The question of the legality of foreign disinformation campaigns; navigating troubled waters

LETRÔNE William Axel *

Abstract

Cyberspace has made our world smaller. But the resulting state of interconnectivity has come at a price; a greater vulnerability to foreign influences and interferences. Online disinformation is one inappropriate influence tool that can be used by foreign actors to undermine a state. The problem is that lasting legal uncertainties call into question the very capacity of international law to address state-led disinformation, and thus to provide adequate protection against foreign influence activities that may prove particularly detrimental to the receiving state. Then, as a growing number of states has been legislating about foreign influence activities relying on information manipulation tactics, a significant gap has emerged between municipal and international regulations pertaining to international communications. This article explores the relevant law and finds out whether international law is up to the task of regulating state-sponsored online disinformation. In doing so, it clarifies the law and identifies the boundaries of an international regulation of online disinformation. It concludes by suggesting a safe way forwards to help restoring accountability and end impunity in the context of disinformation.

Key words: Disinformation, international law, influence, interference

* MEXT doctoral student at the Graduate School of International Cooperation Studies, Kobe University.

Introduction

"Malign influence" has become an important subject in security debates. Since the advent of cyberspace, states have learned to take advantage of new information and communication technologies to influence the politics of their counterparts. Regrettably, ill-intentioned states have leveraged cyberspace and free speech to conduct protracted influence activities with the aim to destabilise foreign societies. This paper argues that cyber-enabled foreign influence may require legal answer if the communications in question can be attributed to a state, and if the effort is ostensibly detrimental to the interests of the receiving state, which is arguably the case of some online disinformation campaigns. However, due to its elusive nature, disinformation is notoriously difficult to address by law.[1] As with most cyber-enabled conduct, conceptual, legal, and practical difficulties encountered when addressing online disinformation inhibit reactions, complicate legal assessments and foster an overall sense of insecurity. Meanwhile, states have started to address the problem through domestic legal means.[2] Even so, many national legislations geared towards tackling online disinformation are controversial.[3] Some are based on unclear conceptual frameworks that may lead to over securitization whereas others use broad formulations that do not necessarily guarantee moderate approaches to the problem, and risk causing a chilling effect on freedom of speech. Fortunately, there is room for improvement when it comes to preventing abusive regulations, while at the same time ensuring that states can be held accountable for conducting or condoning harmful disinformation activities. One way is to enhance legal certainty in the context of the international regulation of disinformation by clarifying the application of some cardinal principles of international law to illegitimate influence activities. Indeed, international law needs clarity to deter inappropriate conduct. State-sponsored disinformation activities will persist until malicious actors are sufficiently aware that their misbehaviours could give rise to claims of international law violation, and subsequently trigger sanction. In addition, clarifying the application of international law to disinformation activities could help states to better calibrate their own national regulations. This article explores the relevant law, and clarify its application to state-sponsored disinformation campaigns. The first section conceptualizes state-led

disinformation. The second section discusses the Law applicable to online disinformation activities. The third section wraps up the discussion and suggests a safe direction that future developments in the law should take to help restoring accountability and end impunity in the context of disinformation.

I. Online disinformation and the concept of malign influence

Disinformation is a particularly appalling method of influence that is reminiscent of the Cold War era. It is generally defined as "[…] false, inaccurate or misleading information designed, presented and promoted intentionally to cause public harm or make a profit."[4] The same way as private actors, some states have mounted disinformation campaigns for their own benefits. The problem is that states possess the financial, technical and organisational means to conduct greater far-reaching campaigns than private actors acting on their own. Furthermore, what states seek to achieve through disinformation activities is directly connected to real strategic considerations. State-sponsored disinformation campaigns are therefore wider in scale, more sophisticated, and more impactful than strictly private disinformation campaigns.[5]

The question of the regulation of international communications in peacetime, which concerned bellicose, subversive (revolutionary) and defamatory transmissions, has been a common topic of discussion among international lawyers since the invention of the radio.[6] In spite of this, the subject remained a contentious issue and only timid steps were taken towards stronger regulation. Basically, the many downsides that would have come with a new regulatory framework on international propaganda outweighed the benefits.[7] The advent of cyberspace has refocussed the debate regarding the regulation of international communication on disinformation and added a sense of urgency. State-led online disinformation activities have disrupted democratic processes and heightened social tensions in multiple states. In the long term, disinformation may negatively affect the rights and interests of individuals and cause irreparable damages to the political and/or social integrity of target states. Meanwhile the information spreads faster and reaches farther in cyberspace than if circulated *via* traditional medium.[8] In addition, malicious actors are disinhibited online, thanks notably to the relative anonymity internet offers. As a consequence, detecting a

disinformation campaign and tracing back a dissemination to the initiator generally requires long and expansive investigation work that makes timely reaction difficult.[9] Overall, cyberspace has enhanced manipulative tactics to the point that knowing how to manage disinformation has become a security imperative for democracies, whose well-functioning depends in a large part on the active participation of civil society in political affairs, the free flow of information and the protection of fundamental rights.

