Kobe University Repository : Kernel

PDF issue: 2026-07-15

“PSYOPS”, State-sponsored online disinformation in international law

Letrone William, Axel

(Degree)
博士（法学）

(Date of Degree)
2023-03-25

(Date of Publication)
2025-03-25

(Resource Type)
doctoral thesis

(Report Number)
甲第8673号

(URL)
https://hdl.handle.net/20.500.14094/0100482421

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。

Submitted the 20/12/2022

“PSYOPS”, State-sponsored online disinformation in international law
(“PSYOPS”, 国家が関与するサイバー空間での情報操作：国際法の観点から)

Graduate School of International Cooperation Studies
Department of International Cooperation Policy Studies
Academic adviser: Professor HAYASHI MIKA
170I802I
LETRONE WILLIAM AXEL

Summary of the Dissertation

"Psychological warfare", as the act of manipulating the information to change the mind of a predetermined target for strategic purposes, is now undergoing a major evolution: it is becoming "cyber-enabled." This is an opportunity to revisit the international law applicable to international communications and to elaborate under-developed areas of international law.

Since the advent of cyberspace, information power, or the capacity to influence behaviours through the manipulation, control or denial of information has become an indissociable aspect of state's national power. Meanwhile, interstate competition unfolds online as it does in the real world. States have learned to leverage information power to influence foreign populations with the aim of gaining strategic advantages. They have done so, notably, through conducting psychological operations online (hereinafter PSYOPS).

There are acceptable and unacceptable PSYOPS. Public diplomacy, unobtrusive and overt communications campaigns intended to promote a policy or an ideology in good faith may be considered acceptable PSYOPS. In contrast, disinformation campaigns, as coordinated releases of false, misleading or harmful information to induce negative feelings among the audience constitute unacceptable PSYOPS.

Unfortunately, states are not always using their information power the right way. Some were accused of sponsoring online disinformation activities to interfere in foreign democratic elections, other were accused of turning a blind eye on disinformation industries operating from within their respective territories. Yet, despite a noticeable surge of attention for online disinformation and the threat the phenomenon represents for democratic institutions around the world, the safety of individuals, and international peace in general, it remains unclear whether any of the accused has breached the law. This state of affairs is both unsurprising and frustrating. Indeed, one the one hand, psychological warfare has always existed. The question of the regulation of subversive, bellicose and defamatory propaganda, including disinformation campaigns, was already an important topic during the Cold War. However, whether by lack of commitment from the part of states, or because abuses of law would have been difficult to avoid in the context of international communications, regulations were kept at a minimum. On the other hand, today cyberspace provides a new interconnected space where states and non-state actors can reach and manipulate audiences of unprecedented size while still maintaining plausible deniability. This new paradigm is highly conducive to information misuses.

Therefore, in a context where truth and authentic information have never been so essential to the well-functioning of our modern information societies, and where "the homeland is no longer a sanctuary"[1], the question inevitably arises as to whether international law provides appropriate protection to the legitimate interests of individuals and states when these are threatened by foreign PSYOPS.

Thus, this work is a study of the international law applicable to PSYOPS, with a focus on online disinformation. It gives an overview of the relevant laws and identifies salient points in the law that should require more consideration. Overall, this work is based on two general ideas; First, the current legal framework for the conduct of psychological warfare perpetuates impunity for information misuse. There is a need to clarify the law in light of new and foreseeable developments in the domain of information and communication technologies. Second, despite their relevance to the subject at hand, too little attention is given to human rights and the correlative due diligence duties of prevention when it comes to regulate state-sponsored disinformation campaigns. And yet, international human rights law contains particularly interesting tools for setting adequate boundaries to psychological warfare and preventing information misuses that can have dramatic consequences on the rights of the audience.

This work is divided into five chapters. The first two chapters set the stage for the legal assessments. The first chapter focusses on defining disinformation. It does so, by commenting on the defining features of the phenomenon, by comparing disinformation with the different types of propaganda, and by highlighting the difficulties commonly associated with analysing disinformation campaigns, such as the questions of intent and causality.

The second chapter takes a historical approach to the question of the regulation of international communication. It is shown that, while pre-cyber propaganda activities may contain disinformation aspects, the phenomenon as it is currently understood by states and practitioners mostly escapes straight-forwards legal characterizations. The dearth of official statements referring to the legal status of foreign disinformation campaigns shows that, the same way as espionage activities, PSYOPS in general, and disinformation campaigns in particular, belong to the grey areas of international law. In other words, state-sponsored PSYOPS, including disinformation campaigns, are presumed lawful until proven otherwise. But when

1 Lt. Gen. Lori Reynolds in Andrew Milburn & Shawna Sinnott, "Competing for Influence: Operations in the Information Environment" (16 January 2021), Modern War Institute Irregular Affairs podcast, <https://mwi.usma.edu/competing-for-influence-operations-in-the-information-environment/>.

exactly would state-sponsored disinformation campaigns breach international law? The next chapters provide an answer to this question.

A third chapter looks at state-sponsored disinformation activities through the lenses of relevant international rules, principles and treaties safeguarding the interests of states against the behaviours of foreign actors. In doing so, it elicits the circumstances under which a foreign disinformation campaign could exceed the boundaries set by said rules. It is first shown that the 1936 Broadcasting Convention and the 1953 Correction Convention, albeit interesting in their approach to international communications, are difficult to operationalize in presence of covert, cyber-enabled activities. The assessment then turns to the customary principles of the non-use and threat of force and non-intervention. It is explained that these principles could find an application to the context of foreign disinformation activities in very limited circumstances. Note is for example taken of the contemporary debate pertaining to the need to re-interpret the non-intervention principle. It is however cautioned against tinkering the rules for the sole purpose of encompassing more modern forms of PSYOPS, as doing so risks rendering them meaningless, making them prone to instrumentalization by states.

The fourth chapter focusses on human rights law. In covering the relevant rights, the chapter delineates the scopes of the freedom of thoughts and opinion, the right to internal self-determination and the right to life. It is thus shown that state-sponsored disinformation also infringes fundamental human rights. In fact, human right law seems the better approach to state-sponsored disinformation. More specifically, is it noted that each one of the covered right could compensate for the *lacunae* of the aforementioned state-centric rules without entailing excessive risks of instrumentalization by states. The chapter concludes by suggesting an interpretation of human rights extraterritoriality that could give effectivity to the framework thus defined.

The fifth and last chapter deals with the question of attribution and negligence. Notably, it is shown that classic attribution regimes may not provide an adequate path for the attribution of disinformation campaigns to a state, even when a connivence between a state and the private actors involved in the spread of disinformation is obvious. The means of responses in case of successful attribution remain extremely limited anyway. As a solution, it is suggested to look at state accountability from the perspective of the positive duties held by states with respect to the prevention of transboundary international law violations arising from activities perpetrated by private actors in their territories. In other words, information misuses should be looked at through the lenses of due diligence. Due diligence obligations must first be grounded in the law in order to exist. That is to say, duties of due diligence with respect to disinformation arise only

if it is attested that disinformation campaigns are susceptible of breaching the law. On this point, this work shows that PSYOPS including disinformation may impair fundamental human rights, and even, in rare occasions, infringe on the rights of states. Hence, it is argued that a state of emission should incur liability for transboundary international law violations, including violations of human rights, arising from disinformation activities conducted by non-state actors operating from within its territory, or under its control, if it is apparent that said state was negligent in preventing information misuses it was entitled and in capacity to act upon. Additionally, it is recalled that under the consensual principle of territorial jurisdiction, recipient states are bound by obligations of prevention of human rights violations *vis à vis* their own population. This translates, in the disinformation context, in various due diligence duties to ensure that the collective will is protected against undue attempts to corrupt the authenticity of democratic processes in times of democratic elections, that adequate laws are in place to regulate egregious speeches, and, most importantly, that independent fact checking networks are not repressed and that sufficient resources are devoted to the development of information literacy. One may also think about an obligation of the receiving state to ensure that the largest service providers are diligent in preventing human right violation arising from information misuses in their platforms.

The burden of prevention should be shared between states and digital platforms, as the latter's services enable psychological warfare. Thus, key social media platform owners should implement corporate due diligence duties in the context of their activities. Of course, even non-state actors would be required to act diligently in prevention abuses arising from information misuses. This means, above all, that the platforms' own prevention efforts are not done to the detriment of the rights of the users. Therefore, social media platforms should devote more attention to designing recommender systems that are less manipulative and less prone to instrumentalization by an external actor. With that being said, corporate due diligence should not deviate attention from improving state accountability in the disinformation context.

Too often psychological warfare is conducted in complete disregard for human rights. Thus, not only is it time to consolidate human rights in light of users' ever-growing vulnerability, but opening the possibility of engaging the responsibility of states and platforms for blatant negligence in dealing properly with disinformation campaigns seems a promising way forwards.

Table of content

General Introduction……..……1

Chapter I. Conceptualizing state-sponsored online disinformation campaigns

1. Introduction…….20

2. Disinformation defined……....20

3. The cyber factor…….…….26

4. Covert or overt……..…27

5 The structure of state-sponsored PSYOPS……...31

6. The question of intent……..……34

Chapter II. Disinformation in the grey zones of international law

1. Introduction…….……41

2. Online disinformation and clearly illegal propaganda types……...43

3. The grey zone and its relevancy to online disinformation…….……50

4. States' views on the legal status of peacetime transboundary disinformation campaigns ……...56

Chapter III. State-sponsored cyber-disinformation campaigns and the state-centric rules of international law

1. Introduction…….. 67

2. The 1936 International Convention concerning the Use of Broadcasting in the Cause of Peace…….68

3. The 1953 Convention on the International Right of Correction…….72

4. State-sponsored cyber-disinformation campaigns as uses or threats of the use of force……..74

4.1 International communications and the maintenance of international security and peace……..75

4.2 Disinformation as illegal “force”…….……79

4.3 Disinformation as illegal threats to use force……..82

4.4 Disinformation and the concept of “armed attack”…….82

4.5 Remarks on international humanitarian law and the concept of “attack”…..…85

4.6 State-sponsored disinformation campaigns and the specific regime of warmongering propaganda……86

4.6.1 War propaganda and disinformation……87

4.6.2 The legal regime of war propaganda……88

4.6.3 The scope of the prohibition ……...90

5. State-sponsored cyber disinformation campaigns as prohibited interventions……..…91

5.1 International relations, foreign influence and foreign interference ……....93

5.2 The function of non-intervention ……...95

5.3 The ubiquity of the principle of non-intervention in international instruments…..96

5.4 The customary rule of non-intervention……100

5.5 Unproblematic applications of the principle of non-intervention to state-sponsored disinformation campaigns……...105

5.6 Cyber-enabled disinformation campaigns and the concept of reserved domain... 107

5.7 Cyber-enabled disinformation campaigns and the concept of coercion…….....110

5.7.1 Dictatorial coercion and online disinformation campaigns…….....113

5.7.1.1. The intent to coerce in cyber-enabled disinformation operations …….113

5.7.1.2 The dissemination of problematic content as generator of credible pressure……..….136

5.7.2 Non-dictatorial coercion and online disinformation campaigns……..…121

5.7.3 Remarks on the incidence of “coercive” message on the overall legality of a campaign…….……..……131

5.8 Alternative views on non-intervention……..…..132

6. State-sponsored cyber-enabled disinformation campaigns as breaches of state sovereignty……..137

6.1 The principle of state sovereignty…….…….137

6.2 The content of the principle of sovereignty …….…..…138

6.2.1 The uncertainties surrounding the content of the principle of sovereignty……...……...139

6.2.2 States' views regarding the content of the principle of sovereignty……..141

6.3 State-sponsored disinformation campaigns and the two functions of state sovereignty……..143

6.3.1 Disinformation campaigns as violations of territorial sovereignty……..144

6.3.2 Disinformation campaigns as interferences with/ or usurpations of/ the inherently governmental functions of a state……...…...145

6.3.3 Conceptual obstacles to the operability of "secondary" non-interference principle stemming from the principle of sovereignty……149

Chapter IV. State-sponsored disinformation campaigns in international human rights law

1. Introduction……..…….152

2. The freedom of thought and the right to hold opinion without interference……....154

3. The right to self-determination……...160

4 The right to equality and non-discrimination……..163

5. The right to life……...171

6. Freedom of expression and speech regulations…….……...……..173

7. The question of the extraterritorial application of human rights treaties…….……178

Chapter V. Attribution, negligence, and due diligence in the context of online disinformation activities

1. Introduction ……..……185

2. Technical aspects of the attribution of cyber-disinformation activities to a state……....186

2.1. The peculiarities of the attribution of cyber-enabled disinformation campaigns……187

2.2 Evidentiary standards for the attribution of disinformation to a state…....189

2.3 The question of the disclosure of electronic communications……....192

3. The attribution of unlawful disinformation campaigns to a state…….....194

3.1 Disinformation activities conducted by a state directly……..…195

3.2 Attribution of a conduct acknowledged and adopted by the state as its own……..196

3.3 Attribution of unlawful private conduct to a state……..…200

3.3.1 The de facto agency test, or "strict control test"……201

3.3.2 Responsibility of a state for specific acts perpetrated by private entities and individuals, or "effective control test"…….203

3.3.3 Subsequent developments regarding the effective control test…205

4. International liability for negligence in the disinformation context……..…….....210

4.1 The no-harm principle, or state responsibility for negligence in preventing transboundary harm arising from the physical consequences of activities not prohibited by international law …….....…211

4.2 Limitations to the enforcement of the no-harm principle with respect to disinformation activities ……..……216

4.3 State liability for negligence in preventing and mitigating infringement on the rights of other states and of persons located outside the territory of the state of emission ……..……218

4.3.1 Due diligence and the prevention and abatement of transboundary disinformation activities infringing on the rights of other states…...……219

4.3.2 Due diligence and the prevention of human rights impairments arising from transboundary disinformation activities ……..……...……...….223

4.3.2.1 Due diligence and the prevention and abatement by states of transboundary conduct infringing on the rights of persons located within their territory or jurisdiction …….…….……...224

4.3.2.2 The problematic question of the prevention and abatement by states, of human rights violations affecting persons located outside their effective overall control but arising from activities initiated in their territory…….……...228

5. The need for a regime of corporate liability of large social media platforms and online news providers for human rights impairments arising from online information misuses…….……..……..……...……...…….233

Concluding remarks……...238

Select bibliography

List of abbreviations

ARSIWA: International Law Commission's draft Articles on Responsibility of States for Internationally Wrongful Acts

ASEAN: Association of Southeast Asian Nations

COVID: Novel Coronavirus

DSA: Digital Service Act

EU: European Union

FT&RHO: Freedom of Thoughts and Right to hold opinion without interference

ICC: International Criminal Court

ICJ: International Court of Justice

ICCPR: International Covenant on Civil and Political Rights

ICRC: International Committee of the Red Cross

ILC: International law Commission

IO: Information operation

IRD: British Information Research Department

NATO: North Atlantic Treaty Organisation

ODNI: Office of the Director of the National Intelligence

OSCE: Organization for Security and Co-operation in Europe

PSYOPS: Psychological operations

UDHR Universal Declaration of Human Rights:

UN: United Nations

UNGA: United Nations General Assembly

UK: United Kingdoms

US: United States

USSR: Union of Soviet Socialist Republics

General Introduction

"*A new type of warfare... is starting to appear. I call it, for the sake of argument, mental war. It's when the aim of this warfare is the destruction of the enemy's understanding of civilizational pillars*"[1]

Andrei Ilnitsky

From Massena's quaker guns at the siege of Genoa in 1800 to the English operation *Sea Lion* during the Second World War, states have always used deception strategies involving methods of information control, manipulation and denial in order to take an advantage over their enemies.

But the tactics of information manipulation were not confined to wartime. For example, in 1870 German chancellor Otto von Bismarck intentionally deceived French and Prussian audiences by disseminating domestically and internationally a falsified version of a report of a diplomatic exchange between the French ambassador to Prussia and the King Wilhelm I, during which the former asked political guaranties to the later.[2] In a communication that will be known as "the EMS dispatch", the German chancellor exaggerated the aggressiveness of the French ambassador, and removed the conciliatory attitude of King Wilhelm and the respectful tone of the exchange from the report. As planned by its author, the dispatch infuriated the French and the German public, both perceiving the reported words of the other correspondent as a blatant sign of disrespect towards their respective nations. The event precipitated France in a war against a confederation of German states led by Prussia, which concluded in a bitter defeat for France.

Closer to contemporary operations of information manipulation, declassified documents about the work of the British Information Research Department (IRD) revealed that the United Kingdom (hereinafter UK) was heavily involved in the planning and execution of influence activities during the cold war. A substantial part of the IRD's work consisted in covertly

[1] Andrei Ilnitsky is an advisor to the Russian defense minister. Comments reported by Gabrielle Tétrault-Farber & Andrew Osborn "U.S. wages psychologicial war on Moscow - Russian defense adviser" (25 March 2021) *Reuters*, <https://www.reuters.com/article/us-russia-usa-minds/u-s-wages-psychologicial-war-on-moscow-russian-defense-adviser-idUSKBN2BH1TD>.

[2] Todd H. Hall, "On Provocation: Outrage, International Relations, and the Franco–Prussian War" (2016) 26(1) *Security Studies* 1.

disseminating anti-communism propaganda in third world countries in order to counter the propaganda efforts of the Soviet Union (hereinafter USSR). To that end, the IRD occasionally manipulated the information, thereby purposefully deceiving foreign populations for the purpose of consolidating western influence over target audiences.[3]

The work of the IRD is only one example among many other. The major communist powers of the Cold War disseminated stories and rumours hostile to the West, and in particular to the United States (hereinafter US) in non-aligned territories. In response, the US transmitted anti-communist radio broadcasts into communist territories.[4] Rapidly, propaganda had "become part of modern international relations."[5] In the course of their relentless battles of words, both sides made ample use of propaganda tactics. They also occasionally made use of public deception methods, which are nowadays known as *disinformation operations* and *disinformation campaigns*.

- The rise of information power

Before tackling the question of disinformation, it is useful to keep in mind that in international relations, *power* does not hinge solely on violence and military superiority. In the words of the former French chief of staff, Pierre de Villier, "To win the war is not enough, it is peace that must be won."[6] In other words, while violence alone may lead to victory on the battlefield, it is of little help when it comes to secure stability in the relation between international actors. A nation that places great emphasis on military power will not inspire stability and as such, will struggle to project influence. Asserting authority is best achieved through a combination of various means of influence. Truly, the ability to inspire behaviours is more important than the ability to wage wars. It is the ability of an actor to leverage all means

[3] Rory Cormac, "Techniques of covert propaganda: the British approach in the mid-1960s" (2019) 34(7) *Intelligence and National Security* 1064; Sanchia Berg, "'Fake news' sent out by government department" (19 March 2019), *BBC News,* <https://www.bbc.com/news/uk-politics-47571253>.

[4] See for instance US Congress, Report of the Congressional Committees Investigating the Iran-Contra Affair, (1987); John B. Whitton, "Cold War Propaganda" (1951) 45(1) *American Journal of International Law* 151.

[5] See Gerhard von Glahn, "The Case for Legal Control of "Liberation" Propaganda" (1966) 31 *Law and Contemporary Problems* 553, at 573; Vernon Van Dyke, "The Responsibility of States for International Propaganda" (1940) 34(1) *The American Journal of International Law* 58.

[6] [translated by author]. In french : « Gagner la guerre ne suffit pas, il faut gagner la paix » General Pierre de Villiers, « Gagner la guerre ne suffit pas à gagner la paix » (18 January 2016), *Le Monde,* <https://www.lemonde.fr/idees/article/2016/01/20/gagner-la-guerre-ne-suffit-pas-a-gagner-la-paix_4850136_3232.html>.

at his disposal, including diplomatic, economic and information tools, to "win a battle before it is ever fought" [7] and to influence behaviours, that makes his national power.[8]

Information power is one aspect of national power that refers to an actor's ability to control, manipulate and deny information within and without cyberspace, which includes information intelligible to humans (semantic information) or information intelligible to machine (bytes), in order to influence behaviours, rally public opinions around common aspirations, or to subordinate a competitor. Whether during wartime or peacetime, the ability to control, manipulate and deny information for the purpose of gaining a strategic advantage has always been an important facet of national power.[9] Today some authors speak of "digital power" or "cyberpower" to refer specifically to information power when it involves cyber means.[10] Thus, digital power is understood as "any actor's ability to exploit digital data to help influence the behavior of other actors on the international stage and to achieve its own ends."[11]

At least until the early 1930's, information power was limited. The international context after the second world war and technological developments in the field of international communications have contributed to the rise of information power.[12] An important feature of international relations after the end of the second world war was the shift away from pure violence as a way to resolve disputes. Since the end of the Second World War, states are inclined, and even obliged to resort to non-forcible means of competition in place of brute force

[7] Adage attributed to Sun Tzu, whose writings are believed to have inspired the Chinese "three warfares" doctrine. The Chinese three warfares doctrine perfectly embodies the idea of a multi-dimensional national power that is not confined to military power. See notably Sangkuk Lee, "China's 'Three Warfares': Origins, Applications, and Organizations" (2014) 37(2) *Journal of Strategic Studies* 198.

[8] Dan Kuehl, "Defining Information Power" (1997) 115 *Strategic Forum* 1; 2021-updated version of Strategic vision of the French Chief of Defence staff, which emphasizes the diversity of the "fields of confrontations" in which France is engaged against its opponents. The document explains that there is a triptych of contexts in military strategy; competition, dispute, and confrontation. The first notion "competition" is defined in the paper as a "war before the war" and "takes place in all domains: diplomatic, informational, military, economic, legal, technological, industrial and cultural". See General Thierry Burkhard, "Vision stratégique du chef des état-major des armées" French Ministry of the Armies, October 2021, <https://www.defense.gouv.fr/ema/chef-detat-major-armees/vision-strategique-du-chef-detat-major-armees-fresgb>). See also a definition of influence operation in Eric V. Larson, et al., "Foundations of Effective Influence Operations A Framework for Enhancing Army Capabilities" (2009) Santa Monica, CA: RAND corporation, <https://www.rand.org/pubs/monographs/MG654.html>.

[9] See notably Walter Lippman, *Public Opinion* (Lanham: Start Publishing LLC, 2015 (reissue)). See also among other Heidi Tworek, "Disinformation: It's History" (14 July 2021), Centre for International Governance Innovation, <https://www.cigionline.org/articles/disinformation-its-history/>; Jack M. Balkin, "The First Amendment Is an Information Policy" (2012) 41(1) *Hofstra Law Review* 13; Francois Chauvancy, "Cyberespace, operations d'information, influence" in Olivier Kempf, Stéphane Dossé et Christian Malis (dir.) *Le cyberspace - Nouveau domaine de la pensée stratégique* (Paris: Economica, 2013).

[10] See generally Jean-Christophe Noel, "What is Digital Power?" (November 2019), IFRI Study Series, <https://www.ifri.org/en/publications/etudes-de-lifri/what-digital-power>.

[11] Noel, *Ibid.*

[12] See Nolan Higdon, "What is Fake News? A Foundational Question for Developing Effective Critical News Literacy Education" (2020) 29(1) *Democratic Communiqué* 1.

in order to shape the international landscape to their advantage. This evolution can be attributed to four major factors; a global aversion for violence after the bloodsheds of the two world wars, the deepening of economic interdependencies between states, the developments of nuclear weapons as means of deterrence, and the development of international law. As a result, competition between major powers gained in sophistication. Today, interstate competition takes the form of complex engagements where non-violent tools of destabilisation running afoul of most uncontested principles of international law are prioritized over force.[13] The broad category of non-violent tools used by states to influence their adversaries notably includes economic coercion, low-level cyber threats, propaganda and its subset disinformation.

- On *psychological* war

An actor exercises information power through information operations, (hereinafter IOs). An IO can be defined as the act of manipulating, denying or controlling semantic information and bytes to gain a strategic advantage over a competitor. The term "information warfare" (or information war), is often used to refer to the situation where a variety of actors leverage their respective information powers to confront the legitimacy and the credibility of a rival, or to protect their reputation through the use of IOs.

The term information war gained popularity during the Gulf War, during which the United States made ample use of IOs to assist the troops on the field for the first time, The tactic was so successful that it subsequently initiated a "revolution in military affairs." From the 1990's onward, states incorporated new communications technologies in their military strategies, and the concept of information warfare made its apparition in military doctrines around the world.[14]

Information warfare is not confined to the deployment of strategies by states to "win hearts and minds." Indeed, like IOs, the term "information" in "information war" is to be understood in its broadest sense. Not only does "information" here mean content of semantic nature (images, ideas, speeches and messages) that are intelligible to humans and contained online, but it also refers to the information intelligible to machines, the bytes, codes, and more broadly

[13] Deschaux-Dutard speaks of a global transition from an international context of *Bellum* during the XXth century to a context of *Guerra* in the XXIst characterised by the complexification of modern conflicts and more diffused levels of violence. Delphine Deschaux-Dutard, *Introduction à la sécurité internationale* (Grenoble: PUG, 2018), at 227; Michael J. Mazarr, "Mastering the Gray Zone: Understanding a Changing Era of Conflict" (2015) *US Army War College*, *Monographs, Books, and Publications* 428, at 2.

[14] Herbert Lin & Jaclyn Kerr, "On Cyber-Enabled Information Warfare and Information Operations" in Paul Cornish (ed.) *The Oxford Handbook of Cyber Security* (Oxford: Oxford University Press, 2021).

all information meant for system manipulation and computer-controlled machinery.[15] An early document from the US Air Force for example defines information warfare as "any action to deny, exploit, corrupt, or destroy the enemy's information and its functions; protecting ourselves against those actions; and exploiting our own military information functions."[16] It is specified in the document that information warfare includes "psychological warfare" as well as conduct involving "destruction."[17] The 2011 Russian Conceptual Views on the Activity of the Russian Federation Armed Forces in Information Space similarly defines information war as a "conflict between two or more states in information space with the aim of causing damage to information systems, processes and resources, critically important and other structures, subverting the political, economic and social systems, mass psychological work on the population to destabilise society and the state, and coercing the government to take decisions in the interests of the opposing side."[18]

In light of the above, "information warfare" both refers to the use of typical cyberattacks targeting computer, phone and other connected devices such as hacking operations or virus-planting, and the use of information manipulation tactics targeted at the minds of an audience. While the former exclusively concerns cyber-enabled operations, the latter encompasses cyber-enabled and non-cyber activities.

Libicki indeed remarks that there are "several distinct forms of information warfare."[19] Among these, "psychological warfare," and in particular, what Libicki calls "operations against the national will", [20] are the concepts that relate the most to the subject of the present work. In psychological warfare, "information is used to change the minds of friends, neutrals, and foes." [21] Psychological warfare is thus waged *via* psychological operations (hereinafter PSYOPS.) PSYOPS are a subset of IOs. PSYOPS is the communication of semantic

[15] Martin Libicki, "Cyberdeterrence and Cyberwar" (2009), Santa Monica, CA: RAND corporation, <https://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG877.pdf>, at 12.; Wei Jincheng, "Information war: a new form of people's war" (1996) *Military Forum column, Liberation Army Daily*, Chinese intelligence agency-related documents, <https://irp.fas.org/world/china/docs/iw_wei.htm>.; Stephens, Dale, "Influence Operations and International Law" (2020) 19(4) *Journal of Information Warfare* 1, at 2.

[16] "Cornerstones of Information Warfare" USAF, (Washington, D.C.: Department of the Air Force, 1995), <https://nsarchive.gwu.edu/document/15889-united-states-air-force-cornerstones>, at 4.

[17] *Ibid.*

[18] See "Conceptual Views on the Activity of the Russian Federation Armed Forces in Information Space" (2011), Section 1, Fundamental Terms and Definitions. See also Molly K. McKew, "The Gerasimov Doctrine" (September/October 2017), *POLITICO,* <https://www.politico.com/magazine/story/2017/09/05/gerasimov-doctrine-russia-foreign-policy-215538/>.

[19] See Martin Libicki, "What is information warfare?" (1995) 28 *Strategic Forum*, Center for Advanced Concepts and Technology Institute for National Strategic Studies, <https://apps.dtic.mil/sti/pdfs/ADA367662.pdf>, at x.

[20] *Ibid.,* at 35.

[21] Libicki, *Supra* note 19, at x.

information at scale and in a coordinated manner, to a pre-determined audience, in order to influence behaviours and to realize strategic objectives.[22] A PSYOP involves information control, manipulation and denial methods. The most common type of PSYOPS takes the form of covert or overt communications. In light of this definition, propaganda and disinformation campaigns constitute PSYOPS when waged or condoned by a political entity to realize strategic objectives.

It is fundamental to make a clear distinction between cyber-enabled PSYOPS relying on semantic information and the other IOs that rely on the use and/or manipulation of computer information. Indeed, the two phenomena are so distinct in nature that the legal assessments and the available legal responses will differ dramatically depending on whether PYOPS relying on semantic information or other IOs relying on bytes are scrutinized.[23]

As this study focusses exclusively on the conduct that involve the manipulation, the concealment, or the spread of semantic information, the more specific term of "psychological warfare" will be preferred over "information war" and the more relevant term "PSYOPS" will be preferred over "IOs."

PSYOPS can be conducted at the tactical level, alongside conventional uses of force. When it is the case, PSYOPS may be called wartime PSYOPS. PSYOPS may also be conducted at the strategic level, during peacetime. When it is the case, they constitute peacetime PSYOPS.[24]

Wartime PSYOPS constitute means of hybrid warfare.[25] They are accessory to conventional uses of military forces and as such, conceived so as to maximize the chances of success of an imminent or ongoing military engagement. In other words, they act as "force multipliers". In the field, wartime PSYOPS serve various specific purposes, such as to promote a military intervention in the eyes of the affected population, to sow confusion in the ranks of an opponent or in the adversary's citizenry, to scare or discourage an opponent and/or its population in order to break their moral and will to resist.

---

[22] Interestingly, the US DOD defines PSYOPS as "planned operations to convey selected *truthful* information and indicators to foreign audiences to influence their emotions [and] motives […] to induce or reinforce foreign attitudes and behavior favorable to the originator's objectives" See Ashley C. Nicolas, "Taming the Trolls: The Need for an International Legal Framework to Regulate State Use of Disinformation on Social Media" (2018) 107 *Georgetown Law Journal* 36, at 39. (Quoting Robert J. Kodosky, *Psychological Operations American Style: The Joint United States Public Affairs Office, Vietnam and Beyond* (Lanham: Lexington Books, 2007). This definition has to be put in context. The official US position is that it does not use false information in the context of its PSYOPS. However, PSYOPS may include false and misleading information.

[23] Stephens, *Supra* note 15, at 3.

[24] For a more elaborated presentation of the areas of operations in which PSYOPS may be used, see Peter J. Smyczek, "Regulating the battlefield of the future: the legal limitations on the conduct of psychological operations (PSYOP) under public international law" (2005) 57 *Air Force Law Review* 209.

[25] See the ICRC definition of "Means of warfare", <https://casebook.icrc.org/glossary/means-warfare>.

Wartime PSYOPS exist since the earliest times. They are still common today. For instance, the French *Centre interarmées des actions sur l'environnement* deployed PSYOPS capabilities in Mali to obtain popular support and understanding for the mission of the French armed forces. Before its invasion of the Georgian province of South-Ossetia in 2008, Russia deployed PSYOPS capabilities to convince Russian and Georgian audiences that it intervened to "protect" the ethnic Ossetian. In the same vein, false claims that the Ukrainian president fled Ukraine during Russia's military intervention in Ukraine in 2022 were relayed by the Russian news agency TASS in what appears to have been an attempt to discourage the Ukrainian resistance.[26] During the Persian Gulf War, the US deployed PSYOPS capabilities in the context of the Desert Shield and Desert Storm operations. PSYOPs ranged from the dissemination of aerial leaflets before the ground assault phases to the dissemination of anti-Saddam Hussein broadcasts in Iraq and in Kuwait, to bolster Kuwait's resistance, discourage Iraqi troops, and sow discontent among the Iraqi population.[27]

The objectives of peacetime PSYOPS are way less specific than those of wartime PSYOPS. It also generally more difficult to determine the true instigator of a peacetime PSYOPS. For these reasons, they are particularly difficult to assess and to address. Peacetime PSYOPS go way beyond the purpose of providing support for military activities. Peacetime PSYOPS often take the form of permanent or semi-permanent campaigns initiated so as to generate long term results. When directed at populations, peacetime PSYOPS may serve various purposes such as influencing election outcomes, making specific economic policies acceptable in the eyes of a community, sowing unrest in the territory of another state, remodelling the expectations of the audience about the conclusion of a conflict to which they are not taking part, and broadly, subverting the will of a sovereign power.[28]

Peacetime PSYOPS may be constructed as parts of a broader hybrid strategy involving several non-forcible operations, or as autonomous and persistent efforts exclusively reliant on information manipulation tactics. A famous example of pre-cyber peacetime PSYOPS was the "Nayirah testimony", a false testimony made at the United States Congressional Human Rights

[26] David Klepper, "Russian propaganda 'outgunned' by social media rebuttals" (5 March 2022) *APNEWS,* <https://apnews.com/article/russia-ukraine-volodymyr-zelenskyy-kyiv-technology-misinformation-5e884b85f8dbb54d16f5f10d105fe850>.

[27] For a detailed analysis of the US-led PSYOPS in the Iraq war and an assessment of their efficiency, see notably, Major Peter A. Whitenack, "An Analysis of Gulf War Psyops and Their Applicability to Future Operations" (1993) *GlobalSecurity.org,* <https://www.globalsecurity.org/military/library/report/1993/WPA.htm>.

[28] Questioning the odds of a cyber "war", Rid concluded that cyberspace would rather foster subversion than cyber-enabled violence. Today, the issue of online disinformation seems to prove him right. See Thomas Rid, "Cyber War Will Not Take Place" (2012) 35(1) *Journal of Strategic Studies* 5.

Caucus following Iraqi invasion of Kuwait, about Iraqi soldiers forcibly removing babies from their incubators in Kuwaiti hospitals. The false testimony was widely relied by the US authorities to demonize Saddam's regime and justify their intervention. More recently Russia conducted conventional and cyber-enabled PSYOPS operations to justify its military intervention in Ukraine in March 2022, spreading unsubstantiated claims of an ongoing "genocide" targeting the Russian minorities in East-Ukraine. But it is in cyberspace that peacetime PSYOPS are the most present. Examples of cyber-enabled peacetime PSYOPS are provided in the following paragraphs.

- Cyber-enabled psychological warfare

Although states have always conducted PSYOPS, including propaganda and information manipulation tactics,[29] the advent of cyberspace has improved PSYOPS tenfold. The impact of the information revolution was unprecedented. First, advancements in communication technologies revolutionized human interactions, which became instantaneous, and unbounded by distance. As a result, the information spreads farther and faster online. Second, information power is within the reach of way more actors than during the pre-cyber era. Non-state actors may wage PSYOPS out of pure self-interest, be it financial or political, and may be used as unwitting information spreaders by states. In a nutshell, psychological warfare has gained so much from cyberspace that it has become an indissociable aspect of contemporary power dynamics.[30]

By enabling information plurality and the free access to information, cyberspace positively impacts human rights. Yet, some of cyberspace's features and shortcomings are also conducive to particularly nefarious behaviors.[31] Cyberspace fosters the use of peacetime PSYOPS to the point that they have become ubiquitous.[32] In contemporary psychological wars, the online information space, including the web and social media applications is a place of fierce power competition where information is "weaponized"[33] to achieve political aim, and where public

[29] Megan Ward, Shannon Pierson & Jessica Beyer "Formative Battles: Cold War Disinformation Campaigns and Mitigation Strategies" (2019) Wilson Center Science Technology Innovation Program.

[30] Joseph S. Nye, "The Information Revolution and Soft Power" (2014) 113(759) *Current History* 19; Noel, *Supra* note 10.

