39. *See, e.g.*, DAVID MILLER, ON NATIONALITY (1995) (defending the principle of nationality).

40. MICHAEL WALZER, JUST AND UNJUST WARS 53 (2015). According to Walzer, the world is comprised of distinct communities of individuals. Individuals within a community have an implicit contract among "the living, the dead, and those who are not yet born" to maintain a shared way of life. *Id.* In Walzer's words: "Over a long period of time, shared experiences and cooperative activity of many different kinds shape a common life. 'Contract' is a metaphor for a process of association and mutuality. . ." *Id.* at 54. States are, presumably, although not always in practice, the political expression of these communities and are an effective means of advancing collective self-determination. According to Walzer, it is valuable for these communities to pursue their own life together without the constant involvement of foreign states and their armies. Walzer's argument, however, has been criticized. For example, David Luban argues that the value of self-determination can only be appreciated in a context wherein the community genuinely has the opportunity and the capacity to shape affairs. David Luban, *Just War and Human Rights*, 9 PHIL. & PUB. AFF. 160 (1980). Luban has highlighted, quite rightly, that there is a difference between (1) *a distinct political community* whose right to self-determination we, as outsiders, respect, and (2) *the regime* ruling over this political community, which may or may not respect the rights of its citizens. *Id.*

41. *See, e.g.*, THE DISINFORMATION AGE (W. Lance Bennett & Steven Livingston eds., 2020); YOCHAI BENKLER, ROBERT FARIS, & HAL ROBERTS, NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS (2018).

If this is right, then the kind of foreign influence Ohlin is worried about becomes an integral part of self-determination, rather than an interference with it.[42] Consider this example: a piece of information originates outside of the United States, reaches the United States, and is taken up by Americans and spreads widely. If the information were taken up and spread, it is plausible to say that at least some Americans found it a relevant piece of information to share for the sake of political debate. The quality of information may be poor, which hurts democratic decision-making – a point that Michael Lynch and others, as already noted, have made – but the quality of information is not what concerns Ohlin. What concerns Ohlin is that foreign actors hiding their identities online are undermining the domestic polity's ability to self-determine. But if domestic actors within that polity seek, engage with, and share the material, it is harder to make the case self-determination has been undermined.[43] The polity is less passive than Ohlin suggests, and agency is expressed in decisions about which information to debate and spread.

Ohlin could argue that the polity is being duped, and if there were more transparency about the identity of the foreign speaker when the domestic hearer encountered the information, the information would not have been spread across the polity. This seems possible, but it is not strong enough to show that self-determination has been undermined.

The second problem with Ohlin's account is that it rests on the idea that it is possible to establish and maintain a distinction between foreign and domestic influence. As a conceptual matter, I argue that this is untenable. While it is possible (although difficult) to establish the origin of a piece of information, it is much harder to establish the origin of an *idea* or draw a sharp line between the domestic and the foreign. What if a domestic speaker on Fox News came up with an opinion that was based, at least in part, on arguments made by foreign speakers writing for Sputnik or on a Reddit forum? At what point does the opinion become foreign and not domestic? It is more helpful to consider the exchange of ideas online as part of a dynamical, interconnected system. An idea may be suggested in the United States on Fox News, be amplified by foreign actors who see the power of the idea, and then make its way back into the United States where the idea proliferates across various platforms.

It is difficult to respond to information-based influence campaigns in part because of how information moves across domains and borders. A story, a meme, or an idea can permeate multiple spaces on the internet. To understand influence

42. I am grateful to David Luban for articulating this point to me.

43. There is a separate question about how we should morally evaluate domestic speakers who amplify the messages of foreign governments for their own personal or political gain, and not because they earnestly support the idea. P.W. Singer and Emerson Brooking write, "Those who deliberately facilitate enemy efforts, whether it be providing a megaphone for terrorist groups or consciously spreading disinformation, especially that from foreign government offensives, have to be seen for what they are. They are no longer just fighting for their personal brand or their political party; they are aiding and abetting enemies that seek to harm all of society." P.W. SINGER & EMERSON T. BROOKING, LIKE WAR: THE WEAPONIZATION OF SOCIAL MEDIA 266 (2018).

efforts, we need to recognize, not downplay, the role domestic speakers play in the promulgation of the same information as foreign speakers. Because the internet does not, in an important sense, have strict territorial boundaries, changing the status of information based on whether it is 'foreign' or 'domestic' obscures how information usually spreads online.

This does not mean that we cannot make some useful distinctions; we can, as I explain earlier, show evidence that Russia mounted a coordinated campaign against the United States, and we have strong evidence of many other campaigns waged by various States against one another. This argument also does not mean that participation of foreign actors does not matter. When a State amplifies information that serves its interests abroad, it is important for all various political, epistemic, and legal reasons. But what it does mean is that the paradigm of self-determination, which depends on these strong boundaries between the foreign and the domestic, is not fitting for contexts involving speech online, as Ohlin suggests it is.