Foreign disinformation is often associated with the broader context of destabilisation strategies. Destabilisation strategies are sometimes called "hybrid interferences", in opposition to hybrid warfare, which involve military means.[10] Accordingly, disinformation is but one threat among a complex aggregate of non-military activities geared towards the realization of the same overarching goal, generally, to "harm, undermine or weaken the target."[11] On this point, a recent study from the Joint Research Centre of the European Commission and the European Centre of Excellence for Countering Hybrid Threats explains that disinformation plays a crucial role in the "priming" and "destabilisation" phases of a strategy of subjugation, because it helps, covertly, to weaken the resolve of the opponent state. The success of destabilisation strategies thus lies in the capacity of the influencing state to progressively bend a competitor to its will without raising the threat perception of the addressee.[12] To achieve this result, the influencing states may attempt to create dependencies, to introduce backdoors in computer programs before distributing them, to install malwares in electronic devices in view to trigger their activation on a later date, to manipulate fringes of the population via the injection of harmful or subversive lies and/or misleading information into an information environment, or to corrupt the elites of a state.[13] Because most activity will appear benign at first glance, they are extremely delicate to address by the receiving state, let alone to detect. Parton speaks of the "potential for interference" of in appearance normal interactions.[14] Obviously, it has become difficult to draw a line between legitimate interaction and illegitimate interference.[15]

In general, interference describes an undue intrusion in the sovereign affairs of one state. A foreign activity will be characterised as foreign interference by the addressee if it is perceived as an attempt to undermine its integrity. Whether a specific foreign conduct deserves the "interference" treatment is partly a subjective

judgement. Hence, what is considered illegal interference under municipal law is not always illegal under international law. For reasons inherent to the necessity of maintaining international stability, international law does not necessarily reflect the tolerance threshold of states. Indeed, only the most blatant forms of foreign interferences, what Ziolkowski calls "massive influence"[16] may breach international law. This is a relatively high threshold compared to what may be criminalized in domestic legal systems. Be that as it may, in international law and in municipal law, an illegal foreign interference is always associated with the act of disrupting current or impeding democratic processes, or the act of obstructing a state in the effective exercise of its sovereign functions, (its *prerogatives de puissance publique*). The problem is that disinformation campaigns are protracted efforts. As such, they do not always constitute clear-cut interferences in the sovereign affairs of the target. In reaction, some states have introduced the concept of "malign influence", which appears to extend beyond what is usually considered an interference, and even further beyond than what might constitute an illegal intervention under international law. Australia, for instance, understands "malign influence" as foreign efforts that remain below the threshold of obvious illegality, and that are deniable, integrated and incremental.[17] As such, malign influences usually fall short of clear-cut interferences but are still "inconsistent with – or carry risk to – a democracy's values or interests."[18] The new US Foreign Malign Influence Response Center adopts an even broader definition, referring to malign influence as any state-sponsored effort aimed at negatively influencing "the political, military, economic, or other policies or activities of the United States Government or State or local governments, including any election within the United States; or (B) the public opinion within the United States."[19]

In many regards, the introduction of the concept of "malign influence" shakes up the traditional schemes; influence pure and simple may still require moderation and may even in some circumstances warrant new legal solutions. Of course, not all activities contained in a strategy of destabilisation require punitive measures, far from it. In the case of some online disinformation campaigns sponsored or condoned by a foreign state, however, the question is worth asking. What says international law?

II. Online disinformation as an unlawful instrument of statecraft

First of all, the concepts of "malign influence" and "destabilisation strategies" are not *per see* contemplated by international law. These are relatively recent concepts in security studies that, for reasons inherent to their very abstract nature, as well as a significant risk of instrumentalization, cannot be regulated. Things are more complicated with disinformation. By engaging in disinformation activities, malicious actors exploit the so-called "grey zones" of international law. For this reason, state sponsored disinformation is often called a "grey-zone activity". The broad category of grey zone activities encompasses a variety of unfriendly acts, most of which are often used in conjunction in the context of a destabilisation strategy. Although grey-zone conduct may constitute a crime on municipal legal systems, it is notoriously difficult to address through legal means.[20] This is either due to the fact that relevant legal frameworks are too unclear to apply in a convincing manner, because the rules are yet to emerge, or because practical hurdles render condemnations pointless. These are recuring problems that are also encountered when it comes to address the legality of online disinformation in international law. It should be noted that it is not one unique operation that is scrutinized here, but an ensemble of operations forming a campaign. International law provides for the possibility to consider a composite act, composed of a series of actions or omission, as a breach of international law.[21] This applies to disinformation campaigns, which should systematically be considered as a whole, taking into account all the operations it contains that are clearly following the same reprehensible purpose.

The same way as the aforementioned concepts, "disinformation", as such, is an unknown term in international law.[22] That is not to say, however, that disinformation activities go completely unregulated. Some existing principles of international law simply apply to disinformation by incidence, when a disinformation campaign displays the characteristics required to fall under a rule's scope of application. This is for example the case of war propaganda, which may involve disinformation methods.[23] Similarly, specific treaty provisions might relate to disinformation, not because disinformation or the politically-motivated dissemination of lies and/or misleading information to cause harm is explicitly outlawed by the parties, but because the treaty

prohibits certain practices that share similar features with modern disinformation activities. This is the case for example, with the now-extinct Treaty of Paris signed between France and Russia in 1801.[24] One of the Treaty's requirements was that the parties abstain from propagating principles contrary to the constitution of the receiving state and to foment internal troubles in the territory of the other state. Another more recent treaty, the 1936 International Convention concerning the Use of Broadcasting in the Cause of Peace (hereinafter the Broadcasting Convention) applies to incorrect statements "likely to harm good international understanding", which could also be interpreted as encompassing disinformation.[25] Still, there is no unique comprehensive legal tool to address online disinformation in current international law. In fact, as far as the negative obligations of states are concerned, the legal framework for assessing online disinformation activities is extremely complex. Furthermore, the most intuitive approaches to the problem have proven challenging for a number of reasons. For the purpose of this analysis, the relevant legal rules are divided in two groups: the rules protecting the collective interests of states (or duties of states) and human rights law. Due to length constraints, only the most relevant rules are analysed below.

A) Disinformation campaigns as a violation of states' rights

By virtue of the principle of sovereignty, states bear rights and duties, some of which have been listed by The International Law Commission a tentative codification of the Rights and Duties of States in 1949. As noted earlier, there is no such rule in customary international law that explicitly prohibits the waging of disinformation campaigns against a state during peacetime. However, there are rules and principles that relate to the conduct at hand, and therefore, could provide *locus standi* to a victim state.