[31] Jack M. Balkin "The Information Society from an Antihumanist Perspective" in Ramesh Subramanian and Eddan Katz (eds.) *The Global Flow of Information: Legal, Social, and Cultural Perspectives*, (New York: New York University Press, 2011), 232-246.

[32] See notably Libicki, *Supra* note 19, at 36.

[33] The French Secretary of the Army for example declared at the occasion of the presentation of the new military doctrine for the cyberwar of influence (L2I); *"*[T]he false, manipulated, subverted information is a weapon." In French : « *L'information fausse, manipulée, subvertie, est une arme. »* [translated by author] See

opinion is used as a "proxy for attacking the state."[34] In this context, gaining control over the narrative, and by extension over the information space, be it through the destruction of a target's reputation or the systematic legitimization of one's actions domestically or abroad, becomes a strategic goal.[35] As Chinese President Xi Jinping remarked in 2013, the "[T]he internet has become the main battlefield for the public opinion struggle."[36]

As far as state-sponsored peacetime cyber PSYOPS are concerned, a 2017 Freedom House report notes that amongst the 65 states analysed in the study, half of them were employing "armies of "opinion shapers" to spread government views, drive particular agendas, and counter government critics on social media"[37] adding that "[t]he number of governments attempting to control online discussions in this manner has risen each year since 2009."[38] In the same vein, a 2019 Oxford study notes that the number of countries resorting to organised social media manipulation campaigns to suppress fundamental human rights; discredit political opposition; and drown out political dissent had increased by 150% between 2017 and 2019.[39] More recently, the Oxford Internet Institute found that in 2020 only, a total of 81 states have engaged in propaganda and disinformation operations online, with the vast majority leveraging the online information space for mounting smear campaigns or attacking domestic opposition.[40]

High profile cyber-enabled PSYOPS including disinformation campaigns frequently make the headlines. The Russian meddling into the 2016 US presidential election being hands-down the most cited example of a state-sponsored disinformation campaign waged principally through cyberspace. Contrary to the common belief, this well-documented peacetime PSYOPS did not specifically targeted the election process. Subsequent investigations indeed revealed

« Florence Parly présente la doctrine militaire de lutte informatique d'influence », (22 October 2021), *Defense.gouv,* <https://www.defense.gouv.fr/ema/actualites/armees-se-dotent-dune-doctrine-militaire-lutte-informatique-dinfluence-l2i>.

[34] Michael P. Ferguson, "The Evolution of Disinformation: How Public Opinion Became Proxy" (14 January 2020), *The Strategy Bridge,* <https://thestrategybridge.org/the-bridge/2020/1/14/the-evolution-of-disinformation-how-public-opinion-became-proxy>.

[35] See Noel *Supra* note 10; See some doctrines on psychological operations "Psychological operations" (7 January 2010) US Joint Chief of Staff, Joint Publication 3-13.2; "Allied Joint Doctrine for Psychological Operations" NATO, AJP-3.10.1(A); "Doctrine militaire de lute informatique d'influence (L2I)" (October 2021) Ministère Francais des Armées.

[36] *"网传习近平 8•19 讲话全文：言论方面要敢抓敢管敢于亮剑"* (2013), *China Digital Times,* <https://chinadigitaltimes.net/chinese/321001.html> [translated by author].

[37] Sanja Kelly et al., "Manipulating Social Media to Undermine Democracy" (2017) Freedom House, Freedom on the Net.

[38] *Ibid.*

[39] Samantha Bradshaw & Philip N. Howard, "The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation" (2019) Oxford, UK: Computational Propaganda Project Working Paper.

[40] Samantha Bradshaw, Hannah Bailey, Philip N. Howard, "Industrialized Disinformation 2020 Global Inventory of Organized Social Media Manipulation" (2021) Oxford UK: Computational Propaganda Project Working Paper.

that the election meddling was only a subset of a massive PSYOPS waged by the Internet Research Agency (hereinafter IRA), a Russian company specialized in conducting influence operations online for the account of the Kremlin, against the US between 2014 and 2016. [41] The campaign was allegedly aimed at dividing the American population, taking advantage of pre-existing racial tensions, and to sow discontent towards the democratic institutions. Other examples of cyber-enabled PSYOPS can be cited. For example in 2020 Meta, the parent company of Facebook, uncovered two vast operations allegedly conducted by France and Russia in Mali since 2019, which were intended to discredit the other side in the eyes of the Malian population.[42] More recently in September 2022, Graphika and the Stanford Internet Observatory conducted an investigation on what they qualified "the most extensive case of covert pro-Western IO on social media".[43] The investigation followed the publication in August 2022 of the report from Meta about its removal of large pro-western influence networks which, for several years, covertly advanced anti-China, anti-Russia and anti-Iran narratives online. Following Meta's report, the Pentagon opened an audit of its own PSYOPS. "Our adversaries are absolutely operating in the information domain, […] There are some who think we shouldn't do anything clandestine in that space. Ceding an entire domain to an adversary would be unwise. But we need stronger policy guardrails" reportedly declared a US senior defence official after the event.[44] Earlier in 2022 the American cybersecurity firm Mandiant uncovered a vast pro-China cyber-enabled influence effort. The campaign reportedly included efforts to "reshape the international image of Xinjiang, criticism of the U.S. and its allies, and attempts to discredit critics of the PRC government"[45]

[41] Renee DiResta et al., "The Tactics & Tropes of the Internet Research Agency", (2019), New Knowledge, U.S. Senate Documents Congress of the United States. See also US Office of the Director of National Intelligence, (ODNI), "Background to 'Assessing Russian Activities and Intentions in Recent US Elections': The Analytic Process and Cyber Incident Attribution" (2017), Intelligence Community Assessment. <https://www.dni.gov/files/documents/ICA_2017_01.pdf>. For an early insight into the IRA's activities, see Adrian Chen, "The Agency", (2 June 2015) *The New York Times Magazine*, <https://www.nytimes.com/2015/06/07/magazine/the-agency.html>.

[42] Nathaniel Gleicher, "Removing Coordinated Inauthentic Behavior from France and Russia" (15 December 2020) *META Newsroom*, <https://about.fb.com/news/2020/12/removing-coordinated-inauthentic-behavior-france-russia/>.

[43] "Unheard Voice- Evaluating five years of pro-Western covert influence operations" (2022) Graphika, Stanford internet observatory, Cyber publications.

[44] Ellen Nakashima "Pentagon opens sweeping review of clandestine psychological operations". (19 September 2022) *The Washington Post,* <https://www.washingtonpost.com/national-security/2022/09/19/pentagon-psychological-operations-facebook-twitter/>.

[45] Ryan Serabian, Daniel Kapellmann Zafra, "Pro-PRC 'HaiEnergy' Information Operations Campaign Leverages Infrastructure from Public Relations Firm to Disseminate Content on Inauthentic News Sites" (4 August 2022) *Mandiant,* <https://www.mandiant.com/resources/blog/pro-prc-information-operations-campaign-haienergy>.

PSYOPS abounded during the coronavirus (hereinafter COVID) pandemic. States actors took advantage of the crisis to boast about their own-efforts to contain the virus and devaluate rivals. A 2021 RAND study for example found that state-controlled media outlets in authoritarian regimes were occasionally manipulating the information about the virus and their response for reputation laundering purpose. The narratives were also aimed at fostering anti-West sentiment among the target audiences.[46]

Finally, alongside states, non-state actors are in capacity to wage PSYOPS for their own benefit. Especially since the advent of cyberspace, there are influence campaigns conducted by independent private actors who do not show clear connection to a state. Privately-conducted influence operations are usually more modest in scale compared to state-sponsored PSYOPS. That being said, even modest private influence campaigns can have nefarious consequences. This may be the case of privately-conducted influence campaigns that relate to important subjects such as the cure of a disease.

- Illegitimate PSYOPS and the case of online disinformation

Although peacetime PSYOPS are ubiquitous in the online information space, a crushing majority of them does not appear unlawful. That is not to say, however, that all these in-appearance lawful PSYOPS are acceptable from a moral standpoint. The latter PSYOPS category is the broadest in the spectrum of tools of psychological warfare. On the left end of this spectrum lie inoffensive PSYOPS that are conducted in good faith to pursue lawful ends. This is for example the case of PSYOPS deployed for no other purposes than to counter inimical narratives, or "re-establishing the truth" *via* the spread of true and reliable information. At the other end of the spectrum lie blatantly illegal PSYOPS, such as war propaganda, terrorist propaganda, and any PSYOPS involving the spread of hateful rhetoric. Between these two extremes, there is a wide variety of communication campaigns that, in spite of being unacceptable from a moral standpoint, are not always internationally illegal. Disinformation campaigns often lie in this category of morally unacceptable conduct for which no dedicated legal framework seems to exist.

[46] Christian Johnson & William Marcellino, "Bad Actors in News Reporting Tracking News Manipulation by State Actors" (2021) Santa Monica, CA: RAND Corporation.

Disinformation refers to "false information that is created and spread, deliberately or otherwise, to harm people, institutions and interests".[47] The same way as other PSYOPS, disinformation activities target the minds of an audience to induce changes in the physical world.[48] When they are waged by actors that possess strong information power, disinformation campaigns may cause particularly nefarious consequences.

Peacetime disinformation campaigns constitute a serious threat for liberal democracies, as they rely heavily on mechanisms involving the participation of the society to function properly. Perhaps a less talked-about concern is the seriousness of the impact that disinformation could also have in underdeveloped societies, where information literacy levels may be weaker than average, and where the governing structures may also lack the proper means to mitigate the effects of a surge of "problematic information."

As suggested in introduction to this study, disinformation always existed. It took the forms of covert information manipulation tactics or "active measures" during the cold war, of fascist propaganda during World War II, or of rumours disseminated to harm the public image of kings before the industrial age. Yet, cyber-enabled disinformation activities differ substantially from their non-cyber versions.[49] There are several interrelated factors that justify a reassessment of the legality of PSYOPS relying on disinformation methods. They can be summed up as follows;

- Apart from rare exceptions, information distribution is decentralized in cyberspace. Official and non-official, domestic and non-domestic information sources coexist in the same online space. As a result, authoritative information sources may have a hard time making their voices heard in the hustle and bustle of busy and open online information spaces.
- In cyberspace "low-quality" information, "problematic" and harmful content accumulate alongside reliable, factual and verifiable information. Users with insufficient information literacy may not be able to distinguish between reliable and unreliable information.[50]

[47] United Nations Human Rights Council (hereinafter HRC), *Disinformation and freedom of opinion and expression, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression Irene Khan*, (2021), A/HRC/47/25.

[48] Lin & Kerr, *Supra* note 14.

[49] See André W.M. Gerrits, "Disinformation in International Relations: How Important Is It?" (2018) 29(1-4) *Security and Human Rights* 3.

[50] See notably Romain Badouard, *Le désenchantement de l'Internet : désinformation, rumeur et propagande* (Limoges: FYP Editions, 2017); Sam Wineburg, Sarah McGrew, Joel Breakstone & Teresa Ortega, "Evaluating Information: The Cornerstone of Civic Online Reasoning" (2016) Stanford History Education Group.

- Cyber PSYOPS are more cost-efficient than their conventional iterations. A single message, provided that it is a well-crafted one, can go viral with very little effort required from the part of the injector. In that sense, a widely cited 2018 study found that false news was spreading faster than the truth online.[51]
- The very way the information is organized and consumed online fosters the spread of disinformation. This is due to the fact that the Web has evolved since its inception. Ertzscheid speaks of a "World Life Web" that have replaced the "World Wide Web." In the World Life Web, social media platforms are the most important places.[52] This led to the development of an "attention economy," a term coined by the American economist Herbert Simon in 1971 that is today used to designate the practice of monetizing the attention of online users.[53] Basically, information providers have a vested interest in prioritizing certain types of content for their high engagement potential irrespective of the truth-value or the overall quality of the information conveyed. This is done to better capture users' attention and hook them to a site as long as possible to maximize their exposition to advertisements.[54] Not only does this lead to an overall degradation of the average quality level of the information available online, but newsfeed personalization algorithms used by social platforms may led to the formation of echo chambers and information "bubbles," when platform recommendation algorithms cut unwitting users from the access to a plurality of information online.
- Finally, state and non-state actors enjoy a high degree plausible deniability while acting in cyberspace. Anyone with little knowledge about information technologies can hide his true identity on internet. IP-spoofing tools even allow online users to hide their true location in the world while they connect to internet. Alongside these technical tools, actors may outsource their online activities to private actors, (or proxies). Not only does

[51] Soroush Vosoughi, Deb Roy and Sinan Aral, "The Spread of True and False News Online" (2018) 359(6380) *Science* 1146.

[52] See Olivier Ertzscheid, *Qu'est-ce que l'identité numérique ? Enjeux, outils, méthodologies* (Marseille: OpenEdition Press, 2014), at 11.

[53] Herbert A. Simon, "Designing organizations for an information-rich world" in Martin Greenberger (Ed.), *Computers, Communications, and the Public Interest* (Baltimore: Johns Hopkins Press, 1971), at 40.

[54] See notably 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>, at 13; Vosoughi, Roy and Aral, *Supra* note 51. See also "The Facebook files, a Wall Street Journal Inverstigation", (30 September 2021), *The Wall Street Journal,* <https://www.wsj.com/articles/the-facebook-files-11631713039>; Julia Carrie Wong, "How Facebook let fake engagement distort global politics: a whistleblower's account", (12 April 2021), *The Guardian,* <https://www.theguardian.com/technology/2021/apr/12/facebook-fake-engagement-whistleblower-sophie-zhang>; Craig Silverman, Ryan Mac & Pranav Dixit, *"'*I Have Blood on My Hands': A Whistleblower Says Facebook Ignored Global Political Manipulation*",* (14 September 2020), *BuzzFeedNews,* <https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo>.

the outsourcing of online activities to a cohort of proxies drastically complicate attribution, but it is an efficient way to boost the spreading potential of online PSYOPS. More information spreaders means that PSYOPS cover more ground, and that the instigator will exert more influence.

Overall, the above factors have for effect to make cyber-enabled PSYOPS cheaper to put together and less perilous for the instigator than PSYOPS conducted through traditional mediums. The absence of clear legal framework for the waging or condoning of harmful PSYOPS, added to the problems associated with the detection and attribution of covert cyber-enabled conduct, foster the use of cyber PSYOPS in peacetime in ways that threaten human rights and the legitimate interests of states.

- The consequences of online disinformation

Peacetime PSYOPS relying on disinformation tactics may have a number of consequences, most of which are felt in the long term.[55] This means that assessing disinformation campaigns based on their initial impact rarely yield convincing results. Disinformation campaigns are long-term oriented. They are used to "corrode" values rather than to provoke immediate physical consequences.

At society-scale, disinformation aggravates public dissatisfaction, erodes the trust people put in the information presented to them, erodes their trust in the institutions, in science, contributes to the build-up of inauthentic social expectations, and may undermine a nation's "traditional spiritual and moral values."[56] Disinformation may also negatively affect the functioning of societies and political institutions by fostering social division, triggering protests, indirectly causing harm and destruction, and corrupting democratic processes.[57] The Office of the OSCE Representative on Freedom of the Media notes for example that the prevalence of

[55] See Higdon, *Supra* note 12, at 9.

[56] The Ministry of Foreign Affairs of the Russian Federation fundamental documents, "Doctrine of Information Security of the Russian Federation*",* (2016). See also for an account of the dangers of disinformation in the context of the evolving relationship between online users and the information Herbert Lin "The existential threat from cyber-enabled information warfare" (2019) 75(4) *Bulletin of the Atomic Scientists* 187, at 189.

[57] See notably Carme Colomina, Héctor Sanchez Margalef & Richard Youngs, "The Impact of Disinformation on Democratic Processes and Human Rights in the World" (2021) European Parliament Report; Higdon, *Supra* note 12, at 9.

disinformation can broadly "[…] threaten security in the OSCE region, participating States' sovereignty, political independence, territorial integrity and the security of their citizens."[58]

It bears noting that disinformation campaigns are sometimes used to pursue similar objectives than PSYOPS relying on the dissemination of truthful information. However, unlike the latter, disinformation campaigns are inherently detrimental to freedom of thought and opinion, which are indispensable to the full enjoyment of almost every other human rights. Deception and manipulation, which lie at the core of disinformation, undercut free speech, the mental autonomy of individuals and corrupt decision-making processes.

- Disinformation in international law

For the above-cited reasons, a consensus has emerged in support of the idea that more efforts should be deployed to curb disinformation and to improve the resilience of democracies against disinformation.[59] Echoing these concern, international lawyers have looked at international law in order to assess the legality of disinformation campaigns.

Arguably, through placing greater responsibility on some key players in this area, international law can indeed assert itself as a powerful tool of deterrence. The problem is that there seems to be no comprehensive legal tool in international law to address disinformation. The same way as cyber espionage, psychological warfare, peacetime PSYOPS and disinformation are not *per se* prohibited under international law. Due to the lack of dedicated legal frameworks, lawyers must explore treaty laws and customary law in order to clarify the legal status of peacetime PSYOPS and disinformation campaigns.

Yet, thus far, customary international law has proven wholly inadequate to regulate information misuses by states, and recent attempts to clarify the legal regime of state-sponsored disinformation lack consistency. This can be explained, in part, because regulating PSYOPS generally implies walking a fine line between the preservation of fundamental freedoms, the

[58] Office of the OSCE Representative on Freedom of the Media, "International Law and Policy on Disinformation in the Context of Freedom of the Media", (2021), Brief Paper, <https://www.osce.org/files/f/documents/8/a/485606.pdf>.

[59] See notably OSCE "Joint Declaration on Freedom of Expression and 'Fake news', disinformation and propaganda" (3 March 2017), The representative on the freedom of the media, FOM.GAL/3/17; Michael J. Mazarr, Ryan Bauer, Abigail Casey, Sarah Heintz, and Luke J. Matthews, "The Emerging Risk of Virtual Societal Warfare: Social Manipulation in a Changing Information Environment" (2019) Santa Monica, CA: RAND Corporation, <https://www.rand.org/pubs/research_reports/RR2714.html>; See also generally David M. J. Lazer et al., "The science of fake news Addressing fake news requires a multidisciplinary effort" (2018) 359(6380) *Science* 1094. See also among other "Online Harms White Paper" (2019) UK Department for Digital, Culture, Media & Sport. And Bronner et al., *Supra* note 54; Sara Dillon, "The Propaganda Conundrum: How to Control This Scourge on Democracy" (2022) 23 *Oregon Review of International Law* 123.

maintenance of international dialogues, and the security of individuals and the right held by all states, to pursue their policies free from external pressure.[60] In addition, PSYOPS are non-violent by nature and often have a complex structure. Finally, whilst disinformation poses a major challenge to international law, the scopes and meaning of the relevant rules of customary international law are unclear to begin with.[61] While this would not pose a problem in the context of the regulation of forceful behaviours causing obvious destructive effects, this ambiguity incapacitates reasonings when it comes to assess more elusive conduct such as PSYOPS. In other words, PSYOPS almost systematically escape straight-forwards legal characterisations.

Indeed, as noted earlier, the immediate effect of disinformation occurs in the minds of the target. In contrast, the concrete effects of peacetime PSYOPS may be felt only after a long time period in a society. As a result, effect-based approaches to disinformation are often precluded, and alternative approaches are equally difficult to implement effectively without disfiguring the rules. Therefore, for now, customary international law fails to provide clear boundaries to the waging of peacetime PSYOPS.

Unsurprisingly, a consensus has yet to be reached regarding how to apply international law to PSYOPS involving disinformation tactics. [62] There is also a dearth of state practice pointing towards a specific approach to disinformation.

In the face of the lack of clear legal boundaries set by international law, states are forced to rely on national laws and on subjective approaches to the problem of state-sponsored disinformation, which is moving in the opposite direction from the idea of a global online information space and a global governance of cyberspace. In addition, there is a risk that national approaches to disinformation activities, if not meticulously executed, contravene human rights, especially the freedom of opinion and expression, as well as other correlative principles such as the principle of web neutrality and the free flow of information.[63]

---

[60] See in that sense John B. Whitton & Arthur Larson, *Propaganda Towards Disarmament in the War of Words* (New York: Oceana, 1964); Arthur Larson, "The Present Status of Propaganda in International Law" (1966) 31(3) *Law and Contemporary Problems* 439. John B. Whitton, "The Problem of Curbing International Propaganda" (1966) 31(3) *Law and Contemporary Problems* 601; Dillon, *Ibid.*

[61] See in that sense Duncan B Hollis, Tsvetelina J van Benthem and Talita Dias, "Information Operations under International Law" (2022) 55(597) *Vanderbilt Journal of Transnational Law* 1.

[62] For an overview of different approaches to disinformation legality under international and a description of the many issues associated with the task of legally assessing disinformation, see Henning Lahmann, "Information Operations and the Question of Illegitimate Interference under International Law" (2020) 53(2) *Israel Law Review* 189.

[63] For accounts of the precautions any actor should take with regard to human rights when articulating its national response to online disinformation, see notably Judith Bayer et al., "Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States" (2019), European Parliament, <https://aei.pitt.edu/97042/1/disinformation_and_propaganda.pdf>; Cass R. Sunstein, *Liars Falsehoods and Free Speech in an Age of Deception* (New York: Oxford University Press, 2020); Dillon, *Supra* note 59.

- Research problems and structure of the dissertation

Information societies relying principally on information and communication technologies are emerging everywhere around the globe. The COVID pandemic has led to an increase in people's reliance on modern days tools including social media. Consequently, interconnectivity is exacerbated.

Meanwhile, authoritarian regimes are trying to assert dominance in the online information space. Preoccupied by their weaknesses in this area, western states invest in defensive and offensive cyber PSYOPS capabilities. This creates a security dilemma where one state's PSYOPS capacity-building efforts push other states, regardless of their physical proximity, to seek better control over the global information space in turn. In such a context, the potential for abuse is high.

Although psychological warfare cannot be outlawed as a whole, there is an obvious need to clarify the operational framework for the conduct of online psychological warfare to avoid abuses and escalatory trends. Obviously, there is also a need to engage in a discussion on the capacity of international law to provide adequate boundaries to psychological warfare.

Clarifications with respect to the applicability of international law are needed in order to restore accountability, end impunity, and prevent future abuses in the PSYOPS context. To elicit the circumstances under which disinformation campaigns and other PSYOPS would breach international law could also help states creating new treaties on the regulation of psychological warfare. The task is all the more urgent that psychological warfare is very likely to intensify in the near future. Virtual reality spaces and high-bandwidth brain computer interfaces promise previously unseen levels of interconnectivity.

This work is based on the idea that some PSYOPS are so ostensibly detrimental to the interests of the addressees that they require specific attention, and in some cases even legal solutions. Thus, this work is a contribution to the discussion about the international regulation of cyber-enabled psychological warfare in general, and disinformation activities in particular. It addresses the question of the international responsibility of states for the waging or the condoning of one of the most concerning types of PSYOPS; online disinformation campaigns waged in peacetime and directed at foreign audiences for strategic purposes. In doing so, this work provides an answer to the following questions;

- What is the current state of play regarding the regulation of international communications in general and online disinformation in particular?

- In which conditions could international law, including human rights law, apply to peacetime PSYOPS targeting foreign populations, particularly those involving disinformation tactics?

This work is divided into five chapters. In order to facilitate the legal assessments, the first chapter provides a detailed look at online disinformation. A second chapter underlines the main issues associated with the regulation of PSYOPS relying on disinformation tactics. The three following chapters are aimed at clarifying the conditions of applications of international law to state-sponsored disinformation campaigns. The relevant legal frameworks are organised according to the nature of the interests they safeguard. Thus, the third chapter focusses on the state-centric rules of international law, including treaty laws and customary law. The customary principles of non-use or threat of force, non-intervention, and sovereignty are examined in turn. The fourth chapter focusses on human rights law and highlights the need to consolidate the human rights framework. The fifth and final chapter focusses on the question of the attribution of PSYOPS, and on the applicability of the no-harm principle and due diligence to the disinformation context.

Overall, this dissertation will show that current international law already provides options for an adequate regulatory framework on disinformation activities. These are the principle of non-intervention, some human rights, and due diligence. However, it is also noted that these legal tools still lack clarity to form a readily-applicable framework.

Indeed, studying how the law could provide adequate limitations to psychological warfare reveals a number of issues inherent to the relevant laws themselves. The main unclarities pertain to the scope of non-intervention, the scope of certain human rights, the question of human rights extraterritoriality, and the dearth of indications regarding how states could discharge their obligations of due diligence in the disinformation context.

Therefore, if it is true that psychological warfare ought not be regulated in its entirety, definition problems and jurisdictional constraints drastically reduces the relevancy of international law, even in situations where human rights are at stake. This work thus suggests a way forward for each of these options in order to palliate the *lacunae* of the law when it comes to regulate psychological warfare.

With that being said, it is important to keep in mind that the importance of the law in the process of ensuring the resilience of democracies in face of ever-more insidious threats should not be overblown. More often than not, transboundary PSYOPS would not constitute internationally wrongful acts. Still, this does not alleviate the fact that states should exercise

due diligence when they wage psychological warfare. Repeated failure of the state of origin, or the recipient state, to prevent and mitigate infringements on the rights of other states and human rights arising from the domestic and extraterritorial consequences of information misuses by private persons located under their jurisdiction should give raise to a duty to compensate the prejudice suffered by the audience. But states should not be the only entities to assume the blame for the prejudices caused by information misuses. The main enablers of the information transmission, the social media platforms and information providers, should also assume due diligence duties in that regard.

Chapter I.
Conceptualizing state-sponsored online disinformation campaigns

1. Introduction

Particularly since the Russian meddling into the US election in 2016, the issue of online disinformation has gained remarkable popularity. Still, due to the complexity of the phenomenon, there is a dearth of agreement in the literature regarding how to properly define disinformation.[64]

This introductory chapter provides an exploratory discussion on various aspects of disinformation, as well as some preliminary clarifications to help making sense of the legal assessments to come. Although this chapter occasionally makes reference to legal concepts, its purpose is not to explore the relevant law nor to conduct legal assessments, but to rough out the subject. To that end, each section contained in this chapter covers one important aspect of disinformation, starting with a tentative definition of disinformation (2). The following sections are dedicated, respectively, to the challenges associated with the cyber aspect of online disinformation campaigns (3), the question of covertness (4), the structure of disinformation campaigns (5), and the question of intent (6).

2. Disinformation defined

Plenty of actors have attempted to define disinformation. While some parameters tend to vary from one definition to another, the core meaning of disinformation have mostly remained the same. Hence a majority of the literature defines disinformation as the deliberate spread of

[64] See in that sense, among other, Rasmus Kleis Nielsen & MacKenzie F. Common, "Submission to UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression report on disinformation" (2021) Reuters Institute; Jason Pielemier, "Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?" (2020) 2020(4) *Utah Law Review* 917.

false or misleading information.[65] In addition, it is clear that the rationale behind disinformation is to mobilize rather than to inform.[66] That being said, it is possible to be more specific.

A 2018 report on fake news and online disinformation authored by a group of experts selected by the European Commission adds more granularity to the definition of disinformation. According to the European experts group, disinformation defines "[…] false, inaccurate or misleading information designed, presented and promoted intentionally to cause public harm or make a profit."[67] The report also adds that disinformation includes communications that fall outside the categories of clearly illegal forms of speech such as hate speech or incitation to violence or discrimination.[68] Although not specified in the report, it is clear that many forms of already illegal speeches may simultaneously constitute disinformation. Hate speech, as the dissemination of discriminatory narratives for the purpose of antagonizing individuals due to membership of a political, racial or religious group, almost systematically display disinformation elements. These includes among other, rumours, stereotypes and unsubstantiated accusations made against said group.

To constitute disinformation, the information disseminated must be verifiably wrong, groundless, or true but misleading. Disinformation is not limited to the injection of obvious lies. For disinformation to work, the information must be believable. To that end, disinformation operations are mounted so as to look like a plausible report of authentic news events.[69] For that reason, true facts voluntarily distorted, or presented in a misleading way may also amount to disinformation if spread intentionally to cause harm or to make a profit.[70] On this point, a recent study from the Reuter Institute explains that disinformation is less about constructing false information from scratch, and more about reorganising the news in a misleading way, drawing ungrounded conclusions from a set of unprecise information, and altering the facts to skew perceptions.[71] Some commentators have similarly emphasized that it is useful to "[…] think of active measures generally, and disinformation specifically, not as a quality of a specific piece

[65] See notably Caroline Jack, "Lexicon of Lies" (2017) *Data & Society Research Institute*, <https://datasociety.net/library/lexicon-of-lies/>.

[66] Bayer et al., *Supra* note 63.

[67] See Carme Colomina, Héctor Sanchez Margalef & Richard Youngs, *Supra* note 57.

[68] *Ibid.,* at 10.

[69] *Ibid. see also* Kate Starbird, "Disinformation campaigns are murky blends of truth, lies and sincere beliefs, lessons from the pandemic"*,* (23 July 2020), *The Conversation,* <https://theconversation.com/disinformation-campaigns-are-murky-blends-of-truth-lies-and-sincere-beliefs-lessons-from-the-pandemic-140677>.

[70] Starbird, *Ibid.*

[71] Felix Simon, Philip N. Howard, Rasmus Kleis Nielsen, "Types, sources, and claims of COVID-19 misinformation", (2020), the Oxford Internet Institute, <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>.

of information, but as a collection of information-actions—or a campaign." [72] The commentators add that "evaluating disinformation is less about the truth value of one or more pieces of information and more about how those pieces fit together to serve a particular purpose."[73]

It transpires from the above that disinformation is not necessarily defined by the mere falsity of each single message that composes a campaign, but by the capacity for a set of mutually reenforcing narratives, to deceive an audience. The point of disinformation is therefore to get the recipients into thinking and *in fine*, into behaving the desired way.[74] For deception to be successful, the penetration of the conveyed message in a given information space must be assured. Similar ideas and suggestions must be repeated in a coordinated and consistent manner. Anybody would certainly ignore an isolated piece of offbeat information, yet relentless swarms of problematic information released over a long time period and purposefully crafted so to have a strong hook on the audience and to elicit emotions are definitely capable of creating misgivings among the most lucid users. In order to take roots in the psyche of a population and subsequently trigger behavioural changes, each communication must be crafted in ways that it fits within a broader narrative.

Disinformation bears similarities to propaganda. Propaganda was used extensively by states to maintain their respective spheres of influence, alongside other sophisticated methods of public opinion manipulation long before the term of online disinformation became popular.[75] In the Cold War period, some states had taken the habit of referring to hostile propaganda as "false report."[76] While the use of the terms "false report" to condemn foreign communication efforts is not always justified, it is important to understand that propaganda and disinformation are deeply interrelated. In fact, in its seminal work on propaganda, French philosopher Jacques Ellul noted "Propaganda feeds, develops, and spreads the system of false claims—lies aimed at the complete transformation of minds, judgments, values, and actions (and constituting a frame of reference for systematic falsification)." [77] In a recent study on disinformation and propaganda, Jack observes that whether the covert, deliberate and systematic spread of

[72] Kate Starbird Ajmer Arif and Tom Wilson, "Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations" (2019), 3 *Proceedings of the ACM on Human-Computer Interaction* (CSCW) 1, at 3. See also Starbird, *Supra* note 69.

[73] Starbird, Arif & Wilson, *Ibid.*

[74] See Bayer et al, *Supra* note 63.

[75] Ward, Pierson & Beyer, *Supra* note 29.

[76] See Elizabeth A. Downey, "A Historical Survey of the International Regulation of Propaganda" (1984) 5(1) *Michigan Journal of International Law* 341, at 352; Eric De Brabandere "Propaganda" (2019) Max Planck Encyclopedias of International Law, Encyclopedia entries.

[77] Jacques Ellul, *Propaganda: The Formation of Men's Attitudes* (New York: Vintage, 1965), at 61.

inaccurate information "can rightly be referred to as propaganda — systematic information campaigns that are deliberately manipulative or deceptive — is often a matter of perspective."[78]

Disinformation could indeed be compared to "black" propaganda, a practice that refers to the act of covertly disseminating false and/or misleading information with the aim of opposing an adversary.[79] Black propaganda is traditionally reconciled with subversive intent. In some situations, disinformation can also take the form of "grey" propaganda, a term that is traditionally used to refer to the act of covertly disseminating information of doubtful accuracy for political or ideological purposes.[80] In contrast, white propaganda, which defines the overt dissemination of accurate information for political or ideological purposes does not constitute disinformation, even if some elements of persuasion are involved, as both the information and its provenance are accurate in that case.[81] In other words, disinformation, black propaganda, and grey propaganda serve to deceive the audience while white propaganda merely persuades.

State-sponsored disinformation campaigns can be of domestic or transboundary nature. In the latter case foreign populations constitute the target while domestic audiences constitute the target in the former. In the domestic context, state-sponsored disinformation is often associated with situations of political upheaval, when unscrupulous regimes attempt to secure their power, or when politicians or party members strive to gather support.[82] For example, a case of domestic state-sponsored disinformation occurred in the context of the 2020 Ugandan presidential elections. In reaction to an online investigation conducted by the Atlantic Council's Digital Forensic Research Lab in January 2021, the social media platform Facebook reportedly removed "[…] 32 pages, 220 user accounts, 59 groups, and 139 Instagram profiles working to

[78] Jack, *Supra* note 65.

[79] See generally Garth S. Jowett & Victoria J. O'Donnell, *Propaganda and Persuasion*, (Thousand Oaks: SAGE publication, 2006 (4th edition)), at 21; Bayer et al., *Supra* note 63, at 27; Jack, *Supra* note 65, at 7.

[80] Jack, *Supra* note 65.

[81] It is worth noting that some commentators have argued that the dissemination of factual information could still constitute an international wrongful act, notably if the information spread has been unlawfully acquired by the malicious actor beforehand. See Ido Kilovaty, "Doxfare: Politically Motivated Leaks and the Future of the Norm on Non-Intervention in the Era of Weaponized Information" (2017) 9 *Harvard National Security Journal* 148. Because the act of openly releasing correct information does not constitute disinformation, it is excluded from the reach of this paper. It is enough to recall that the act through which the malicious actor covertly acquired the information may be considered a breach of the sovereignty of the target state, regardless of subsequent leaks. The unlawful acquisition of genuine but secret information may constitute either, an isolated unlawful act, or, if the information is subsequently leaked, the preparatory act of a complex internationally wrongful conduct. See Commentary to the Articles on State Responsibility, International Law Commission (ILC) (hereinafter ARSIWA) (2001) 2 *Yearbook of the International Law Commission* (Part II) 31, commentary to Article 14§13. See also Francois Delerue, *Cyber Operations and International Law*, (Cambridge: Cambridge University Press, 2020), at 249. Yet, this last hypothesis requires that the politically-motivated leaks of non-deceptive information is itself constitutive of an internationally wrongful act. According to a commentator, this may be the case when the leaks are intended to disrupt the internal affairs of the concerned state, or to cause harm. See Kilovaty, *Ibid.*

[82] For a detailed account of the methods of information manipulation employed by partisan media in the US, see notably Dillon, *Supra* note 59, at 160

promote Ugandan president and National Resistance Movement (NRM) leader Yoweri Museveni."[83] It is explained in the report that several networks of information spreaders had recourse to astroturfing tactics to amplify their messages across the Ugandan online information space. The platform Twitter followed suit with the removal of inauthentic accounts similar to those removed by Facebook.