Finally, it is worth critiquing Ohlin's emphasis on foreignness. This is not to straw-man Ohlin's argument; his case is focused on deceptive foreign speech, not foreign speech per se. Moreover, Ohlin ensures his recommendations for policy changes are consistent with the right to receive foreign speech.[44] But a thread connecting multiple points Ohlin poses is that foreigners should not have a say in elections, and the boundary between the foreign and the domestic ought to be robustly maintained.[45] To be sure, it is legally permissible to limit the speech of those in many countries and contexts. In the United States, two prominent examples include the Foreign Agents Registration Act (FARA), which requires foreigners working on behalf of another government to register with the Justice Department, and foreign expenditure laws, which ban foreigners from making campaign donations.[46] These laws rest on a similar proposition to that at the bedrock of Ohlin's argument: we ought to treat foreign speakers differently irrespective of the truth content of their views, since these laws target speakers on the basis of their nationality or relationship to a foreign government.[47]

These laws are important, but it is unclear that FARA and foreign expenditure laws are applicable to all online discourse.[48] Moreover, someone who fully

44. *See generally* Ohlin, *supra* note 6.

45. There is another line of argument (made to me by David Luban) we might pose against Ohlin: why shouldn't foreigners have a say in American elections if these electoral choices are going to impact their country? The argument about self-determination may not make sense if the State in question (such as the United States) has the power to impose huge externalities on people in other parts of the world. Ohlin finds this line of argumentation unpersuasive. *See* Ohlin, *supra* note 6, at 122. I do not pursue this line of argument because I believe it is possible to show Ohlin is wrong even if we accept the claim that the United States, like other countries in the world, has a right to self-determination.

46. *See* FARA, 22 U.S.C. §§ 611-621; Prohibition on Contributions, Donations, Expenditures, Independent Expenditures, And Disbursements By Foreign Nationals, 11 CFR § 110.20 (2004).

47. I am grateful to Ricky Altieri for making this point.

48. For examples of litigation against online discourse, *see, e.g.*, Joshua R. Fattal, *FARA on Facebook: Modernizing the Foreign Agents Registration Act to Address Propagandists on Social Media*, 21 N.Y.U. J. Legis. & Pub. Pol'y 903 (2019).

endorses the existing laws need not be committed to the view that *all* foreign speakers in *all* contexts ought to be marked out for their foreignness. A transparency regime that marked out foreigners precisely for their foreignness is the wrong target. On my view, our focus should be on deception, not foreignness, as I explain below.

The best argument Ohlin can offer for the wrongfulness of the 2016 Russian influence campaign – and one that he does empasize at various points – is that Russia's influence was *deceptive*.[49] That is, Russian speakers hid their identities online, posing as American citizens, and it was by virtue of this deception that they were able to influence the polity. As Ohlin writes, the Internet Research Agency (IRA) social media accounts were made to appear American.[50]

This is a compelling argument. However, the persuasiveness of the argument does not rest on the foreignness of the speaker; it rests on the speaker's deception. Put differently, the objection to speaker-identity deception is not based on the principle of self-determination. To see this, consider that speaker-identity deception can happen among citizens of the same State. Speaker-identity deception by Americans pretending to be other Americans (for example, a right-wing troll pretending to be a liberal democrat) still seems problematic in some cases; but it is not a problem per se for Ohlin because it does not, on his view, undermine self-determination. I will return to the wrongfulness of speaker-identity deception in Part II.[51]

To conclude Part I, influence *is* not, legally speaking, a kind of warfare, an infringement of sovereignty, or an impediment to self-determination. Moreover, as I have argued, influence *should* not be conceived of as a kind of warfare, a kind of sovereignty infringement, or a kind of self-determination impediment. My case is multifaceted and rests on a number of conceptual and normative arguments. Ultimately, the dominant paradigms in international law – warfare, sovereignty, self-determination – are inadequate for capturing a problem like influence, and do not provide us with the right tools for understanding or responding to the phenomenon.

How, then, ought we to evaluate the permissibility of information-influence campaigns? We turn to this question in Part II.

II. BEYOND FOREIGN INFLUENCE: A FOCUS ON NEFARIOUSNESS, NOT FOREIGNNESS

My aim in Part II is to explore approaches to framing information-based influence that de-center foreignness and do not rely on the war, sovereignty, and self-determination paradigms. I begin by articulating distinctions between persuasion, manipulation, and other kinds of influence to underscore the difficulty in drawing bright lines between permissible and impermissible influence. I then sketch a

49. *See generally* Ohlin, *supra* note 6.
50. Ohlin, *supra* note 6, at 21.
51. *See* discussion *infra* text accompanying notes 75-85.

promising framework, provided by Duncan Hollis, and build on this by adding additional normative considerations related to transparency and deception. While there are no simple checklists for us to rely on, we can and should continue developing a richer apparatus for evaluating the ethics of influence.

A. Starting Intuitions: What Counts as Permissible Influence?

In Jane Austen's most poignant novel, *Persuasion*, the protagonist, Anne Eliot, is dissuaded from marrying a person she loves by Lady Russell, the best friend of her late mother. Lady Russell advises Anne against the match with Anne's best interests in mind, and Anne trusts her guidance. Seven years after rejecting the man, Anne finds herself reacquainted with him and, because this is a Jane Austen novel, they fall in love all over again. This time, there is no one standing in the way of their getting married and, even if there were, it is unlikely Anne would be dissuaded again from marriage—she is older, knows her mind better, and relies less on the judgment of advisors, however trustworthy they may be.

Was Lady Russell's influence over Anne permissible? Many might think that persuasion – that is, convincing someone through arguments or reasons to adopt some belief or take some action – is a morally acceptable kind of influence. We can contrast persuasion with other forms of influence that are generally taken to be morally unacceptable: coercion, undue inducement, "no choice" situations, and cases involving deception.[52] Persuasion, unlike these other forms of influence, plays a critical role in everyday interactions, and without it, humans would be without a vital tool. Parents persuade children to finish their homework, friends persuade one another to go out for drinks, employees persuade employers to give them a raise, and so on. On this view, in persuading Anne not to marry, Lady Russell did something that is permissible and very common.