From state sovereignty stem two negative obligations; the fundamental duty not to violate a state's territorial integrity, (or territorial inviolability), and the fundamental duty not to coercively intervene[26] against a state's political integrity, (or independence)[27]. Violations of one or the two aspects will simultaneously constitute a violation of state sovereignty. Hence, online disinformation may engage the

international responsibility of a state if the said conduct is attributable to the said state, and if it is harmful to the target state's right to political independence, and/or territorial integrity.

1) The prohibition of uses or threat to use force

The prohibition of uses or threat to use force enshrined in the UN Charter Article 2§4, protects a state from the most daring conduct, -those that involve *force*. Traditionally the non-use of force principle applies to military or armed force. Much like other grey-zone conduct used in destabilisation strategies, disinformation campaigns are non-destructive and non-deadly by nature. Hence, the dissemination of false and/or misleading news or reports in the territory of a state by another, be it motivated by harmful intentions, does not, *in principle*, constitute an unlawful use of force. On this point, the experts of the Tallinn Manual 2.0 specified in their definition of uses of force, that non-destructive cyber psychological operations intended solely to undermine confidence in a government or economy do not qualify as uses of force.[28]

Be that as it may, it is now well established that some activities that do not involve conventional weapons may still breach the prohibition if they lead to consequences akin to what could have been done through military force.[29] In other words, the prohibition is breached when "there is some element of armed force involved, or at least actions resulting in physical injury or damage."[30] Therefore, state-sponsored disinformation campaigns could still breach the non-use of force principle if it is found that the disinformation has contributed in a major way to significant disruptions equalling the severity of a conventional use of force in the territory of the target state, and that it was clear that such result was expected by the disinformant.[31] The degree of harm resulting from non-deadly and non-forcible conduct would have to be established on a case-by-case basis, taking into account the context of the spread, the characteristics of the communications and the nature of the instigator.[32] However, as the point of destabilisation strategies is to remain under the radar for a sufficiently long time period, the materialization of eventual damages is always delayed. That is to say, that in the eventuality that disinformation has contributed to damage, it is likely to be one contributing factor among many others. In response, some commentators have argued that non-violent conduct *per se* could still breach the prohibition if "the

foreseeable effects of the cyber operation would rise to the level of a use of force."[33] Death and destruction would be foreseeable in the context of disinformation campaigns targeting a state that is battling a pandemic, as it happened during the coronavirus pandemic, or a natural disaster, during which the access to reliable information is sometimes a matter of life and death. Other authors have also hinted at the possibility of treating some international communications as use of force, "if the danger created by the propaganda in question had reached the point at which a condition of 'clear and present danger' could be shown to exist."[34] A similar argument was made by Whitton and Larson, who claimed that retaliations akin to what is traditionally reserved to uses of force should be admitted if "the interference takes the form of psychological warfare threatening the very existence of the government and perhaps the state itself."[35]

In normal circumstances, however, the use of force thresholds may never realistically be crossed by a disinformation campaign. Destabilisation strategies are far more reliant on corrupted forms of soft power (or sharp power) than on hard power tools, and it is unlikely that a state goes as far as to condemn a disinformation campaign as an illegal use of force, be it a particularly debilitating one.[36] As a commentator recently noted, "The worry of counting CDOs [**ndlr: cyber disinformation operations**] as 'use of force,' other than it is counter-intuitive to the plain language, is that the effect of CDOs might be disproportionate to that of belligerent military actions (the traditional understanding of 'use of force')."[37]

Immaterial activities that do not involve undue trespassing of borders, that do not or are not likely to *cause* physical damages, death or injury in the territory of the target state, will not constitute a violation of territorial integrity. They may still violate its political independence if they contain threats to use force.[38] This would be the case when lies and misleading information are used in the context of a scare tactic intended to intimidate, or to obtain some concessions from the bullied state.[39] This might also apply to situations where disinformation is used as a bait tactic. These are however highly hypothetical examples.

2) The non-intervention principle

Another rule that safeguards the territorial integrity and political independence of

states is the non-intervention principle. As it sets less demanding thresholds than article 2 § 4, the non-intervention principle seems a more appropriate framework in the present case. The principle of non-intervention is ubiquitous in international law. It is enshrined in a variety of legal instruments, from bilateral treaties such as the 1954 Panchsheel Treaty signed between China and India, to the founding treaties of regional communities such as the Charter of the Organisation of American States, to General Assembly Resolutions such as the Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and The Protection of their Independence and Sovereignty from 1965, the Friendly relation Declaration from 1970, and the Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States from 1981.

Although non-intervention's ubiquitous presence in past and current international law could signal that states are on the same page when it comes to interpreting non-intervention, the principle has been the subject of an immense amount of discussion ever since the first codification attempts. Debates about the prohibitive power of the principle and, by extension, its meaning, are particularly common among states and international lawyers. Non-intervention tends to be interpreted in vastly different ways depending on the state and the nature of the instrument in which it is proclaimed.[40] To add to the confusion, the principle proved a powerful rhetorical device. As such, it has been abusively invoked by states to "criticize a measure that plainly does not infringe it"; "to justify inaction or passivity" or to "justify an action that it does not necessarily prescribe."[41]

Eventually, the principle made its way into the international jurisprudence. In its oft-cited Nicaragua case, the International Court of Justice held that coercion formed "the very essence" of a prohibited intervention, and that coercion was prohibited when it bears on "matters in which each State is permitted, by the principle of State sovereignty, to decide freely. […]"[42] As it did not specify what kind of conduct could constitute a prohibited intervention without simultaneously breaching the use of force prohibition, the cautious leading statement of the World Court did little to clarify the content of the principle. This left international lawyers wandering about the legal value of the principle.[43] In spite of this, the general understanding is that the scopes of the prohibition of the uses of force and the principle of non-intervention, although

concentrical, do not exactly extend to the same activities. Particularly invasive yet non-forceful foreign conduct such as the financing of political factions abroad, trade embargos, or subversive activities may thus fall under the broader scope of the non-intervention rule without qualifying as uses or threats to use force.[44]