The other disinformation type, state-led transboundary disinformation campaigns, have drawn much attention due to their incompatibility with traditional concepts of international relations theory. In the international context, state-sponsored disinformation is one way to conduct global strategic engagements with foreign populations, which dovetails with public diplomacy and soft power.[84] Contrary to formal diplomatic efforts deployed so as to engage with other states officials with the goal of strengthening inter-states relationships, public diplomacy seeks to obtain support and validation directly from foreign populations. Soft power, it is commonly understood as "the ability to affect others to obtain the outcomes one wants through attraction rather than coercion or payment."[85] Thus, public diplomacy is one way to project soft power.[86]

Through transboundary online disinformation campaigns, state actors take advantage of the cyber medium in order to engage directly with foreign populations, which is reminiscent of public diplomacy. However, disinformation deviates in many aspects from public diplomacy and soft power. An actor who engages in public diplomacy usually seeks to seduce the audience. In the same vein, soft power hinges on the capacity of an actor to influence behaviours through the attraction of its policies, cultural magnetism, or values. In contrast, disinformation is mainly used as a repulsion tool by states. Even in the eventuality that disinformation is used to obtain public validation for certain policies or postures on the international scene for instance, the practice remains very different from soft power and public diplomacy. Indeed, disinformation entails that words alone compensate for the absence of past and current actions and accomplishments from the part of the speaker. Consequently, while soft power is all about

[83] "Facebook removes inauthentic assets linked to Ugandan government" (10 February 2021) *DFRLab*, <https://medium.com/dfrlab/facebook-removes-inauthentic-assets-linked-to-ugandan-government-c3949be9f810>.

[84] *Encyclopædia Britannica,* "Public Diplomacy"*,* <https://www.britannica.com/topic/public-diplomacy>.

[85] Joseph S. Nye, "Public Diplomacy and Soft Power" (2008) 616 *The Annals of the American Academy of Political and Social Science* 94. See also from the same author, "Soft Power and the Public Diplomacy Revisited" (2019) 14 *The Hague Journal of Diplomacy* 1; "Soft power: the evolution of a concept" (2021) 14(1) *Journal of Political Power*, 196. See also Christopher Walker & Jessica Ludwig, "The meaning of sharp power: how authoritarian states project influence" (16 November 2017) *Foreign Affairs*, <https://www.foreignaffairs.com/articles/china/2017-11-16/meaning-sharp-power>.

[86] See Nye (2008), *Ibid.*

leveraging the convincing power of reality, disinformation exploits beliefs and cognitive biases through deception in order to foster adhesion to inaccurate representations of reality.

More often than not, disinformation will be used as a repulsion tool. That is to say, that disinformation is almost systematically used "to undermine, delegitimize, and denigrate a public image rather than working to improve it."[87] This is so, whether disinformation is used to interfere in the politics of another state, or to secure support domestically or internationally.[88] Regrettably, the information spreads faster and farther online when the narratives are calibrated to fuel outrage, hate or resentment.[89] For instance there are good reasons to believe that Russia's successful influence in Africa is partly due to the exploitation of deep-rooted anti-West ressentiment among the local populations.

But if disinformation is not soft power, it is not hard power either. Hard power is traditionally understood as the ability to coerce someone to do or abstain from doing something. However, disinformation is not coercion *stricto-sensu*. In that sense an author has explained 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."[90] Others however argue that disinformation efforts can still be coercive if they involve the communication of "extremely framed information."[91] From that perspective, the ability to wage disinformation campaigns abroad can in theory be a form of hard power.

The two views are correct. However, state-led information manipulation campaigns directed at foreign audiences always follow long term objectives that go well beyond forcing specific behaviours at individual scale. In the eventuality that disinformation has some coercive effects at the individual scale, speaking of hard power deployed against a whole state is already a stretch from the initial meaning of coercion and hard power.[92] With that being said, the relationship between coercion and disinformation deserves more attention. The section dedicated to disinformation and the principle of non-intervention will later provide more in details about coercion in international law.

---

[87] Starbird, Arif & Wilson, *Supra* note 72, at 3.

[88] See also Bayer et al. *Supra* note 63. See also Ibid.

[89] See for example Luke Munn, "Angry by design: toxic communication and technical architectures" (2020) 7 *Humanities and Social Sciences Communications* 1.

[90] Henning Lahmann *Unilateral Remedies to Cyber Operations Self-Defence, Countermeasures, Necessity, and the Question of Attribution* (Cambridge: Cambridge University Press, 2020), at 38.

[91] Nye (2021) *Supra* note 85. At 8. See also in that sense Steven Wheatley, "Foreign Interference in Elections under the Non-intervention Principle: We Need to Talk about "Coercion" (2021) 31 *Duke Journal of Comparative & International Law* 161.

[92] See in that sense, Lahmann, *Supra* note 90. At 38.

Alternative concepts are occasionally used to characterize disinformation activities. Notably, some commentators have introduced the term of "sharp power" to describe disinformation and other unconventional means of destabilization deployed by a state to project influence.[93] Drawing similarities with the practice of *dezinformatsiya* pioneered by the Soviets during the Cold War, Rid speaks of "active measures" designed to "drive wedges into the pre-existing cracks" of societies.[94] In the same vein, Wigell speaks of "hybrid interferences" targeting democracies.[95] Before cyberspace became ubiquitous, Libicki spoke of "operations against the national will", and "PSYOPS" when referring to the use of information by a state "against the human mind."[96]

3.The cyber factor

Disinformation campaigns may be waged outside or within cyberspace. As opposed to conventional disinformation that are waged exclusively through traditional medium, cyber disinformation defines PSYOPS that take advantage of the characteristics of the online information space. As argued in the introduction, global interconnectivity, the proliferation of information online, and the volume of active users of social media make disinformation operations waged in cyberspace way more invasive, flexible, and more harmful than their pre-internet iterations.[97]

While the present study focuses on online disinformation, references will occasionally be made to PSYOPS conducted on radio, television, or paper, primarily because most of the relevant regulation date from before the advent of cyberspace. Still, for the sake of completeness, it is worth noting that disinformation spread through traditional conduits is sometimes brought online concomitantly. In a similar fashion, a story first spread online can later be broadcasted on traditional conduits. This was for example the case in July 2014, during the Russian annexation of Crimea, when the Russian TV channel "Channel one" falsely reported that soldiers of the Ukrainian army had committed atrocities on civilians, citing the crucifixion of a young boy near the Russian border and his mother being tied to the back of a

[93] See Walker & Ludwig, *Supra* note 85; Nye (2021) *Supra* note 85.

[94] Thomas Rid, "Disinformation A primer in Russian active measures and influence campaigns" (2017), Hearings before the select committee on intelligence, United States Senate, 115th congress.

[95] Mikael Wigell, "Hybrid interference as a wedge strategy: a theory of external interference in liberal democracy" (2019) 95(2) *International Affairs* 255, at 262.

[96] See Libicki, *Supra* note 19. At 35.

[97] Jack, *Supra* note 65. Soroush Vosoughi, Deb Roy, and Sinan Aral, *Supra* note 51. See also Gerrits, *Supra* note 49. See also Balkin, *Supra* note 31.

driving tank. An intriguing post, dated from shortly before the controversial report was aired would later be found on Facebook. Its author, a pro-kremlin politician, depicting a nearly identical story.[98]

Another point to keep in mind is that all the reflexions made about online disinformation in the course of this study also concern conventional disinformation. The focus on online disinformation is essentially due to the fact that cyber-enabled PSYOPS appear to have more pernicious effects that conventional PSYOPS. The question of whether this means that cyber-enabled PSYOPS are more likely to fall under the scopes of the rules of international law than conventional PSYOPS will be answered in this work.

4. Covert or overt

The same way as other PSYOPS, disinformation operations may be executed covertly or overtly. An overt activity is an activity done in the open. The identity of the author of a covert act is known to the public and his behaviour plainly visible. In contrast, a covert activity defines an act that is hidden from the public, or an act conducted anonymously. By fear of denunciation or retaliation, the authors of an internationally wrongful act often hide their true identities. Because doubts subsist as to their legality under international law, and in order to maximize the efficiency of disinformation campaigns, state-sponsored disinformation activities often take the form of covert acts, executed by non-states actors or "proxies" for the account of the contractor. The proxies are then called "executors", "information spreaders" or disinformants. The states that outsource a campaign of disinformation may be called "instigators" or "contractors".

From state-owned media, to public relations firms, to paid private individuals, non-state actors involved in disinformation campaigns can be of various types, and it is not uncommon to see different types of non-state actors involved in a same sophisticated disinformation campaign. For instance, in February 2022, the Stanford Internet Observatory found that similar false narratives surrounding the Russian military build-up at the border of Ukraine were simultaneously spread offline, through state-controlled media and online, *via* unattributed

[98] Andrew E. Kramer "To Battle Fake News, Ukrainian Show Features Nothing but Lies*"*, (26 February 2017), *TheNewYorkTimes*, <https://www.nytimes.com/2017/02/26/world/europe/ukraine-kiev-fake-news.html>; Jane Wakefield, "TED 2018: Ukrainian journalist fights fake news", (11 April 2018), *BBC News*, <https://www.bbc.com/news/technology-43568238> video of the conference in question available online on Youtube <https://www.youtube.com/watch?v=yue6t5UmP4o>.

Telegram channels.[99] As mentioned in the introduction, the outsourcing of disinformation activities may benefit the contractor in three ways; First, his identity is kept secret, which allows plausible deniability. Second, outsourcing disinformation multiplies the number of relays and injection points of disinformation and thus allows the operation to spread faster and further. Third, relying on private actors to spread a narrative helps making the information more believable because it creates an impression of popularity.

The proxies themselves may also have recourse to identity-concealment methods. Social media platforms' content moderation policies provide a useful account of the different identity concealment methods that may be used in the disinformation context by the information spreader. Twitter's Help Center for example lists "stock or stolen profile photos, particularly those depicting other people"; the "use of stolen or copied profile bios"; and the "intentionally misleading profile information, including profile location"[100] as deceptive information-sharing practices. Together, the outsourcing of disinformation activities to proxies and the deployment of identity-concealment methods at individual scale render technical and legal attribution difficult.

In the context of cyber-enabled disinformation, false accounts can be used for generating artificial engagements, thereby giving the reader the impression of broad social mobilization around a specific topic. Content spreaders may also impersonate entities instead of persons. An interesting case study in that sense is the disinformation campaign attributed to India and uncovered by EU DisinfoLab in December 2020.[101] The campaign named "Indian Chronicles" by the investigatory website is said to have been pursued to realise long-term objectives domestically and abroad, notably, "[I]n India, to reinforce pro-Indian and anti-Pakistan (and anti-Chinese) feelings" and "internationally, to consolidate the power and improve the perception of India, to damage the reputation of other countries and ultimately benefit from more support from international institutions such as the EU and the UN."[102] The campaigners reportedly "resurrected" dead peoples, dead NGOs, dead media, while also stealing the

[99] Shelby Grossman, Catarina Buchatskiy, Benjamin B.-B., Kate D., Renee DiResta, Christina H. and Julia Steinberg, "Full-Spectrum Pro-Kremlin Online Propaganda about Ukraine" (22 February 2022) Freeman Spogli Institute for International Studies.

[100] "Misleading and deceptive identities policy", Twitter Help Center, <https://help.twitter.com/en/rules-and-policies/twitter-impersonation-and-deceptive-identities-policy>.

[101] Gary Machado Alexandre Alaphilippe, Roman Adamczyk & Antoine Grégoire, "Indian Chronicles: deep dive into a 15-year operation targeting the EU and UN to serve Indian interests" (9 December 2020) *EUDisinfoLab*, <https://www.disinfo.eu/publications/indian-chronicles-deep-dive-into-a-15-year-operation-targeting-the-eu-and-un-to-serve-indian-interests>.

[102] *Ibid.*

identities of authentic figures and organisations in order to boost the efficiency of a fifteen years-long influence campaign waged both online and in the real space.[103]

The mobilization by state actors, of astroturfers (or "trolls", "cybertroops" or "online commentators") to spread narratives at scale is often reconciled with the concept of *astroturfing*, or its online equivalent; "cyberturfing",[104] which refers to the act of coordinating mobs to inundate a predetermined online information space with messages promoting the same cause, thereby simulating spontaneous grassroots movements, in exchange for payment. Both real persons and fake accounts or bots can be involved in astroturfing.[105]

Astroturfing is a well-documented practice.[106] In the context of state-sponsored cyber PSYOPS, astroturfers are used to mislead a population into thinking that the messages disseminated are authentic opinions of nationals while they are, in fact, published by paid posters or even machines. Of course, private actors may also have recourse to astroturfing tactics. Scandals involving private individuals buying followers or using "sock puppet" accounts to build a reputation or to boost their credibility abound. In the same vein, commercial actors may have recourse to astroturfers to manipulate consumers.[107]

The same way as other influence methods, astroturfing is way more effective when supported by state-level funds and capabilities.[108] When organised by a state, astroturfing can be used both for offensive and defensive purposes.[109] A regime may target its own citizens to stifle unfavourable opinions domestically or steer public attention away from unwanted topics or scandals.[110] Astroturfing tactics may also be used against foreign populations, sometimes for the same purposes, but more often to interfere in the politics of the target state.[111] Both Russia[112]

---

[103] *Ibid.*

[104] Mark Leiser "AstroTurfing, 'CyberTurfing" and other online persuasion campaigns" (2016) 7(1) *European Journal of Law and Technology* 1.

[105] *Ibid.* at 4.

[106] See for instance Paul Charon & Jean-Baptiste Jeangène Vilmer « Les Opérations d'Influence Chinoise Un moment Machiavélien » (2021) études de l'IRSEM. At 356. See also Cheng Chen, Kui Wu, Venkatesh Srinivasan, and Xudong Zhang, "Battling the internet water army: Detection of hidden paid posters" (2013) IEEE/ACM *International Conference on Advances in Social Networks Analysis and Mining* 116; Machado et al. *Supra* note 101. See Nick Fielding & Ian Cobain "Revealed: US spy operation that manipulates social media" (17 March 2011) *TheGuardian,* <https://www.theguardian.com/technology/2011/mar/17/us-spy-operation-social-networks>.

[107] See generally Leiser, *Supra* note 104.

[108] Leiser*, Supra* note 104 at 3

[109] Charon & Vilmer, *Supra* note 106.

[110] *Ibid.*

[111] See ODNI, *Supra* note 41.

[112] Clint Watts, "Disinformation: A Primer in Russian Active Measures and Influence Campaigns", Statement Prepared for the U.S. Senate Select Committee on Intelligence hearing, (30 march 2017).

and China[113] have been accused of using astroturfing methods in the context of their PSYOPS, but they are not alone, far from it.[114]

It bears noting that the patriotism displayed by astroturfers may sometimes be genuine, providing, of course, that they are real persons.[115] Nevertheless, it is not too much of a stretch to assume that covertly financing communication campaigns by default annihilates the authenticity of the messages they contain.[116]

The act of buying information spreaders with the view to disseminate information covertly and at scale amounts to disinforming the public, regardless of the truth value of the content. The misleadingness of the astroturfing operation is constituted when the financial incentive behind the message, or the sponsorship, are hidden to the audience. The deceptive and manipulative character of the campaign is then verified, not necessarily because the ideas proffered are false or misleading, but because the audience is misled into thinking that the opinions disseminated are authentic and widely supported in the society.[117] Transparency, in the context of one-to-many communication campaigns, be they of political nature or confined to marketing communication is key to the proper enjoyment of the right to information and the freedom of expression and opinion.[118]

Of course, the level of deception involved in covert communication campaigns containing accurate information will always be inferior to the level of deception encountered in the context of black propaganda. Covertness alone seldom leads to measures or sanctions for the instigator. From an international law perspective, the contractor state that has made use of astroturfing tactics or front organisation for the purpose of disseminating information in a covert manner, does not incur responsibility under international law unless the whole effort is found to be harmful to the interests of the target. In other words, signs of malice are not sufficient to give raise to a violation of international law.

Admittedly, lack of transparency with regard to the true identity of the instigator is generally a strong indicator of malicious intent. The widespread use of bots, false profiles and other online identity concealment tactics as well as the mobilisation of virtual mobs for disseminating specific narratives abroad are staples of an ill-intentioned communication campaign. The possibility however remains that covert and manipulative practices do not cause harm to the

[113] See notably Charon & Vilmer, *Supra* note 106. At 362.
[114] See notably Bradshaw & Howard, *Supra* note 39.
[115] Charon & Vilmer, *Supra* note 106. At 365.
[116] *Ibid.*
[117] See Leiser, *Supra* note 104. See also Zhao Wenqing, "Cyber Disinformation Operations (CDOS) And A New Paradigm of Non-intervention" (2020) 27(1) U.C. *Davis Journal of International Law & Policy* 35.
[118] International Covenant on Civil and Political Rights, 999 UNTS 171 (hereinafter ICCPR). Articles 18, 19.

legitimate interests of the target, or at least not in a way that justifies legal treatment.[119] Take the example of a government paying non-state actors exclusively to promote the benefit of its policies internationally without explicitly attacking the reputation of other states. Disinformation campaigns of this kind were notably used by China throughout the last decade.[120]

5. The structure of state-sponsored PSYOPS

Although PSYOPS may be initiated by state and non-state actors, they are not operating on the same scale. Privately-conducted disinformation is usually reconciled with rather limited and well-defined goals.[121] In contrast, state-sponsored PSYOPS are complex and long-term oriented. Disinformation, when conducted by private actors, is usually financially motivated whereas they constitute foreign policy tools when conducted by states. State-sponsored disinformation may be motivated by security imperatives including the safeguarding of state sovereignty, the survival of a government or an ideology, or the enabling of military success State-sponsored PSYOPS also serve to "influence the 'correlation of forces'" into the advantage of one state or more.[122] Because what state actors seek to achieve through disinformation activities is connected to concrete strategic considerations, state-sponsored disinformation campaigns are usually wider in scale, more sophisticated and more impactful than privately-conducted disinformation campaigns.[123]

With regard to their structure, state sponsored PSYOPS are not articulated around isolated communications upon which depends the success of an entire influence effort. Rather, state-sponsored disinformation takes the form of protracted campaigns organised around "[…]

[119] There is an exception to the above, some commentators have recently hinted on the possibility that some human harm generated by disinformation could be interpreted in terms of the harm inflicted to the rights to freedom of thought and to hold opinions without interference. Accordingly, it is not only the intent to cause physical harm, destruction or to subjugate the will of a target state that renders online disinformation illegitimate and potentially illegal, but also the high manipulative value of a communication campaign. Therefore, being subject to highly manipulative methods, such as astroturfing, could in itself justify a legal claim for a violation of human rights law regardless of other harm inflicted. In other words, the harm to the mental autonomy of the target users constitutes a harm susceptible of legal sanction in itself. Although the possibility has for now only been explored for holding social media companies accountable, the argument is worth raising in the context of state responsibility. See Kate Jones "Online Disinformation and Political Discourse: Applying a Human Rights Framework" (2019) *Chatham House Research Paper*, at 52. The question of the responsibility of states for the violation of the rights to freedom of thought and to hold opinions without interference will be studied in more details later in this work. See Chapter V "State-sponsored online disinformation campaigns in human rights law" below.

[120] See in general Charon & Vilmer, *Supra* note 106.

[121] See Bayer J. et al. *Supra* note 63.

[122] See Gerrits, *Supra* note 49.

[123] See Starbird, Arif, and Wilson, *Supra* note 87.

planned group of especially political, business, or military activities that are intended to achieve a particular aim."[124] Disinformation campaigns are constituted by an aggregate of operations of communications which share enough characteristics to be considered as parts of the same effort.

Assessing state-sponsored disinformation activities thus entails assessing a significant number of communications displaying suspicious levels of coordination and that appear to have served the same premeditated design. [125]

From a legal perspective, state-sponsored disinformation campaigns may constitute composite breaches of international law. That is to say, that once multiple disinformation operations or any multi-faceted conduct involving disinformation activities clearly appear to follow a pattern of misconduct leading up to a potential violation of international law, the violation can be said to have arisen from a composite act made of a variety of communications aimed at the realization of the same malicious, premeditated design.[126]

The International Law Commission' Articles on State Responsibility (hereinafter, ARSIWA) provides for the possibility of considering an aggregation of actions as forming a unique unlawful act if all acts are connected by the same premeditated design. Article 15 provides; "[T]he breach of an international obligation by a State through a series of actions or omissions defined in aggregate as wrongful occurs when the action or omission occurs which, taken with the other actions or omissions, is sufficient to constitute the wrongful act." And "[I]n such a case, the breach extends over the entire period starting with the first of the actions or omissions of the series and lasts for as long as these actions or omissions are repeated and remain not in conformity with the international obligation."[127]

State-sponsored disinformation campaigns will always constitute wrongful aggregate. Once a violation of international is established, the prejudice is considered to extend from the moment the campaign is initiated, that is, when the first operations of communications are conducted "to lay the groundwork" for the rest of the effort. Also, each disinformation operation deployed in the context of an overarching campaign does not need to be itself wrongful. Taken independently, each release may appear benign, what counts is that each operation of

[124] "Campaign" Online Cambridge dictionary, <https://dictionary.cambridge.org/dictionary/english-japanese/campaign>.

[125] See Giovanni Distefano, "Fait continu, fait composé et fait complexe dans le droit de la responsabilité" (2006) 52 *Annuaire français de droit international* 1, at 15. (Speaking of *même dessein prémédité.)*

[126] Manuel Rodriguez, "Disinformation Operations Aimed at (Democratic) Elections in the Context of Public International Law: The Conduct of the Internet Research Agency During the 2016 US Presidential Election" (2019) 47(3) *International Journal of Legal Information* 149, At 169.

[127] See ARSIWA, *Supra* note 81. Article 15.

communication is "part of a series" of acts committed with the same malicious intent.[128] In other words, the acts in question must be connected by the same premeditated design.[129]

The amount of disinformation operations aggregated into a campaign informs both the duration and the invasiveness of the campaign. Isolated and short-lived operations to which it is not possible to ascribe a clear unlawful purpose may therefore not fall under international law's scope.

It is entirely possible that complex breaches of international law are made of a combination of disinformation campaigns with other types conduct, thereby forming "heterogeneous" aggregates. Indeed, foreign disinformation is sometimes associated with the broader context of destabilisation strategies or "hybrid interferences", where a variety of non-violent conduct is used against a competitor to realize a range of short-term and long-term objectives.[130] Destabilisation strategies usually include cyber-attack, economic coercion, corruption, and disinformation.[131]

Yet, caution should be exercised when assessing "homogeneous" aggregates of disinformation operations (disinformation campaigns) and "heterogenous" aggregates (destabilisation strategies or hybrid interferences) involving disinformation activities. More often than not, the objectives of disinformation campaigns and destabilisation strategies are long-term oriented. Destabilisation strategies involving disinformation activities and disinformation campaigns are often associated with vague objectives such as "challenging the core values of democratic societies"[132], undermining public trust in democratic institutions,[133] the "corrosion of democracy"[134], or "dividing a target country or coalition, thereby weakening its counterbalancing potential."[135] None of these goals constitute an actual violation of international law. In a similar vein, hybrid interference or destabilisation strategies are not *per se* contemplated in international law. These are relatively recent concepts in security studies

[128] ARSIWA, *Supra* note 81, at 63

[129] Distefano, *Supra* note 125. See also *Phosphates in Morocco,* (1938) Permanent Court of International Justice, (PCIJ), Judgement,

[130] Wigell for example explains that complex hybrid strategies may include economic coercion, corruption of elites, disinformation campaigns. Wigell, *Supra* note 95. See also Patrick Cullen & al. "The Landscape of Hybrid Threats: A Conceptual Model (Public Version)", Giannopoulos, G., Smith, H. and Theocharidou, M. (eds.), Publications Office of the European Union, Luxembourg, 2021.

[131] Cullen, *Ibid.*

[132] *Ibid*.

[133] See ODNI report, *Supra* note 41.

[134] Nina Jankowicz, "How Disinformation Corrodes Democracy" (30 November 2021) *Foreign Affairs*, <https://www.foreignaffairs.com/articles/world/2021-11-30/how-disinformation-corrodes-democracy>.

[135] Wigell, *Supra* note 95; Cullen, *Supra* note 130.

that, for reasons inherent to their very abstract nature, as well as a significant risk of instrumentalization, simply cannot be regulated.

With that being said, it is shown in this work that it might still be possible sometimes, to make the point of a violation of international law arising from disinformation campaigns or hybrid interferences, whatever overarching goal can be attributed to them.[136]

6. The question of intent

The question of intent is essential to the assessment of disinformation. First of all, intent allows to differentiate between disinformation and misinformation, the latter being sometimes mistaken for the first. Thus, disinformation defines the *intentional* dissemination of false or distorted facts while misinformation implies an *unintentional* act.[137] "Misinformation" refers to the spread of wrong information by individuals acting in good faith. Misinformation occurs when individuals share information that they misinterpreted, or when they failed to independently verify a source's claims before sharing. Misinformants are not aware of the fallacious character of the content they share online, they trusted the information they received and spread it to "inform" their followers, friends or families. Therefore, misinformants do not intend to deceive the public, let alone to cause harm. Widespread misinformation is the indirect and expected consequence of a disinformation operation. In a sense, misinformants are the first victims of the disinformant and the unwitting vehicles contributing to the dissemination and the amplification of PSYOPS.

It bears noting that misinformation is not always connected to an episode of disinformation. Misinformation simply occurs when individuals share information that they misinterpreted or misunderstood, irrespective of whether said information initially comes from a disinformation campaign.

The distinction between disinformation and misinformation is fundamental because the question of whether an information spreader can be held legally accountable for his communications depends primarily upon the presence of an element of intention, a deceptive intent. The user who misinforms in good faith is not liable for the consequences of his actions. Only the act of disinforming at scale, or the act of knowingly turning a blind eye to harmful

[136] See Delerue, *Supra* note 81, at 254. (Arguing that the ultimate objective of disinformation activities is often too obscure to be discerned, but that it may still be possible to ascribe more short-term and more definite goals to disinformation campaigns).

[137] See notably Jack, *Supra* note 65.

disinformation conducted by private nationals can possibly account for a violation of international law for the concerned state. International responsibility is likely to arise for the state that disinformed with the intention to cause harm, or when gross negligence in preventing harm or mitigating risk arising from the spread of information is involved.

The treatment of the question of intent in the disinformation context is of outmost importance, not only because it is the deceptive intent that distinguishes disinformation from legitimate opinion-sharing, but also because in disinformation, the nature of the intent matters more than the truth-value of the information spread. In addition, the element of intention is central to the identification of a breach of international law, and the determination of the structure of a breach.[138]

An important preliminary step before conducting a legal assessment on disinformation thus consists of discerning an element of intention from an information spread. (*The intention to manipulate and deceive)*. This first step allows to distinguish between innocent misinformation and disinformation.

- The intention to manipulate and deceive

There are several indicators to take into account to clarify the element of intention. They pertain to the content of the communication, the identity of the information conveyer, and the way the information is presented.[139] For example, non-verifiable assertions presented as facts[140], innuendoes, sweeping assumptions, references to values, moral, out of context information, inaccurate information, or information presented without reference are strong signs of partially. They may also reveal the information conveyer's bad faith. Similarly, one should expect from a legitimate information conveyer that the source of the information is clearly stated, and that neither the title or the cover image is misleading. In addition, the use of bots, fake and stolen accounts, and other location and identity concealment tactics by a communicator is usually a strong indicator of a deceptive intent.[141] Of course, all these criteria are not decisive in themselves when it comes to establish the malicious intent to deceive. It is

[138] See ARSIWA, *Supra* note 81, Article 15. Commentaries; Distefano, *Supra* note 125.

[139] Jacques Ellul, *Supra* note 77, at 56.

[140] False information presented as factual statement is usually appreciated through the lenses of defamation law. See for instance Pedersen and Baadsgaard v. Denmark, (2003) ECHR judgement.

[141] See Twitter Help Center Platform manipulation and spam policy, (as of 2020), <https://help.twitter.com/en/rules-and-policies/platform-manipulation>.

nonetheless safe to assume that transparency-related indicators are crucial in the assessment of a deceptive intent.[142]

Beyond the suspicious inaccuracy of a message, or the lack of transparency surrounding the sharing of information, the quantity of messages and the question of their origin are of paramount importance to the determination of malicious intent because they inform the level of coordination of a campaign. Consequently, the malicious intent is best established when the same entity is found to be involved in the repetitive injection of false information, fallacious statements, or misleading content. In other words, the greater the volume of problematic content attributed to the same actor, the more apparent the deceptive intent.

It is worth noting that, due to the speed with which problematic information travel online,[143] the withdrawal of fallacious online messages by the initial spreader following foreign denunciation should not suffice in itself, to disprove deceptive intent. It is safe, and furthermore necessary, to assume that every authority or media who engages in communication activities online is well aware of the spreading potential of a controversial piece of information. Furthermore, debunking false or inaccurate information takes an incredible amount of time and effort. The effects of disinformation will be felt in a society long before problematic information is debunked.[144] Therefore, there should be a presumption of malice when state-run media outlets repetitively amplify false or ungrounded statements through the publication of online reports featuring problematic information. In such case, the right reaction would be for the concerned entity or eventually the owner state, to make an official statement rectifying the information previously shared, or to simply ensure that such mistakes remain isolated. Indeed, the failure for a media outlet, to verify the authenticity of the document, testimony, or other materials relied upon to write problematic reports, as well as the failure to rectify verifiably false pieces of information may constitute a violation of journalism ethics and standards.[145] If repeated, lack of diligence in the process of reporting shall constitute admissible evidence of an intention to manipulate and deceive the audience. This also applies to individuals and entities that, while not officially entitled to inform the public, nonetheless portray themselves as reliable sources or trustworthy persons.[146] At the same time, the mere choice to report on only one

[142] Transparency-related indicators will also be particularly crucial in the legal characterisation of the communication efforts. See for example Jens David Ohlin, "Election Interference: The Real Harm and The Only Solution" (2018) *18(50) Cornell Legal Studies Research Paper* 1.

[143] Vosoughi, Roy & Aral, *Supra* note 51.

[144] Ellul, *Supra* note 77.

[145] See for instance international federation of journalists (IFJ), "Global Charter of Ethics for Journalists" (2019); IF, "Declaration of Principles on the Conduct of Journalists" (Bordeaux Declaration), (1954); "Declaration of rights and duties of journalists" (Munich Charter), (1971).

[146] Dillon, *Supra* note 59, at 157.

aspect of an issue, albeit misleading in itself[147], does not suffice in establishing the deceptive intent of the concerned media. Caution should then be taken so as not to confuse biased reporting with disinformation.[148] The opposite may stretch the definition of disinformation way beyond reason.

With that being said, the absence of an element of intention does not exclude the possibility that a state is liable for a failure to prevent harm or mitigate a risk arising from disinformation campaigns initiated by non-state actors placed under its jurisdiction.[149] In addition, proving deceptive intent is not sufficient to legally assess disinformation. Indeed, the act of lying is not the finality of disinformation. Disinformation campaigns are waged by humans, and humans are rational entities. Thus, one does not deceive for the satisfaction of deceiving, but in order to obtain a particular result, usually to harm or to make a profit.[150] This applies to individuals, and to an even greater extent to states. States' actions are systematically the result of calculus oriented towards the realization of concrete advantages.

Thus, the term "disinformation," as understood in this study, does not *merely* qualify the spread of false or misleading information, but the dissemination thereof in order to attain a particular goal. The expected goal can be of various natures and depends upon the context and the nature of the instigator. For instance, individuals initiating a disinformation campaign in their personal capacity will tend to seek political or economic gains, whereas a regime that undertakes to deceive its own population may seek to secure its grip over a territory. In contrast, state actors that purposefully disinform foreign populations at scale may do so for gaining strategic advantages such as superiority in battle or destabilising neighbouring states in order to better project its influence upon them. Thus, for any of the following legal assessments to make sense, it has to be borne in mind that disinformation is above all a method to achieve strategic aims. In other words, disinformation is a means rather than an end. State-sponsored disinformation is either a part of a calculated strategy, or a calculated strategy on its own. In both cases, the disinforming state seeks to trigger specific events or cause specific damages domestically, or internationally.

---

[147] Sandrine Baume, Véronique Boillet, and Vincent Martene, *Misinformation in Referenda* (London: Routledge, 2020), at 21. (Quoting Judge Damon Keith in *Detroit Free Press vs. Ashcroft*, (19 April 2002) US Court for the Sixth Circuit.

[148] See in that sense, Baade Björnstjern, "Fake News and International Law" (2018) 29(4) *European Journal of International Law* 1357.

[149] See a treatment of the question of state responsibility for the failure of preventing harm and mitigating risk in the disinformation context at Chapter V.

[150] See notably Ben Nimmo, "The case for information defence, a pre-emptive strategy for dealing with the new disinformation wars" in Legatum Institute, *Supra* note 306, at 29.

It will be shown throughout this study that malicious intent can take several forms in the disinformation context. Some state-sponsored disinformation campaigns may serve interventionists ambitions, other may have been waged with the intent to harm individuals.

The above clarifications are fundamental because deception is not illegal *per se* under international law. If the act of spreading disinformation was *on itself* illegal under international law, then the mere proof of an intent to deceive could trigger legal consequences for the instigator. As disinformation is not expressly outlawed in current international law, assessing disinformation campaigns with the view to apply legal concepts requires that an important investigation work is done beforehand.

- The intention to undermine protected interests

The second step entails looking deeper into the nature of the inferable intent in order to determine which specific result is intended by the disinformant, and the extent to which said result is expected. This second step is as important as the first. A clear insight into the intent of the instigator is mandatory for deciding the offender's degree of culpability, as well as to the determination of the appropriate framework. For instance, an actor that manipulated and deceived for profit is unlikely to incur responsibility under international law if its communications did not cause substantial harm to the interests of a state or to the interests of the individuals that compose its audience. Similarly, an actor making use of identity concealment tactics to denounce the wrongdoings of another state by fear of retaliation would not necessarily be in breach of its international obligations, even though the audience was deceived as to the real identity of the author of the communication. Conversely, the actor who disseminates false information to vilify individuals in a systematic manner may very well be in violation of several rules of international law, as the intention to cause harm would be obvious in that case.

Various degree of malice may be encountered in the context of state-sponsored disinformation campaigns. In order to determine whether a disinformation campaign has breached international law, several factors have to be taken into account. These range from the nature of the content of the communication, to the quantity of mutually reenforcing messages attributed to the same entity, the context of the spread, the alleged motivation of the instigator, and if any, the consequences arising from the act.

With regard to the consequences of disinformation, an impact-based approach to the legality of international communications is generally of little help when it comes to apply legal concepts to disinformation campaigns.

First of all, the same content may be experienced differently depending on individual-specific variables such as the education level of the recipients, or the knowledge they have about the subject they are reading about. At society-scale, problematic information may be absorbed by the information environment of the recipient state without causing any visible effect. Conversely, unattended nefarious consequences could arise from messages communicated in good faith, without intention to harm.[151]

Second, more often than not, PSYOPS would not generate quantifiable harm or physical destruction, at least at the stage of an information spread. For instance, it is impossible to prove that the Russian meddling into the US 2016 elections succeeded into bringing about political change.[152]

The causation problem is incapacitating when it comes to assess a variety of behaviours. Many rules of international law require clear and convincing evidences of a causal nexus between the act and a result.[153] This is the case for some forms of incitements to violence, coercion, or uses of force, as the effects of the problematic conduct must be apparent in both cases. In the first case, causation serves as proof of incitement. In the second, defendants must prove that the conduct in question resulted in them feeling compelled to act in a certain way.[154] In the third, defendants must prove that substantial damages occurred and were caused by the conduct in question. In both cases, the impossibility to know how problematic information is processed by the audience, and whether a given content effectively led to the taking of the expected behaviours precludes the applicability of the relevant rules. The causation problem also affects the determination of international liability. Indeed, a prerequisite to international liability is that visible and quantifiable harm has arisen from a conduct.[155]

There is currently no workaround for the speech act causation problem that is widely accepted in the legal literature. On the one hand, basing legal assessments on the projected consequences of disinformation is a slippery slope that could have dramatic chilling effects on

[151] Baume, Boillet, and Martene, *Supra* note 147, at 23. (Arguing that the element of intention is not relevant when considering the impact of misinformation on the audience.).

[152] Henning Lahmann, "Infecting the Mind: Establishing Responsibility for Transboundary Disinformation" (2022) 33(2) *European Journal of International Law* 411; Pielemier, *Supra* note 64.