Others might disagree with this framing and argue that persuasion is a kind of tool that can be used for a variety of purposes, and whether the persuasion is permissible depends on the end pursued. Lady Macbeth persuading her husband to kill King Duncan is a morally bad use of the tool, whereas Harry Potter persuading Horace Slughorn to share a painful memory because it will help defeat

52. Coercion, roughly defined, is the threat to make someone seriously worse off than they are or should be, unless they consent. *See* ALAN WERTHEIMER, COERCION (1987). Because there is something in the act of coercing that we might find morally objectionable, even if the ends of coercion might be constructed as good – for example, a child is coerced into being kinder to her sibling or trying harder – the act of coercion is morally wrong. Undue inducement is "a term of art meaning that something is being offered that is alluring to the point that it clouds rational judgment, for instance cash in hand or airline tickets in return for kidney donation or risky study participation." Nir Eyal, *Informed Consent*, STANFORD ENCYCLOPEDIA PHIL. (2019), https://perma.cc/Q9W7-2QZ8. In "no choice" situations, the lack of decent alternatives to accepting a bad offer are said to compel someone to choose the offer; these are "your money or your life" situations. *Id.* This kind of influence is taken to be impermissible, and meaningful consent cannot be given if someone is put into a no choice scenario. In cases involving deception, one is deceived in some way and, had the deception not occurred, one would not have consented. For example, person A is deceived into believing that person B is wearing a condom and, on the basis of this deception, consents to have sex. Person B has acted impermissibly (and, as a side matter, has engaged in an activity that many scholars and activities constitute as rape).

Voldemort is a morally good use of the tool. Turning back to Austen, we might say that irrespective of how well-intentioned she was, Lady Russell was giving bad advice to Anne, and the end pursued (to convince Anne not to marry for love) was not a good end. The persuasion was not acceptable, let alone admirable.

Still others may think that we have not yet fully articulated the complexity of the issue. Determining whether A's influence over B is permissible depends not just on the tool used or end pursued, but on a variety of other factors. What was the power relationship between A and B? Was B in a position or mental state to understand the kind of influence being exerted by A? Austen hints that Lady Russell is using her status as an authority, at least in Anne's eyes, just as much as a reason to convince Anne not to marry. She is using ethos as well as logos, to use Aristotle's terminology, to persuade.[53] Given this state of affairs, it is reasonable to conclude that Anne was persuaded not merely because of reasons, but because of appeals to authority.

There are still other questions we may ask to ascertain the permissibility of the influence. Did A stand to gain something by influencing B, and was B aware of this gain? Is A primarily appealing to reasons or to emotions? Does A's influence over B constitute a kind of manipulation?

Manipulation is, generally speaking, a kind of influence over someone's beliefs, desires, and behaviors without their conscious awareness. According to Daniel Susser, Beate Roessler, and Helen Nissenbaum, manipulation involves the exploitation of a target's cognitive or affective weaknesses for the purpose of steering the target's decision-making process toward the manipulator's ends.[54] By eroding our ability to steer ourselves towards the ends we would critically endorse for ourselves, manipulation undermines our "self-authorship," to use their terminology. Here, the authors build on Joseph Raz's work on autonomy; according to Raz, "the ruling idea behind the ideal of personal autonomy is that people should make their own lives."[55] This account of manipulation is useful because it draws attention to the fact that the target of manipulation is often being influenced in a manner that serves someone else's (the manipulator's) ends. It

53. *See* Aristotle, Rhetoric (350 B.C.E). Persuasion is a function of logos (reason), pathos (emotion), and ethos (assessment of the credibility and character of the speaker). *Id.*

54. Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in A Digital World*, 4 Geo. L. Tech. Rev. 1, 3 (2019). Our vulnerability to manipulation is deeply human. As Susser, Roessler, and Nissenbaum explain:

> Famously, Daniel Kahneman and Amos Tverksy demonstrated that people are often influenced by irrelevant information (so-called "anchoring effects") and give more weight to evidence they can easily recall (the "availability heuristic"). People draw different conclusions from the same information depending on how it is presented ("framing effects"), and so on. Because these cognitive biases are widespread and predictable, manipulators can easily treat them as vulnerabilities to exploit. Manipulators can remind targets of unimportant facts so that they give them undue weight. They can point out that their targets' friends believe certain things in hopes that they will believe them too. They can frame information in misleading ways. Manipulation, therefore, need not involve outright deception; the truth can also be used to control our decision-making. *Id.* at 21–22.

55. *Id.* at 35.

also stresses the particular ethical harm of manipulation: the diminishment of autonomy.

If advertisements, for example, are surreptitiously appealing to emotions rather than offering reasons, they are manipulation rather than persuasion on this definition.[56] Not all cases are so clear-cut. Separating manipulation from persuasion proves to be difficult in some contexts. Was Lady Macbeth persuading Macbeth or manipulating him when she provided both reasons to kill *and* appeals to emotion? Are advertisements manipulative when they appeal to certain emotions or desires (without making it transparent that they are doing so) while also offering reasons? In these cases, it is difficult to separate persuasion from manipulation.