At this point, a lot has been written about the principle of non-intervention and its ability (or inability) to provide legal protection against non-forceful activities such as foreign disinformation campaigns.[45] A recurring problem is the difficulty to define interstate coercion. As regard to disinformation activities, questions were raised as to whether disinformation was not closer to persuasion than coercion, whether the intensity and the methods of a disinformation campaign could make up for the impossibility to infer a coercive intent, or whether disinformation activities were only contributing to the coerciveness of subsequent, more brazen undertakings. Some authors eventually rejected the argument that disinformation campaigns are coercive.[46] Lahmann, for example, argues that, because "the term implies compulsion with some degree of forcible conduct in the broader sense, deceptive manipulation by way of a disinformation campaign cannot be conceived as coercion."[47]

Foreseeably, legal assessments relying on the principle have differed dramatically in their methodologies. Attempts to reconcile the insidious phenomenon of disinformation with coercion are often too convoluted to make a persuasive argument susceptible of being used by a state as a legal defence, when there are not outrightly reinventing coercion.[48] In spite of this, recent submissions indicate an emerging *convictio juris*, that among all hypothesis, disinformation campaigns directed at democratic processes should fall under the scope of the non-intervention principle.[49] More generally, there seems to be a potential for consensus around the necessity to move on from coercion, and to focus on more concrete criteria such as the context or the means used by the influencer and the inferable intent.[50] Not only is such a development intuitively appealing, but it is highly reminiscent of past discussions about unlawful interferences, notably those that led to the adoption of the above-cited UN resolutions by the General Assembly, which were not restricted to condemning coercive interferences.[51] And even though the legal value of these resolutions is disputed, one should note that some interventions were deemed unlawful, not because they involved the use of coercive methods, but because they impaired, or were taken

with the clear intention to impair the sovereign prerogatives of the target state. It results that the focus on coercion in the context of insidious threats such as disinformation is misplaced. Past debates about non-intervention have shown that knowing what falls under its scope does not exclusively hinges on the coerciveness of a conduct. Thus, in the context of non-coercive destabilisation strategies, and disinformation activities in particular, observers should always inquire whether the influencer intended negatively affect the right of the target. In other words, the litigious conduct must be interventionist in nature to be capable of breaching the norm.[52] The interventionist intent will be obvious when specific protected prerogatives are clearly targeted by the influencing state. For example, in the case of the Russian meddling into the 2016 US elections, the impaired prerogative was the capacity of the population to freely choose the way it is governed. It might be difficult to draw a line between internationally illegal activities and protected interactions that may still go against the interests of the recipient. For that reason, other factors should be considered, such as the context, the means used by the influencer, the truth-value of the content, the duration of the act and its degree of invasiveness.

In the context of destabilisation strategies, there will often be situations where the inferable intent is too ambiguous to make the point of a prohibited intervention, even if malicious motives have been imputed to the influencer. For example, and unless the litigious communications take the form of explicit encouragements for sedition, it is unclear whether disinformation activities broadly aimed at causing internal disturbances and tensions are impairing a sovereign right. On this point, perhaps past discussions on subversive propaganda are worth revisiting, but it is very unlikely that the non-intervention principle applies to disinformation activities that merely "corrode" democracy.

B) Disinformation campaigns as a violation of human rights

Disinformation activities can also be looked at through the lenses of international human rights law. As individuals are the first victims in the disinformation context, human rights law seems a more logical approach. Jones, for instance, argues that individuals possess a right not to be subjected to deceptive or manipulative practices,

which includes disinformation. Taken together, "the right to hold opinions without interference," and the "freedom to seek, receive and impart information and ideas of all kinds," as enshrined in 1966 International Covenant on Civil and Political Rights (ICCPR) and the European Convention on Human Rights (ECHR), protect the mental autonomy of individuals. Doing so, they establish a correlative duty not to subject someone to manipulative or deceptive tactics. Accordingly, a state that is using manipulative methods against selected audiences at scale, could be liable, for the sole reason that it engaged in disinformation. The relevancy of this argument will only grow stronger with the development of virtual-reality spaces such as the Metaverse, which will enhance the manipulative value of disinformation activities. The problem is that the obligation to secure the freedom of thought, opinion and expression does not extend beyond the territories of the signatories of the aforementioned treaties, and does not apply at all to non-signatories. Also, the freedom of thought and opinion perspective has not yet been considered as an approach to sanctioning disinformation activities.[53] This argument is however worth considering, notably because freedom of thought and opinion are indispensable to the full enjoyment of almost every other human rights.

Online disinformation could also violate customary human rights that are binding upon all states, regardless of jurisdictional constraints.[54] These are the right to self-determination, the right to life, and the non-discrimination principle. The first confers on all peoples the ability to freely determine their political status and freely pursue their economic, social and cultural development.[55] Arguably, the right to self-determination is what disinformants are infringing upon when they target foreign voters during election processes or *referenda*.[56] This proposition is premised on the internal aspect of the right to internal self-determination, which empowers individuals to participate in the public affairs of their country. On this point, Ohlin argues "Outsiders are free to express their opinions but covertly representing themselves as insiders constitutes a violation of these political norms, [**the membership rules for political decision-making**] which are constitutive of the notion of self-determination, just as much as covertly funnelling foreign money to one candidate."[57] Accordingly, the impersonation of nationals by foreign agents via the use of fake accounts in order to influence foreign public opinions in times of elections is an usurpation of the right to

self-determination of the audience. It may also simultaneously violate the principle of non-intervention.[58] However, because self-determination has been brandished only in the context of decolonization, states and individuals may be reluctant to use the argument to condemn electoral disinformation. Also, as disinformation is a protracted effort, care should be taken not to overstretch the scope of the principle of self-determination beyond the periods of open political participation.