[153] Lahmann, *Ibid.,* at 439.

[154] Lahmann, *Supra* note 90, at 35. (Arguing that coercion is necessarily a combination of the consequences and intent.)

[155] See in general Sompong Sucharitkul, "State Responsibility And International Liability Under International Law" (1996) 18 *Loyola of Los Angeles Int'l. & Comp. L.J.* 821.

freedom of speech.[156] On the other hand, one cannot ignore the danger that disinformation represents in modern societies. Assessment exclusively on the visible consequences of a disinformation campaign risks excluding ill-intentioned behaviours that might still cause severe and irreversible harm in the long term. [157]

There are ways to circumvent the speech act causation problem. Specifically, formal and contextual parameters could in certain situations make up for the impossibility to establish causality when no immediate quantifiable effects arise from an information spread. Approaches could for instance focus on revisiting the conditions of application of some rules or principles of international law so that they better apply to disinformation campaigns made during periods when the access to reliable and accurate information is fundamental to the safety of the target audience and the territorial and political integrity of the target state.

The point could for example be made that the act of purposefully targeting foreign voters with disinformation during national election or referenda could amount to an interference in the affairs of another state, thereby falling under the scope of the principle of non-intervention regardless of coercive effects. In the same vein, it could be argued that the conscious promotion of unreliable, false and misleading information during critical events when the lives of individuals are at stake should engage the international responsibility of the influencer on the basis of a violation of the fundamental rights to health and the right to life of the audience.

This approach is not new. Communication campaigns containing incitements to revolt, communications amounting to threats to use force, or hateful rhetoric *de facto* constitute wrongful acts under international law, regardless of their effects in a society, provided that they are of sufficient scale.

At any rate, legal assessments should give primacy to the element of intention, the context of the spread, the characteristics of the act including its scale, and the foreseeability of the effects instead of the occurrence of effect. Whether international law truly provides for the possibility to engage the responsibility of a perpetrator for the sole reason of the content and scale of a communication campaign it instigated will be examined in this work.

[156] See in that sense Pielemier, *Supra* note 64; Lahmann, *Supra* note 152.
[157] See among many other Bayer et al, *Supra* note 63; Dillon, *Supra* note 59.

Chapter II.
Disinformation in the grey zones of international law

1. Introduction

Once suspicious levels of coordination, blatant disregard for truth and transparency are observed in a communication campaign, and if the finality of the act is sufficiently manifest and enough indicators point towards the involvement of a state, then the question of the legal characterisation of the act under international law arises. This second chapter sets the stage for the legal assessments by providing an initial insight into the international law of transboundary communications. In doing so, it shows that more often than not, disinformation activities in particular, and PSYOPS in general, can be accurately referred to as grey zone conduct.

As a starting point to a study on the legality of cyber-enabled disinformation, a few lines should be dedicated to setting out the current state of the law regarding cyber-enabled threats. On this point, it is clear that international law applies to cyber-enabled conduct. On four occasions, the Governmental group of Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security (hereinafter GGE) recognised that international law applies to cyberspace and the cyber cyberoperations.[158] In 2021, The Open-ended Working Group on Developments in the Field of Information Communication Technologies in the Context of International Security, (hereinafter OEWG) took exactly the same position on its first Final Substantive Report: "International law, in particular the Charter of the United Nations, is applicable and essential to maintaining peace, security and stability in the ICT environment."[159] Its open membership model providing the declarations of the OEWG greater authoritative weight than those of the GGE, the former's explicit endorsement of the

[158] *Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security,* , (2013), A/68/98; *Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, (2015), A/70/174; *Developments in the field of information and telecommunications in the context of international security,* (2021), A/RES/76/135; See also *Developments in the field of information and telecommunications in the context of international security*, (2015), A/RES/70/237.

[159] *Open-ended working group on developments in the field of information and telecommunications in the context of international security*, Final Substantive Report, (2021), A/AC.290/2021/CRP.2.

GGE's statements definitely brushes aside any remaining uncertainties regarding the lack of international agreement about the question of the applicability of international law to cyberspace and cyber operations.

Whilst several states still profoundly disagree on the way some rules should apply to cyberspace,[160] and even though some shared interpretations must be considered with caution, no actor departs from the initial presumption that international law applies to cyberspace.[161] In addition, scholarly works of great legal rigor have multiplied in recent years. Among the most noticeable, the Tallinn Manuals 1.0, 2.0, and the highly anticipated and upcoming Tallinn Manual 3.0, are, and will be, based on the same axiomatic view that international law applies to cyberspace and cyber-operations.[162] Therefore, whichever medium is used by state actors to wage psychological warfare, they will not be operating in a legal vacuum.[163]

In spite of this, states and legal practitioners are still confronted to an area of great legal uncertainty when it comes to assessing and responding to cyber PSYOPS, a so-called legal grey zone. Indeed, much like most cyber-enabled conduct, state-sponsored online disinformation campaigns seldom exceed the limits set out by the most commonly-accepted rules of international law. In the event that a rule of customary international law, or a particular treaty provision can be interpreted so as to apply to online disinformation campaigns, chances are high that the interpretation does not make consensus. As a result, state-sponsored cyber-enabled disinformation activities often defy straight-forward legal characterisation.

This state of affairs precludes legal characterisation, attribution, and inhibits the taking of legal responses by the affected actors. Added to the fact that online disinformation campaigns are on themselves complex to analyse[164], relying on international law might *in fine* be of little help for the victim state.

---

[160] France, for example, in a 2019 White Paper, has expressed several dissident views with regard to some of the concepts supported by the majority of the experts who participated to the writing of the Tallinn Manuals. See in that sense French Ministry of the Armed Forces, *Droit International Appliqué Aux Opérations dans le Cyberespace*, (2019), Position paper; William Letrone, "Droit du cyberespace : désillusion dans les débats" (1 March 2020) *LesYeuxduMonde*, <Https://les-yeux-du-monde.fr/actualites-analysees/42931-droit-cyberespace-desillusion-debats>.

[161] For an elaborated account on states' position regarding the applicability of international law to cyber conduct in general, see notably Marco Roscini, *Cyber Operations and the Use of Force in International Law* (Oxford: Oxford University Press, 2014), at 19-23; *Official compendium of voluntary national contributions on the subject of how international law applies to the use of information and communications technologies by States*, (2021), A/76/136.

[162] Michael N. Schmitt (ed.) *Tallinn Manual on the International Law applicable to Cyber Warfare,* (Cambridge: Cambridge University Press, 2013), *and* Michael N. Schmitt (ed.) *Tallinn manual 2.0 on the international law applicable to cyber operations,* (Cambridge: Cambridge University Press, 2017).

[163] *How* exactly international law applies to cyber conduct is still a highly debated issue. There is no consensus on this subject.

[164] See *Supra* chapter I.

The first section introduces the different categories of illegal propaganda in order to pinpoint the problems associated with the task of applying pre-existing frameworks to contemporary PSYOPS, including disinformation (2). The second section then defines the legal "grey zone," and explains why state-sponsored disinformation campaigns constitute grey zone conduct (3). The chapter concludes with a *tour d'horizon* of recent state practice. Notably, it will be seen whether states have made reference to the legal status of online disinformation activities in recent position papers and declarations, and where they seem to have set the thresholds of illegality (4).

2. Online disinformation and clearly illegal propaganda types

As deceptive and non-deceptive communication campaigns aimed at fomenting tensions abroad or between states have always existed, states have been well aware of the threat they could represent for their integrity. At least since the Napoleonic wars states have protested against the dissemination of subversive ideas by a foreign power or foreign individuals inside their territories. Eventually, some states considered setting limitations to the capacity of external powers to remotely spread problematic information beyond their territories. The advent of radio broadcasting shortly after the First World War accentuated the pressure on governments.[165] It then comes as no surprise to see many references to harmful propaganda in bilateral and multilateral agreements dating from before the advent of cyberspace.[166]

It is interesting to note that few treaties made explicit mention to propaganda before the first world war.[167] The reasons for this were multiple; One issue that remains relevant today is the difficulty of balancing communications regulations and the control of private information conveyers with the right of free speech. Other reasons were more specific to this historical period, such as the limited means available to states at the time for remotely influencing populations abroad, the ease with which a state could repel foreign propaganda, and the general acceptability of the recourse to conventional wars between states to resolve international disputes.[168]

[165] See Downey, *Supra* note 76;Van Dyke, *Supra* note 5. See also for more example of treaties on international communications Wilson & Larson, *Supra* note 60, at 124.

[166] See notably Downey, *Supra* note 76; see also in general Whitton & Larson, *Supra* note 60.

[167] To provide a general background, the Soviet Union was among the first states to include references to international propaganda in the bilateral treaties it concluded with the other powers, see Downey, *Supra* note 76, at 343.

[168] See *infra*, General Introduction.

It is only from the 1930's that states began to consider setting an international framework to safeguard the interests of states and individuals against certain types of transboundary communications. A cursory look at the way states used to deal with the question of harmful international communications before the advent of cyberspace shows two things. First, that most of the initiatives at the time were in the form of bilateral treaties. Second, that despite a noticeable incentive towards regulating certain forms of international communications, explicit references to the act of disseminating lies at scale against another state for political purpose were scarce. In general, most of the attention was devoted to international communications that constitute a treat to the security of a state, or a threat to the peace, without mention of the truth-value of the communications in question.

Legally characterising PSYOPS has puzzled lawyers for a long time already. Debates pertaining to the question of why and how to regulate international communications while at the same time preserving diplomacy and human rights date back from before the two world wars and received renewed interest throughout the 1960's, in the dramatic moments of the Cold War.[169] Information manipulation back then mainly followed the same conceptual scheme as today, with protracted destabilisation strategies akin to modern hybrid strategies deployed in peacetime with the aim of achieving or maintaining regional or global hegemony. Much like disinformation, propaganda often entail causing long term societal damages abroad, and eventually triggering regime change.

Many pre-cyber disinformation campaigns can be cited such as the well-documented AIDs disinformation campaign (or operation "Denver"), allegedly fabricated by the Soviet secret services and disseminated in the mid 1980's. The operation "Denver" consisted of accusing the Pentagon for being responsible for the creation of the human immunodeficiency virus (HIV) which was then wreaking havoc in the African continent. The alleged aim of the campaign was to discredit the US.[170] The defensive measures deployed by the US in response of the conspiracy are interesting and bear relevance to the way online disinformation should be countered even today. At first, the US sought to expose the campaign and spread factual information about HIV and the campaign massively, handing the corrections directly to foreign journal editors and

[169] Richard A. Falk "On Regulating International Propaganda: A Plea for Moderate Aims" (1966) 31(3) *Law and contemporary Problems* 622, at 624.

[170] The US Department of State argued that a possible goal of the PSYOPS was to divert the attention from the accusations made by the US, of the USSR's own violations of the 1925 Geneva Protocol on Chemical Weapons and the 1972 Biological Weapon Convention. See United Department of State Washington, D.C., "The USSR's Aids Disinformation Campaigns" Foreign Affairs Note (July 1987); Thomas Boghardt, "Soviet Bloc Intelligence and Its AIDS Disinformation Campaign" (2009) 53(4) *Studies in Intelligence* 1.

foreign representatives.[171] US officials then met with their Soviet counterparts to request the end of the disinformation campaign.[172] While the US did not formulate formal condemnations about the HIV disinformation campaign on the basis of international law, diplomacy and a successful attempt to promote transparency were relatively efficient at curbing the conspiracy even though attempts to revive the HIV conspiracy are still observable even today.[173] It is worth noting that the handling of the defamation campaign by the US was remarkably similar to the procedure of response provided in the 1953 Correction Convention, to which the US were not, and are still not, signatories.[174]

Another example of cold-war disinformation campaign can be found in the broadcasts of Radio Moscow beamed to Iran in the 1950's. The campaign was mainly aimed at fostering anti-US sentiment within the Iranian population.[175] Such radio activities were common during the cold war. They also often directly involved the intelligence services of the sponsoring state. Not explicitly denounced as "disinformation", past commentators used to refer to radio broadcasts aimed fomenting distrust and discontent within a population as "defamatory propaganda."[176]

US-sponsored disinformation efforts before the advent of cyberspace also deserve attention. Especially in Latin America, the US engaged in numerous subversive activities in order to quell the rising tide of communism. In the 1950's, during the Eisenhower administration, fabricated stories, rumours and posters were conveyed by the CIA to the Guatemalan population and the international audience to sow dissent and confusion in Guatemala and tarnish the international reputation of the government in order to overthrow the President Arbenz and his government. Part of the operation "PBSUCCESS" consisted of the broadcast of a clandestine radio program named Voice of Liberation, with the purpose of fomenting rebellion in the Guatemalan population. In a similar vein, revolutionary propaganda and false news were disseminated by the US in the 1980's to support the activities of the contras against the Sandinista government of Nicaragua.[177]

Adopting a historical perspective to the international regulation of PSYOPS could be enlightening as to how contemporary PSYOPS involving disinformation could be approached

[171] See US Foreign Affairs Note, *Ibid.* at 9.
[172] *Ibid.*
[173] Selvage & Nehring, *Supra* note 170.
[174] The Correction Convention is analysed later in this work. See infra Chapter III.
[175] See some written extracts of the programs in question in Whitton & Larson, *Supra* note 60, at 107-110.
[176] *Ibid.*
[177] See Report of the Congressional Committees, *Supra* note 4.

today.[178] Similarly, pre-cyber PSYOPS tell a great deal about what sort of strategic goals state actors could be aiming for, and how to mitigate their impact and assess their legality under international law.[179]

The subversive activities of the great powers during the Cold War largely contributed to the international impetus for the regulation of international communications. Minimal steps were however taken in the direction of a comprehensive international framework especially dedicated to transboundary communications. Be that as it may, some PSYOPS were so ostensibly inadmissible that their legal status did not give raise to disagreements amongst states and lawyers. Examples of blatantly unlawful propaganda were (and still are) international communication that violate *jus cogens*, such as incitation to genocide akin to what was witnessed prior and during the Rwandan genocide in 1994, or brazen incitement to war, or revolution, against which jamming activities may even be ordered by the Security Council, in virtue of UN Charter Chapter VII, Article 41. As such, they may accurately be defined as "clearly illegal propaganda types", as opposition to grey zone propaganda.[180]

Lawyers who worked on the question have in the past singled out warmongering propaganda, defamatory propaganda and subversive (or revolutionary) propaganda, as international communications of concern.[181] Falk enumerates the categories distinguished by Whitton and Larson and summarizes them as follows; "1. Warmongering - the incitement to war and uses of force in international relations. 2. Subversive intervention - the incitement of audiences to take illegal action against their own government. 3. Defamation - the vilification of foreign governments and of their political leaders."[182] To this list, discriminatory (or hate) propaganda including genocide propaganda, and terrorist propaganda were later added.[183] The latter types of propaganda have drawn particular attention in the 1990's and the early 2000's following the Rwandan and Bosnian genocides and the rise of international terrorism.

[178] See in that sense 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.

[179] *Ibid.*

[180] Un Charter article 41 provides that "The Security Council may decide what measures not involving the use of armed force are to be employed to give effect to its decisions, and it may call upon the Members of the United Nations to apply such measures. These may include complete or partial interruption of economic relations and of rail, sea, air, postal, telegraphic, radio, and other means of communication, and the severance of diplomatic relations." See also in that sense Jamie Frederic Metzl, "Rwandan Genocide and the International Law of Radio Jamming." (1997) *91(4) The American Journal of International Law* 628.

[181] See Whitton, *Supra* note 60.

[182] Falk, *Supra* note 169; Whitton & Larson, *Supra* note 60.

[183] De Brabandere, *Supra* note 76.

The categories of prohibited propagandas were derived from the principles of non-intervention, non-use and threat of force, and more broadly, the necessity to promote good understanding among civilisations and the maintenance of peace. The question of the regulation of *deceptive* communication campaigns was thus contained within the provisions of documents governing the recourse to international propaganda with the aim of harming other states or endangering the peace.[184] Among all the different illegal propaganda types, war propaganda, genocide propaganda, hate speeches, terrorist propaganda and the most egregious subversive propaganda are relatively easy to legally characterise. These are acts that, because they contravene the most basic norms of morality, are already expressly outlawed.[185]

Yet, looking at how propaganda activities were regulated before the advent of cyberspace and how jurists attempted to tackle the question also reveals profound difficulties inherent to the challenge of correctly placing the cursor of illegality for PSYOPS. An important characteristic of past discussions pertaining to the legal status of international communications was the difficulty in drawing clear legal and conceptual lines against which to assess their legality under international law. In fact, already in the 1960's Falk noted, in the course of an analysis on the regulation of transboundary communications that; "it is virtually impossible to regulate the behavior of nations unless the boundaries of impermissible conduct can be specified in fairly objective terms; as such boundaries are impossible to provide in relation to international propaganda, the prospect is that any regulating scheme would be arbitrarily interpreted and applied, if it were effective at all."[186] Obviously, this reflexion does not apply to unquestionably illegitimate propaganda activities. Rather, it concerns more ambiguous forms of propaganda. Defamatory propaganda, non-blatantly subversive communications, and more generally, every insidious PSYOPS that remain too subtle to constitute brazen violation of international law are extremely difficult to regulate in an adequate way.[187]

Today, the most common disinformation campaigns seem to pose similar problems to lawyers. At least to the extent that the falsities conveyed in the communications are not outrageously inappropriate, it seems that *disinformation campaigns are not, in principle, illegal. They may, however, by way of their formal characteristics and/or, the objective they serve, breach the law*.

[184] Mark Leiser, "Regulating Computational Propaganda: Lessons from International Law" (2019) 8(2) *Cambridge International Law Journal* 218.

[185] This is the case of genocide propaganda, which constitutes a crime under international criminal law.

[186] Falk, *Supra* note 169.

[187] Falk notably mentions false allegation of "germ warfare" as one particularly inacceptable information to spread internationally to which the adequate answer would be a right to reply, as opposed to forceful retaliation. Falk, *Supra* note 169, at 625.

As argued in the first chapter, an appropriate approach to disinformation would be to focus on the intent of the instigator rather that the truth-value of the information spread. After all, there can be no assurance that malicious actors will not use the disinformation argument to unjustifiably condemn an international communication campaign they deemed inappropriate or embarrassing.[188] The same way as other grey zone conduct, the concept of disinformation carries with it a powerful potential for instrumentalization. This is primarily due to the fact that disinformation is not confined to obvious, verifiable lies.[189] Legal claims arising from the obvious falsity of the information conveyed, and/or the non-transparency of the communication in question are conceivable, basing a claim on the overall misleadingness of a whole communication campaign, or the fact that the audience had been deceived to believe in lies, in contract, is untenable as far as the burden of proof assumed by the victim state is concerned.[190] A similar approach was already adopted during the cold war. Regulators were not concerned with deceptive tactics *per se*, but with propaganda activities carried out in peacetime to undermine the interests of the target state or the international community at large. In other words, the purpose of the communications always mattered more than the truth-value of the content.[191] Outside the hypothetical case of verifiably wrong information repeated over a considerable period of time, reasoning in terms of what is true and what is untrue would have set unpracticable parameters to work with for states and lawyers.[192]

Disinformation is not confined to one category of international communication. Disinformation can take many forms and thus sometimes constitutes blatantly unlawful act. This would concern incitement to war, genocide propaganda, terrorist propaganda and revolutionary propaganda. While these thresholds do not directly hinge on the truth-value of a communication, its manipulative nature, its deceptiveness, or even the disruption they may cause, they have the advantage to be less hazardous than any imaginable rule created to uniquely sanction international communication efforts based on the accuracy of the information conveyed.[193] Hence, disinformation campaigns amounting to the aforementioned types of international communication may clearly contravene international law. This may however concern only a minority of conduct. This means that principles of customary law such as the non-intervention principle or the non-use of force principles may apply to certain forms of

[188] See notably Baade, *Supra* note 148.
[189] See Starbird, Arif & Wilson, *Supra* note 72.
[190] See in that sense Whitton, *Supra* note 60, at 619.
[191] Whitton & Larson, *Supra* note 60, at 177.
[192] See a similar point at Lahmann, *Supra* note 62.
[193] De Brabandere, *Supra* note 76.

online disinformation campaigns, regardless of their qualification as subversive or defamatory propaganda. The problem is that even in these cases, there is no straightforward answer to the question of their legality under customary international law.[194] This is because none of the cited principles is breached by an international information campaign for the mere reason of the deceptive character of its content, nor for the mere reason that there is a possibility that the conduct has detrimental effects on the authority of the victim state in the long term.

Traditionally, it is when the alleged goal of the disinformant is sufficiently unequivocal to leave no doubts as to the coerciveness of its intent that the international communication in question can amount to an illegal intervention and eventually a use of force. This would undoubtedly concern disinformation activities carried out to provoke the overthrow of a government, i.e., subversive propaganda, or propaganda of war.

However, most state-sponsored PSYOPS either bear resemblance to defamatory or subversive propaganda, or are outright impossible to classify in existing propaganda categories. In practice, this does not make much of a difference. There is no consistent regulatory framework in contemporary international law for defamatory and subversive propaganda. As regards defamatory statements, state practice often contradicted with potential prohibitions of defamatory propaganda.[195] States generally refused to take responsibility in the waging of subversive or defamatory propaganda against another state on the basis that the communications came from non-state actors upon which they are not entitled to exercise control.[196] At best the accused states would get away with diplomatic protests from the aggrieved party.

Examples of defamatory and subversive propaganda abound. Although contradictory practice to a principle would not necessarily empty a principle from its legal substance, the existence of a customary principle forbidding states from engaging in defamatory propaganda remains debatable, contrarily to what some authors have argued in the past.[197] In the eventuality that such a principle exists, its scope of application remains insufficiently elaborated, and so are the legal consequences that a contravention to that principle may give raise to. With regard to subversive propaganda, it is generally admitted that communications clearly intended to

[194] See in that sense Stephens, *Supra* note 15.

[195] See Whitton, *Supra* note 60, at 606; Larson 60, at 447.

[196] See notably Falk, *Supra* note 169. As for the connected principle of non-intervention, which is supposed to apply to subversive propaganda, some authors have even questioned its status of rule of international law due to the consistency with which it was violated by states. See in that sense Glahn, *Supra* note 5, at 562. The argument was however later (partly) rebutted by the ICJ in *Case Concerning Military and Paramilitary Activities In and Against Nicaragua* (Nicaragua v. United States of America) (1986) ICJ, judgement. An in-depth analysis of the principle of non-intervention is submitted in Chapter III.

[197] See notably Whitton & Larson, *Supra* note 60, at 110.

overthrow a government are illegal.[198] Yet again, although the possibility that contemporary PSYOPS activities take such a form should not be excluded, this would concern a crushing minority of them. That is to say, that the approach of a disinformant is more subtle than the approach of a revolutionary group. Even if the ends are similar, the means are not. Indeed, a common feature of most online disinformation campaigns is specifically that they are sufficiently subtle to not appear blatantly illegal under current international law.[199] Online disinformation is above all a tool for creating disruption in an incremental manner, a way to deflect the attention of the victim state and eventually, to align the interests of the victim with those of the disinformant. In essence, state-sponsored PSYOPS have *all* features of a grey zone conduct; they are non-forcible, deniable, incremental and hard to qualify and to attribute, at least in the legal sense of the terms.[200]

3. The grey zone and its relevancy to online disinformation

Complaints about the lack of clarity of the law have dramatically increased in the recent years. Some were notably made in connection to assessments of transboundary state-sponsored PSYOPS.[201] The reason is relatively simple: an increasing number of actors has recourse to sophisticated destabilization methods that do not include any military aspect. Blatant and unjustified coercion, be it cyber-enabled or not, would generally leave no doubt as to its illegality. But as far as the contemporary security threat landscape goes, blatant coercion is rare. In addition, malicious actors increasingly rely on proxies to carry out condemnable conduct. In the eventuality that a breach of the law occurs, illegal conduct may not even be legally attributable to a state.

Some conduct have proven so elusive that states and legal practitioners are often forced to venture into the extremities of a rule's scope of application in order to make up for the lack of dedicated framework. In such a situation, however, invoking the law against the purported "wrongdoer" is generally of little help for the alleged victim. Contentious interpretations of a rule might be easily rebutted and attribution might not even be possible. Whether because the conduct does not fit well within existing legal frameworks, or because non-state actors have

[198] See infra chapter III, "unproblematic application of the principle of non-intervention"
[199] See Lyle J. Morris, *Supra* note 207, at 9.
[200] See in that sense Hoffman, *Supra* note 214; Lahmann, *Supra* note 62; Zhao, *Supra* note 117, at 46.
[201] Hollis, van Benthem and Dias, *Supra* note 61.

been called upon to carrying out the conduct for the account of a state, instigators evolve in a situation of relative impunity.

Early December 2021, at the occasion of a conference on influence operations and disinformation in cyberspace, Colonel Pascal Ianni, chief of the French Communication Department and spokesperson of the Chief of Staff deplored a gradual "erosion" of international law since the second world war. More specifically, Colonel Ianni, deplored the difficulty of legally characterizing the disinformation campaigns targeting the French forces engaged in Mali and central Africa.[202] According to the French colonel, the resulting lack of clear operational framework hobbles the capacity for the country to defend, legally, against the act.

The PSYOPS that Colonel Ianni was referring to in his keynote speech fall under a specific and widely-acknowledged category of "grey zone" conduct[203] that usually take advantage of cyberspace. "Grey zone conduct" refers to activities falling short of blatant illegality while still proving particularly harmful to the interests of the target.[204] In other words, these are illegitimate conduct that defy straight forward legal categorization. From the perspective of international law, grey zone activities lay well outside the scopes of the most uncontested rules such as the prohibitions of the use of force, the right to self-defence and the effective control test for the attribution of private conduct to a state.

The types of conduct the grey zone incorporates can be defined as acts "[…] which seek to gain advantage without provoking a conventional military response […]"[205] Some scholars have provided more elaborated definition of conduct typically falling into the grey zone; "covert or illegal activities of non-traditional statecraft that are below the threshold of armed organized violence; including disruption of order, political subversion of government or non-governmental organizations, psychological operations, abuse of legal processes, and financial corruption as part of an integrated design to achieve strategic advantage."[206] A 2019 Rand study adds that one defining feature of grey zone activities is their incremental nature. That is

[202] Remarks of Colonel Pascal Ianni at the Cyberdefense and Strategy conference held in Paris, (6 December 2021).

[203] The terms of "grey zone activity, tactics, conduct", and "under-the-threshold conduct" are used interchangeably to refer to the same category of conduct.

[204] See Wigel, *Supra* note 95. See also Katherine Mansted, "The Domestic Grey Zone: Navigating the space between Foreign Influence and Foreign Interference" (2021) National Security College, Australian National University; Australian Government, Department of Home Affairs, "Countering Foreign Interference" (27 February 2020), <https://www.homeaffairs.gov.au/about-us/our-portfolios/national-security/countering-foreign-interference>.

[205] "Competing in the Gray Zone Countering Competition in the Space between War and Peace" Center for Strategic and International Studies, (CSIS) International Security Program, <https://www.csis.org/features/competing-gray-zone>.

[206] See Frank G. Hoffman, "Examining Complex Forms of Conflict: Gray Zone and Hybrid Challenges" (2018) 7(4) *PRISM* 31.

to say, the completion of the definitive goal of grey zone activities and at a greater extent, the completion of whole strategies relying on grey zone activities is almost systematically delayed.[207] This allows for grey zone conduct to go unchecked and undetected while gradually gaining traction within the territory of the target.[208] Because most activity will appear benign at first glance, they are extremely delicate to address by the recipient state, let alone to detect. On that point, commentators have warned about the "potential for interference" of in appearance normal interactions.[209]

From scathing provocations to tactics of *fait accompli,* to economic coercion, political corruption, to protracted PSYOPS involving disinformation methods, grey zone activities can take a variety of forms.[210] The great powers of the Cold War have conducted grey zone election interference throughout the twentieth and the twenty first century to gain influence or protect their respective spheres of influence. More recent, the targeted drone executions conducted by the US in sovereign territories would qualify as a grey zone activity. Grey zone activities may also take the form of state-sponsored election meddling.[211] Another example of grey zone activities relates to the activities of the Chinese Coast Guard and fishing vessels in the East China Sea, around the Senkaku Islands over which both China and Japan claim sovereignty.[212] Economic coercion, when not provided for by law, might also qualify as grey zone activity. The decision taken by the Chinese authorities, to cut trade with Lithuania after the latter's decision to strengthen ties with Taiwan, and to pressure other European countries to refrain from trading with Lithuania can be considered a form of grey zone economic coercion.

Unsurprisingly, most cyber-enabled threats constitute grey zone activities. This concerns, notably, data-stealing for the account of a state, the introduction of backdoors in computer programs before their distribution to the governmental services of a competitor, the installation of malwares in the electronic devices of foreign agents in view to trigger their activation on a later date, or cyber disinformation conducted against foreign populations.

---

[207] See notably, Lyle J. Morris et al., "Gaining Competitive Advantage in the Gray Zone Response Options for Coercive Aggression Below the Threshold of Major War" (2019) Santa Monica, CA: RAND Corporation. See also Mansted, *Supra* note 204.

[208] See Lyle J. Morris et al., *Ibid.,* at 9; Wigell, *Supra* note 95.

[209] 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.

[210] See Wigell, *Supra* note 95.

[211] Michael N. Schmitt, "'Virtual' Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law" (2018) 19(1) *Chicago Journal of International Law* 30, at 38.

[212] Lyle, J. Morris et al., *Supra* note 207; Céline Pajon, "Japan's Coast Guard and Maritime Self-Defense Force in the East China Sea: Can a Black-and-White System Adapt to a Grey-Zone Reality?" (2017) 23 *Asia Policy* 111.

With that being said, the category of grey zone conduct is inherently fluid and may grow or shrink over time, depending on technological advancements or the state of the law. For these reasons, it has been recognised that any attempts to enumerate every grey zone conduct is of limited help in terms of indicating an approach to the issue.[213]

If grey zone activities do not fit well within the categories of the law, it is partly for the reasons that the scopes of the applicable rules are in themselves ambiguous.[214] The main characteristics of grey zone conduct is specifically the difficulty of applying legal concepts to them. The "legal grey zone" can thus be described as the "international law principles and rules that are poorly demarcated or are subject to competing interpretations."[215] Specifically in this context, the grey zone refers to the rules that are supposed to apply to non-forcible and non-forcible conduct. These are mainly the sovereignty and the non-intervention principles. They also concern secondary rules governing the consequences of a breach of primary rules of international law such as the regime of legal attribution of the state responsibility and the circumstances precluding wrongfulness contained in the ARSIWA.[216] What all these rules have in common is their ambiguity, and the fact that observers may have great difficulty making sense of their scopes.

The elusiveness of grey zone activities transpires from recent statements made by state officials, organs or state-affiliate organisations about acts that do not fit in a straight-forward way in available frameworks. For example, in a 2018 statement, the British National Cyber Security Centre condemned some cyber-enabled acts that it attributed to the Russian military intelligence services as "flagrant violation of international law," without specifying on what way the act in quesiton had contravened the law.[217] Another illustration was the public statement made by the then-US President Obama in 2015, about a series of cyber-operations

[213] Lyle, J. Morris et al., *Supra* note 207.

[214] See notably Kubo Macak, "Is the International Law of Cyber Security in Crisis?" in N. Pissanidis, H.Rõigas, M.Veenendaal (Eds.) 2016 *8th International Conference on Cyber Conflict: Cyber Power* (Tallinn: NATO CCDCOE Publications), at 127; Elizabeth K. Kiessling, "Gray Zone Tactics and the Principle of Non-Intervention: Can "One of the Vaguest Branches of International Law" Solve the Gray Zone Problem?" (2021) 12 *National Security Journal* 116; Michael N. Schmitt, "Grey Zones in the International Law of Cyberspace" (2017) 42(2) *Yale Journal of International Law Online 1*; Frank G. Hoffman, "Examining Complex Forms of Conflict: Gray Zone and Hybrid Challenges" (2018) 7(4) *PRISM* 31; Gen Michael V. Hayden, "The Future of Things Cyber" (2011) 5(1) *Strategic Studies Quarterly* 3. Andrew C. Foltz, "Stuxnet, Schmitt Analysis, and the Cyber "Use-of-Force" Debate" (2012) 67(4) *Joint Force Quartely* 40. See Rebecca Crootof, "International Cybertorts: Expanding State Accountability in Cyberspace" (2018) 103 *Cornell Law Review* 565; Hollis, van Benthem and Dias, *Supra* note 61; Nicolas, *Supra* note 22.

[215] See Schmitt, *Supra* note 214, at 1.

[216] See Schmitt, *Supra* note 211.

[217] "Reckless campaign of cyber-attacks by Russian military intelligence service exposed" (3 October 2018) National Cyber Security Centre. <https://www.ncsc.gov.uk/news/reckless-campaign-cyber-attacks-russian-military-intelligence-service-exposed>; See Macak, *Supra* note 214; Kubo Mačák, "On the Shelf, But Close at Hand: The Contribution of Non-State Initiatives to International Cyber Law" 113 *AJIL Unbound* 81, at 83.

affecting government staff and subsequently attributed to China. The then-American president qualified the acts as "not acceptable" and a "core national security threat" without legally qualifying them.[218] In a similar note, no reference to an existing rule of international law was made by President Obama following the Russian meddling in the 2016 US elections, the then-president only referring the meddling as "efforts to harm U.S. interests in violation of established international norms of behavior."[219] One may also see in the public condemnations following the low-intensity cyber-attacks allegedly conducted by Russia against the information systems of Georgia on October 2019 an illustration of the elusiveness of grey zone conduct. Following the events, the state of Canada indeed deplored a "blatant disregard for international law, norms and agreements" without specifying a particular rule.[220] Apart from the references to the vague concept of state sovereignty, other actors such as the EU, Estonia, and the Netherlands have mostly condemned a violation of the "norms of responsible state behaviour in cyberspace"[221], a concept with modest legal significance.

The disinformation campaign targeting the French forces deployed in Mali, the *Opération Barkhane*, is a case in point to illustrate the issue raised by grey zone conduct. The Wagner paramilitary group, in collaboration with other pro-Kremlin actors, is accused to have mounted a disinformation campaign in order to foster distrust and discontent towards the French presence in Mali among the locals. Fake online identities mimicking those of locals were used to propagate anti-French and anti-European sentiment, making ample use of anti-colonialist rhetoric. The disinformation campaign also involved the spread of the absurd rumour according to which France was supporting terrorist groups in the region. Shortly before France's official withdrawal from the region, fabricated videos showing civilian corpses purportedly murdered

[218] "Obama: China cyber-attacks 'unacceptable'" (12 September 2015), BBC, <https://www.bbc.com/news/world-us-canada-34229439>.

[219] "Statement by the President on Actions in Response to Russian Malicious Cyber Activity and Harassment" The White House President Obama (29 September 2016) Statement available at: <https://obamawhitehouse.archives.gov/the-press-office/2016/12/29/statement-president-actions-response-russian-malicious-cyber-activity> See also Schmitt, *Supra* note 211. at 39.

[220] "Canada condemns Russia's malicious cyber-activity targeting Georgia" (20 February 2020) Global Affairs Canada, <https://www.canada.ca/en/global-affairs/news/2020/02/canada-condemns-russias-malicious-cyber-activity-targeting-georgia.html>.