How we draw the line between the permissible persuasion and impermissible manipulation matters crucially for political life. The kind of influence governments, corporations, and media organizations are permitted to yield affects how people think, vote, and engage with one another. Given this, determining what falls beyond acceptable influence – perhaps a tobacco company is no longer allowed to advertise its products, or Donald Trump is banned from X – is an important political question. In liberal democracies, where there exist robust protections for expression,[57] the debate about influence is inseparable from the issue of how much and what kinds of speech should be permitted in different fora.

As we saw in Part I, the dominant frameworks in international law are inadequate for addressing the phenomena. How, then, should we conceptualize and evaluate information-based information? Scholars have developed different frameworks for assessing the permissibility of influence campaigns, and I now turn to and build on a promising account that articulates relevant factors for assessing influence.

B. Hollis' Framework

A strong account is provided by Duncan Hollis who, in his paper "The Influence of War; the War of Influence," seeks to answer the question, where do we draw the line between influence operations[58] that "are normal parts of human interaction" and those that are legally and morally controversial?[59] Hollis provides five factors that can be used to distinguish between influence operations (IOs): transparency, deception, purpose, scale, and effects.[60] One of the virtues of

56. Vance Packard has argued that the advertising industry has appealed to non-rational and subconscious mental processes—that is, the industry has exploited human cognitive weaknesses to sell us things. *See* Vance Packard, The Hidden Persuaders (1957).

57. The extent of protection for expression, of course, varies on a country-to-country basis. Hate speech, for example, is illegal in a number of liberal democracies, in a variety of contexts, but is tolerated in others. More generally, many states regulate expression on the internet more forcefully than the United States does.

58. Hollis uses the term "influence operations" while I more regularly use "information-based influence campaign." We are tracking, more or less, the same phenomenon. *See* Duncan Hollis, *The Influence of War; The War of Influence*, 32 Temp. Int'l. & Comp. L. J. 1, 2 (2018).

59. *Id*. at 36.

60. Other scholars have proposed helpful frameworks for analyzing influence operations. For example, Wanless and Pamment analyze influence operations in terms of "*intent* behind communication; the *truth* of

Hollis' framework is that it does not matter whether the speaker is foreign, as we will see.

The variables Hollis introduces help us isolate those aspects of influence that are of greatest concern. These variables also provide us with the vocabulary and conceptual tools to talk about influence in a fruitful way. Concerning transparency, Hollis writes that we can separate "white," "gray," and "black" influence operations.[61] In white operations, "the State is open and transparent in its authorship or responsibility for the resources deployed."[62] In black operations, the author of the influence operation "false flags" the source, most often "to a hostile adversarial state, group, or individual."[63] Gray operations are ones in which the origins of the source are ambiguous or unknown. According to Hollis, "[t]he question is whether such differences – alone, or in combination with other factors – should help us delineate acceptable from unacceptable [influence operations]. Might 'white' IOs be per se more acceptable than 'black' ones?"[64]

Hollis then addresses the second factor, deception, and considers whether "the veracity of the information resources deployed might thus be another ground for differentiating among IOs."[65] He writes that while some influence operations leak accurate information (for example, WikiLeaks released John Podesta's authentic emails), other influence operations rely on disinformation.[66]

The third factor is purpose. Some influence operations "target public attitudes and *dispositions generally*; they attempt to *shift sentiments* rather than particular positions."[67] Other influence operations are more targeted, seeking to generate specific behavioral outcomes: for example, casting a particular vote. Hollis recognizes that a purpose-based criterion for delineating between influence operations invites a broad range of new questions on what purposes warrant segregation: "Is it just IOs that seek violent outcomes? What about those designed to impact elections? What about IOs that target a shift in a specific legislative outcome or foreign policy? In other words, when should we delineate an IO as inappropriate based on what it appears designed to do?"[68]

Hollis writes that the fourth way we can differentiate is by looking at scale. Influence operations were once "relatively costly, requiring substantial resources to generate modest effects."[69] Now, the information age has "significantly scaled up what can be done."[70] Hollis' examples of "low scale" influence operations also fall into the "old school/pre-information age" era, while his examples of "high scale"

---

the communication; the *origin* of the communicating actor; and the *legitimacy* of the communication techniques used." Wanless & Pamment, *supra* note 9, at 8.

61. Hollis, *supra* note 58, at 36.
62. *Id.*
63. *Id.*
64. *Id.* at 37.
65. *Id.*
66. *Id.*
67. *Id.* at 37.
68. *Id.*
69. *Id.*
70. *Id.*

influence operations sit comfortably in the "contemporary information age" era.[71] It is unclear, based on the text alone, whether this is intentional or accidental. It is entirely possible, I suggest, that looking at only contemporary influence operations, we may still distinguish between low-scale and large-scale ones.

The fifth variable for distinguishing between influence operations is "actual effects." Hollis posits that "the greater the guarantee that an IO will involve a measurable loss of human agency or free will, the more problematic the IO becomes."[72] Influence operations that successfully leverage the cognitive and emotional biases of individuals, leaders, groups, and networks are more dangerous.

Hollis' framework is helpful for understanding and comparing various kinds of influence operations, and it also offers a starting point for thinking about the permissibility of certain kinds of information-based influence campaigns. Going forward, I will build on Hollis' framework to add further substance to our evaluation of information-based influence campaigns.[73]

I will introduce a set of normative considerations that can help us think through what makes certain kinds of information-based influence more or less ethical. I focus on speaker-identity deception (whether a speaker is being forthright about their identity, hiding their identity, or lying about their identity), and quality of information (whether information is true, accurate, and non-misleading, or whether it is false, inaccurate, and misleading).[74]

---

71. *Id.*

72. *Id.* at 38.