Another human right of interest is the right to life, and its correlative right to health. The right to life is enshrined in the ICCPR, article 6 and in the EU Charter of Fundamental Rights, article 2. It has been observed that online disinformation could lead to violations of the rights to life and health if there is a sufficient causal link between a disinformation campaign and the materialization of harm or casualties.[59] Although the critical issue of causation in the disinformation context may incapacitate legal assessments,[60] Schmitt and Milanovic have argued that purposefully creating the conditions for the loss of life through the dissemination of false or misleading information suffices in establishing a causal nexus.[61] Regardless of the effective materialization of harm, the conscious promotion of unreliable, false and misleading information during critical events when the lives of individuals are at stake could thus engage the international responsibility of the influencer. This would particularly concern the communication of unreliable information during natural disasters, or the spread of disinformation that has for intended effect to cripple the management of national health crisis abroad, as it allegedly happened during the COVID-19 pandemic.[62] It bears noting that foreign-led disinformation campaigns that caused substantial damages to life abroad also engage the international responsibility of the influencing state on the basis of the above-cited Article 2 § 4.

A last human right of interest is the principle of non-discrimination, which is enshrined in a number of international instruments such as the ICCPR and the International Convention on the Elimination of All Forms of Racial Discrimination. Non-discrimination means, *inter alia*, that states are under the obligation to "declare illegal and prohibit organizations, and also organized and all other propaganda activities, which promote and incite racial discrimination [⋯]"[63] This especially applies to incitation to genocide, war propaganda and hate speech in general, both of which are often at the same time constitutive of disinformation campaigns. An example is the

online disinformation campaign conducted and supported by Armenia that Azerbaijan recently brought before the ICJ.[64] In its request for provisional measures, Azerbaijan denounced "cyber disinformation operations to incite and stir ethnic hatred and violence against Azerbaijanis"[65]

III. Remaining uncertainties and future developments

So, is state-sponsored disinformation legal? Not always. The previous section has shown that there is no unique rule that governs the waging of online disinformation campaigns by a state, but a constellation of interrelated laws that may or may not apply, and sometimes overlap, depending on the specificities of each campaign. It bears noting that while the truth-value of the litigious communications may be an aggravating factor in an international offence, it is never the determining factors in the choice of a legal framework. Various elements will factor differently in the assessments, depending on the selected framework. The duration and the scale of an operation are very important factors to take into account, whatever the rule relied upon, but it is unclear whether duration and invasiveness alone could make up for the difficulty of inferring a specific, reprehensible intent from an insidious conduct. Sometimes, it will not seem much of a stretch to consider some disinformation activities as falling under the scope of traditional principles if the analogy does not contradict with the object and purpose of the principle in question. This holds particularly true for non-intervention, whose very purpose is to protect states from external subjugation. A purposive interpretation of non-intervention should better align international law with the expectations of states *vis à vis* disinformation targeting democratic processes.

Then, when assessing a high-scale disinformation campaign, international lawyers should always inquire whether:

- The disinformation campaign is clearly aimed at corrupting democratic processes abroad
- The disinformation campaign has directly or indirectly caused quantifiable harm
- The disinformation campaign contributes to a situation of clear danger

At the same time, destabilisation strategies are long-term oriented. Malign influence specifically designates conduct to which the effects might be felt during periods of political participation, but the threat that it poses extends far beyond elections or referenda. In other words, although malign influence intensifies during these periods, foreign-led efforts at manipulating the minds of an audience do not start during elections, nor do they stop after a candidate is elected. In fact, as the US homeland security Council explains; "Foreign influence and disinformation should be seen as a continuous, ongoing assault on the United States, rather than a series of discrete, targeted, event-specific campaigns."[66] Most existing traditional legal concepts do not clearly accommodate the long-term perspective that should be adopted while approaching disinformation activities, let alone their persistent nature. In spite of this, creating new laws might not be the right solution. Expressly outlawing online disinformation in general is practically impossible without relying on obscure requirements. For example, confining the question of illegality to the truth-value of the communications misses the mark, because disinformation is more about "blending misleading rhetoric with accurate and inaccurate content as well as inaccurate sourcing informations" rather than about spreading easily-debunkable lies.[67]

Similarly, relying on loose concepts such as the manipulative value of a communication, the corrosive nature of a conduct, or the *potential* for harm, would at best render the law pointless, at worse, would make the law susceptible to political exploitation, as states will undoubtedly disagree on the matter. Caution should thus be taken not too cast wide nets that would outlaw most inter-state interactions, undermine human rights, and foster escalatory trends amongst states. Also, it should be kept in mind that human rights considerations will systematically stand in the path of heavy-handed regulation. Therefore, although minimal efforts could be deployed to readjust the relevant rules so that they fit a broader category of malicious foreign state-sponsored conduct aimed at periods of open political participation abroad, disinformation campaigns, and destabilisation strategies in general can hardly be outlawed on the sole basis that they are aimed at broadly weakening the target state. Then, disinformation campaigns that do not fit in one or more of the three aforementioned hypotheses will constitute inadmissible yet internationally legal conduct, against which developing critical thinking and improving media information

literacy of users, as recommended by the French Bronner Commission in its 2022 report on disinformation[68], is the best defensive measures.