[221] Republic of Estonia ministry of foreign affairs, "Statement of the Foreign Minister of the Republic of Estonia Urmas Reinsalu" (20 February 2020), <https://vm.ee/en/news/statement-foreign-minister-republic-estonia-urmas-reinsalu>; Government of the Netherlands, "The Netherlands considers Russia's GRU responsible for cyber-attacks against Georgia" (20 February 2020), <https://www.government.nl/ministries/ministry-of-foreign-affairs/documents/diplomatic-statements/2020/02/20/the-netherlands-considers-russia%E2%80%99s-gru-responsible-for-cyber-attacks-against-georgia?fbclid=IwAR0gmhX1vWkbss7KMe6Id7tFjuKTRVBpl4nbt60rZgTyQZ4jILS3A31_PTg>.

by the French forces were disseminated in Mali's online information space.[222] France officially condemned the act and accused the Wagner group of "staging mass graves."[223] Wagner's actions may have contributed to tensions between the French forces and the locals. Several incidents implicating locals occurred before the French withdrawal from Mali. It was notably reported that a French convoy was attacked by locals who were convinced that the French were arming terrorist groups, a rumour that was amply disseminated inside Mali's own online information space.[224] Overall, it is alleged that the PSYOPS had for effect to facilitate the actions of the Russian private military group Wagner in Mali and reenforced Russia's influence in the region.[225]

Several questions however arose from the case; did the deception work constitute an internationally illegal act? On the basis of which rule? How to respond legally to it? What about the enabling state, Mali? And is it possible to attribute the acts of the members of the Wagner group to the Russian government? These are recurring questions in the context of grey zone conduct.

When facing grey zone conduct such as disinformation, the room for manoeuvre is quite narrow for the victim state.[226] Notably, it is particularly difficult for the target to sanction an illegitimate behaviour as an undue conduct when the contours of the rule called upon to sanction the said act are yet to be indisputably established. In a similar vein, engaging in the same behaviour against the wrongdoer expose the victim to condemnations, which may damage its credibility in the long term. France learnt it at its own expense.[227]

Some states might also be reluctant to legally qualify the conduct they suffered from for the reason that they seek to engage in similar unacceptable behaviours for their own benefits. The potential victims that possess the means to act covertly through the cyber medium, or to engage

[222] Nicolas Barotte, "Au Sahel, la France veut faire tomber le masque du groupe Wagner" (21 Avril 2022), *Le Figaro International,* <https://www.lefigaro.fr/flash-actu/mali-des-mercenaires-russes-filmes-en-train-d-enterrer-des-corps-20220421>.

[223] Morgane Le Cam, Cyril Bensimon, Arthur Carpentier, Delphine Bernard-Bruls & Service vidéo du Monde, "Charnier filmé dans le nord du Mali : l'armée française accuse les mercenaires russes du Groupe Wagner de manipulation" (22 April 2022), *Le Monde*, <https://www.lemonde.fr/afrique/video/2022/04/22/charnier-filme-au-nord-du-mali-l-armee-francaise-accuse-les-mercenaires-russes-du-groupe-wagner-de-manipulation_6123290_3212.html>.

[224] Colonel Ianni, *Supra* note 202.

[225] *Ibid.*; Gleicher, *Supra* note 42; Jared Thompson, Catrina Doxsee & Joseph S. Bermudez Jr., "Tracking the Arrival of Russia's Wagner Group in Mali" (2 February 2022), *Center for Strategic and International Studies,* <https://www.csis.org/analysis/tracking-arrival-russias-wagner-group-mali>.

[226] See in that sense, European Federation of Journalists (EFJ), "Fighting disinformation with censorship is a mistake" (2022), <https://europeanjournalists.org/blog/2022/03/01/fighting-disinformation-with-censorship-is-a-mistake/>; Schmitt, *Supra* note 211; Mačák, *Supra* note 214. Lyle, J. Morris et al., *Supra* note 207.

[227] See for example Gleicher, *Supra* note 42.

in other grey zone conduct may prefer to maintain an ambiguity rather than condemning conduct that they may later adopt to secure valuable advantages.[228]

Finally, there is the possibility that what a state considers a "grey zone" conduct is simply not meant to be treated as an internationally wrongful conduct. An important aspect of the grey zone is that the concept is often evoked in military studies. The grey zone serves primarily as a conceptual framework for military planners. Hence one should not understand the "grey zone" as a domain where every activity deserves legal remedies.

Be that as it may, it is probable that at least a small portion of state-sponsored disinformation activities that usually belong to the grey zone should receive legal treatment, and thus be removed from the grey zone. Doing nothing would normalize the use of disinformation activities for the purpose of waging psychological warfare, which could have dramatic consequences on the long term, both for human rights, states sovereignties, and even for the future of information and communications technologies.

4. States' views on the legal status of peacetime transboundary disinformation campaigns

No international treaty expressly regulating transboundary harmful communications campaigns have been adopted since the advent of cyberspace. Similarly, no treaty in international law specifically outlaws the waging of disinformation campaigns by a state, and it is unlikely that new *lex specialis* on malicious use of information meet universal approval in the near future.[229] Meanwhile, as remarked earlier, most disinformation campaigns hardly fit in the category of illegal propaganda types. However, a number of international instruments can be interpreted as applying to disinformation campaigns. Specific treaty provisions may relate to disinformation, not because disinformation or the politically-motivated dissemination of lies and/or misleading information to cause harm was expressly outlawed by the parties, but because the treaty prohibited certain practices that share similar features with modern disinformation activities. Similarly, many principles of customary international law may apply to transboundary disinformation campaigns when said conduct undermine the values protected by

[228] Macak, *Ibid.*

[229] Several authors have called upon states to set forth a new regulation on disinformation activities. See notably Nicolas, *Supra* note 22.

the principle in question. The crux of the matter being to clarify *how* protected values are undermined by disinformation.

Disinformation is attracting a great deal of interest. Many states and scholars are currently working in clarifying the legal status of state-sponsored disinformation activities. Therefore, it is interesting to inquire on the way states have so far approached state-sponsored disinformation campaigns.

So far, official statements regarding the legal status of foreign disinformation campaigns have focussed on the principles of customary law, mainly, the principle of non-use or threat of force, the prohibition of illegal intervention, and the principle of sovereignty. Customary law is more flexible that treaty law. Although some customary rules have been codified, their meaning may still evolve over time depending on state practice and *opinio juris*. In turn, subsequent international treaties may modify or reject an existing international custom.[230] Thus, the extent of the protection offered by the above-cited principles vary depending on what states do and say. In turn, state practice in the context of international security is influenced by real-life events such as the emergence of new security threats.

The advent of cyberspace has given rise to new security threats that have challenged a number of long-established rules and principles of customary international law, including those subsequently enshrined in international treaties such as the prohibition to use or threaten force or the principle of non-intervention. As state-sponsored disinformation may not always fit into traditional categories of illegal propaganda types, it is interesting to look at the practice of state in order to establish whether a particular existing framework is favoured over another when it comes to assess state-sponsored disinformation, or whether states are considering law creation.[231]

Article 38 of the ICJ statute defines international custom as "evidence of a general practice accepted as law." An objective element, a general practice, and a subjective element, the belief to be acting in conformity with the law, are needed for an international custom to emerge.[232] That is to say, to create law, a substantial number of states must first have adopted and consistently engaged in the same conduct over a sufficiently long time period.[233] While there are no clear requirements as to the scale of the practice, it is generally expected that a

[230] See notably International Law Commission, "Ways and Means of Making the Evidence of Customary International Law More Readily Available" (1950) 2 *Yearbook of the International Law Commission* 367, A/1316, at 368.

[231] See also an examination of state practice on disinformation at Lahmann, *Supra* note 62.

[232] See in that sense *North Sea Continental Cases* (Germany v Denmark and the Netherlands) (1969) ICJ, Reports, §77, at 44; ICJ (1986), *Supra* note 196, §207.

[233] See for example *Right of Passage over Indian Territory* (Portugal v. India), (1960) ICJ, Judgment, at 40.

sufficiently diverse group of states engage in the same conduct to form a custom. Also, the more widespread is a practice, the less stringent the duration requirement.[234]

Looking at states practice before the advent of cyberspace with regard to propaganda in general, some scholars have noted that despite persistent derogatory practice, official protests were often lodged against the state from which defamatory propaganda was emanating.[235] The same way, non-aligned states have persistently protested against the imperialistic propaganda of the great powers.[236]

Looking at how states have behaved since the advent of cyberspace does not bear much light as to potential evidences of state practice and *opinio juris* regarding the legal status of state-sponsored disinformation campaigns in general.[237] With regard to disinformation specifically, it is well-established that a number of states are engaging in PSYOPS against each other. Disinformation campaigns have been identified and some have already been formally attributed.[238] The states that were suspected in engaging in disinformation campaigns against foreign populations have either expressly denied their involvement in disinformation campaigns, remained silent in face of accusations, or argued that the information they disseminated was at least truthful.[239] Therefore, states are unwilling to consider disinformation campaigns as an admissible type of PSYOPS. Whether states are considering a pre-existing legal framework when it comes to condemn foreign disinformation campaigns can only be verified through an examination of written statements and declarations made by state officials.

Acts performed through the medium of spoken or written words may also be considered state practice.[240] The ILC indeed established in 1950 that evidence of international custom may also be found, *inter alia*, in international treaties, the decisions of national courts, diplomatic correspondence, national legislations and opinions of national legal advisers.[241] While it is certain that everything that state officials say and write about state-sponsored disinformation

[234] See ICJ (1969), *Supra* note 232, at 42, §73.
[235] Whitton & Larson, *Supra* note 60. at 111.
[236] Downey, *Supra* note 76.
[237] Lahmann, *Supra* note 62.
[238] See notably Statement by Mr. Nicolas de Riviere, Permanent Representative of France to the UN To the Security Council, "Russia is making unfounded accusations", (13 May 2022), <https://onu.delegfrance.org/russia-is-making-unfounded-accusations>; ODNI report, *Supra* note 41. It bears noting that formal accusations, whether justified by the facts or not, are to be distinguished from legal attribution.
[239] Shaun Walker, "'Read my lips – no': Putin denies Russian meddling in US presidential election" (30 March 2017), *TheGuardian,* <https://www.theguardian.com/world/2017/mar/30/vladimir-putin-denies-us-election-interference-trump>; Mark Scott & Elisa Braun, "France feuds with Facebook over disinformation claims", (17 December 2020), *Politico*, <https://www.politico.eu/article/france-facebook-disinformation/>; Ellen Nakashima, *Supra* note 44.
[240] ILC 1950 report, *Supra* note 230; ILC Report of the 70th Session UN Doc A/73/10 (30 April–1 June and 2 July-10 August 2018), at 126.
[241] See ILC 1950 Report, *Ibid.*

campaigns should not be taken as face value when it comes to assess the legal status of the conduct, they may at least indicate a general trend towards an approach to the limitations set by customary international law to psychological warfare.[242]

Regarding municipal laws, it is interesting to note that, not only states and representatives of the civil society regularly denounce disinformation and warn against abusive regulations, but defamation, libel or slander, which may integrate disinformation aspects, are also penalized in most domestic systems. This is for example the case in the French *loi du 29 juillet 1881 sur la liberté de la presse*, the Chapter 181 of the Code of Laws of the United States of America, Articles 87 and 246 of the Criminal Law of the People's Republic of China, and the German criminal code.

As a general principle, it is also expected that large news providers comply with the norms of media ethic, such as exercising "due impartiality" when reporting about particular events. For this, independent organisations and national regulators have been instituted in most states to ensure media literacy, a "safe" media environment, and in some cases, information pluralism. National Telecommunication regulators also provide sanction mechanisms in case of non-compliance. Sanctions span from fines to the revocation of broadcasting licences. An example of application is UK's OFCOM broadcasting Code Communications Act of 2003 (2003 Act), which has found RT and Sputnik, the two major Russian news providers in Europe, in violation of their duty of "due impartiality in relation to: matters of major political or industrial controversy; and major matters relating to current public policy", finning them for repeated failures to comply with the regulation.[243] Among the other national regulators can be cited the Croatian Agency for Electronic Media (AEM) or the American Federal Communications Commission (FCC) or the French Regulatory Authority for Audiovisual and Digital Communication (ARCOM). In recent years, the aforementioned authorities have been concerned with the rise of foreign state-sponsored disinformation activities. Early 2022, the European Regulators Group for Audiovisual Media Services (ERGA) called for a unified response to foreign disinformation and information manipulation among European media regulators through the elaboration of a European Code of Practice on Disinformation and its monitoring.[244] These efforts show that existing regulatory frameworks are used to address disinformation, and that new regulations are being put in place in order to improve states'

[242] Roscini, *Supra* note 161, at 30.
[243] See "Broadcast standards cases", (2018), 369 *Ofcom Broadcast and On Demand Bulletin*.
[244] "ERGA's united response to foreign disinformation and information manipulation: European media regulators strengthen their cooperation" (10 February 2022) ERGA, Press Release.

response to disinformation. However, the work of these entities only concerns traditional media and official journalists, and does not indicate much in terms of the legal status of state-sponsored disinformation activities under international law.

Many states have strengthened their posture *vis à vis* foreign malign influence involving the manipulation of online information.[245] Citing the necessity of ensuring the protection of national security objectives,[246] public interests,[247] sovereignty[248] and domestic electoral processes,[249] many states have reappraised and revised their legislation on foreign influence and interferences in order to better address foreign disinformation. Some of these positions may have a legal resonance, but they can hardly on themselves account for genuine expression of legal opinion. In fact, in general, the question of illegality in municipal law is a different issue than the question of illegality in international law.[250] Thus, in order to gauge the extent of state practice and *opinio juris* in the context of online disinformation, it is required to go beyond municipal law. This means assessing whether and how states have publicly and officially justified their postures by references to international law.

With regard to declarations by officials, some states have admitted having recourse to PSYOPS to influence foreign populations, which may include disinformation, on a regular basis. In fact, declarations from defence officials and military doctrines reveal that psychological warfare is part and parcel of international relations.[251] Yevgeny Prigozhin, a Russian businessman suspected of having played a central role in organizing PSYOPS, including disinformation campaigns, for the account of Russia recently declared "[W]e have interfered, we are interfering and we will continue to interfere. Carefully, accurately, surgically and in our own way, as we know how to do."[252] A few years ago, Loch Johnson, former staff director of the US House Subcommittee on Intelligence Oversight former similarly admitted

[245] See Funke & Flamini, *Supra* note 733.

[246] See for instance the 2016 US Countering Information Warfare Act: Section 2.

[247] See for instance Singapore's Foreign Interference (Countermeasures) Act, (FICA) (2021).

[248] See for instance Australian government Department of Home Affairs, "Countering foreign interference", <https://www.homeaffairs.gov.au/about-us/our-portfolios/national-security/countering-foreign-interference>.

[249] See for instance *LOI organique n° 2018-1201 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information*.

[250] See Brian J Egan, "International Law and Stability in Cyberspace" (2017) 35 *Berkeley Journal of International Law* 169, at 174; Lahmann, *Supra* note 62, at 211.

[251] See for instance the updated doctrine on influence orations, as well as NATO and US's respective doctrines on psychological operations, *Supra* note 35; See also Lahmann, *Supra* note 62, at 210.

[252] Comment reportedly posted on the Russian social media VKontakte, Pjotr Sauer, "Putin ally Yevgeny Prigozhin admits interfering in US elections" (7 November 2022) *The Guardian*, <https://www.theguardian.com/world/2022/nov/07/putin-ally-yevgeny-prigozhin-admits-interfering-in-us-elections>.

that the US engaged in foreign election meddling activities against their adversaries, notably *via* disinformation.[253]

As regards disinformation in particular, many state officials have been vocal about the threat posed by foreign disinformation. In spite of this, very few have made references to actual rules of international law.[254] Some relevant national positions are listed in the 2021 Official Compendium submitted by the experts of the GGE[255] and the eventual subsequent views can also be consulted on the updated list of national positions available on the "international cyber law interactive toolkit" rubric of the website of NATO's Cooperative Cyber Defence Centre of Excellence (CCDCOE). Below is a list of recent state declarations about state-sponsored online disinformation and the possible frameworks:

- Switzerland's position paper on the application of international law in cyberspace briefly mentions online disinformation and propaganda. Reiterating the need to clarify the contours of the principle of sovereignty, Switzerland inquired about the possibility to consider as an interference with or usurpation of an inherently governmental function the "situations in which a state has sought to influence, disrupt or delay democratic decision-making processes in another state through the coordinated use of legal and illegal methods in cyberspace e.g., propaganda, disinformation and covert actions by intelligence services."[256]
- In a 2020 position paper, New Zealand determined that "a prolonged and coordinated cyber disinformation operation that significantly undermines a state's public health efforts during a pandemic" could constitute violation of the non-intervention principle.[257]
- In its 2019 position paper, the Netherlands makes a very brief reference to disinformation, merely warning that the increase in online disinformation activities is conducive to the taking of freedom-destroying policies.[258] The Netherlands added in its 2020 response to the pre-draft report of the OEWG that "The Netherlands is

[253] Scott Shane, "Russia Isn't the Only One Meddling in Elections. We Do It, Too", (17 February 2018), *The New York Times*, <https://www.nytimes.com/2018/02/17/sunday-review/russia-isnt-the-only-one-meddling-in-elections-we-do-it-too.html>.

[254] See also Hollis, van Benthem and Dias, *Supra* note 61, (deploring that "there is an absence of legal rhetoric in the rising use of accusations about IOs.")

[255] See *Official compendium, Supra* note 161.

[256] Federal Department of Foreign Affairs, "Switzerland's position paper on the application of international law in cyberspace" (May 2021) 2-3.

[257] The Government of New Zealand "The Application of International Law to State Activity in Cyberspace" (2020), <https://dpmc.govt.nz/publications/application-international-law-state-activity-cyberspace>.

[258] Government of the Kingdom of the Netherlands, "Appendix: International law in cyberspace" (2019).

appalled by the abuse of the COVID-19 crisis by States to conduct or effectively control non-state actors in launching cyber operations, including the disruption of the healthcare sector, and cyber enabled information operations to interfere with the crisis response in times of urgent crisis. Not only are these operations highly deplorable examples of irresponsible state behaviour; in many instances, they constitute violations of international law."[259]

- More recently in 2021, Germany published a well-documented position paper on the applicability of international law to the cyber-operations. The paper notably suggests that online disinformation may violate the principle of non-intervention; "malicious cyber activities targeting foreign elections may – either individually or as part of a wider campaign involving cyber and non-cyber-related tactics – constitute a wrongful intervention. For example, it is conceivable that a State, by spreading disinformation via the internet, may deliberately incite violent political upheaval, riots and/or civil strife in a foreign country, thereby significantly impeding the orderly conduct of an election and the casting of ballots. Such activities may be comparable in scale and effect to the support of insurgents and may hence be akin to coercion in the above-mentioned sense. A detailed assessment of the individual case would be necessary."[260]
- Finally, Iran in 2020 made explicit reference to online-information based interference, stating that "sending mass messages in a widespread manner to the voters to affect the result of the elections in other states is also considered as the forbidden intervention."[261]

Alongside official papers, states membership in relevant multilateral initiatives may also indicate a particular stance regarding the legal status of cyber-enabled operations. For instance, the Deputy Director-General of the International Legal Affairs Bureau of the Ministry of Foreign Affairs of Japan is a signatory of all five *Oxford Statements on International Law Protections in Cyberspace*. The Oxford Statement on International Law Protections in Cyberspace is a program initiated by several international lawyers to clarify the application of

[259] The Kingdom of the Netherlands' Response to the Pre-draft Report of the OEWG, (2020), §2.

[260] Federal Government of Germany "On the Application of International Law in Cyberspace" (2021), Position Paper, at 5.

[261] Declaration of General Staff of the Armed Forces of the Islamic Republic of Iran Regarding International Law Applicable to the Cyberspace, (August 2020) Translated version retrieved from NATO CCDOE's International Cyber Law in Practice: Interactive Toolkit.

the law to cyber activities. [262] Coincidently, the fourth "Oxford Statement on International Law Protections in Cyberspace" adopted in June 2021 addresses transboundary cyber-enabled information operations, which, as specified in the text of the statement, encompasses cyber disinformation.[263] The fourth statement notably provides in its second point that "[s]tates must refrain from conducting information operations and activities when they would violate the principles of sovereignty and non-intervention in a State's internal or external affairs."[264] The document then makes reference to war propaganda, incitement to discrimination and human rights. While caution should be taken not to overinterpret such commitments, Mr. Mikanagi's support may nonetheless be informative about Japan's official position on the question.

Another useful indicator of state's legal opinion are the publications, studies and reports submitted to and/or adopted by the relevant international organisations. With regard to the European Union, while European institutions and officials have been traditionally cautious about the question of the application of international law to cyber-enabled disinformation campaigns, a noticeable change was the 2021 report on the impact of disinformation on democratic processes and human rights in the world requested by the European Parliament and its Subcommittee on Human Rights, which contains several references to international law concepts; "When the instigator or agent of disinformation is – or has the backing of – a foreign state it may be breaching the public international law principle of non-intervention."[265] The report also indicates; "[f]reedom of thought entails a right not to have one's opinion unknowingly manipulated or involuntarily influenced."[266]

Another example of a legal document adopted by an international institution under the impulsion of the members is the "International Code of Conduct of Information Security" submitted by the Shanghai Cooperation Organization (SCO) to the United Nations in 2015. The Code notably reiterates the application in cyberspace of the right and freedom to seek, receive and impart information of expression, in accordance with the ICCPR article 19, and calls upon states "[n]ot to use information and communications technologies and information and

[262] See "The Oxford Process on International Law Protections in Cyberspace" initiated in 2020 by Oxford Institute for Ethics, Law and Armed Conflict (ELAC), <https://www.elac.ox.ac.uk/research/the-oxford-process-on-international-law-protections-in-cyberspace/>.

[263] "The Oxford Statement on International Law Protections in Cyberspace: The Regulation of Information Operations and Activities".

[264] See "The Oxford Statement on International Law Protections in Cyberspace: The Regulation of Information Operations and Activities", *Supra* note 263.

[265] Colomina, Margalef, and Youngs, *Supra* note 57, at 6.

[266] *Ibid.,* at 10.

communications networks to interfere in the internal affairs of other States or with the aim of undermining their political, economic and social stability."[267]

Several resolutions adopted by the United Nations General Assembly (hereinafter UNGA) touch upon disinformation, specifically in the context of the protection of election processes.[268] A 2019 resolution notably expressed deep concerns that "the spread of disinformation by both internal and external actors on traditional and social media, as well as efforts to manipulate voting systems and the blocking of the Internet and social media in the framework of elections, constitute an increasing problem for democracies around the world" and strongly condemned "any manipulation of election processes, coercion and tampering with vote counts, particularly when done by States, as well as by other actors, and calls upon all Member States to respect the rule of law and the human rights and fundamental freedoms of all persons, including the right to vote and to be elected at genuine periodic elections"[269] In a similar way, the draft resolution On The Role Of States In Countering The Negative Impact Of Disinformation On The Enjoyment And Realization Of Human Rights authored by fourteen states, mostly western, denotes a certain commitment to approach state-sponsored disinformation through the lenses of human rights law.[270]

Finally in 2022, building upon the past works of the G7 and G20 on the matter, the US initiated the "Declaration for the Future of the Internet." The Declaration, which gathers 60 western and like-minded states, singles-out online disinformation as one "state-sponsored or condoned malicious behaviour" that undermines "respect for and protection of human rights and democratic institutions."[271]

The above investigation shows that states have made reference to a number of different rules. It is possible to observe two general tendencies in the way states deal with the issue of the legal status of disinformation in international law. As other authors have remarked, states that share a statist approach to international law through defending a vision of "digital sovereignty" are likely to defend broad interpretations of the principles of non-intervention and

[267] See *Letter dated 9 January 2015 from the Permanent Representatives of China, Kazakhstan, Kyrgyzstan, the Russian Federation, Tajikistan and Uzbekistan to the United Nations addressed to the Secretary-General*, (2015), A/69/723.

[268] See for instance *Strengthening the role of the United Nations in enhancing periodic and genuine elections and the promotion of democratization*, (2017), A/RES/72/164; *Respect for the principles of national sovereignty and non-interference in the internal affairs of States in their electoral processes,* (1989), A/RES/44/147.

[269] UNGA, *Strengthening the Role of the United Nations in Enhancing Periodic and Genuine Elections and the Promotion of Democratization*, (2019), A/RES/74/158.

[270] *Role of States in countering the negative impact of disinformation on the enjoyment and realization of human rights*, (2022) A/HRC/RES/49/21.

[271] Declaration available at <https://www.state.gov/wp-content/uploads/2022/04/Declaration-for-the-Future-for-the-Internet.pdf>

sovereignty when dealing with what they view as inappropriate international communications, which obviously includes foreign disinformation.[272] In contrast, liberal democracies tend to approach the issue in a more restrictive manner, side-lining state sovereignty in favour of human rights, self-determination and the regimes of particularly grave forms of communications such as propaganda of war and incitement to genocide. Other human rights, such as the freedoms of thought and opinion and the right to privacy are also frequently cited in western-sponsored documents and studies dealing with online disinformation.

Be that as it may, concepts such as sovereignty, non-intervention and human rights at large are powerful rhetorical devices. As they may encompass drastically different normative propositions, unspecified references to these principles do not allow for drawing useful conclusion about their legal significance to the present context. It is so, even when they are contained in documents intended to reflect the genuine legal position of their supporters.[273]

Although references to the principles of non-intervention and sovereignty are common and defensible, especially with respect to the protection of election processes against foreign disinformation, the lack of specification regarding *when* and *how* online disinformation effectively violate these principles greatly complicates the issue. In fact, because international law does not apply straight-away to disinformation activities, establishing under which circumstances disinformation campaigns undermine the interest of a state is often an essential prerequisite to the assessment.

None of what states have declared so far regarding foreign disinformation activities is inherently wrong. But insofar as states do not agree upon the same approach, and insofar as all do not share the same interpretation of the relevant rules, it cannot really be said that a particular *opinio juris* is gaining more traction than another. Hence, achieving legal certainty in the context of state-sponsored disinformation, for now remains in the domain of wishful thinking. This makes the treatment of online disinformation activities particularly susceptible of arbitrariness.

To conclude on states' stances about disinformation, it is possible to observe some emerging trends in the way states have approached the issue. In spite of this, legal positions on the status of peacetime disinformation in international law are, if not ambiguous, not (yet) sufficiently stable to constitute a uniform state practice pointing towards a specific approach to the problem.[274] The crux of the matter therefore remains to clarify the circumstances under which

[272] Lahmann, *Supra* note 62, at 209.
[273] See the same point made about the principle of sovereignty at Lahmann, *Supra* note 62, at 91.
[274] Lahmann, *Supra* note 62, at *216.*

disinformation campaigns would fall within the ambit of a given instrument, or principle of international law.

Instigators of online disinformation campaigns excel at taking advantages of the blind spots of the law. Therefore, until a sufficiently widespread *opinion juris* emerges, notably with regards to the legality of cyber-enabled conduct, the lack of clear legal framework will continue to be exploited by states.

However, while there is an obvious need to avoid broadening the scope of the existing thresholds with the sole attention of outlawing disinformation as a whole,[275] it will be shown in the following chapters that there are ways to do better with current international law in terms of adequate legal protection. Of course, the point is not to consider as illegal a disinformation campaign on the sole basis that it may foreseeably produce detrimental effects on political integrity or territorial inviolability. Doing so would unquestionably result in freedom-destroying laws, as well as making the law prone to instrumentalization.[276] States and lawyers must avoid providing legal grounds to malicious actors to condemn legitimate communications. Instead, it is argued in the following chapters that current international law already contains the tools to provide adequate protection to states and individuals against disinformation, and that minimal developments are needed in order to achieve satisfactory legal coverage.

[275] *Ibid.*
[276] See in that sense Falk, *Supra* note 169.

Chapter III.
State-sponsored cyber-disinformation campaigns and the state-centric rules of international law

1. Introduction

This chapter assesses state-sponsored disinformation campaigns from the perspective of the protection accorded by international law to the rights of states. In doing so, the chapter provides a review and an interpretation of the main state-centric rules of international law, and elicits the circumstances under which PSYOPS in general, and state-sponsored cyber disinformation campaigns in general may violate the concerned rules.

This chapter only focusses on state-sponsored communications that have for purpose to underme the rights of other states. Thus, before getting into the hearth of the subject, it is important to understand what the notions of "rights" or "legitimate interests" of states are referring to in the context of this legal analysis.

The rights or legitimate interests of states are connected to the concept of state sovereignty. By virtue of the principle of sovereignty, states bear rights and duties. Some rights and duties of states have been listed by The International Law Commission in a tentative codification of the Rights and Duties of States in 1949.[277] From state sovereignty stem two central negative obligations; the fundamental duty not to violate a state's territorial integrity, (or territorial inviolability), and the fundamental duty not to coercively intervene[278] against a state's political integrity, (or political independence).[279] The first obligation protects the physical dimension of a state, its territory, and the second protects the moral dimension of the state, the collective will.

[277] ILC Draft Declaration on Rights and Duties of States (with commentaries), (1949), adopted by the UNGA in *Draft Declaration on Rights and Duties of States,* (1949), A/RES/375.

[278] *See ICJ*, *Supra* note 196, at 212.

[279] 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, Encyclopedia entries.

Violations of one or the two aspects will simultaneously constitute a violation of state sovereignty. Hence, a state may bear international responsibility for instigating online disinformation campaigns when the conduct is harmful to the target state's right to political independence, and/or territorial integrity.

The first two sections are dedicated to the analysis of important legal instruments that may be of use in setting boundaries to psychological warfare. The 1936 International Convention concerning the Use of Broadcasting in the Cause of Peace is analysed in the first section (2). The second section is dedicated to the 1953 Convention on the Right to Correction, which provides an interesting legal tool to mitigate the effects of false or distorted news reports (3). Then, three sections are dedicated to customary international law. From the most specific regime to the broadest instrument, the chapter contributes an analysis of the principle of non-use and non-threat of force, including the prohibition of war propaganda (4); the principle of non-intervention (5); and the principle of sovereignty (6). As a result of this examination, it is concluded that the legal boundaries set by the prohibition of war propaganda and the principle of non-use and non-threat of force may simply never be exceeded in the context of psychological warfare. Likewise, the broader principles of non-intervention and sovereignty may never be breached by disinformation campaigns alone.

2. The 1936 International Convention concerning the Use of Broadcasting in the Cause of Peace

Post-war developments in terms of the regulation of international communications have drawn substantial inspiration from the oft-cited 1936 International Convention concerning the Use of Broadcasting in the Cause of Peace, (hereinafter the Broadcasting Convention).[280] This League of Nations Treaty is particularly interesting for the subject at hand in that some of its provisions make explicit reference to the transmission of incorrect information abroad. As such, the sub-section will devote some lines to the appreciation of the relevance and practicability of the Broadcasting Convention in the context of state-sponsored online disinformation campaigns.

[280] International Convention concerning the Use of Broadcasting in the Cause of Peace (1936), (17 December 1954), A/RES/841.

The Broadcasting Convention came into effect in 1938 and is still in force today. The Broadcasting Convention was directly inspired by the concept of "moral disarmament." [281] It represents an important segment in the constellation of rules that regulate international communications and more specifically, online disinformation. A scholar has described the Convention as the "strongest anti-propaganda instrument written up to that time."[282] Some of its provisions even directly concern state-sponsored disinformation. In fact, the document is based on the premise that inaccurate and incorrect information are detrimental to good national understanding and may as such lead to interstate frictions. In order to prevent this, the convention lays down several obligations of conduct. Hence, pursuant to the Convention, states have due diligence duties to ensure that harmful communications do not emanate from their territories.

The Convention was adopted in a context of deepening international tensions, just before the second world war. It was also the time when radio waves communications technologies were rapidly developing around the world. As information was now able to travel anywhere without physical restraints, radio communications could finally prove a precious ally to states in the competition for power, which prompted states to discuss regulations on international communications.[283] The United Kingdoms (hereinafter UK), which quickly learned to master propaganda for peacetime power projection in the 1930's, drafted and submitted a proposal to the Council of the International Broadcasting Union in 1933 advocating for an international regulation on "[…] systematic diffusion of programmes or communications which are specially intended for listeners in another country […]."[284]

Regrettably, the international context within which the Convention came into force greatly diverted the attention of states from the document. Indeed, Hitler issued the Godesberg Memorandum the same year the Broadcasting Convention came into force. Claiming control over a portion of the Former Czechoslovakia, Hitler's ultimatum sparked outrage and the fear of an upcoming war among European leaders. The Convention was thus completely eclipsed by the dire security context. The imperialist views of the Nazis and the alleged ignorance of the German public regarding the "true state of affairs" following the German Chancellor's decision to absorb its neighbours convinced the UK, an original signatory of the Broadcasting

[281] Nordenstreng & Seppä, *Supra* note 311, at 16.
[282] Downey, *Supra* note 76, at 344.
[283] *Ibid.*
[284] See Nicholas Pronay & Philip M. Taylor, "'An Improper Use of Broadcasting…'" The British Government and Clandestine Radio Propaganda Operations against Germany during the Munich Crisis and after" (1984) *19(3) Journal of Contemporary History* 357*,* at 361.

Convention, to foment clandestine propaganda operations against Germany in an attempt to diminish domestic support for the Nazi regime, or as brilliantly put by Pronay and Taylor; to "drive a wedge between Volk and Fuhrer."[285] It is interesting to note that, while in the minds of the British, all the broadcasts were about "re-establishing the truth" for German citizens, the broadcasts in question would certainly have breached the Convention.[286]

Quite foreseeably, the convention had very limited success. The reasons behind the Convention's failure to live up to the expectation are obvious. In fact, first problems arise during its first years of existence. The absence of Italy, Germany and Japan among the original signatories of the Treaty was a major let-down because these states were pushing an imperialistic agenda and made extensive use of deceitful propaganda in support of their policies at the time.[287] While Japan and Germany left the League of Nations before the signature of the Convention in 1933, Italy did not sign the Convention and left the organisation in 1937. But even after the downfall of the League of Nations, many powerful states did not ratify the convention. The US, which were reluctant to exercise any form of control over the American private broadcasting companies declined to ratify the convention.[288] Alongside other western states, they maintained this posture consistently throughout the years, in the many multilateral *fora* dealing with the regulation of international communications.

Today, around thirty states are signatories to the Convention. The Russian Federation as well as several European NATO member states are still parties. The Convention has been denounced by four of its original signatories, France, Australia, Netherland and the UK in the 80's, allegedly in reaction to USSR's accession, out of concern that the latter may call upon the Convention to take measures against their own emissions.[289] So far, the Convention has never been invoked by the parties, and the attempts of the UNGA to revive and update the Broadcasting Convention in 1954 failed to get enough traction from member states.[290] Two main causes for this inertia are significant enough to be detailed below.

First, the scope of the Convention remains relatively unclear when it comes to international communications that are not already illegal under customary international law. Indeed, although

[285] *Ibid.,* at 358.
[286] *Ibid.,* at 377.
[287] "Reich and Japan Shun Radio Treaty; They Are Only Great Powers Remaining Silent on the Proposed League Pact. Its aim is to aid peace United States Says It Cannot Join Convention but Will Seek to Carry Out Principles" (27 March 1936) *TheNewYorkTimes*, <https://www.nytimes.com/1936/03/27/archives/reich-and-japan-shun-radio-treaty-they-are-only-great-powers.html>.
[288] Von Glahn, *Supra* note 5, at 567; Whitton & Larson, *Supra* note 60, at 70.
[289] This was not the case of France, which advanced the reasons that its laws already provided sufficient protection against foreign communications regulated by the Convention. See also in that sense Baade, *Supra* note 148.
[290] Broadcasting Convention, *Supra* note 280; Baade, *Supra* note 148.

the Broadcasting Convention merely reaffirmed the then-emerging international consensus pertaining to the inadmissibility of war propaganda and revolutionary propaganda, the third article is slightly more obscure than the other. Article 3 of the Broadcasting Convention indeed provides, *inter alia*, "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."[291] Most disinformation campaigns would easily fall under the scope of this provision. However, it is unclear how this provision would operate in practice. Obvious difficulties might indeed arise regarding what exactly ought to be understood as "incorrect statements" falling under the scope of article 3, and there is no doubt that states would disagree on that matter.[292] With that being said, disinformation campaigns may sometimes simultaneously constitute war propaganda and revolutionary propaganda. If this is the case, the disinformation campaign in question would unquestionably fall under the scope of the Convention.[293]

Second, reservations to the compromissory clause of the Broadcasting Convention have weakened its force. More specifically, its article 7 provides a compromissory clause that attributes compulsory jurisdiction to the ICJ (replacing the CPIJ) or an arbitral tribunal to decide upon a dispute arising between the parties regarding the interpretation of the treaty that was not successfully settled through diplomatic channels. However, several of its signatories entered a reservation to the compromissory clause upon ratification. In doing they rejected the possibility that disputes are brought before the ICJ unilaterally, by one single party.[294] The reservations

[291] Broadcasting Convention, *Supra* note 280, article 2.