73. There are, however, significant limitations to this framework. First, Hollis distinguishes between "normal parts of human interaction" and influence that is legally and morally controversial. *Id.* at 36. This is a false distinction: many normal parts of human interaction can be controversial. Consider espionage and political advertisements that target cognitive biases. Both are normal parts of human interaction and yet they remain controversial because they involve deception and manipulation. Second, there seems to be a missing distinction between those factors that are morally dubious even if used for good purposes (like a lack of transparency and deception) and those that are not (like scale). Deception is often an attempt to fool other people into trusting something that is not trustworthy. Even if its scale is small, or its purpose good, we can say that it is intrinsically morally questionable, while truthfulness is not.

74. Ultimately, we should build toward a theory of influence, one that is perhaps similar in structure to how we think about ethics of warfare. For those in other disciplines, here is a sketch of the Just War literature. The publication of Michael Walzer's *Just and Unjust Wars* in 1977 marked the renaissance of scholarly focus on Just War Theory, a moral theory that denies both pacifism and realism. *See* MICHAEL WALZER, JUST AND UNJUST WARS (1977); *see also* William E. Murnion, *A Postmodern View of Just War*, *in* INTERVENTION, TERRORISM, AND TORTURE: CONTEMPORARY CHALLENGES TO JUST WAR THEORY 23, 29-30 (Steven P. Lee ed., 2007). While classical Just War Theory has its origins in Christian theology, the post-Walzer literature has recast the basic question of "why and how are wars justified" in secular terms. *See* Murnion, *supra* note 74 at 26-28, 35. The literature on Just War Theory, as conventionally conceived, ranges over two main fields of inquiry. The first, known as *jus ad bellum*, specifies the conditions that must be satisfied in order for the resort to war to be morally justified. *See* WILLIAM V. O'BRIEN, THE CONDUCT OF JUST AND LIMITED WAR 13 (1981). The second, *jus in bello*, addresses the moral permissibility of conduct in war by individual participants. *See, e.g.*, Murnion, *supra* note 74, at 28. In addition to *jus ad bellum* and *jus in bello*, two more fields of Just War Theory have been brought to the fore as a result of military failures in the past few decades: justice in ending wars (*jus ex bello*) and justice following war (*jus post bellum*). Accounts of the content of *jus ad bellum* vary in their details, but the following six conditions represent a general consensus: (1) *just cause:* the

C. Capturing Nefarious Influence

In de-centering speaker foreignness in the influence debate, we are left with room to decide what should be at the center of our discussion. I argue it should be the actual features of influence that we take to be normatively questionable because, in and of themselves, they are morally dubious. I focus on speaker-identity deception and information quality, which Hollis refers to as transparency and deception, respectively, because these are most normatively charged.

The factors I introduce are not exceptionless principles; there will be, as we will soon see, various contexts where exceptions must be made. For this reason, there is no simple set of rules; it is not possible, using these factors alone, to determine whether a given campaign is permissible or impermissible.[75]

However, it is possible, using these factors, to get a better grip on how strong a normative justification we can give for the campaign. There may be other normatively relevant considerations – for example, a speaker's intentions – that bear on these other factors but which, I suggest, we ought to sideline because of their slippery nature. I will discuss them where relevant.[76] I focus, however, on the normative factors that are most tractable.[77]

---

war must have a just cause; (2) *right intention*: it must be fought with the right intention(s); (3) *proportionality*: the harm caused by the war must be proportionate to the good achieved; (4) *last resort*: it must be the last resort; (5) *probability of success*: it must have a reasonable prospect of success; and (6) *proper authority*: it must be initiated and waged by a proper authority. *See* O'BRIEN, *supra* note 74, at 16-36. These conditions are generally taken to be individually necessary and jointly sufficient for justifying the resort to war: should one of the criteria not be satisfied, the war is deemed unjust, and if all criteria are satisfied, the resort to war is justified. *See id.* at 35-36. As noted above, the *jus in bello* conditions concern permissible conduct during war. Typically, the *jus in bello* list comprises: (1) *discrimination*: belligerents must always distinguish between military objectives and civilians, and intentionally attack only military objectives; (2) *proportionality*: foreseen but unintended harms must be proportional to the anticipated military advantage; and (3) *necessity*: the least harmful means feasible must be used. *See id.* at 38-43, 64-66. To gain further insights about historical Just War Theory, see the works of Greg Reichberg, Pablo Kalmanovitz, Daniel Schwartz, and Rory Cox. In particular, *see* Rory Cox, *The Ethics of War up to Thomas Aquinas*, *in* THE OXFORD HANDBOOK OF ETHICS OF WAR 99 (Seth Lazar & Helen Frowe eds., 2015); Pablo Kalmanovitz, *Early Modern Sources of the Regular War Tradition*, *in* THE OXFORD HANDBOOK OF ETHICS OF WAR 145 (Seth Lazar & Helen Frowe eds., 2015); Gregory M. Reichberg, *The Historiography of Just War Theory*, *in* THE OXFORD HANDBOOK OF ETHICS OF WAR 59 (Seth Lazar & Helen Frowe eds., 2015); Daniel Schwartz, *Late Scholastic Just War Theory*, *in* THE OXFORD HANDBOOK OF ETHICS OF WAR 122 (Seth Lazar & Helen Frowe eds., 2015).