But perhaps an even greater issue when addressing disinformation in particular, and grey-zone conduct in general, is the fact that they are rarely executed by the sponsoring state itself. Often, the conduct is outsourced to private entities, or proxies that will carry out the conduct for the sponsor. Consequently, attribution processes are never straight-forward. This holds true, as said above, for the technical process leading up to the discovery of the territory of origin, as well as for the legal process leading up to the indictment of the state that ordered the campaign, if any. The main hurdle is encountered in the second process, because the conditions set forth by the law in order to attribute the conduct of private entities to a sponsor state are very difficult to satisfy. In addition, disinformation activities are often executed remotely. The target state is thus often unable to proceed with criminal indictments because the perpetrator is not placed under its jurisdiction. Accountability relies entirely on the host state, who may be unwilling to cooperate in matters of extradition.[69]

A potential remedy to the attribution problem is to be found in various treaties regulating international communications and subversive activities.[70] The 1936 Broadcasting Convention from 1936, for example, sets a general obligation to prevent and cease harmful communications from emanating from their respective territories.[71] Its article 3 notably provides that "The High Contracting Parties mutually undertake to prohibit and, if occasion arises, to stop without delay within their respective territories any transmission likely to harm good international understanding by statements the incorrectness of which is or ought to be known to the persons responsible for the broadcast." Expanding upon this model, and considering the well-documented threat that disinformation represents today, it is reasonable to expect that states are taking appropriate steps to prevent and cease harmful communications from emanating from their respective territories, and from spreading in their own territories. In addition, states should demonstrate due diligence in implementing this no-harm principle in their respective territories. This means, above all, that any steps taken to prevent harmful disinformation operations should be proportionate to the harm caused, or likely to be caused.

But if the recognition of a no-harm principle in the context of online disinformation

could release the victim state from the burden of establishing direct legal attributions in order to engage the international responsibility of an enabling state,[72] a remaining issue lies in the difficulty to quantify the harm caused or likely to be caused by disinformation. On this point, and drawing upon the observations made earlier, the harm should be measured on a case-by-case basis, taking into account the relevant factors that weighted in all the above-presented assessments. These are, *inter alia*, the volume of the audience reached by the disinformation, the degree of coordination with which the transmitters are acting, the scale and the duration of the spread, the context in which the messages are transmitted, the degree of openness of the online information space of the target state or the average education level of the target audience.

The harm can be interpreted exclusively in terms of the harm to human rights without neglecting other aspects of the question. In fact, this article has shown that there is a significant overlap between situations where the rights of states are abused, and situation where substantial damage is inflicted to human rights. Basically, disinformation campaigns that corrupt domestic or foreign political processes are harming the right to internal self-determination of the voters, and may simultaneously breach the non-intervention principle. Similarly, disinformation campaigns that cause harm or contribute to a clear situation of danger are harming the right to life or the non-discrimination principle, and may simultaneously breach the prohibition of the use or threat of force. A human-centric approach to harm thus seems a better approach. It also allows observers to address domestic disinformation. Take the situation where an authoritarian regime makes systematic use of censorship methods and other *astroturfing* tactics to stifle domestic criticism, or when the ruling regime consents to foreign disinformation campaigns against its own population.

To conclude, there is no denying that disinformation is a new security threat that should not be taken lightly. But as it might prove difficult for states not to succumb to over-securitization in a world where influence rimes with interference, one will have to tread carefully in the troubled waters of the international law applicable to disinformation activities. This article has shown that international law already provides the tools to regulate the most dangerous disinformation campaigns. It has also shown that minimal developments could improve legal certainty in the disinformation context

while maintaining a reasonable approach to the problem. In order to ensure the resilience of democracies in face of ever-more insidious threats, it seems more promising to focus on consolidating human rights, which better covers the aspects of disinformation, rather than state-centric rules. Future developments should also focus on elaborating mechanisms that take full advantage of the recent inroads made by due diligence in legal debates.

Acknowledgements
The author would like to express his sincere appreciation to the EGUSA Foundation for International Cooperation In the Social Science for their grant, which helped supporting the writing of this paper. An early version of this paper was presented at the 2022 ISA Conference, at the occasion of a panel session about the capacity of international law to meet the challenges of a "smaller world."

Notes

1 Jason Pielemeier, "Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?" (2020) 4(1) *Utah Law review* 917.

2 For a guide on state legislations recently taken in response to disinformation, see Daniel Funke and Daniela Flamini, "A guide to anti-misinformation actions around the world" *Poynter.* <https://www.poynter.org/ifcn/anti-misinformation-actions/>

3 See for example "Singapore passes 'fake news' legislation that threatens press" (9 May 2019) Committee to protect Journalists (CPIJ), <https://cpj.org/2019/05/singapore-passes-fake-news-legislation-that-threat/>.; "Belarus moves to prosecute 'fake news,' control the Internet" (8 June 2018) CPIJ, <https://cpj.org/2018/06/belarus-moves-to-prosecute-fake-news-control-the-i/>; Samy Magdy "Egypt tightens restrictions on media, social networks" (20 March 2019) AP News, <https://apnews.com/article/1540f1133267485db356db1e58db985b>.; Guy Faulconbridge, "Russia fights back in information war with jail warning" (4 March 2022) Reuters, <https://www.reuters.com/world/europe/russia-introduce-jail-terms-spreading-fake-information-about-army-2022-03-04/>

4 Carme Colombia, Héctor Sanchez Margalef & Richard Youngs, "The Impact of Disinformation on Democratic Processes and Human Rights in the World" (2021) European Parliament Report.

5 See in general, Kate Starbird Ajmer & Tom Wilson, "Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations" (2019), 3 (CSCW) Article 127 *Proceedings of the ACM on Human-Computer Interaction* 1-26.

6 See in general John B. Whitton & Arthur Larson, *Propaganda Towards Disarmament in the War of Words* (New York: Oceana, 1964).

7 See notably John B. Whitton, "The Problem of Curbing International Propaganda" (1966) 31 (3) *Law and Contemporary Problems* 601.

8 See notably Sorough Vosoughi, Deb Roy & Aral Sinan, "The spread of true and false news online" (2018) 359(6380) *Science* 1146. See also Herbert Lin, "The existential threat from cyber-enabled information warfare" (2019) 75(4) *Bulletin of the Atomic Scientists* 187.; Gerald Bronner, *Apocalypse cognitive*, (Paris, PUF, 2021).; Romain Badouard, *Le désenchantement de l'Internet : désinformation, rumeur et propagande* (Limoges: Fyp Editions, 2017).; and Thomas Rid, "Cyber War Will Not Take Place" (2012) 35(1) *Journal of Strategic Studies* 5. At 22.