[292] See notably, Baade, *Supra* note 148.

[293] It bears noting that disinformation campaigns amounting to war propaganda or subversive propaganda also violate customary international law. This point will be elaborated later in the study.

[294]See notably "Ratification of the International Convention Concerning the use of broadcasting for the cause of peace" (23 September 1936) deposited on the 3rd February 1983 by the Soviet Union, Reservation to Article 7, 1299 *United Union Treaty Series*, C-4319; Declaration of the United Kingdom of Great Britain and Northern Ireland with respect to the reservation and declaration made by the Union of Soviet Republics upon ratification, received on 9 December 1983, 1342 *United Union Treaty Series*, C-4319; Accession and reservation to the International Convention Concerning the use of broadcasting for the cause of peace (23 September 1936) of the People's Republic of Bulgaria, instrument deposited on 17 May 1972; Objections to the reservations made by the Government of Bulgaria in its instrument of accession to the above-mentioned Convention by the United Kingdom of Great Britain and Northern Ireland, notification received on 12 May 1972, *United Nations Treaty Series*. ; Objection of United Kingdom of Great Britain to the reservation and declaration made by the German Democratic Republic upon accession received on 5 December 1984, 1381 *United Nation Treaty Series*, C-4319; See also a recent discussion on the convention and its applicability in the context of Russia's propaganda effort regarding Ukraine; Talita De Souza Dias "Russia's "genocide disinformation" and war propaganda are breaches of the International Convention Concerning the Use of Broadcasting in the Cause of Peace and fall within the ICJ's jurisdiction" (March 2022), *EjilTalk*, <https://www.ejiltalk.org/russias-genocide-disinformation-and-war-propaganda-are-breaches-of-the-international-convention-concerning-the-use-of-broadcasting-in-the-cause-of-peace-and-fall-within-the/>.

made to Article 7 were opposed by UK and other western states. A prominent reason for these reservations was the outrageous result that these reservations would produce: an aggrieved party is left with no means to ensure the respect of the Convention against the reserving states in case of a contravention.

3. The 1953 Convention on the International Right of Correction

Another document adjoined in the above-cited Final Act of the 1948 Conference on Freedom of Information requires a separate mention here. It is the draft Convention Concerning the Institution of an International Right of Correction, a French-sponsored project that culminated in with the adoption of the Convention on the International Right of Correction in 1953, (hereinafter the Correction Convention). The Correction Convention establishes an international right of correction.[295] As it precisely applies to incorrect communications, this Convention might eventually be of use in the context of state-sponsored disinformation.[296] Its approach to disinformation was particularly innovative for its time. Indeed, the text fully embraces the concept of information plurality, inviting states to publicize the views of the disagreeing state rather than empowering them to put an end to contentious communications. It is then up to the individuals to make up their minds regarding the quality of the problematic communications.

The Correction Convention has been in force since 1962 and has so far been ratified by 17 states. From the outset the Correction Convention proves a particularly attractive mechanism for third world countries, which tend to be concerned about the fact that the narratives of a few major powers with better information systems could easily dominate the information spaces of the weaker states that do not necessarily possess the same capabilities in terms of information media and infrastructures.[297] Consequently the overwhelming majority of the parties of the Convention, with the notable exception of France, the original sponsor of the Convention, are South-American and African countries.

Article 2 of the Convention entitles a contracting party that deemed a news dispatch "capable of injuring its relations with other state or its national prestige or dignity transmitted

[295] Convention on the International Right of Correction, (1952) A/RES/630(VII). The term "correction" here might be a misnomer. As Baade indeed underlines, it is in fact a right to reply that the convention provides, not a right to correction *stricto-sensu*. Baade *Supra* note 148.

[296] See another analysis of the relevancy of the Correction Convention in the context of cyber-enabled disinformation activities at Baade, *Supra* note 148.

[297] Downey, *Supra* note 76, at 349.

from one country to another by correspondents or information agencies of a Contracting or non-contracting State and published or disseminated abroad is false or distorted," to "submit its version of the facts (communiqué) to the Contracting States within whose territories such dispatch has been published or disseminated." Article 3 stipulates that the recipient state has five days from the day of receiving the communiqué to transfer the *communiqué* to the information agency responsible for the dispatch for eventual correction. However, the usefulness of the Correction Convention is very limited in the context of state-sponsored disinformation campaigns.[298] Firstly, the convention only counts 17 states parties. Secondly, the convention applies only to the communications made by a foreign agency and intended for dissemination in the territory of a third contracting state. Hence, the most common situations of a disinformation campaigns intended for the disinformant's own population and negatively affecting the reputation of a foreign state, and a disinformation campaign intended to tarnish the reputation of a state in the eyes of its own population fall outside the scope of the Convention.[299]

However, the Convention would still apply to a disinformation campaign intended to be read by a foreign audience but having for purpose the defamation of a third state party to the convention. Take the example of State A targeting the population of State B with disinformation about State C with the aim of denigrating the policies of a State C. In that case, provided that States C and B are parties to the convention, State C could send a *communiqué* of its version of the facts to State B for domestic dissemination. This would for example apply to the disinformation campaign allegedly conducted by Russian agents in Mali intended to tarnish the reputation of the French forces engaged in anti-terrorist activities in the eyes of the local populations.[300] Comically, this would also apply to the French disinformation operations deployed in the same time to counter the Russian influence in Mali. Regrettably, while France is party to the convention, Mali and Russia are not. Assuming that Mali was a party and its news agencies reported the contentious facts, it would have been obliged to disseminate the French version of the facts within five days if France made the demand. In absence of action from its interlocutor, France would have been entitled to transmit its communiqué to the Secretary General of the UN for publicity, thus exposing the campaign.[301]

[298] See Baade, *Supra* note 148.
[299] Baade, *Supra* note 148.
[300] For a detailed explanation of the facts, see Gleicher, *Supra* note 42.
[301] Correction Convention, *Supra* note 295, Article 3

One last potential hurdle may render the application of the Convention to modern malign influence activities more uncertain. It is the fact that the sources that relay online disinformation are often private individuals assuming no press-related activities whatsoever. Hence, as the Convention applies only to problematic communications disseminated by an information agency or a correspondent which is "regularly engaged in the collection and the reporting of news material, and who when outside its State is identified as a correspondent by a valid passport or by a similar document internationally acceptable", their dissemination efforts will systematically fall outside its scope.[302] Notwithstanding, heavy disinformation spreaders often conceal their true identity online.

The previous two sections looked at treaty law in order to establish whether existing international instruments provide adequate protection to the interests of a state victim of state-sponsored disinformation campaigns. From now, the chapter turns to the foundational principles of customary laws safeguarding the rights of a state.

4. State-sponsored cyber-disinformation campaigns as uses or threats of the use of force

Militaristic terms and misnomers are frequently employed when it comes to referring to politically-driven information manipulation activities. Some have compared the information to a "weapon" [303] others have campaigned for the inclusion of certain international communications in the definition of aggression. [304] Yet, with regards to international communications, terms such as "war" or "weapon" do not necessarily corroborate a reality. The fundamental concept of violence lying at the hearth of both notions of "armed attack" and, at a lower extent, the notion of "force," do not translate well in the context of cyber-enabled disinformation.

Contrarily to what the wording appears to suggest, "information war" and "psychological war" do not relate to a situation of armed conflict, nor does the term imply the use of forceful activities. Although information manipulation, control and denial activities may sometimes be waged in the context of conventional warfare, they do not constitute violent activities on their

[302] Correction Convention, *Supra* note 295, Definition of a "correspondent", at 3.
[303] See for instance Parly, *Supra* note 33.
[304] See *travaux preparatoires* of the *Draft Resolution Concerning Aggression*, (Definition of Aggression A/RES/29/3314 (14 December 1974)), Draft Resolution Submitted by the Union of Soviet Socialist Republics, (A/AC.66/L.2/Rev.1).

own. Information war and psychological war in particular describes a state of competition that is "not warfare in the Clausewitzian sense (nor in any sense presently recognized under the laws of war or armed conflict.)"[305] On this point, commentators have observed "[F]or the West, war is the opposite of peace, and information warfare is seen as part of a response to a conflict situation, specific in time and defined in goals. In Russia and China, information warfare is seen as a permanent activity which is to be practised regardless of the absence of immediate conflicts, open-ended and widely applicable."[306]

Bet that as it may, the preservation of peace and good understanding between nations is intrinsically linked to the question of the regulation of international communications during peacetime. This begs the question, as to whether state-sponsored disinformation campaigns could ever constitute unlawful uses or threats of force, whether the conduct could ever receive the "armed attack" treatment, and whether state-sponsored disinformation could constitute warmongering propaganda.

4.1 International communications and the maintenance of international security and peace

The non-use of force principle is a customary rule of international law. It is so important that many authors consider it to be a peremptory norm of international law (*Jus Cogens*), understood as "[…] a norm accepted and recognized by the international community of States as a whole as a norm from which no derogation is permitted and which can be modified only by a subsequent norm of general international law having the same character."[307] The prohibition of uses or threat to use force is enshrined in Article 2§4 of the UN Charter, which prohibits the "[…] the threat or use of force against the territorial integrity or political independence of any state, […]."[308] The prohibition therefore protects a state from the most daring conduct, -those that involve *force*. As such, the principle operates simultaneously, as a safeguard to the integrity of a state and its independence.

Throughout history, states were mostly concerned with propaganda campaigns waged with the aim of causing harm or disruption abroad, or propaganda of the nature to undermine "good understanding between nations." These were bellicose transmissions that would accurately be

[305] Lin & Kerr, *Supra* note 14, at 4.
[306] Ben Nimmo, cited in "Information at War: From China's Three Warfares to NATO's Narratives" (2015), Legatum Institute, Beyond Propaganda series, <https://li.com/reports/information-at-war-from-chinas-three-warfares-to-natos-narratives/>, at 3;
[307] Vienna Convention on the Law of Treaties (1969), Article 53.
[308] UN Charter, Article 2§4.

defined as threats to the peace likely to fall under the scope of the prohibition to use or threaten force.

While disinformation was common in the context of bellicose transmissions,[309] the legal status of said communications hinged on other characteristics than the levels of deception involved. Therefore, from a legal perspective, the distinction between persuasion *via* white propaganda and deception *via* disinformation, black propaganda or grey propaganda was of little importance, at least to the extent that bellicose propaganda susceptive of endangering the peace were concerned.

A considerable amount of effort has been deployed to prevent wars between nations. This entailed, among other, setting fairly generous yet strict limitations to what could be said by a state, and communicated by national media. The question of the limits to be set to communications in the context of the relations between states was for example a central topic in the exchanges that led to the conclusion of the 1933 *Roosevelt-Litvinov Agreements,* through which the US and the young Soviet Union officially established diplomatic relations. After heated negotiations, the representative of the Soviet Union agreed to refrain or restrict all person "from any act tending to incite or encourage armed intervention, or any agitation or propaganda having as an aim, the violation of the territorial integrity of the United States, its territories or possessions, or the bringing about by force of a change in the political or social order of the whole or any part of the United States, its territories or possessions."[310] This declaration in particular related to conflict-prevention between the two nations. It was essentially an insurance that the Soviet Union would not start a war against its new partner, or do anything that could lead to a war between the two nations, including the transmission of subversive or bellicose messages into the territory of the other party.

The work of the League of Nations in the 1930's is also worth mentioning for its focus on international communication regulation in the context of conflict-prevention. In 1925, multiple states emphasized the importance of the role of the press in bringing about the necessary conditions for international peace at the occasion of the 12th plenary session of the League of Nations Assembly.[311] The initiative spurred the creation of international conferences dedicated

[309] See notably Whitton & Larson, *Supra* note 60, at 104-132.

[310] "Exchange of Communications Between the President of the United States and Maxim M. Litvinov People's Commissar for Foreign Affairs of the Union of Soviet Socialist Republics" (1934) 28(1) *The American Journal of International Law* 2, at 3.

[311] For a detailed account of the discussions held about mass media and their role in enabling international peace under the auspices of the League of Nations, see Kaarle Nordenstreng & Tarja Seppä, "The league of nations and the mass media: The Rediscovery of a Forgotten Story", Transcript of a presentation made at the XV Conference of the International Association for Mass Communication Research (IAMCR/AIERI), Section of

to the study of the role of mass media in international relations. In its final resolutions, the first Conference of Press Experts underlined the need to mitigate the distribution of "Tendencious News" including "obviously inaccurate, highly exaggerated, or deliberately distorted news or articles." Another resolution focussed on "moral disarmament", and the problem of bellicose transmissions.[312] The use of accurate information to counter the nefarious effects of the transboundary spread of falsehood, the value of an international right of correction and the importance of information pluralism were mentioned in subsequent Conferences and Committees. The concept of "moral disarmament" was further considered at the occasion of 1932 The Conference for the Reduction and Limitation of Armaments. As commentators noted, "[T]he aim of moral disarmament was not merely to disarm men's minds but to transform them with a view to establishing a firm psychological basis for the future development of the international community."[313] The rapid degradation of the international context at the time would regrettably stall the discussions. The League would still manage, in its last moments, to work out the Broadcasting Convention, which has previously been analysed in this chapter.

Efforts to regulate warmongering communications resumed shortly after the second world war. The 1947 Resolution titled *Measures To Be taken Against Propaganda And The Inciters Of A New War* (Resolution 110 (II)) requested states to "encourage the dissemination of all information designed to give expression to the undoubted desire of all peoples for peace."[314] In 1951, the UNGA condemned, alongside "incitements to conflicts or acts of aggression", "measures tending to isolate the peoples from any contact with the outside world, by preventing the press, radio and other media of communication from reporting international events, and thus hindering mutual comprehension and understanding between peoples." [315]

As the Cold War went on, the strong ideological polarizations of the time prevented any substantial developments in the direction of the development of norms and standards for the conduct of harmful transboundary communications. The main protagonists, the US and the USSR systematically disagreed about the very idea of regulating speech.[316] As a result, only a

---

International Communication session on "Communication and Peace; The Role of the Media in International Relations", (1986).

[312] League of Nations document Conf. E.P.13 (Geneva, 25 October 1927) cited in *Ibid.,* at 6.

[313] Nordenstreng & Seppä, *Supra* note 311, at 13.

[314] *Measures to be taken against propaganda and the inciters of a new war*, (1947), A/RES/2/110; Article 4 of the 1966 Tashkent Declaration signed between India and Pakistan is an example of application of the Resolution. Article 4 encourages the deployment of positive propaganda between the two parties; "both sides will discourage any propaganda directed against the other country, and will encourage propaganda which promotes the development of friendly relations between the two countries."

[315] *Condemnation of Propaganda Against Peace*, A/RES/381(V), (1951).

[316] See for instance, Metzl, *Supra* note 180.

handful of multilateral treaties dating from the cold war period contains a reference to internationally harmful communications.

Article 19 of the 1982 United Nations Convention on the Law of the Sea (hereinafter UNCLOS) for instance provides, *inter alia*, that passage of a foreign ship shall be considered to be prejudicial to the peace, good order or security of the coastal State if in the territorial sea it engages in "any act of propaganda aimed at affecting the defence or security of the coastal State."[317] The UNCLOS does not provide a definition of what exact type of propaganda falls within the scope of Article 19. It is however more than probable that revolutionary propaganda, war propaganda or any international communications advocating for hate and violence towards the members of a group and beamed in the territory of the coastal state from a foreign ship during its passage would qualify. State-sponsored disinformation campaigns displaying the requisite characteristics of revolutionary, hate or war propaganda would therefore qualify as an act of propaganda affecting the security of the coastal state in the sense of UNCLOS Article 19 if disseminated by foreign ships engaging in safe passage.

Finally, a mention to Resolution 110 (II) was made in the preamble of the 1967 Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (Outer Space Treaty).[318] Several non-aligned states and Eastern powers advocated for the inclusion of a specific provision prohibiting bellicose propaganda *via* satellite during the drafting of the treaty. The drafters ended up including only a mention of the above-cited resolution in the preamble. As a result, the use of satellites for the purpose of inciting wars or genocide is not expressly prohibited by the Treaty.[319] With that being said, it is common sense that communication satellite should be used exclusively to promote peace between nations.

With no mention of disinformation campaigns in past work on the regulation of international communications, it is left to international lawyers to determine whether and under what circumstances disinformation campaigns would constitute threats to the peace susceptible of engaging the responsibility of the instigator on the basis of Article 2§4 of the UN Charter.

[317] UN Convention on the Law of the Sea, (1982), at article 19, §d.

[318] Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies, (1967).

[319] That is not to say that the customary prohibitions of war and genocide incitements would not apply to communications distributed *via* satellites.

4.2 Disinformation as illegal "force"

Article 2§4 of the UN Charter provides that "All members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations."

While it makes no doubt that the non-use of force principle applies to military or armed force, that is, the use or the threat of deployment of troops, kinetic energy weapons and other military equipment, it is now well established that some conduct not involving conventional kinetic energy weapons or military personnel are still capable of breaching the prohibition.[320] In that sense, some states and commentators have advocated for an effect-based approach to the use of force.[321] Such a view was for example endorsed by the experts of the Tallinn Manuals 2.0; "In the cyber context, it is not the instrument used that determines whether the use of force threshold has been crossed, but rather, [**…**], the consequences of the operation and its surrounding circumstances."[322] Accordingly, the experts advance the criteria of scale and effect of a conduct in order to determine whether the thresholds of the principle are being crossed.[323] Hence, the prohibition is breached when "there is some element of armed force involved, or at least actions resulting in physical injury or damage."[324]

Disinformation campaigns are non-destructive and non-deadly by nature. That is to say, that disinformation does not *directly* cause death or injury. Therefore, the systematic dissemination of news or reports in the territory of a state by another, be it motivated by particularly harmful intentions, *in principle* cannot constitute an unlawful use of force. The two Tallinn Manuals for instance provide in their definitions of the use of force that "non-destructive cyber

[320] See for example Russel Buchan, "Cyber Attacks: Unlawful Uses of Force or Prohibited Interventions?" (2012) 17(2) *Journal of Conflict and Security Law* 212. See also Schmitt (2017), *Supra* note 162, Rule 68, at 162.; ICJ, *Supra* note 196. §228; *Legality of the Threat or Use of nuclear weapons, (hereinafter Nuclear Weapon advisory*), (1996), ICJ, Advisory Opinion., §39.

[321] Ian Brownlie, *International Law and the Use of Force by States* (UK, Oxford University Press 1963). at 362. Harold Hongju Koh, "International Law in Cyberspace" Remarks as Prepared for Delivery by to the USCYBERCOM Inter-Agency Legal Conference Ft. Meade, MD, (18 September 2012) The Remarks are available online, < https://2009-2017.state.gov/s/l/releases/remarks/197924.htm>.

[322] Schmitt (2017), *Supra* note 162, Rule 69. at 328. See also Foltz, *Supra* note 214. See also Michael N. Schmitt "The Use of Cyber Force and International Law" (2015) in. Marc Weller (ed) *The Oxford Handbook of the Use of Force in International Law* (UK, Oxford University Press 2015), at 5.

[323] Schmitt (2017), *Supra* note 162, at 331.

[324] Gary Corn, "Coronavirus Disinformation and the Need for States to Shore Up International Law" (2020), *Lawfare,* <https://www.lawfareblog.com/coronavirus-disinformation-and-need-states-shore-international-law>. See also Henning Lahmann "Protecting the global information space in times of armed conflict" (2020) 102 (915) *International Review of the Red Cross* 1227, at 1241.

psychological operations intended solely to undermine confidence in a government or economy do not qualify as uses of force."[325]

Yet, Schmitt and Milanovic argued in a recent paper that non-violent conduct *per se* to which multiple deaths "are the foreseeable effect of the cyber operation would rise to the level of a use of force." [326] This opens up the possibility consider PSYOPS as violations of the prohibition to use force when the communications could have foreseeably caused consequences of similar level than what could have been achieved through forceful means. Both the experts of the Tallinn Manuals and Milanovic note that 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 a number of factors, including *inter alia* "the severity of the consequences, the invasiveness of the operation, the measurability of the effects, the causal directness of the operation, and the entity that mounted the operation."[327] For example, it might be relatively easy to demonstrate that the act of disseminating massively inaccurate information with regard to the cure of a virus during a pandemic, or false information with regard to first aid measures during natural disaster or terrorist attacks could foreseeably provoke deaths. Schmitt and Milanovic for example explain that a cyber-operation causing significant increase of virus-related cases and deaths could satisfy the use of force threshold, provided that the malicious motive behind the conduct can be inferred and that the operation in question is of sufficiently high magnitude to be comparable to non-cyber uses of force.[328]

Furthermore, some states and scholars also understand the use of force prohibition as extending to acts that, while not resulting in death and destruction, generate disruptions of the level of what can be obtained through conventional force.[329] For instance, Roscini notes that "[a] week long cyber-attack that shuts down the national grid, and thus leaves millions of people without electricity, cripples the financial market and the transport system, and prevents government communications is likely to be treated as a use of force, whether or not physical damage ensues."[330] Schmitt similarly argues that cyber operations causing substantial loss of functionality to the infrastructures of the target state would qualify, the crux of the issue being

[325] Schmitt (2013), *Supra* note 162, Rule 11; Schmitt (2017), *Supra* note 162, Rule 69, at 331.

[326] Michael Schmitt & Marko Milanovic, "Cyber Attacks and Cyber (Mis)information Operations During a Pandemic" (2020), 11 *Journal of National Security Law & Policy* 247, at 259.

[327] *Ibid.* see also Schmitt (2017), *Supra* note 162, at 334-35

[328] Schmitt & Milanovic, *Supra* note 326, at 258.

[329] This view is for example shared by the French government. See notably French Ministry of the Armed Forces, *Supra* note 160.

[330] Roscini, *Supra* note 161, at 62.

for the defendant to be able to prove that a similar level of disruption could have been achieved through conventional force.

In light of this, state-sponsored disinformation campaigns, albeit non-forcible and non-deadly *per se*, could still breach the non-use of force principle if the disinformation has contributed in a major way to significant disruptions, including deaths and injury, 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.[331]

In normal circumstances, however, the use of force thresholds may never realistically be crossed by a disinformation campaign. There are very few imaginable scenarios where a disinformation campaign may satisfy the requirements of the use of force prohibition and even in situations when the access to reliable information is literally a matter of life and death for the audience, other requirements will have to be fulfilled. If mere nuisance will surely not suffice, it appears as implausible that a disinformation campaign causing unquantifiable damages, or merely contributing to non-foreseeable damage, will ever cross the use of force threshold. This thus excludes from the scope of the prohibition the most common protracted disinformation campaigns aimed at sowing doubts, distrust and discontent among the audience without explicitly calling for civil strife.

In general, the wise position is to set the threshold of causality very high for all adverse events happening within a states' territory to which no unique contributing cause can be singled out. All in all, as Roscini recalls "[i]t cannot be overemphasized that it is only cyber-attacks that go beyond mere inconvenience and significantly disrupt the functioning of critical infrastructure that can potentially fall under the scope of Article 2§4."[332] Thus, states should be cautioned not to overestimate the disruptive potential of disinformation campaigns.

It is anyway unlikely that states would go as far as to denounce a disinformation campaign, be it a particularly debilitating one, as an illegal use of force. On this point an author recently noted, "The worry of counting CDOs as 'use of force,' other than it is counter-intuitive to the plain language, is that the effect of CDOs [cyber disinformation operations] might be disproportionate to that of belligerent military actions (the traditional understanding of 'use of force')."[333]

[331] Schmitt & Milanovic, *Supra* note 326, at 260.
[332] Roscini, *Supra* note 161. at 62- 63.
[333] Zhao, *Supra* note 117. At 47.

4.3 Disinformation as illegal threats to use force

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 violate its territorial integrity. However, they may still violate the political independence of the target state if they contain threats to use force, which arguably, is a way more plausible scenario.[334]

It is evident that communications campaigns containing explicit threats to use force would contravene article 2§4 of the UN Charter. This is so, regardless of the nature of the content. Whether the communication incorporates disinformation elements or merely contains facts does not matter, it is the utterance of credible threats to use force that matters. Hypothetical instances of prohibited threats to use force may arise 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, or when disinformation is used as a bait tactic. Take for example a state falsely claiming to be about to use force against another, or purposefully provoking the target via false declarations in order to cause the latter to make a first move that the instigator may later use as an excuse for engaging in an apparently lawful use of defensive force.

4.4 Disinformation and the concept of "armed attack"

A discussion on the use of force prohibition would not be complete without mentioning the question of "armed attack." An armed attack is commonly considered "the most grave forms of the use of force."[335] The term of "aggression" is sometimes employed to refer to an "armed attack."[336] An armed attack, (or an aggression), is the starting point of an armed conflict and as such, the condition *sine qua non* for the exercise of the right of self-defence pursuant to Article 51 of the UN charter.[337] Hence a state that has fallen victim of such a grave conduct is legally entitled to use kinetic force to secure its interests and put an end to the illegal act.

[334] Duncan B. Hollis "The Influence of War; The War for Influence" (2018) 32(1) *Temple International & Comparative Law Journal* 31, at 39.

[335] *Case Concerning Oil Platforms* (Islamic Republic of Iran v. United States of America), (2003), ICJ, judgement; Karl Zemanek, "Armed attack" (2013) Max Planck Encyclopedias of International Law, Encyclopedia entries.

[336] See notably *Definition of Aggression, (1974), A/RES/29/3314.*

[337] See ICJ, *Supra* note 196, at 191.

As an armed attack implies the use of conventional force, disinformation campaigns, be they particularly invasive in the victim's information space, will not in themselves amount to an aggression justifying the use of force by the victim in response.[338]

With that being said, the Soviet Union has on several occasions since 1953, attempted to introduce the rather obscure notion of "ideological aggression" during the discussions on the definition of aggression at the UN.[339] A similar unsuccessful attempt to incorporate subversive interventions through the concept of "indirect aggression" under the definition of aggression was made. [340] Both notions were broad enough to include undesirable international communications directed at a state.[341] It is probable that disinformation campaigns targeting foreign audiences would have qualified if of sufficiently high-scale to be noticed by the recipient state.

Adopting the concepts of ideological or indirect aggression could have meant that the recipient state is legally entitled to forcibly respond to some particularly violent episodes of foreign propaganda. As underlined by the French and Sweden delegations at the time of the negotiations, the concept of "ideological aggression" is so broad and subjective that its recognition would contradict the very purposes of the Charter. The two aforementioned concepts were subsequently rejected by the majority.

In spite of this, some authors have argued that certain state-led communication campaigns could justify the use of force in reprisal. Glahn for example considered that the use of force in defence of propaganda would be justified "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."[342] Whitton and Larson similarly argued that "more forceful measures of defense or even retaliation must be expected if the interference takes the form of psychological warfare threatening the very existence of the government and perhaps the state itself."[343]

Although maintaining the possibility to act forcibly in self-defence against disinformation activities of interventionist nature would certainly have strong deterrence effect, it is unclear when and how online disinformation could invite military response in practice.[344] Also, as US

[338] Quincy Wright, "Subversive Intervention" (1960) 54(3) *The American Journal of International Law* 521.

[339] Draft Resolution on the Definition of Aggression Submitted by the Union of Soviet Socialist Republics, *Supra* note 304.

[340] Wright, *Supra* note 338.

[341] Russia for example argued that the term would encompass communications that "Promotes the propagation of fascist-nazi views, of racial and national exclusiveness, and of hatred and contempt for other peoples" see *Ibid.,* submission by the Union of Soviet Socialist Republics, A/C.6/L.332/Rev.118 (1954).

[342] Glahn, *Supra* note 5. At 583.

[343] Whitton & Larson, *Supra* note 60. At 84.

[344] One imaginable instance where forceful retaliation to a communication campaign could be justified is when a state engages in war propaganda activities with the aim of justifying an imminent aggression against another

Major Kinslow recently argued, there is a fundamental difference between trying to obtain a more favourable position in world politics through espionage, data thief or disinformation, and "hacking to undermine the game itself", for example, through hacking voting machines in order to change the vote counts, the latter being of such a gravity in terms of damage to the political

state. Without going into the details of the question of self-defence in case of an imminent armed attack, it is necessary to clarify that self-defence will not be allowed in absence of abnormal military display in the territory of the disinformant suggesting an upcoming aggression. The determination of whether an armed attack is about to occur being subjected to a criterion of reasonableness.

Another legal mechanism that may apply in that situation is necessity. Accordingly, a state is entitled to resort to the plea of necessity as an exceptional circumstance precluding the wrongfulness of his subsequent actions if this is "the only way for the State to safeguard an essential interest against a grave and imminent peril." See ARSIWA, *Supra* note 81. Article 25 "necessity." It matters not in that case to legally attribute the communications to a state. The imperilled state can take action against the immediate source of the peril. (See Lahmann, *Supra* note 92. "Necessity" 201-258.)

However, while this seems appropriate in a context where legal attribution of international communications is complicated by the use of proxies, or other private auxiliaries, outside the very rare case of war propaganda announcing an imminent aggression, it is hardly conceivable that even the most bellicose language employed in international communications could give raise to a situation of necessity posing "a grave and imminent peril" to its preservation for the victim state. Another uncertainty relates to the possibility for an imperilled state to take down satellites or undersea cables for the purpose of interrupting the communications that pose an imminent peril to its preservation. The obvious reason is that such actions risk inadvertently affecting the interests of other states and as such, may not satisfy the proportionality requirement. See ARSIWA, commentaries, *Supra* note 81. at 84.

At any rate, with regard to forceful measures of preservation, the International Law Commission has been reluctant to recognize the possibility for a state to use force on the basis of the plea of necessity as a means of self-preservation. The law seems unsettled at this point. See ARSIWA, commentaries, *Supra* note 81. at 84. at least it can be noted that retaliation *in kind*, in the form of direct calls for revolt by the affected state addressed to the population of the state that engaged in similar activities was not a rare practice during the cold war. (See Whitton & Larson, *Supra* note 60.)

Aside from self-defence and necessity, and pursuant to Articles 34 and 35 of the *Constitution and Convention of the International Telecommunication Union*, states are able to stop or suspend transmissions that "may appear dangerous to the security of the State or contrary to its laws, to public order or to decency" See *Constitution and Convention of the International Telecommunication Union* (1992); See also *International Radiotelegraph Convention* 84 LNTS 97. This may be done at the technical level and does not require physical destruction of assets. Yet again, such measures may contradict other obligations owned by the state, notably the human rights obligations related to the preservation of freedom of speech and opinion enshrined in the ICCPR. (See in that sense Lahmann, *Supra* note 92, at 203-204.)

With regard to human rights considerations, article 19 provides exceptions to the freedom of expression, which "shall only be such as are provided by law and are necessary: (a) For respect of the rights or reputations of others; (b) For the protection of national security or of public order (*ordre public*), or of public health or morals." Article 20 of the ICCPR provides "1. Any propaganda for war shall be prohibited by law.; 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law." In the context of particularly bellicose international communications, chances are that the incriminated communications fall under the scope of the cited articles. Hence, when a communication displays clear aspects of war propaganda or hate speech, the affected state is entitled to act online to ban the concerned information sources. (See ICCPR, *Supra* note 118.)

Less drastic responses are imaginable in face of less grave violations that do not involve bellicose or hate rhetoric. In this most likely situation, Baade suggests "[W]hen false news violates the principle of non-intervention, its correction and acknowledgment as false could be demanded as restitution and satisfaction" (Baade, *Supra* note 148.) Other soft means of responses such as publicising the disinformation campaign and the formal attribution of the act in question, when it does not amount to a clear-cut breach of international law but nonetheless impede on the interests of the affected state, may be envisaged.

independence of a state (as it is basically choosing the leadership of the recipient *in lieu* of the legitimate electorate), that the self-defence argument would make sense.[345]

To conclude on the use of force, it should be cautioned that, as for the notion of "ideological aggression," opening the possibility to respond forcibly to international communications is an untenable approach to modern international communications landscape. Such an uncompromising attitude towards disinformation activities, be they particularly intrusive into the politics of the target state, could only be justified if it was not for the existence of a principle of non-intervention, which extends beyond uses of force *stricto-sensu* and as such, constitutes a more appropriate framework to apply to state-sponsored disinformation campaigns.

4.5 Remarks on international humanitarian law and the concept of "attack"

Although the present work addresses state-led disinformation activities waged during peacetime, the question of disinformation campaigns targeting foreign populations in the context of an armed conflict deserves a few comments.

The 1977 additional Protocols to the Geneva Conventions relating to the Protection of Victims of International Armed Conflicts define an armed conflict as the situation in which "[…] one or more States have recourse to armed force against another State."[346] International humanitarian law regulates the conduct of war. As such, it protects the persons that are not participating in hostilities.

A state conducting an "attack" in the humanitarian law sense of the term is subject to an obligation of precaution, proportionality and distinction. The use of terms such as "methods or means of warfare," "weapons, projectiles and material," "armed force," suggests that the Geneva Conventions are not intended to regulate the coordinated release of problematic information by a state with a view to influence or manipulate the opponent or its population. In that sense, it is important to recall that, while disinformation campaigns are not constitutive of armed attacks in themselves, they may still be waged by states alongside conventional means of warfare in the context of an ongoing armed conflict.

In the particular context of an armed conflict raging between states, both the Geneva Conventions, and most military manuals treat disinformation as permissible ruse of war, as

[345] Christopher H. Kinslow, "Game of Code: The Use of Force Against Political Independence in the Cyber Age" (2019) July-Aug. 2018 *Army Lawyer* 28, at 33.

[346] "How is the Term 'Armed Conflict' Defined in International Humanitarian Law?" (2008) International Committee of the Red Cross (ICRC) Opinion Paper.

opposed to perfidy which is impermissible.[347] It is also recalled in rules 97 and 98 of the Tallinn Manual 2.0, that "non-violent operations, such as psychological cyber operations and cyber espionage, do not qualify as attacks [in the humanitarian law sense of the term.]"[348] This also concerns the dissemination of false news in order to sow panic amongst the population.[349] Although a reappraisal of the said regime might be needed, especially with regard to the peculiarities of the online information space and the high disruptive potential of disinformation,[350] whether a state actively involved into an armed conflict that is conducting cyber-disinformation campaigns in support of traditional military technics to sow panic, fear or to cause death within the adversary's population is liable for violations of humanitarian law is still a pending question.[351]

Drawing upon the jurisprudence of International Criminal Tribunal for the former Yugoslavia and the International Criminal Tribunal for Rwanda regarding instigation to commit a crime, Lahmann suggests applying a causality test to gauge the lawfulness of a disinformation campaign waged during wartime. Therefore, a strong causal nexus between a communication campaign and a resulting damage (destruction, death, injury), for instance, when a communication substantially contributed in the realization of the said damage, could make it an "attack" subjected to international humanitarian law and the principles of precaution, distinction and proportionality.[352] Further discussion is needed on this topic.

4.6 State-sponsored disinformation campaigns and the specific regime of warmongering propaganda

In 1898, a US navy ship, the USS Maine, sank in Havana Harbour in Cuba. Following the incident, many in the American press started to propagate the rumour that the Spanish were responsible for the sinking. The allegations were often accompanied by incendiary statements and cartoon images depicting the Spanish as the enemies. Many authors have since established that the unsubstantiated accusations, repeated over a relatively short period of time inflamed

[347] Additional Protocols to the Geneva Conventions of 12 August 1949 relating to the Protection of Victims of International Armed Conflicts (1977) 1125 UNTS 3, Article 37 "Prohibition of perfidy", §2. See also the Lieber Code (1863), article 15. For states practice on the mater, see the ICRC's IHL database: "Practice Relating to Rule 57. Ruses of War" ICRC, <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule57>.