75. I distinguish my approach from Ohlin's because his is categorical while mine is contextual: Ohlin suggests that speaker-identity deception by a foreign actor is per se impermissible, whereas I suggest that analysis should proceed on a case-by-case basis. *See* Jens David Ohlin, *A Roadmap for Fighting Election Interference*, 115 AM. J. INT'L L. UNBOUND 69, 69-71 (2021). I also diverge from traditional Just War Theory which suggests the fulfillment of the relevant *jus ad bellum* criteria renders a war permissible.

76. *See* discussion *infra* text accompanying note 86.

77. I also put to the side questions relating to censorship and deletion because the warrant separate attention. Briefly, my view is that liberal democracies ought to favor a rich information environment, although there will be many exceptions and various circumstances in which deletion is appropriate, such as taking down revenge porn and stopping the livestreaming of a murder.

Let's begin with speaker-identity deception, which Hollis discusses in "transparency" category. In the wake of Russian meddling in the 2016 U.S. presidential election, many commentators, including Ohlin, pointed to speaker-identity deception as a problematic feature of the campaign.[78]

Some states also regularly engage in speaker-identity deception to shape what their own populace thinks and talks about. The Chinese government, for example, pursues a misdirection strategy on social media: the overarching goal is to change the subject or crowd out negative and critical posts with positive messages.[79] The government orchestrates the broader message while "[i]nternet commentators," commonly known as the "50c army" (because, reportedly, that is how much they are paid per post), do the grunt work.[80] This is a type of "astroturfing", a deceptive practice of posting large numbers of social media comments ghost-written by the government as if they were genuine opinions.[81]

Speaker-identity deception is, in most cases, a morally dubious kind of influence because it involves lying. One need not be committed to a strong deontological stance to recognize that lying, generally speaking, is morally suspect.[82]

Lying is also epistemically troublesome. A speaker's identity is epistemically important because it can be a proxy for other factors that we need to know in

78. *See generally* Ohlin, *supra* note 6.

79. *See* Gary King, Jennifer Pan & Margaret E. Roberts, *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument*, 111 AM. POLI. SCI. REV. 484, 484-85 (2017).

80. *See id.*

81. *See* Adam Bienkov, *Astroturfing: What Is It and Why Does It Matter?*, GUARDIAN, (Feb. 8, 2012, 10:17 AM), perma.cc/6E8V-5U8L.

82. What is wrong with lying? One answer is that lying harms its victims: when B relies on false information, B raises the chances of incurring various bad consequences. Arguably, B is worse off if she believes something false even if this does not translate to any bad consequences. As Claudia Mills explains, "In getting B to believe something false, A brings it about that one item of B's mental furniture is seriously flawed–even if it never collapses on [B], [B] has been injured by the transactions. Claudia Mills, *Politics and Manipulation*, 21 SOC. THEORY & PRAC. 97, 103 (1995). Moreover, lying "increases the chances we will be led to believe false things, develop bad desires, or make wrong choices . . . the harm here lies not in the end consequences . . . but in the sheer fact of our holding false beliefs and unworthy desires." *Id.* at 103-04. A second reason that lying is wrong is because in many cases it is a violation of trust. *See id.* While "anyone who purported to be surprised by deceit and manipulation in high office would show herself to be preposterously naïve," it is nevertheless the case that "we ought to be able to trust our elected representatives" and "public accountability lies at the heart of representative government." *See id.* at 105. A third reason is that lying reveals bad character. Character is "relevant to fitness to govern," and "[t]he habitual dishonesty of the liar is a serious character flaw" that may render a person an inappropriate choice for a political position. *Id.* We may extend Mills's point and say that in cases of speaker-identity deception, lying conceals bad character. There is a final reason that merits emphasis. This one is not articulated by Claudia Mills but by Hannah Arendt (although not in the context of the wrongfulness of lying). Arendt writes, "It is . . . [human] fragility that makes deception so very easy up to a point, and so tempting." HANNAH ARENDT, *Lying in Politics: Reflections on the Pentagon Papers*, *in* CRISES OF THE REPUBLIC 1, 6 (1972). Lying is therefore not only a violation of trust, but an attack on a natural vulnerability—a feature of humanity that makes us fragile and, in cruder terms, easy targets.

order to assess the reliability of the information.[83] If I am an online user who encounters information on gun control by an NRA lobbyist, the speaker identity justifiably shapes how I encounter the information: *Do I dismiss the information immediately? Do I check with another source? Am I the target of manipulation?*

Knowing a speaker's identity helps us understand the epistemic standards they might be employing, whether they are biased in particular ways, and so on. To be sure, many speakers online choose to engage anonymously. Anonymity, however, is not the same as speaker-identity deception. If a hearer knows that a speaker is intentionally hiding their identity, the hearer knows they are missing a critical piece of information: who the speaker is. But if a hearer thinks they know who a speaker is but in fact does not because the speaker has lied about their identity, the hearer is epistemically worse off.