9 See for example Ben Nimmo, "11. Network analysis and attribution" Verification Handbook

3, *DataJournalism.com*, <https://datajournalism.com/read/handbook/verification-3/investigating-platforms/11-network-analysis-and-attribution>.

10 Mikael Wigell, "Hybrid Interference as a Wedge Strategy: A Theory of External Interference in Liberal Democracy" (2019) 95(2) *International Affairs* 255.

11 Cullen Patrick & al., "The Landscape of Hybrid Threats: A Conceptual Model (Public Version)" in Giannopoulos, G., Smith, H. and Theocharidou, M. (Luxembourg: Publications Office of the European Union, 2021).

12 See Wigell, *supra* note 10. and *ibid.*

13 See Wigell, *supra* note 10.

14 Charles Parton, "China–UK Relations Where to Draw the Border Between Influence and Interference?" (2019) Royal United Services Institute for Defence and Security Studies (RUSI), Occasional Paper. At 3.

15 Kristine Berzina & Etienne Soula "Conceptualizing Foreign Interference in Europe" (2020) Alliance for Securing Democracy.

16 Katharina Ziolkowski, "General Principles of International Law as Applicable in Cyberspace" in Katharina Ziolkowski, (ed.) *Peacetime Regime for State Activities in Cyberspace: International Law, International Relations and Diplomacy* (Tallinn: NATO CCDCOE Publication, 2013). At 165.

17 Katherine Mansted, "The Domestic Grey Zone: Navigating the space between Foreign Influence and Foreign Interference" (2021) National Security College, Australian National University, Occasional paper. At 3.

18 *Ibid.* at 14.

19 50 US Code § 3059. Foreign Malign Influence Response Center: (e) definition of malign influence, Title 50-War and National Defense Chapter 44- National Security Chapter I-Coordination for National, <https://uscode.house.gov/view.xhtml?req=(title:50%20section:3059%20edition:prelim>.

20 See Michael N. Schmitt, "Grey Zones in the International Law of Cyberspace" (2017). 42(2) *Yale Journal of International Law Online* 1.; Duncan B. Hollis, "The Influence of War; The War for Influence" (2018) 32(1) *Temple International & Comparative Law Journal,* 31.; Michael N. Schmitt, "Virtual" Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law" (2018) 19(1) *Chicago Journal of International Law* 30.

21 See "Draft Declaration on Rights and Duties of States" (1949) International Law commission.

22 Wenqing Zhao, "Cyber Disinformation Operations (CDOS) And A New Paradigm of Non-Intervention" (2020) 27 *U.C. Davis Journal of International Law & Politics,* 35. At 47.

23 Björnstjern Baade, "The EU's 'Ban' of RT and Sputnik A Lawful Measure Against Propaganda for War" (8 March 2022) *Verfassungsblog on matters constitutional,* <https://verfassungsblog.de/the-eus-ban-of-rt-and-sputnik/>

24 Alexandre de Clercq, (ed.) "Traité de paix conclu à Paris le 8 octobre 1801 entre la France et la Russie" 1 *Recueil des traités de la France.* (Paris: Durand et Pedone-Lauriel, 1880). At 467.

25 International Convention concerning the Use of Broadcasting in the Cause of Peace (1936), adopted by UN General Assembly in A/RES/841 (1954)

26 *Case Concerning Military and Paramilitary Activities In and Against Nicaragua* (Nicaragua v. United States of America) (1986) ICJ, judgement. At 212.

27 See, in general, Henning Lahmann, "On the Politics and Ideologies of the Sovereignty Discourse in Cyberspace" (2021) 32 *Duke Journal of Comparative & International Law* 61.; Samuel K. N. Blay, "Territorial Integrity and Political Independence" (2010) *Max Planck Encyclopedias of International Law* [MPIL] Encyclopedia entries.

28 "The use of force" in Michael N. Schmitt (ed.) *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Cambridge: Cambridge University Press, 2017). At 328-356

29 Ian Brownlie *International Law and the Use of Force by States* (UK; Oxford University Press, 1963). At 362.;Russel Buchan, "Cyber Attacks: Unlawful Uses of Force or Prohibited Interventions?" (2012) 17(2) *Journal of Conflict and Security Law,* 212.; Schmitt (2017), *supra* note 28. Rule 68.

30 Gary Corn, "Coronavirus Disinformation and the Need for States to Shore Up International Law" (2 April 2020) *Lawfare*, <https://www.lawfareblog.com/coronavirus-disinformation-and-need-states-shore-international-law>

31 Marko Milanovic & Michael N. Schmitt, "Cyber Attacks and Cyber (Mis)information Operations During a Pandemic" (2020) 11 *Journal of National Security Law & Policy.* At 259.

32 *Ibid.*

33 *Ibid.*

34 See in that sense, Gerhard von Glahn, "The Case for Legal Control of "Liberation" Propaganda" (1966) 31 *Law and Contemporary Problems* 553.

35 Whitton & Larson, *supra* note 6. At 84.

36 Hitoshi Nasu, "The 'Infodemic': Is International Law Ready to Combat Fake News in the Age of Information Disorder?" 39(1) *The Australian Year Book of International Law Online* 65.

37 Zhao, *supra* note 22. At 47.

38 Hollis, *supra* note 20. At 39.

39 Whitton & Larson, *supra* note 6. At 85.

40 Denitsa Raynova, "Towards a Common Understanding of the Non-Intervention Principle" (2017) European Leadership Network, Post-Workshop Report.; Michael N. Schmitt, "Foreign Cyber Interference in Elections" (2021) 739(97) *International Law Studies.* At 744. See also Dissenting Opinion of Judge Schwebel to Nicaragua v. US), *supra* note 26.