[348] Schmitt (2017), *Supra* note 162. Rule 92 §2. at 415.

[349] Schmitt (2017), *Supra* note 162, at 433, commentary 3.

[350] See Robin Geiss & Henning Lahmann, "Protecting Societies: Anchoring A New Protection Dimension In International Law In Times Of Increased Cyber Threats" (2021) Working Paper Geneva Academy.

[351] Lahmann, *Supra* note 324. See also Nicolas, *Supra* note 22.

[352] *Ibid.* at 1240.

American public opinion and despite Spain's denial of involvement, pushed America to enter a war against Spain.[353] By communicating biased rumours and poorly-researched information, the US newspaper engaged in what is sometimes called, "yellow journalism." In the eyes of contemporary lawyers, the yellow journalism of some US newspaper following the sinking of the USS Maine would certainly constitute illegal warmongering propaganda.

Arguably one of the least contentious of the principles outlined in this study, the customary prohibition of warmongering propaganda stems from the prohibition to use or threaten force. Although state-sponsored disinformation may rarely attain such a level that it breaches the customary prohibition of war propaganda, the recent events in Ukraine provide a strong reason to present the legal regime of war propaganda, the scope of the prohibition, and explain how disinformation may relate to the prohibition.

4.6.1 War propaganda and disinformation

The prohibition of war propaganda is applicable to online disinformation, as it does with all international and domestic communication activities conducted within or outside cyberspace. In some rare occasions, state-sponsored online disinformation may amount to illegal war propaganda. War propaganda has been called "the most serious form of illegal propaganda."[354] Yet, quite paradoxically, states and lawyers have not devoted much attention to it.[355]

Almost systematically, war propaganda will involve spreading disinformation. War propaganda can be understood as pervasive communication campaigns designed to convince the audience of the necessity and correctness of a war whether or not it is justified. When promoted by the ruling regime and conveyed to the home population, war propaganda aims at obtaining the consent and the support of the audience for a specific military undertaking. War propaganda can also be directed abroad, to foster hostility between two or more states in order to get them to wage war against each other.

Methods used by a state for promoting wars are various and include the enforcement of censorship policies against dissenting voices such as alternative media and foreign journalists located in the territory of the instigator, the constant and systematic repetition of the same

[353] See a detailed depiction of the event here Louis Fisher, "Destruction of the Maine (1898)" (2009) *The Law Library of Congress* 1.

[354] Larson, *Supra* note 60.

[355] See Michael G Kearney, "Propaganda for War, Prohibition of" (2009) Max Planck Encyclopedias of International Law, encyclopedia entry..

images, particularly those charged with strong ethnocentric connotations, the systematic use of hateful rhetoric and of course, disinformation. Indeed, as war propaganda constitutes a continuous and relentless effort to build a narrative, unverified and unfounded claims will often compose most of the communications. Whitton and Larson say it clearly "[t]he concept of war-mongering propaganda does include one particularly vicious and effective kind of communication: false news."[356] Hence, a state-sponsored disinformation campaign directed at the national population or at a foreign audience so as to justify an illegal high scale military engagement may violate the prohibition of warmongering propaganda and as such, engage the international responsibility of the state and eventually the criminal responsibility of the main disseminators.

4.6.2 The legal regime of war propaganda

Warmongering propaganda is capable of bringing about wars and international crimes. Warmongering propaganda often precedes wars and is a common phenomenon in periods of international and domestic tensions and conflicts.

The prohibition of propaganda of war by a state is a fairly uncontroversial principle of customary international law.[357] Many rules make up the legal regime of incitement to war.[358] The prohibition of propaganda of war can be inferred from the whole UN Charter, and even more generally, from every legal instrument that expressively condemns the waging of war between states, such as the Kellogg-Briand Pact of 1928 or Protocol Additional to the Geneva Conventions of 1949.[359] The prohibition is also enshrined in numerous legal instruments, from UN resolutions to international treaties and the criminal codes of some states and can be inferred from a an even greater number of documents.[360] For instance, the unanimously-adopted 1947 UN resolution titled *Measures to be taken against propaganda and the inciters of a new war* expressly condemns "all forms of propaganda, in whatsoever country conducted, which is either designed or likely to provoke or encourage and threat to the peace, breach of the peace, or act

[356] *Ibid.* at 65.
[357] Whitton & Laron, *Supra* note 60. at 69-82.
[358] Kearney, *Supra* note 355.
[359] Whitton & Larson at 60. See also Quincy Wright "The Crime of "War-Mongering." (1948) 42(1) *The American Journal of International Law* 128.
[360] Whitton & Larson, *Supra* note 60. at 156. And at 72-73.

of aggression;"[361] Since then, the principle has been reaffirmed consistently in multiple UN resolutions.[362]

The prohibition is also contained in several international treaties. For example, the 1936 Broadcasting Convention cited earlier provides a duty of due diligence for states to ensure that their respective territories are not used by private actors to engage in war propaganda activities; "The High Contracting Parties mutually undertake to ensure that transmissions from stations within their respective territories shall not constitute an incitement either to war against another High Contracting Party or to acts likely to lead thereto."[363]

Also, a non-negligible part of the legal regime of propaganda of war stems from the jurisprudence of the international criminal tribunals.[364] Finally, Article 20 of the ICCPR famously prohibits "any propaganda for war and any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (…)"[365]

The legal regime of war propaganda substantially overlaps with the regime of hate speech. The prohibition also relates to a number of different areas of international law, touching upon many bilateral and multilateral treaties, the law of state-responsibility, criminal law and human rights law. For example, a state-sponsored subversive disinformation campaign inciting insurgents located in a foreign territory to revolt against the territorial state may simultaneously breach the prohibition of warmongering propaganda, the principle of non-intervention, and the principle of the non-use of force, and may also engage the international criminal responsibility of the main mouthpieces if violence ensue.[366] The sponsoring state may also be in breach of the 1936 Broadcasting Convention and other international treaties regulating international communications between the parties if it is itself party to such a treaty, alongside the victim state. War propaganda may also trigger the correction mechanisms enshrined in the 1952 Correction Convention.

---

[361] *Measures to be taken against propaganda and the inciters of a new war*, *Supra* note 314.

[362] See notably UNGA (1951), *Supra* note 315; *Strengthening of peace through the removal of barriers to free exchange of information and ideas*, (1954), A/RES/819, (IX); *Declaration on the Promotion Among Youth of the Ideals of Peace, Mutual Respect, and Understandings between Peoples,* (1965), A/Res/2037, (XX); *Declaration on Principles of International Law Friendly Relations and Cooperation Among States in Accordance with the Charter of the United Nations* (The friendly Declaration), (1970), A/RES/2625, (XXV); *Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States*, (1981), A/RES/36/103.

[363] Broadcasting convention, *Supra* note 280, Article 2.

[364] See *The Nuremberg Trial*, (1946) IMT, at 162; *The Prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza, Hassan Ngeze* (2003) ICTR judgement; *The Prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza, Hassan Ngeze* (2007) ICTR Appeal Judgment.

[365] See ICCPR, *Supra* note 118; *General Comment No. 11: Prohibition of propaganda for war and inciting national, racial or religious hatred (Art. 20): 29/07/1983.CCPR General Comment No. 11. (General Comments)* (1983), Office of the United Nations High Commissioner for Human Rights.

[366] See Wright, *Supra* note 359.

In general, states are under the obligation to proscribe warmongering propaganda in their respective legal systems, and to prevent their nationals from engaging in warmongering propaganda activities against other states. When affected by foreign war propaganda activities, the victim state is entitled to act in self-defence using force against the transmitters and the warmongering state, only if it is resolutely clear that the latter is about to turn words into action against it. States other than the imperilled state do not have the right to intervene in the territory of the warmongering state, or its online information space, for putting an end to warmongering communications.

4.6.3 The scope of the prohibition of war propaganda

The prohibition does not apply to communications that are merely producing tensions. There is an intensity requirement that hinges on how sustained the communication effort is, the vehemence with which the messages are relayed, their explicitness, and the volume of the target audience. In other words, to fall under the scope of the prohibition, the incriminated messages must clearly advocate for a war of aggression, or promote hatred against individuals for the reason of their membership in a group. They must also be reiterated over a sufficiently long time period.

But if it is certain that explicit incitements to war are prohibited under international law, the telling of lies in order to justify a future or imminent military intervention might be harder to assess. In that regard, Kearney cites the rejected proposition of the Brazilian delegation during the discussions on Article 20 of the ICCPR; "the prohibition of propaganda for war would encompass not only incitement to war, but also 'the repeated and insistent expression of an opinion for the purpose of creating a climate of hatred and lack of understanding between the peoples of two or more countries, in order to bring them eventually to armed conflict.'"[367] If adopted by all drafters, this interpretation would have applied to substantially more disinformation campaigns than the current formulation.

Nevertheless, the Office of the United Nations High Commissioner for Human Rights clarified in 1983 that the prohibition "extends to all forms of propaganda threatening or

[367] Kearney, *Supra* note 355.

resulting in an act of aggression or breach of the peace contrary to the Charter of the United Nations."[368] This has to be interpreted as a relatively narrow scope of application.

Among many other states, the US, Germany, Yugoslavia and Russia had recourse to disinformation campaigns for the purpose of fostering hatred and bellicose attitude against another nation, notably during the Spanish-American war, the first and second world-wars, the Iraqi war, the Yugoslavian war, and during the conflict between the Russian Federation and Ukraine. In light of the above, even the most recent of these disinformation episodes would amount to an illegal propaganda for war. Specifically, Russia's recent transboundary disinformation efforts to justify the war in Ukraine seem to qualify.[369]

With that being said, the regime of war propaganda is rarely called upon by states themselves to condemn hostile foreign communications. As Kearney recalls, the international law on propaganda of war is insufficiently elaborated. Only few references to the prohibition have been made since the end of the Cold War and the definition of warmongering propaganda is subject to contention.[370] It should also be noted that article 20 ICCPR is itself opposed by thirteen states.

5. State-sponsored cyber disinformation campaigns as prohibited interventions

Terms such as "influence" and "interference" have occasionally been employed to qualify state-led transboundary disinformation efforts. This vocabulary is important because the concepts of influence, interference and intervention have concrete legal implications. The term "influence," when unqualified, is generally used to describe harmless foreign conduct, or *benign* foreign influence activities. "Interference" is more specific than "influence" in that it describes an undue intrusion in the sovereign affairs of a state.

Although some states strive to remain objective with regard to their respective definitions of foreign interference, the concept of "interference" is inherently subjective. Foreign conduct

[368] *Ibid* and General Comment No. 11: of the Office of the United Nations High Commissioner for Human Rights, *Supra* note 365.

[369] Björnstjern Baade, "The EU's "Ban" of RT and Sputnik: A Lawful Measure Against Propaganda for War" (8 March 2022), *Verfassungsblog*, <https://verfassungsblog.de/the-eus-ban-of-rt-and-sputnik/>; Talita de Souza Dias, "Russia's "genocide disinformation" and war propaganda are breaches of the International Convention Concerning the Use of Broadcasting in the Cause of Peace and fall within the ICJ's jurisdiction" (4 March 2022), *Ejil:Talk*!, <https://www.ejiltalk.org/russias-genocide-disinformation-and-war-propaganda-are-breaches-of-the-international-convention-concerning-the-use-of-broadcasting-in-the-cause-of-peace-and-fall-within-the/>.

[370] Kearney, *Supra* note 355.

will be characterised as "interference" if they are known to the passive party and disagreed upon by the latter. Thus, a foreign interference by definition violates domestic law.

But if foreign interferences constitute illegal acts under the law of the recipient state, an interference into the affairs of a state may also contravene international law, provided that it can be attributed to a state. In general, only egregious foreign interferences, what Ziolkowski calls "massive influence"[371] may constitute internationally wrongful acts. Internationally unlawful foreign interferences are called "prohibited intervention." A prohibited intervention traditionally describes coercive foreign efforts that breach the customary principle of non-intervention.

The principle of non-intervention in international law has been the focus of a considerable amount of work in past decades. Yet, critics often point out the ambiguity of the principle. Noel wrote about the notion that it was "not only imprecise and ambiguous, but also still difficult to analyse."[372] For Little, the principle of non-intervention "[…] has always been and remains an imprecise and extremely ambiguous concept […]"[373] More recently, Jamnejad and Wood referred to the principle as being "One of the most potent and elusive of all international principles."[374] The list goes on.[375] Arguably, the ambiguity of the principle reflects its dynamism and underlines its flexibility. However, one cannot ignore the questions its content and nature systematically raise, and the sheer number of biased political references the principle suffers from.

These first observations give rise to an uncomfortable paradox; despite its fundamental importance in the current horizontal system of heterogenous states, the meaning of the principle of non-intervention still eludes practitioners. This is problematic because some states have referred to the possibility of applying the non-intervention paradigm to state-sponsored PSYOPS, particularly those targeting election processes. Recent literature on state-sponsored disinformation similarly reveals a preference for the non-intervention paradigm when it comes to determine the legality of state-sponsored disinformation campaigns.

---

[371] 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, Estonia: NATO CCD COE Publication, 2013), at 165.

[372] Jacques Noël, *Le principe de non-intervention : théorie et pratique dans les relations interaméricaines* (Bruxelles : Édition de l'Université de Bruxelles - Bruylant, 1981), at 3. [translated by author]

[373] Richard Little, "Revisiting Intervention: A Survey of Recent Developments" (1987) 13(1) *Review of International Studies* 49, at 49.

[374] See Maziar Jamnejad & Michael Wood, "The Principle of Non-intervention" (2009) 22 *Leiden Journal of International Law* 345.

[375] See also Olivier Corten et al., *A Critical Introduction to International Law* (Bruxelles : Éditions de l'Université de Bruxelles, 2019), at 156.

This section thus establishes whether and under what circumstances state-sponsored disinformation campaigns can amount to a violation of the non-intervention principle. To that end, the section first introduces the principle of non-intervention, its function, source, and the two constituting elements of a prohibited intervention; the *domaine reserve* of states and the concept of coercion.

5.1 International relations, foreign influence and foreign interference

The non-intervention principle applies to egregious foreign interference. The issue of foreign influence, interference and intervention occupies a prominent place in contemporary security-related debates. Since the advent of cyberspace, and in a significantly faster pace after the annexation of Crimea in 2014 and the Russian meddling into the 2016 US elections, states, organisations and think thanks have laid down extensive delineation works in order to revisit the foreign influence /interference /intervention taxonomy in light of the mutations brought forth by the digital revolution.[376] Alongside this theoretical effort, many states have solidified their policies regarding undue foreign influence and interference over the last decades. A noticeable development was the systematic inclusion of disinformation as a particularly common and allegedly efficient statecraft tool of influence against which states should defend. A guide from the Poynter Institute for Media Studies for example shows that in the span of a couple of years, many states have enacted laws to combat online disinformation, most of which are targeting disinformation disseminated by foreign actors.[377]

The crackdown on "foreign influence" including foreign disinformation does not concern a handful of states. This is a global phenomenon that is symptomatic of the contemporary digital world. The unprecedented state of proximity that resulted from global digitalization multiplied the interactions between domestic and non-domestic actors and heightened states' respective senses of threat perception. As cyber-enabled foreign influence, by definition, enjoys better reach and impact than the conventional non-cyber forms of influence, it has become much more difficult to distinguish acceptable foreign influence from illegitimate foreign interference, and

[376] See notably and among other Robert O.Keohane, & Joseph S. Nye, "Power and Interdependence in the Information Age." (1998) 77(5) *Foreign Affairs, Council on Foreign Relations* 81; Kristine Berzina & Etienne Soula, "Conceptualizing Foreign Interference in Europe" (2020), Alliance for Securing Democracy; Mansted, *Supra* note 204.

[377] Daniel Funke & Daniela Flamini "A guide to anti-misinformation actions around the world" *Poynter*, <https://www.poynter.org/ifcn/anti-misinformation-actions/>.

to distinguish illegitimate foreign interference from internationally illegal foreign interventions.[378]

Notably, some states have introduced the concept of "malign influence" to better apprehend foreign influence *via* disinformation. A "malign influence" 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.[379] 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."[380] 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."[381]

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 under municipal law. 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.

Of course, not all "malign influence" activities are illegal under international law, far from it. But as argued earlier, the question is worth asking for some online disinformation campaigns sponsored or condoned by a foreign state. The fact that "malign influence" was designed to apprehend disinformation campaigns however highlights an important issue with disinformation. Indeed, disinformation campaigns are protracted efforts. As such they rarely constitute clear-cut interferences in the sovereign affairs of the target. This may prove incapacitating when it comes to apply the principle of non-intervention to state-sponsored disinformation campaigns. Nevertheless, submissions about the non-intervention principle have also shown that the non-intervention principle is more flexible than it appears.

---

[378] See in that sense, Crootof, *Supra* note 214, at 620-628.

[379] Mansted, *Supra* note 204, at 3.

[380] Mansted, *Supra* note 204, at 14.

[381] 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>.

5.2 The function of non-intervention

The principle of non-intervention prohibits states from unlawfully intervening into the domestic affairs of other states. Much like Article 2§4, the principle of non-intervention protects the territorial integrity and the independence of states against unsolicited interventions. It does so, however, in a different way.

First of all, prohibited uses or threats of force will always simultaneously constitute violations of the principle of non-intervention. In the same way, "acts constituting a breach of the customary principle of non-intervention will also, if they directly or indirectly involve the use of force, constitute a breach of the principle of non-use of force in international relations."[382]

The general understanding is that the scopes of the prohibition of the uses of force and the principle of non-intervention, albeit 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. The principle of non-intervention thus sets less demanding thresholds than Article 2§4.

The prohibition of intervention is a basic principle of international law. It is based on the premise that all sovereign states are equal and independent from each other. A premise on which rests the whole structure of the international edifice. The prohibition owes its existence to the concept of state sovereignty. Consequently, its centrality in the current international order is undiscussed. President Nagendra Singh for example stressed in a separate opinion to the famous *Nicaragua* judgement; "I cannot conclude this opinion without emphasizing the key importance of the doctrine of non-intervention in the affairs of States which is so vital for the peace and progress of the international community."[383]

The principle of non-intervention is essential to the highly heterogeneous arrangement that is the international community. Ideological, political, economic, cultural and military differences are as many sources of potential frictions between states. Weaker actors are *de facto* put in a vulnerable position while strong powers compete to expand or maintain their respective spheres of influence. In that sense, not only the principle of non-intervention acts as a legal equalizer for reducing power inequalities between states, but it also helps ensuring the strict observance for the correlative rights of independence and territorial inviolability owned by states. In Oppenheim's words, the principle is the "corollary of every state's right to

[382] ICJ, *Supra* note 196, at 109, §209
[383] ICJ, (1986), *Supra* note 196, Separate Opinion of Judge Nagendra Singh, at 146.

sovereignty, territorial integrity and political independence."[384] Therefore the principle assumes a pivotal role in the maintenance of international cohesion.

In its oft-cited Friendly Declaration, the UNGA proclaims that "The strict observance by States of the obligation not to intervene in the affairs of any other States is an essential condition to ensure that nations live together in peace with one another […]"[385] The same insistence on the importance of the rule in contemporary international law is observable throughout the whole work of the UNGA and in a variety of bilateral and multilateral treaties.

5.3 The ubiquity of the principle of non-intervention in international instruments

References to the principle of non-intervention can be found as far back as in the 18th century, in the seminal work of Vattel.[386] An early oft-cited illustration of a prohibited intervention can be found in the period of the French revolution. The intervention consisted in "revolutionary decrees" transmitted in 1792 by the French revolutionary regime to the populations of European monarchies. The decrees were calling upon the European people to "recover their liberty" and embrace the concepts of "liberty and equality" by bringing down monarchies, promising to bring order and social justice to the liberated nations.[387] The vigour and consistency with which European states protested about the French revolutionary propaganda, together with the fact that the French revolutionary government hastened to withdrew the decrees and clarify its position following the political backlash provoked by its calls for revolution could be interpretated as strong evidence of *opinio juris* regarding the existence of a customary rule of international law prohibiting external calls for civil strife.[388] Not only that, but the French clarification also reaffirmed the primacy of a principle of non-interference applying, *inter alia*, to particularly brazen forms of revolutionary propaganda.[389] In a way, the event constituted an application of the principle of non-intervention.

The principle of non-intervention has since acquired the status of customary rule of international law. Notably, the prohibition has been incorporated in a variety of legal instruments throughout history.

---

[384] Lassa Oppenheim, *Oppenheim's International Law* in Arthur Watts & Robert Y. Jennings (Ed.), 9th edition, Vol. 1: Peace, (London: Longman Group, 1996), at 428.
[385] The Friendly Declaration, *Supra* note 362.
[386] Emer de Vattel, *Droit des gens ou principes de la loi naturelle*, (1758), vol 1, §37.
[387] Whitton & Larson, *Supra* note 60, at 18.
[388] *Ibid.,* at 97.
[389] See in that sense Van Dyke, *Supra* note 5.

Article 2§7 of the UN Charter contains one application of the non-intervention principle. Article reads as follows, "Nothing contained in the present Charter shall authorize the United Nations to intervene in matters which are essentially within the domestic jurisdiction of any state […]" Article 2§7 is exclusively intended to govern the relationship between the organisation and the member states. As such, it sets a vertical application of the principle. The present work however focusses on the principle of non-intervention as it applies between states, to which no explicit references are made in the Charter. Yet, the *travaux preparatoires* of the Charter and the Friendly Declaration *inter alia* show that most member states opined that the non-intervention principle was to be deduced from the Charter as a whole.[390] In the same vein, a majority of the doctrine acknowledges that the principle is implied throughout its article 2, without agreeing on a specific paragraph.[391] Vincent would later observe; "If the principle of non-intervention between states is nowhere explicit in the charter, it is, in the Statement of Principles everywhere implicit."[392]

Perhaps the most commonly recalled form of the principle is the one enshrined in the Charter of the Organisation of American States (hereinafter OAS). Its article 3 provides; "[…] Every State has the right to choose, without external interference, its political, economic, and social system and to organize itself in the way best suited to it, and has the duty to abstain from intervening in the affairs of another State.*"* In its fourth chapter, the principle is also listed among the *Fundamental Rights and Duties of States*. Articles 16 and 19 provide even more precision with regards to the principle. Article 19 defines the principle as follows; *"*No State or group of States has the right to intervene, directly or indirectly, for any reason whatever, in the internal or external affairs of any other State. The foregoing principle prohibits not only armed force but also any other form of interference or attempted threat against the personality of the State or against its political, economic, and cultural elements." Regretfully, these references do not say much about what acts should in practice be understood as a prohibited intervention. This lack of precision is endemic to the various recognitions of the principle of non-intervention.

In another regional setting, the Association of Southeast Asian Nations (hereinafter ASEAN), put the principle of non-interference at the hearth of its security community

[390] See Noel, *Supra* note 372. at 62.; *Special Committee on principles of international law concerning friendly relations and co-operation among states: summary record of the 29th meeting,* (1964), A/AC.119.SR-2.

[391] See for instance Dire Tladi, "The Duty Not to Intervene in Matters within Domestic Jurisdiction" In Jorge E. Viñuales (Ed.), *The UN Friendly Relations Declaration at 50: An Assessment of the Fundamental Principles of International Law* (Cambridge: Cambridge University Press, 2020), at 89.

[392] R.J. Vincent *Non-Intervention and International Order*, (Princeton University Press, Princeton, 1974), at 236.

process.[393] Article 2-points b and c of the *Treaty of Amity and Cooperation in Southeast Asia* of 1976 signed between the founding states of the ASEAN respectively recognize "The right of every State to lead its national existence free from external interference, subversion or coercion" and the "Non-interference in the internal affairs of one another" as fundamental guiding principles.[394]

The principle of non-intervention is also explicitly referred to as a "Principle Guiding Relations between Participating States" in the Helsinki Final Act of 1975[395], although it is not clear whether the signatories agreed on the same concept at the time of the signature of the Treaty.[396] It also listed in the Sino-Indian 1954 agreement (also known as the Panchsheel Treaty), as one of the "The Principles of Peaceful Coexistence", in which the Soviet doctrine of peaceful coexistence of states endorsed by Khrushchev found guiding principles to build upon during the Cold War.

Many UNGA resolutions contain references to the principle of non-intervention. In an early attempt to codify existing international law, the International Law Commission dedicated the third article of its Draft Declaration on Rights and Duties of States of 1949 to the principle of non-intervention, noting in commentary that "the substance of this text, which was derived from article 5 of the Panamanian draft, has already found place in various international conventions."[397] The Draft would be adopted by the UNGA in a resolution dating from December 6th 1949. Ample consideration would be given to the principle from 1961, in the context of the work of the UNGA on developing the values and principles tailored to the maintenance of peaceful interstate relations. The prohibition is for example mentioned as one of the seven principles considered fundamental in the "progressive development of international law and in the promotion of the rule of law among nations" in the ensuing 1962 resolution 1815, (XVII).[398] Shortly after, the UNGA Special Committee on Principles of International Law concerning Friendly Relations and Co-operation among States further elaborated the content of the principle of non-intervention as an interstate doctrine.[399]

---

[393] See ASEAN Security Community Plans of Action.

[394] Treaty of Amity and Cooperation in Southeast Asia (1976).

[395] Conference on Security and Cooperation in Europe, Final Act (1975).

[396] See Denitsa D. Raynova "Towards a Common Understanding of the Non-Intervention Principle" (2017) European Leadership Network, Post-Workshop Report.

[397] ILC, *Supra* note 277.

[398] *Consideration of principles of international law concerning friendly relations and co-operation among States in accordance with the Charter of the United Nations*, (1962), A/RES/17/1815.

[399] See notably *Proposals submitted to, and decisions of, various United Nations organs relating to the question of State responsibility: supplement prepared by the Secretariat to document*, (1969), A/CN.4/165; *Proposals submitted to, and decisions of, various United Nations organs, relating to the question of State responsibility: Supplement to document A/CN.4/165*, (1969), A/CN.4/209.

Plenty of resolution adopted by the UNGA relate to the principle of non-intervention. For example, the widely-approved *Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the protection of their Independence and Sovereignty* of 1965 notably condemns "armed intervention and all other forms of interference or attempted threats against the personality of the State or against its political, economic and cultural elements."[400]

The highly-political 1976 *Declaration on Non-Interference in the Internal Affairs of States* proclaimed *inter alia* the duties of a state "to refrain from the exploitation and the distortion of human rights issues as a means of interference in the internal affairs of States, of exerting pressure on other States or creating distrust and disorder within and among States or groups of States" and "to abstain from any defamatory campaign, vilification or hostile propaganda for the purpose of intervening or interfering in the internal affairs of other States."[401]

Finally, the contentious *Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States* of 1981 proclaimed the duty of states to combat "within their constitutional prerogatives, the dissemination of false or distorted news which can be interpreted as interference in the internal affairs of other states, or as being harmful to the promotion of peace, co-operation and friendly relations among States and nations." [402]

Interestingly, the possibility that the dissemination of disinformation constitutes a prohibited intervention is reflected in most resolutions pertaining to the principle of non-intervention. Yet, as far as the principle of non-intervention is concerned, it has been consistently affirmed in the legal literature that these resolutions can hardly be considered declaratory of customary law. [403] This raises the inevitable question of which exact representation of the principle of non-intervention has the status of customary rule of international law.

[400] *Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and The protection of their Independence and Sovereignty,* (1965), A/RES/2131(XX).

[401] *Declaration on Interference in the Internal Affairs of States*, (1976), A/RES/31/91.

[402] *Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States,* (1981), A/RES/36/103.

[403] See Jamnejad & Wood, *Supra* note 374; Benedetto Conforti "Le principe de non-intervention" in Bedjaoui, Mohammed (Ed.) *Droit international*: *bilan et perspectives*, (Paris: Pedone, 1991), 489; Lahmann. *Supra* note 15. See also Von Glahn, *Supra* note 5. at 558-560. See also Ori Pomson "The Prohibition on Intervention Under International Law and Cyber Operations" (2022) 99 *International Law Studies* 180, at 199.

5.4 The customary rule of non-intervention

The above inventory did not pretend to be exhaustive by any means and other relevant regional agreements recall the primacy of the principle in the organisation they institute, both in its vertical (between an organization and its member states) and horizontal (between states) versions. Still, few are the sources that properly elicit the scope and content of the rule in question.[404] Recognitions of the principle of non-intervention will often not say much on the attributes of the rule, or will betray regional specificities in their interpretations that would hardly fit the whole international system of heterogeneous states the principle is expected to govern.[405] Consequently, the question of what exactly constitute a prohibited intervention still results in uncertainty and inconsistent interpretations.

Defining precisely the content of the principle of non-intervention as it currently applies to states can be a challenging task. Indeed, depending on how its components are interpreted, it may appear either too broadly stated or on the contrary, too restrictive.[406] Some authors have even expressed doubts concerning the legal force of the prohibition, notably by pointing that "apart from the particular instances of military measures which have been judged with respect to the specific regime of the use of force, we know no precedent in which a state has been condemned by a court on the basis of a violation of the principle of non-intervention."[407] Other scholars further argue that the functions usually attributed to the principle of non-intervention are already assumed by other correlative obligations such as the prohibition of the use or the threat of force and the principle of self-determination.[408]

Departing from the above, most commentators however view the principle as an autonomous binding rule of international law that is susceptible of being violated on its own. Today the majority of the legal literature indeed agrees that, although the non-intervention principle encompasses and overlaps other correlative obligations that are more specific such as the principle of self-determination and the use of force[409], a prohibited intervention constitutes a specific type of conduct that can occur in a wide variety of situations, even outside the context

[404] See Corten et al. *Supra* note 375.

[405] See Philip Kunig, "Intervention, Prohibition of" (2008) Max Planck Encyclopedia of Public International Law, Encyclopedia entries.

[406] See Vincent, *Supra* note 392, at 8.

[407] See Corten, *Supra* note 375, at 163.

[408] Conforti *Supra* note 403. See also Corten, *Supra* note 375.

[409] See for instance Nicholas Tsagourias "Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace" in Dennis Broeders and Bibi van den Berg (eds.) *Governing Cyberspace: Behaviour, Power and Diplomacy* (Rowman & Littlefield, 2020).

of the use or the threat of force. In arguing so, many commentators consider non-intervention to be an autonomous rule of international law.

In fine, the extensive reliance on a principle of non-intervention by states and experts may be reflective of a widespread willingness to consider non-intervention as an obligation in its own right, understood as an autonomous rule of international law that is breached without the need for correlative sub-rules to be individually violated. The current discussions pertaining to the non-existence of a rule of sovereignty in cyberspace could similarly indicate that non-intervention, which by definition applies to violations of sovereignty, has its own function. Nevertheless, it should be kept in mind that the principle has so far been given inconsistent meaning.[410]

If non-intervention as a customary and autonomous rule of international law indeed exists, the crux of the matter is to determine its scope. In that sense, the ICJ in its *Advisory Opinion on the Legality of the Threat or use of Nuclear Weapons* noted "General Assembly resolutions, even if they are not binding, may sometimes have normative value."[411] Hence a natural reflex would be to rely on the above-cited resolutions and their *travaux preparatoires* to get a sense of the extent to which states were willing to consider the formulations enshrined in the text of the resolution as reflecting the scope and content of the non-intervention principle.[412] However, it is worth noting that the Court also held that a substantial numbers of negative votes and abstentions, while "clear sign of deep concern regarding the problem of nuclear weapons, […] still fall short of establishing the existence of an *opinio juris*"[413] More recently, the Court pointed out that "a State's vote on such resolutions [UNGA resolutions] cannot by itself be taken as indicative of the position of that State on each and every proposition within that resolution, let alone of the existence of a legal dispute between that State and another State regarding one of those propositions."[414]

In that regard, the question of the content to give to the prohibition was a significant point of contention between the member states. Russia and China were reluctant to compromise on their long-standing statist traditions by defending the view that state sovereignty and non-interference rank at the top of the normative pyramid. In the same vein, the non-aligned states that were previously subject to the yoke of colonialism sought to preserve at all cost their own

[410] See notably Xuan, *Supra* note 178.
[411] ICJ, *Supra* note 320, §70.
[412] See Pomson, *Supra* note 403, at 187.
[413] Nuclear Weapon Advisory, *Supra* note 320. See also *Ibid.*
[414] *Obligations Concerning Negotiations Relating to Cessation of the Nuclear Arms Race and to Nuclear Disarmament* (Marshal Islands vs India), (2016) ICJ, Judgement, at 855, §56.

sovereignties and independence, submitting absolutist interpretations of the prohibition.[415] Western liberal democracies, in contrast, deplored that some broad formulations could be easily instrumentalized by unscrupulous states, thus serving as legal safeguards against even the most benign conduct. Australia. Canada, France, Italy, the United Kingdom and the US for instance submitted in a 1966 joint proposal that states should enjoy a freedom to "seek to influence the policies and actions of other states",[416] a position firmly maintained through the discussions. Predictably, the bolder attempt of the resolution of 1981, to provide more details with regard to the conduct that ought to be considered a prohibited intervention or interference did little to advance the discussion. Overall, despite a significant number of states willing to consider the content of the resolutions as law, a "bloc of States had clear reservations on the subject and refrained from recognizing forms of prohibited intervention other than essentially two categories of acts."[417]

Against this background, affirming that the non-intervention principle has the status of customary law is perplexing, and raises the question of which exact representation of the principle does, in fact, enjoy customary nature. On this point, neither the principle as it is spelled out in the founding treaties of regional organisations and the deliberations at the UN seems to bring a satisfactory answer.[418]

Judge Ago, in a separate opinion to the Nicaragua case expressed doubts about its rather expeditious handling of the question of the customary nature of the discussed principle, principally regarding the nature of the resolutions cited above.[419] Judge Schwebel, in a famous dissenting opinion similarly disagreed that the "sweeping provisions of the OAS Charter concerning intervention constitute customary and general international law."[420] Jennings added that the principle of non-intervention is "a principle of law which in the inter-American system has its own peculiar development, interpretation and importance."[421] More recently Schmitt reiterated; "caution is merited in applying the treaty rules [of non-intervention] because their parameters may differ from their customary counterpart, […]"[422]

---

[415] *Ibid.* at 193.

[416] *Joint proposal by Australia. Canada, France. Italy, the United Kingdom of Great Britain and Northern Ireland and the United States of America* (A/AC.125/L.13) in *Report of the 1966 Special Committee on Principles of International Law Concerning Friendly Relations and Cooperation Among States*, (1966), A/6230, §279.

[417] Pomson *Supra* note 403*,* at 199.

[418] See generally *Ibid.*

[419] ICJ, *Supra* note 196, Separate Opinion of *Judge* Ago, at 146.

[420] ICJ, *Supra* note 196, Dissenting Opinion of Judge Schwebel, at 295.

[421] *ICJ, Supra* note 196, Separate Opinion of Sir Robert Jennings, at 534.

[422] Michael N. Schmitt "Foreign Cyber Interference in Elections" (2021) 97 *International Law Studies* 739, at 744.