There is a difference between hiding one's identity and fabricating one's identity. To hide is not necessarily to deceive. For example, I could tell you online, truthfully, "I would prefer not to reveal my identity." Similarly, I could engage online in communities where anonymity, not disclosure, is the dominant norm. Determining whether hiding one's identity is deceptive becomes intricately linked to context. On Reddit, it may not be considered deceptive because it is a community norm to engage anonymously or with few identity markers. In other online communities or speech arenas where a speaker's identity is considered important – such as X or Facebook – hiding one's identity can be deceptive. Moreover, concealing one's identity may be deceptive when the speaker knows that the audience is likely to attribute the information to another particular actor. One therefore need not pretend to be that actor in order to achieve this effect.

But is being deceptive about one's identity always morally problematic? What are the exceptions to this rule? Consider these two cases:

1. ***Anti-Racialization.*** Government employees are asked to infiltrate violent online extremist communities in an effort to persuade members of the community, who are located in different parts of the world, to question their beliefs. The government employees only share true, accurate, and non-misleading information, but they are deceptive about their identities. If the members of the online extremist group knew they were talking to government employees undercover, the extremist group members would not engage with the employees at all.[84]

---

83. I am grateful to Megan Ritz for emphasizing this point to me.

84. This case is based on real practices. The "77th Brigade [in the United Kingdom] maintains a small presence on Facebook and Twitter under its own name"; the accounts and content are "clearly marked as government operated." *See* Bradshaw & Howard, *supra* note 78, at 10-11. The purpose of the unit is to shape public opinion through the use of "dynamic narratives that undermine political propaganda disseminated by terrorist organizations." *See id.* at 4 (internal quotation marks omitted). But there are other accounts operated by the British government under aliases. This means that their identity is kept hidden, and they engage with other users online who do not know that they are speaking with British government officials. Building on other scholarship and reporting in this area, Bradshaw and Howard write that the British government has created and uploaded YouTube videos that "contain

2.[85] ***Well-Intentional Foreign Government Messaging.*** The U.S. government knows that the Chinese government is engaging in activities in Latin America that are a cause for concern and believes the public should know. Because of the U.S. government's history of violence and nefarious influence in Latin America, the United States now has a well-deserved reputation for being a nefarious actor in the region, and anything that is said by the U.S. government or one of its employees will be regarded with suspicion and possibly disregarded altogether. U.S. government employees decide to hide their identities and share true, accurate, and non-misleading information about Chinese government activities in Latin America with media outlets in Latin America. The employees know that if they are forthright about their identity, the information they are sharing may not get uptake.[86]

In both of these cases, the speaker's deception about their identity is arguably justified.

The speakers are seeking an arguably morally justified end that would be impossible for them to pursue if not for the speaker-identity deception. In both cases, the speakers are trying to get out the truth. We can also imagine speakers being deceptive about their identities for other benevolent intentions and for praiseworthy reasons. Whistleblowers and individuals who share information anonymously with the press also fit into this category; they are being deceptive about their identities for good reasons and are acting on good intentions. For example, they may aim to share information in the interest of the public. To be sure, it is worth noting that whistleblowers may not always act for good reasons or on good intentions—they may be neutral or even bad. We do not, however, have to posit that all whistleblowers act on good reasons or on good intentions in order to see that there need some caveats to the blanket condemnation of speaker-identity deception.

---

persuasive messages." *See, e.g.*, *id.* at 12. Note, too, that this case brings to mind a now-notorious 2009 article published by Cass Sunstein and Adrian Vermeule that endorsed the use of identity-deceptive government counter-speech. Sunstein and Vermeule wrote that:

> Government agents (and their allies) might enter chat rooms, online social networks, or even real-space groups and attempt to undermine percolating conspiracy theories by raising doubts about their factual premises, causal logic, or implications for action, political or otherwise. In one variant, government agents would openly proclaim, or at least make no effort to conceal, their institutional affiliations . . . In another variant, government officials would participate anonymously or even with false identities . . . the two forms of cognitive infiltration offer different risk-reward mixes and are both potentially useful instruments. Cass R. Sunstein & Adrian Vermeule, *Symposium on Conspiracy Theories Conspiracy Theories: Causes and Cures*, 17 J. POL. PHIL. 202, 224-25 (2009).

85. I am grateful to David Luban for raising this example.

86. There is one difference between the two cases that merits stating: in case one, the speaker hides their identity and speaks directly to the hearer. In case two, the speaker hides their identity, speaks to an intermediary (the media), and the intermediary speaks directly to the hearer. While this may be an important distinction in some contexts, it is not for our purposes.

There is another way of framing the speaker-identity deception point so that we tease out differences between the permissible and impermissible cases: by focusing on manipulation. The core question for us in this context is whether the speaker's deception about her identity is sufficient for rendering the hearer the object of manipulation. If it is, then we must also conclude that the government employees in case one (anti-radicalization) and case two (well-intentioned foreign government messaging) are being manipulative.

We are left with a few options. First, we can bite the bullet and say that in all the cases in which the speaker is deceptive about her identity, the hearer is manipulated to some degree. This option, however, is unpromising. If it were right, then all cases in which the press received true, accurate, or non-misleading information from a source who hid or lied about her identity would be instances in which the hearers (readers of the press) were being manipulated to some extent. Moreover, if this notion were right, then each and every time someone shared true, accurate, and non-misleading information online while hiding behind another identity, they would be acting manipulatively. This, clearly, is too broad a conceptualization of manipulation.

Our second option is more promising. We can say that in those cases that deception about one's identity disrupts or undermines the hearer's decision-making capacities, the speaker is being manipulative. If the hearer's decision-making capacities are not undermined when the speaker hides or fabricates her identity, then the hearer has not been manipulated. It will, of course, be difficult to create clear and stable boundaries between kinds of deception that are non-manipulative and kinds that are manipulative, but this need not raise conceptual problems for us if we can, in theory, create those lines.