41 Olivier Corten et al., *A Critical Introduction to International Law* (Bruxelles: Éditions de l'Université de Bruxelles, 2019). At 156.

42 ICJ, *supra* note 26.

43 See Corten, *supra* note 41. And Benedetto Conforti, "Le principe de non-intervention" in Bedjaoui, Mohammed (Ed.) *Droit international: bilan et perspectives,* (Pedone,: Paris, 1991). At 489.

44 Michael Wood, "Non-Intervention (Non-interference in domestic affairs)" *Encyclopedia Princetoniensis.*

45 See, *inter alia* Schmitt & Milanovic, *supra* note 31, Schmitt (2021), *supra* note 40,; Zhao, *supra* note 22.; Henning Lahmann, *Unilateral Remedies to Cyber Operations Self-Defence, Countermeasures, Necessity, and the Question of Attribution* (Cambridge: Cambridge University Press, 2020). At 38.; Thibault Moulin, "Reviving the Principle of Non-Intervention in Cyberspace: The Path Forward" (2020) 25(3) *Journal of Conflict and Security Law* 423.; Xuan W. Tay "Reconstructing the Principle of Non-Intervention and Non-Interference – Electoral Disinformation, Nicaragua, and the Quilt-work Approach" (2022) 40(1) *Berkeley Journal of International Law 39.;* Gary P. Corn, "Covert Deception, Strategic Fraud, and the Rule of Prohibited Intervention" in Jack Goldsmith (ed.) *The United States' Defend Forward Cyber Strategy: A Comprehensive Legal Assessment* (New York: Oxford Academic, 2022). See finally Henning Lahmann, "Information Operations and the Question of Illegitimate Interference under International Law" (2020) 53(2) *Israel Law Review* 189.

46 See notably Zhao, *supra* note 22.; Ido Kilovaty, "Doxfare: Politically Motivated Leaks and the Future of the Norm on Non-Intervention in the Era of Weaponized Information" (2018) 9 *Harvard National Security Journal.* At 171.

47 Lahmann (2020) *supra* note 45. At 202.

48 See Kilovaty, *supra* note 46.

49 See notably "Official compendium of voluntary national contributions on the subject of how international law applies to the use of information and communications technologies by

States", UNODA, A/76/136, (August 2021).

50 Henning Lahmann "Infecting the Mind: Establishing Responsibility for Transboundary Disinformation" (2022) *European Journal of International* Law 411.

51 Eric David, "Portee et Limite du Principe de Non-Intervention," (1990) 23(2) *Revue Belge de Droit* 350.

52 Kilovaty, *supra* note 46

53 Kate Jones, "Online Disinformation and Political Discourse: Applying a Human Rights Framework" (2019) *Chatham House Research Paper.* At 52.

54 See Malcolm N. Shaw, *International Law* (UK,: Cambridge sixth edition, 2008). At 291.; Yearbook of the ILC, 1988, vol. II, Part 2, at 64. See also *Case Concerning East Timor* (Portugal v. Australia), (1995) ICJ judgement, §29.; *Advisory Opinion on the Legal Consequences for States of the Continued Presence of South Africa in Namibia,* (1971), ICJ.

55 International Covenant on Civil and Political Rights, (1966), UNGA Res2200A(XXI), 999 UNTS, article 1.

56 See in that sense Lahmann (2020), *supra* note 45.

57 Jens Ohlin, "Election Interference: A Unique Harm Requiring Unique Solutions" 18(50) *Cornell Legal Studies Research Paper* 1.

58 Nicholas Tsagourias, "Electoral cyber interference, self-determination and the principle of non-intervention in cyberspace" in Dennis Broeders and Bibi van den Berg, *Governing Cyberspace: Behaviour, Power and Diplomacy,* (Lanham: Rowman & Littlefield, 2020)

59 Schmitt & Milanovic, *supra* note 31. At 255.

60 Lahmann (2022), *supra* note 50.

61 Schmitt & Milanovic, *supra* note 31. At 262.

62 See for instance Miriam Matthews, Katya Migacheva, and Ryan Andrew Brown, "Superspreaders of Malign and Subversive Information on COVID-19: Russian and Chinese Efforts Targeting the United States" (Santa Monica, CA: RAND Corporation, 2021).

63 International Convention on the Elimination of All Forms of Racial Discrimination, (1965), article 4.

64 Application of the international convention on the elimination of all forms of racial discrimination (Azerbaijan v. Armenia) (2021) ICJ, Order.

65 Interpretation and application of the international convention on the elimination of all forms of racial discrimination (Azerbaijan v. Armenia) (2021) ICJ, Request for indication of provisional measures of protection. At 13.

66 US Homeland Security Advisory Council Interim Report of the Countering Foreign Influence Subcommittee (2019), <https://www.dhs.gov/sites/default/files/publications/ope/hsac/19_0521_final-interim-report-of-countering-foreign-influence-subcommittee.pdf>.

67 Starbird, *supra* note 5.

68 Gérard Bronner et al. (Bronner Commission), "Enlightenment in the Digital Age" (2022), <https://www.elysee.fr/en/emmanuel-macron/2022/01/11/enlightenment-in-the-digital-age-report-1>.

69 See Benjamin Wittes et al., "Russian Electoral Interference: 2018 Midterms Edition" (19 October 2018), *Lawfare* <https://www.lawfareblog.com/russian-electoral-interference-2018-midterms-edition> See also Zhao, *supra* note 22. At 65.

70 Quincy Wright, "subversive intervention" (1960) 54(3) *The American Journal of International Law.* At 84.

71 *Supra* note 25.

72 See in that sense Coco A. Talita Dias "'Cyber Due Diligence': A Patchwork of Protective Obligations in International Law" (2021) 32(3) *European Journal of International Law* 771. See also Lahmann (2022), *supra* note 50.