There is also a third option, which is perhaps the most promising. We might acknowledge that in nearly all instances of speaker-identity deception, the hearer is being manipulated but add that manipulation is not a morally bad outcome or morally impermissible behavior in all cases. We can argue that manipulation of a violent extremist into becoming a non-extremist is a morally acceptable outcome, or suggest that using manipulation as a tool in some cases is morally permissible. To be sure, whether the manipulation is morally permissible will depend not only on whether the purposes were benign, but also on the severity and invasiveness of the manipulation. For example, while it may be permissible to manipulate the violent extremist using some means (for example, strong emotional appeals), it will not be permissible to engage in brainwashing or to induce Stockholm syndrome. Nudges – positive reinforcement and indirect suggestions as ways to influence an agent's decision-making – are another example, although not all agree that nudges are manipulative per se.[87] Suffice it to say that we can morally justify certain kinds of manipulation in some contexts.

87. According to Susser, Roessler, and Nissenbaum, not all nudges are manipulative; only nudges that are hidden and exploit cognitive vulnerabilities count as manipulative. Susser, Roessler & Nissenbaum, *supra* note 54, at 27. Nudges that are transparent and seek to correct cognitive vulnerabilities should not be characterized as manipulative. *See, e.g.*, *id.* at 25.

In options two or three, both of which are promising, the distinctions that need to be made between the permissible or impermissible are always made in context by some kind of authority. For this reason, no hard-and-fast rule can be given that applies in all cases. What remains true is that speaker-identity deception allows for an impermissible influence in cases where it enables the erosion of autonomy.

Like speaker-identity deception, information quality – what Hollis refers to in the "deception" category – is a critical factor to consider in evaluating information-based influence. If information shared is true, accurate, and non-misleading – its quality is high – we should be less concerned with where it originates. Recognizably, speaker-identity deception makes it harder to verify that information is true, accurate, and non-misleading because speaker identity can be a proxy for epistemically important information.

What we should care about is the sharing of mis/disinformation and fake news. It is now well-understood that disinformation has epistemic, political, and social harms, although the full extent of those harms is still being explored by scholars. Because of mis/disinformation, there is increased skepticism about well-supported facts as well as greater uptake of false beliefs in some communities. The result is that there is no set of common facts we all believe and can refer to as we engage in disagreement and discourse. Mis/disinformation also leads to other kinds of social harms: we are more polarized as a society, there is decreased trust in public institutions, and many of us believe in dangerous conspiracy theories. The presence of mis/disinformation, irrespective of the source or the intentions behind which the information was communicated, is problematic. True, accurate, and non-misleading information is not concerning in the same manner.

This is not to deny that true, accurate, and non-misleading information can still be used for nefarious purposes. During the 2016 U.S. presidential election, factual information was leaked by Russian government-affiliated agents to serve political purposes. In early 2022, the U.S. government also shared true, accurate, and non-misleading information to achieve particular political ends related to Russia's aggression toward, and subsequent invasion of, Ukraine.

It is worth pausing to consider whether there are exceptions to the general acceptance of true, accurate, and non-misleading information. One could argue, for example, that the sharing of true, accurate, and non-misleading information can often lead to discord or exacerbate social tension. Introducing true, accurate, and non-misleading information (for example, the Earth revolves around the Sun) may directly cause disagreements, protests, divisions, or worse. It is also possible to argue that democracies, while built for and resilient to disagreement, may still have a "disagreement threshold point" which marks when healthy disagreement turns into dangerous disagreement.[88] After this threshold is surpassed, violence, civil war, and other kinds of socially harmful consequences might ensue. Indeed, this may be precisely what hostile foreign powers who seek to influence U.S. discourse are aiming for. Arguably, the sharing of true, accurate, and non-

88. I am grateful to David Luban for making this point.

misleading information that aims to push a democracy past this threshold point ought not to be considered ethically permissible.

The problem with this argument is that it is difficult to predict the consequences a piece of information will have. Moreover, we may be willing to tolerate a certain degree of risk of bad consequences for the sake of fostering an environment in which the truth is told. In summary, I suggest that we should be concerned with speaker-identity deception when it undermines someone's decision-making capacity; other instances, especially those online, are less or not concerning at all. With respect to information quality, we should be concerned with preventing the spread of mis/disinformation and fake news irrespective of who the speaker is; high-quality information, regardless of the reasons it is shared, is less concerning because high-quality information generally improves rather than hinders the flourishing of individuals and societies. This is the start to a longer examination of how we might build a robust apparatus to evaluate the ethics information-based influence campaigns.

CONCLUSION

This article contains both an argument against using dominant frameworks in international law to articulate the harms of information-based influence and an argument for reframing the conversation so that we can develop an ethics of influence that de-centers "foreignness." In Part I, I argued against the war, sovereignty, and self-determination paradigms on conceptual and normative grounds, and concluded that they are inadequate for conceptualizing the problem of influence. In Part II, I explored promising approaches for evaluating the ethics of information-based influence that did not rely on the dominant paradigms in international law, nor on the foreignness of a speaker's identity. There is much more to be done in developing a framework for assessing the ethics of influence. It is my hope we can continue to do so by focusing on the normatively troublesome aspects of influence without too relying on legal paradigms ill-suited to the task.