In reality, making sense of a customary rule of non-intervention entails taking into account three factors; first, that the principle, as enshrined in regional and bilateral agreements tends to be formulated broadly for reasons inherent to the very nature of the organisation they regulate.[423] This is particularly noticeable in the founding treaties of economic organisations, for which the absence of friction between their parties is an essential precondition to their proper functioning. Second, that the two resolutions with the most substantive content about the principle of non-intervention did not achieve unanimity, and as such, should not necessarily be taken at face value when it comes to establishing the customary prohibition of intervention.[424] Third, that despite its elusive character, the principle of non-intervention is a powerful rhetorical device. As such, it may be invoked simultaneously 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."[425]

In view of the above, it is rather unsurprising to see the ICJ articulating a "highly formal interpretation" of the prohibition.[426] Indeed, the Court held that the principle prohibits "[…] interventions 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. Intervention is wrongful when it uses methods of coercion in regard to such choices, which must remain free ones."[427] Then, according to the Court, an unlawful intervention is one that deprives the target of free will on matters not subject to international regulation, which converges with Oppenheim's own oft-cited definition of an unlawful intervention.[428]

The ICJ's rather restrictive interpretation of the principle of non-intervention remarkably aligns with the second paragraph of the unanimously approved formulation of the 1965 Declaration on the Inadmissibility of Intervention, which forbid states to "use or encourage the use of economic, political or any other type of measures to coerce another State in order to obtain from it the subordination of the exercise of its sovereign rights or to secure from it advantages of any kind", and provides that "no State shall organize, assist, foment, finance, incite or tolerate subversive, terrorist or armed activities directed towards the violent overthrow

[423] ICJ, *Supra* note 196, at 38, §175; Sean Watts, "Low-Intensity Cyber Operations and the Principle of Non-Intervention" in Jens David Ohlin, Kevin Govern, and Claire Finkelstein (eds), *Cyber War: Law and Ethics for Virtual Conflicts* (Oxford: Oxford University Press, 2015), 249.

[424] See notably Jamnejad & Wood, *Supra* note 374, at 355.

[425] Corten, *Supra* note 375, at 156.

[426] Corten, *Supra* note 375, at 161. See also *Case Concerning Armed Activities on the Territory of the Congo* (Democratic Republic of the Congo v. Uganda), (2005), ICJ, Judgement.

[427] ICJ, *Supra* note 196. §204, at 107.

[428] Oppenheim, *Supra* note 384.

of the regime of another State, or interfere in civil strife in another State."[429] However, the Court's formulation does not appear to encompass most of the more contentious provisions contained in the other aforementioned resolutions. Nevertheless, such a cautious attempt to clarify the meaning and the content of the customary prohibition of intervention remains until today a widely acknowledged (yet perhaps imperfect) reading of the principle.

With that being said, by restricting the scope of the prohibition to coercive state conduct while providing little to no indications as regards as to which conduct, other than obvious uses of force or ultimatums ought to be considered coercive, the Court did little to clarify a principle that is said to have been derogated so many times by states.

As remarked earlier, such a strictly framed formulation led some commentators to cast doubts about the legal value of the principle.[430] This is all the more confusing that the fact that the Court attempted to clarify the content of the principle tends to suggest that the judges really intended to make non-intervention a *standalone rule* susceptible of being breached in its own.

Nevertheless, especially since the advent of cyberspace and the subsequent multiplication of "low intensity" cyber operations, a growing portion of the legal literature is concerned with the quantity of illegitimate state conduct clearly affecting the political integrity of the target that slip through the prohibition's net for the reason that they are not coercive.[431]

In face of this, some have argued that the Court left the door open to more fluid interpretation of coercion that do not presuppose a use or a threat of force.[432] Due to the difficulty of defining interstate coercion outside the context of uses or threats to use force, other authors attempted to find alternatives to the Court's formulation. In doing so, they have called for a redefinition of the two constituent elements of a prohibited intervention in the sense of the current principle of non-intervention; the concept of coercion and, to a lesser extent, the notion of reserved domain.[433] Remarking that the criteria proposed by the Court ultimately failed to clarify the contours of a *rule* of non-intervention, other scholars more bluntly argued "PNIVIF [principle of non-intervention and interference] is an empty container principle which has never stopped evolving."[434] Although their approaches vary from one to another, most commentators

[429] See Resolution 2131, *Supra* note 400.

[430] See notably Corten, *Supra* note 375; Conforti, *Supra* note 403.

[431] It bears noting that some commentators were already advocating for a reappraisal of the rule before the advent of cyberspace. See notably Lori Fisler Damrosch, "Politics across Borders: Non-Intervention and Nonforcible Influence over Domestic Affairs" (1989) 83 *American Journal of International Law* 1.

[432] See for instance Marcelo Kohen, "The Principle of Non-Intervention 25 Years after the Nicaragua Judgment" (2012) 25(1) *Leiden Journal of International Law* 157, at 161.

[433] See notably Thibault Moulin "Reviving the Principle of Non-Intervention in Cyberspace: The Path Forward" (2020) 25(3) J*ournal of Conflict and Security Law* 423; Tsagourias, *Supra* note 409.

[434] Xuan, *Supra* note 178, at 29.

seem to converge on the necessity to move on from the ICJ's restrictive formulation. With that being said, it is worth noting that there are rare situations where it is relatively easy to conclude on the applicability of the principle of non-intervention. These situations may eventually involve state-sponsored disinformation campaigns.

5.5 Unproblematic applications of the principle of non-intervention to state-sponsored disinformation campaigns

If there are still doubts regarding the scope of the principle of non-intervention, some state-sponsored PSYOPS may unquestionably contravene the principle. Thus, the formulation by a state of overt threats to use force against another in order to compel a particular behaviour *vis à vis* its reserved domain would give rise to an unproblematic application of the principle of non-intervention and would simultaneously constitute a breach of Article 2§4.

Other state-sponsored disinformation campaigns might breach the non-intervention principle by reason of their subversive character, regardless of the occurrence of effects on the territory of the target state. For instance, the World Court considers the provision of support to foreign insurgent groups as a prohibited intervention.[435] By analogy, disinformation campaigns may take the form of blatantly illegal intervention "if they aim to foment revolt or civil strife in another State or are devoted to assisting illegal and violent activities", and if they can be attributed to a state.[436] Fauchille in the past referred to such conduct as "*intervention de propagande*".[437] These would give raise to particularly uncontentious application of the principle of non-intervention. The communications in question do not necessarily have to target election or other internal processes abroad. The fact that a foreign power intended to substantially alter the outcome of an internal struggle generally suffices.

Specifically, in its Nicaragua case, the Court considered that the financing of the *contras* by the US amounted to a violation of the principle of non-intervention because the latter had knowledge of the activities conducted by the formers. In light of this, the argument could be made that disinformation campaigns waged with the view to support the overthrow of a regime are similar to the application of "indirect coercion" on the target state.[438] In practice, this could

[435] See ICJ, *Supra* note 196. See also ICJ, *Supra* note 426.
[436] Kunig, *Supra* note 405; Jamnejad & Wood, *Supra* note 374, at 374.
[437] Paul Fauchille, *Traité de droit International Public,* Vol 1, Sec 300(3) (Paris, Rousseau, 1921), cited in Whitton & Larson, *Supra* note 60, at 96.
[438] See ICJ, *Supra* note 196, at 108, §228.

concern disinformation campaigns especially conducted by a state so as to persuade insurgent groups already present on the territory of the target, to pursue their subversive activities. The PSYOPS could for example involve encouragements, vocal support, the beaming of inflammatory information to the groups in question, or the provisions of instructions on how to operate. Provided that the activities and the overall aim of the target revolutionary group could not have been ignored by the disseminator, and that an extensive study of the problematic information conveyed by the foreign state reveals that the campaign was specifically waged for the purpose of stimulating existing revolutionary activities within the territory of the target, then the communications in question may constitute an "indirect involvement in coercion" and thus amount to a prohibited intervention and eventually an indirect use of force.[439]

More generally, direct calls for sedition relying on false or misleading facts made by a foreign state and addressed to the population of another to initiate a revolt would give rise to a particularly uncontentious application of the non-intervention principle.

In presence of such "unproblematic" intervention cases, there will be no need to establish a causal link between the communications in question and subsequent internal disturbances. Also, material involvement into the affairs of a foreign state is not required. Each of the aforementioned communications may be considered internationally unlawful by the mere fact of their nature, for that they are unquestionably interventionist. As such, they contravene the customary prohibitions of subversive/revolutionary propaganda and simultaneously constitute prohibited interventions. On that point, state practice and *opinio juris* are sufficiently consistent and uniform, as evidenced in the numerous resolutions taken on the subject during the cold war, the antiquity of the rules themselves, and more generally, the consistency with which states have denounced such conduct, to leave no doubts as to the illegality of the conduct[440] Therefore, it makes little doubts that overt or covert communications clearly intended to encourage sedition abroad are violating the prohibition of unsolicited interventions.

Nevertheless, that is a long way from claiming that all coordinated releases of misleading information and forgeries aimed at causing discontent and distrust amongst the target audience are encouraging sedition in one or more foreign territories. In fact, online disinformation campaigns are mostly used as corrosive and subtle disruptive tool. As such, they will rarely involve open calls for subversion. Consequently, it is necessary to expand upon the constitutive elements of prohibited interventions in more details in order to properly assess whether international communication involving information manipulation tactics that do not constitute

[439] See Watts, *Supra* note 423; ICJ, *Supra* note 196, at 48, §99; Stephens, *Supra* note 15, At 6.
[440] Whitton & Larson, *Supra* note 60, at 96-103.

open call for revolt or war are capable of breaching the principle of non-intervention. The following sub-sections will first deal with the principle as formulated by the ICJ. In doing so, it will establish whether the principle, in its current form, provides adequate limitations to psychological warfare. Then, in a second step, the legality of state-sponsored disinformation campaigns will be appreciated using alternative interpretations of the principle.

5.6 Cyber-enabled disinformation campaigns and the concept of reserved domain

A first step in assessing the legality of conduct against the principle of non-intervention is to determine whether the operations in question are susceptible of affecting what is safeguarded by the principle of non-intervention; the sovereign competences of a state.

To be deemed illegal, foreign interventions must bear "[…] on matters in which each State is permitted, by the principle of State sovereignty, to decide freely."[441] These matters do not constitute a "target" *per se*, but are to be understood as the competences allegedly affected by the coercing state's behaviour. In addition, the category does not necessarily require for specific infrastructure or competences to be affected. Rather, it is the sovereign authority of the victim state over a matter that must be weakened.[442]

Hence, foreign interference that negates the ability of a state, to freely decide on matters for which it is entitled by international law, to be the exclusive deciding force, is breaching the non-intervention principle. The ICJ, in its Nicaragua judgement, specifies that matters related to "the choice of a political, economic, social and cultural system, and the formulation of foreign policy" obviously come under the scope of the prohibition.[443] The ICJ also correlates these matters with the notion of sovereignty, recalling that it is in the name of the principle of sovereignty that the state is the single entity to decide on certain matters.[444]

Diverse, yet rather consistent language is used to refer to these matters. The ICJ employed the term "matter within the domestic jurisdiction of a state."[445] This term usually defines an area where a state has competence "to exercise governmental power through its own organs, for the solution of any situation, conflict or dispute insofar as no obligation of general

[441] ICJ, *Supra* note 196, at 108, §205.
[442] Schmitt (2017), *Supra* note 162, at 315.
[443] ICJ, *Supra* note 196.
[444] *Ibid.*, at 101, §212.
[445] *Ibid.,* at 144, §288.

international law or treaty affects the scope or manner of such exercise."[446] Some authors use the term "sovereign prerogatives" to refer to the content of this area of exclusive jurisdiction.[447] More generally, the literature reconciles the matters in question with the concept of *domaine reserve*, or reserved domain of a state, a category that encompasses competences that the PCIJ, in the *Nationality Decrees in Tunis and Morocco advisory opinion*, recognised to be evolving depending on the development of the conventional engagement of a state.[448] A set of resolutions adopted by the Institute of International Law in 1954 similarly indicates that "the *domaine reserve* is that of states activities where the competences of the state is not bound by international law." Adding that its extent "depends on international law and varies according to its development."[449] Therefore, because international law regulates a growing number of issues, the extent of the *domaine reserve of* states tends to shrink over time.[450] Consequently, only through careful examination of states practice, *opinion juris* and identification of *jus cogens* norms can the extent of the *domaine reserve* of a state be accurately captured.[451] The content of the *domaine reserve* of a state is thus assessed on a case-by-case basis, at the time of a claim.

It transpires from the above that the terms "sovereign prerogatives," "reserved domain" and "domestic jurisdiction" are used interchangeably to refer to the category of prerogatives protected by the principle of non-intervention. This study is not intended as a thorough discussion of the relevance of one term over another. It is nevertheless worth noting that, whatever the term adopted, they all undisputedly boil down to the autonomy of states. In other words, the non-intervention principle is, above all, a safeguard to independence in internal and external matters that have not yet been submitted by the state to international oversight.[452] Matters submitted to international regulation then fall outside the scope of the principle of non-intervention. Hence, economical engagements with other states, ratification of human rights treaties, or extradition agreements, are as many examples of limitations set to the independence of a state, as they subsequently become matters on which the state is not anymore the sole entity to decide.

---

[446] Quincy Wright, "The Prevention of Aggression*"* (1956) 50(3) *The American Journal of International Law* 514.

[447] Nicholas Tsagourias, *Supra* note 409; Eric David, "Portée et limite du principe de non-intervention" (1990), 23 *Revue Belge de Droit International* 350.

[448] Nationality Decrees Issued in Tunis and Morocco, (1923) PCIJ. Advisory Opinion, at 23.

[449] "Résolutions Adoptées Par l'institut de Droit International à La Session d"Aix-En-Provence 22 Avril-1er Mai 1954" (1953-55), 52 *Die Friedens-Warte, Berliner Wissenschafts-Verlagat,* at 254–57.

[450] Katja S Ziegler, "Domaine Reserve" (2013), Max Planck Encyclopedias of International Law, Encyclopedia entries.

[451] Schmitt (2017), *Supra* note 162, at 314.

[452] Ziegler, *Supra* note 450.

Persuasive arguments can be made with regards to the inappropriateness of the relative notions of reserved domain, or domestic jurisdiction, to define what the Court considers as the subject of the principle of non-intervention.[453] This also calls into question the resilience of the non-intervention principle. In that regard, admitting than the scope of the principle tends to decline over time does contradict its alleged function.[454] Both the vagueness of the examples provided by the Court, and the apparent non-limitative character of its list of examples does not lead to clear-cut answers. With that being said, and as underlined by an author, the "displacement of a matter or issue from the *domaine reserve* does not constitute an overall eradication or waiver of the principle of non-interference, nor an open season on influencing conditions in another state's territory."[455]

The debate surrounding the content of a state's *domaine reserve* is inconsequential to the present assessment. Indeed, whilst some questions may lie at the edge of the scope of the reserved domain of a state and thus raise some doubts as to their classification, there are some prerogatives that can be considered *typical* parts of a state's *domaine reserve*.[456] These matters, upon which states are *in principle* enjoying exclusive jurisdiction, are clearly falling under the scope of the principle of non-intervention. These comprise, for instance, matters related to the choice of political leadership, to which disinformation campaigns are particularly susceptible of inducing change, the planning and execution of health policies, the choice of extending diplomatic and consular relations, the choice of abrogation or formation of treaties, or the deployment of measures intended to restore national safety.[457]

Therefore, and in principle, state-sponsored disinformation campaigns having for intended and verified effects to severely disrupt national epidemic management policies or election processes will qualify as interference bearing on a matter contained in its reserved domain. State-sponsored disinformation campaigns to which it is impossible to ascribe a particular goal, or state-sponsored disinformation campaigns affecting non-protected competences are thus excluded from the scope of the prohibition.

It bears noting that proving that a foreign disinformation campaign was specifically aimed at its reserved domain may prove a difficult task for the victim state.[458] Indeed, this entails, as

---

[453] *See* Watts, *Supra* note 423.

[454] See in that sense Moulin, *Supra* note 433.

[455] Sean Watts, "International Law and Proposed U.S. Responses to the D.N.C. Hack" (14 October 2016), *Just Security,* <https://www.justsecurity.org/33558/international-law-proposed-u-s-responses-d-n-c-hack/>.

[456] See notably Watts, *Supra* note 423.

[457] See regarding health policies Schmitt & Milanovic, *Supra* note 326, at 247. See Schmitt (2017), *Supra* note 162, at 317, Rule 66, commentary 16; ICCPR, *Supra* note 118.

[458] Corten, *Supra* note 375, at 160.

far as online disinformation is concerned, to establish that the intention of the disinformant was specifically to affect the sovereign prerogatives of the target state, or at least, to prove that the enabling or sponsoring state could not have ignored that its behaviour would lead to detrimental effects on the protected competences of the victim state. This may not be an issue in the context of electoral disinformation, but may pose a problem when it comes to disinformation campaigns occurring outside the very specific context of elections. At any rate, the burden of proof is placed on the defendant, who is required to investigate all the characteristics of a problematic communication campaign to make sense of the conduct and to infer the intent of the instigator.

5.7 Cyber-enabled disinformation campaigns and the concept of coercion

Coercion is a highly debated concept that is even more complex to grasp in an interstate setting than in its interpersonal dimension. Regrettably, there is no indications as to whether the concept of coercion in the context of international relations is the same as in the interpersonal context.[459]

Still, the ICJ considers that the element of coercion "forms the very essence of a prohibited intervention."[460] As such, coercion serves as a threshold in discriminating between mere influence and persuasion, which are admissible under international law, and legally reprehensible intervention. Coercion is equivalent to "pressure" or "intimidation" in meaning. It is traditionally said to encompass threat and compulsion. For example, a treaty is traditionally considered void if concluded under coercion by the threat or use of force.[461] But whereas forcible actions are inherently coercive, there seems to be a consensus in the legal literature around the idea that a conduct which is not breaching the non-use or threat of force principle can theoretically be coercive.[462]

In 1905 Oppenheim recalled that "[...] intervention proper is always dictatorial interference, not interference pure and simple."[463] The adjective *dictatorial* defines the imposition of external will on a subject. It implies the loss of control over a situation and thus, presupposes the absence of willingness on the part of the coerced, to act in the way an exterior authority is pushing it. In

[459] Jens David Ohlin, "Did Russian Cyber Interference in the 2016 Election Violate International Law?" (2017) 95 *Texas Law Review* 1579, at 1581. (Noting that there is no theory of coercion in international law.)

[460] ICJ, *Supra* note 196, §205, at 108.

[461] See VCLT, *Supra* note 307, article 52.

[462] See generally Schmitt (2017), *Supra* note 162, at 318, Rule 66, commentary 20. See also the government of the Netherland, "the application of international law to cyber operations" (14 October 2019), Position paper, <https://dig.watch/updates/netherlands-position-international-law-cyberspace-made-public>.

[463] Oppenheim, *Supra* note 141, at 182.

contrast, a subject under mere influence is not conscious of the impact that the influencer has on its choices, nor does it need to comply with any sort of demand to extricate itself from a situation it considers disadvantageous. Then, as opposed to an influencer, who seeks to guide the addressee towards the taking of wilful actions, a coercer seeks to take away the possibility of free and conscious choice from the victim.[464] The commentaries under ARSIWA Article 18 makes reference to coercion between states as a "[...] conduct which forces the will of the coerced State [...], giving it no effective choice but to comply with the wishes of the coercing State."[465] In other words, coercion is forcing, or attempting to force a target to do something they would not otherwise do.[466] Thus, the experts of the Tallinn Manual explain that, to amount to a prohibited intervention, the coercive act must be "designed to compel the victim State to change its behaviour with respect to a matter within its *domaine reser*vé."[467] As a result, they add that "persuasion, criticism, public diplomacy, propaganda, retribution, mere maliciousness, and the like" do not represent coercive conduct.[468]

Many scholars have criticized the above interpretation for its rigidity, especially in a cyber context.[469] As an alternative, they have articulated a more flexible approach to the element of coercion. Thus, some authors have argued that the coercion criterion should be considered fulfilled once a foreign conduct "depriv[ed] the target, of its ability to freely decide on sovereign matters."[470] Accordingly, the "very inability of the target state to exercise control over its sovereign functions, with the harmful effects that are likely to ensue within the target state as a result, is the outcome that the perpetrating state is seeking to compel."[471]

This interpretation departs from Oppenheim's in that it leaves aside the requirement of a "dictatorial" interference. Consequently, there is no need to investigate whether the coercer intended to get the victim state to act in a certain way, nor is it required for the target state to have knowledge of the intervention to be "coerced."[472] According to this interpretation, coercion materializes when the victim state is effectively deprived of its authority *vis à vis*

[464] Schmitt (2017)*, Supra* note 162, at 318, Rule 66, commentary 21.
[465] ARSIWA, *Supra* note 81, at 69, rule 18.
[466] See Corten, *Supra* note 375. See Wheatley, supra note 91.
[467] Schmitt (2017), *Supra* note 162, at 318-319.
[468] *Ibid*.
[469] See *inter alia* Moulin, *Supra* note 433; Harriet Moynihan, "The vital role of international law in the framework for responsible state behaviour in cyberspace" (2021) 6(3) *Journal of Cyber Policy* 394; Kilovaty, *Supra* note 81; Schmitt & Milanovic, *Supra* note 326; Xuan, *Supra* note 178; Damrosch, *Supra* note 431.
[470] See notably Moynihan, *Ibid;* Australian Government, "Supplement to Australia's Position on the Application of International Law to State Conduct in Cyberspace", (2019), Department of Foreign Affairs and Trade; Harriet Moynihan, "The Application of International Law to State Cyberattacks" (2019) Chatham House Research Paper.
[471] Moynihan, *Ibid.*
[472] Tsagourias, *Supra* note 409, at 20.

sovereign areas. Simply put, dictatorial coercion originally defined an attempt to override the sovereign will of a state whereas non-dictatorial coercion refers to an attempt to override the sovereign authority of the state.

In theory, this formulation of coercion makes the principle more easy to apply. Yet, it is not an optimal formulation either.[473] For example, it remains relatively unclear from which point in time it may be possible to consider a non-forceful yet protracted attempt at undermining a state's sovereignty "coercive" in the sense of a non-dictatorial conduct, as the approach still requires considerably more than mere disruptions. It also seems that only *ex post facto* assessments would allow for such conclusion to be drawn. Also, it is simply unclear why such a restatement of coercion should still be called "coercion."

Regarding state practice and *opinio juris*, states such as Australia[474], New Zealand[475], and Iran[476] appear to have subscribed to the relaxed interpretation of coercion. Others, like Japan and Italy, have stressed the need for articulating a version of the principle of non-intervention which better captures the specificities of cyber-enabled interferences.[477] In contrast, Norway, Estonia, and Switzerland have, in 2021, articulated traditional interpretations of the principle of the prohibition of dictatorial intervention.[478]

With that being said, the concept of coercion, and by extension the principle of non-intervention remain sources of significant confusions. Indeed, some authors appear to endorse yet other interpretations of the non-intervention principle. These alternative interpretations of what should amount to a prohibited intervention are slightly different from one to another. Yet, they all have in common the fact that they downplay the criterion of coercion in favour of considerations of scale and intent. In doing so, they argue that the point of a prohibited intervention could be made if the conduct in question displays the required characteristics irrespective of the level of pressure generated on impact.[479]

---

[473] For an analysis of how cyberspace has changed coercion, see Miguel Alberto Gomez "Coercion and Cyberspace" (2018) Real Instituto Elcano.

[474] See Australian Government, *Supra* note 470.

[475] The Government of New Zealand, *Supra* note 257.

[476] Declaration of General Staff of the Armed Forces of the Islamic Republic of Iran Regarding International Law Applicable to the Cyberspace, *Supra* note 261.

[477] Italian Ministry for Foreign Affairs and International Cooperation, "International law and cyberspace" (2021), position paper; Ministry of Foreign Affairs of Japan, "Basic Position of the Government of Japan on International Law Applicable to Cyber Operations", (2021).

[478] Federal Department of Foreign Affairs of Switzerland, "Switzerland's position paper on the application of international law in cyberspace" (May 2021). See also *Official compendium, Supra* note 161.

[479] See notably Mohamed S. Helal, "On Coercion In International Law", (2019), 52 *International Law and Politics* 1; Gary P. Corn, "Covert Deception, Strategic Fraud, and the Rule of Prohibited Intervention" (2020), 2005, *Aegis Series Paper.*

As neither states or the doctrine seem to be on the same page when it comes to the meaning of coercion, the following will begin with an analysis of dictatorial coercion in online disinformation campaigns before addressing the question of the relevancy of the non-dictatorial reading of coercion in the assessment of the legality of online disinformation. That way, it can be established whether the relaxed, non-dictatorial reading of non-intervention really makes a difference in the legal assessments of state-led disinformation activities. Then, drawing upon alternative submissions about the non-intervention principle, a sub-section sketches a possible way forward when it comes to interpret non-intervention in an era where coercion *stricto-sensu* is becoming rare. Notably, it will be seen whether the resulting formulation could really make a difference regarding the assessment of disinformation campaigns.

5.7.1 Dictatorial coercion and online disinformation campaigns

Distinctive end and means define coercive intervention as a dictatorial interference;

- An *end*: It carries a goal that involves coercing the victim to do or not do something *vis à vis* its reserved domain
- *Means*: It relies on instruments of pressure to get the target to act in a certain way

5.7.1.1. The intent to coerce in cyber-enabled disinformation operations

All coercive interventions follow the same rationale of reshaping the status-quo at the advantage of the coercer, and each dictatorial coercive intervention aims at achieving compliance.[480] What specific behaviour is expected by the coercer varies depending on subjective and contextual factors. A state may for instance intervene with the intention of getting a population to oust their leader, or to get a neighbour to refrain from dropping off a denuclearisation treaty. In any case, the active party's action towards a passive third party is said coercive when it involves getting the target *to act or refrain from acting*.[481] Thus, coercion is particularly obvious in conditional threats.

States rarely publicly unveil the goal behind the coercing force. International law and strategical calculus prevent states from revealing too much about their deeds, and unjust

[480] See notably Kunig, *Supra* note 405.
[481] Wheatley, *Supra* note 466; Moulin, *Supra* note 433; ICJ, *Supra* note 196, at 114, §241.

requests for compliance are seldom uttered publicly, as it would be particularly unacceptable for the international community, to witness a state overtly blackmailing another. Yet, coercion may occur even in absence of overt threats.[482] In theory, the coercer can act covertly if its intent is sufficiently inferable from its actions and the context.[483] Coercion in its traditional meaning requires at least that the target perceives or is able to perceive what appears to be the action the coercer expects it, or not, to undertake. Whether the true identity of the coercer is hidden to the target or whether the demand is implicit is inconsequential.[484]

As the most rational entities, states never act wantonly. Disinformation operations are not waged by a state without purpose. Yet, without any insights into the decision-making processes that led to the taking of disinformation campaigns, the target is often forced to base its judgement solely on the "exhaust" portion of the operations and the context within which they are taken, which, as far as cyber-enabled conduct are concerned, may lead to misinterpretation and ultimately, non-compliance.

Assuming that cyber disinformation campaigns are designed to force particular policy changes abroad, finding out what result is expected by the disinformation is made harder by the interconnectivity of the online information spaces and the incremental and corrosive nature of disinformation campaigns. Also, in the same way as for most terrorist activities, the aims of cyber disinformation operations are way less specific than what coercion traditionally entails.[485] Disinformation propagates, reverberates and is constantly repelled in an information space, which makes it rather hazardous for the coercer to realistically expect disinformation campaigns to force a behaviour. Also, online disinformation is incremental and corrosive. The effects are only felt over a certain period of time. Of course, the disinformant may aim for the same kind of results as a coercer. That is to say, the disinformant may aim for regime overthrow without explicitly inciting civil strife, or it may seek economic or territorial concessions from the target state. But these are incredibly long-term goals in the context of disinformation campaigns, and coercion entails by definition that the victim does not speculate on the expected outcome. In other words, arguing that disinformation operations are designed to force change at the state level is at best difficult, at worst, too speculative to demonstrate the existence of coercive intent. Notwithstanding, the action thus induced must bear on matters within the reserved domain of the target.

---

[482] See Watts, *Supra* note 423.
[483]Gomez, *Supra* note 473.
[484] See notably Schmitt (2017), *Supra* note 162, at 321, Commentary 26.
[485] Scott Anderson, "Coercion", (2011), The Stanford Encyclopedia of Philosophy Archive (Summer 2021 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/sum2021/entries/coercion/>.

With that being said, the states involved, the political context, the nature of the conveyed messages and the interests at stake might provide sufficient indications as to the reasons underlying the waging of disinformation operations, and eventually give a glimpse into the interventionist intent of the acting state.[486] However, whereas the rationale behind disinformation might be guessed with relative clarity, the question of whether a disinformation campaign is especially designed to coerce the target state *to do or refrain from doing something specific* is not easily answered. Coercion will have to be appreciated in relation to the perception of the victim state, as well as factual elements regarding the scale and duration of the operation appreciated from outside the party's mind.

For instance, a smear campaign relying on false or misleading information (or defamation propaganda) waged by a state against another in response to a conduct taken by the latter, and aimed clearly at forcing the latter to cease the said conduct might very well constitute a prohibited intervention if the conduct in question relates to the reserved domain of the addressee. Yet, when "the pressure is such that it could reasonably be resisted, the sovereign will of the target state has not been subordinated."[487] What counts is that "the coercive act must have the potential for compelling the target state to engage in an action that it would otherwise not take" *vis à vis* its reserved domain.[488] Hence it remains to be established whether a disinformation campaign taken with the aim of getting the target to engage in a particular conduct is really capable of generating credible pressure to that end. In that regard, it could be that the more protracted and invasive the campaign, the more credible the pressure. The problem is that online disinformation campaigns rarely, if ever, generate credible pressures to act or abstain from acting in a certain way.

Also, if no answer whatsoever can be given to the preliminary question that is; which act the active party expect, or not, the passive party to undertake, as it is arguably the case with most disinformation campaigns, then the communication will not constitute dictatorial coercion. It is nonetheless interesting, for the sake of argument, to appreciate how disinformation can constitute credible pressure.

[486] Noel, *Supra* note 372, at 74.
[487] Jamnejad & Wood, *Supra* note 374, at 348.
[488] Schmitt (2017), *Supra* note 162, at 319.

5.7.1.2 The dissemination of problematic content as generator of credible pressure

For coercion to have any chance of success, the addressee must *feel* that complying with an overt or implicit demand is the least damaging decision. To that end, a prohibited intervention relies on instruments of pressure.[489] The instruments, the means, or methods deployed to apply the pressure can take various forms in international relations. The pressure could be exercised through diplomatic, economic or military means, what matters is that the means of pressure are capable to *force* the actions of the target.[490] Also, Noel underlines quite clearly, the means do not need to be themselves illegal for an intervention to be deemed illegal; "[...] what Law condemns, is the desire to impose its will, even by means in principle lawful. It is the end that corrupt the means."[491] Consequently, the deployment of instruments of pressure with the aim of forcing a state to comply with its international obligations is lawful.[492]

More than ever, non-forcible means of pressure represent the crushing majority of state interactions.[493] The marked preference of states for non-forcible intervention was a foreseeable phenomenon in an international system where the use of force is notoriously framed by law and in principle centralized in the hands of a unique universal organization. Graber already deplored in 1959, that because these interventions were; "[...] less crude and apt to cause less resentment in the victim, and are less likely to lead to war, interventionist states favored them increasingly over use of military force."[494] This also applies to disinformation campaigns, as they do not involve the use of military force.

Cyber disinformation, as the dissemination of false and/or distorted news across cyberspace, can in theory be used as a means of pressure against a state to win concession if the purpose of a disinformation campaign is sufficiently obvious to the target. The fact that lying is not *per se* an unlawful act under international law makes no difference either, in the characterisation of an illegal intervention. An ultimatum issued by a superpower that is aware of its superiority over a dependant state is illegal because it is inherently coercing the recipient to take unwanted decisions regarding its reserved domain, yet the acting state would have had recourse to nothing

[489] See Corten et al. *Supra* note 375, at 162; Moynihan, *Supra* note 470; Jamnejad & wood, *Supra* note 374.

[490] The status of economic coercion remains disputed in international law. Generally, as the ICJ observed in its Nicaragua case, the decision to halt trade, even if it is felt as an undue pressure by the other party, will not be constitutive of a prohibited intervention. See in that sense ICJ, *Supra* note 196, at 126, §245. See also Corten, *Supra* note 375, at 161.

[491] Noel, *Supra* note 372, at 82. [Translated by author]

[492] Ohlin, *Supra* note 459, at 1589; Kohen, *Supra* note 432, at 161.

[493] See generally Damrosch, *Supra* note 431.

[494] Graber D. A. *Crisis diplomacy: an history of U.S. intervention: policies and* practices (Washington: Public Affairs Press, 1959), at.3, cited in Noel, *Supra* note 372, at 82.

else than the usual diplomatic channels to verbally exteriorise the bargain. What does matter in such situation is the context in which the ultimatum is communicated. A flagrant imbalance of power between the active and the passive party being conducive of pressure, the mere formulation of an ultimatum addressed at a weaker state that is obviously not in a position to resist is the same as to imposing its will on the weaker state and thus, is coercive.

Evidence of credible pressure will be obvious in the presence of direct threats addressed by a state to another. This is the case when the targets of coercion are the individuals vested with decision-making powers *vis à vis* the reserved domain of the state. Take for instance a state formulating a proposal to another and pretending to have the means and the determination to take non-forceful yet damaging actions against the addressee in case of non-compliance.[495] Also, in that case, the threat needs not be successful to amount to a prohibited intervention.[496] In contrast, the communication of lies and misleading information to a state in order to persuade it not to sign a treaty or withdraw a particular national policy does not constitute a prohibited intervention. Indeed, although the intent of the disinformant may be perceivable in that case, the addressee is not compelled but persuaded to take a particular decision.[497]

Yet, in the same way as terrorism and propaganda, disinformation campaigns mostly rely on the feelings of populations.[498] As the principle of non-intervention applies to states, the exposition of the population to lies and misleading information must be at least capable of producing coercive effects on the capacity of the individuals vested with decision-making powers to take decisions *vis à vis* the reserved domain.[499] Two situations are therefore distinguishable:

[495] As argued above, such a communication may also violate the use of force prohibition. See generally Corten et al. *Supra* note 375.

[496] Schmitt (2017), *Supra* note 162, at 323.

[497] See Watts, *Supra* note 423; Lahmann, *Supra* note 92.

[498] Anderson, *Supra* note 485.

[499] Schmitt and a majority of the Tallinn Manual experts argued that in absence of direct coercion stemming from a cyber-enabled activity, reasoning in terms of "indirect causation" of coercive effects. (See Schmitt; *Supra* note 211, at 51; Schmitt (2017), *Supra* note 162. At 320. See also Xuan, *Supra* note 178.) The concept of "indirect causation" of coercive effect has however been criticized in the literature, notably because this would drastically stretch the scope of the non-intervention principle. (See Zhao, *Supra* note 117, at 56.) A way to avoid casting overly wide nets in this case is to require that the pressure generated by an act is at least credible. As shown in the following pages, this remains unhelpful to assess disinformation campaigns.

- When disinformation targets voters during national election processes or *referenda* (hereinafter electoral disinformation)

In democracies, the representatives of a government are not the only individuals vested with decision-making powers to take decisions *vis à vis* the reserved domain. Voters are given similar responsibility at the occasion of *referenda* and, most obviously, national elections. By virtue of the principle of self-determination, all people are free to "determine their political status and freely pursue their economic, social and cultural development." [500] The right to self-determination has an external and an internal aspect. The latter establishes a right for "the people of a state to govern themselves without outside interference."[501] National elections are the dedicated institutional processes through which peoples are asked to exercise their right to internal self-determination by electing the government that will accomplish the next mandate. Doing so, they lend legitimacy to the representative they judge best suited to govern the nation. It is by definition through national election processes that peoples are exercising their sovereignty. From that perspective, the principles of sovereignty and self-determination are inherently intertwined. [502] Hence, at least to the extent that democracies are concerned, individuals are sovereign, and express this sovereignty through various ways. They may do so, through exercising their right to protest, to political participation, and by exercising their rights to vote. Among these expressions of sovereign will, the choice of political organisation through the vote is by definition, akin to a decision-making power to take decision *vis à vis* the reserved domain.[503]

Therefore, a communication campaign containing messages explicitly threatening voters of the consequences of their choice if they pick the wrong candidate may be seen as a coercive intervention if the scale and coordination of the campaign reached such a level that it attained a substantial quantity of voters, and if the campaign mostly contained explicit and credible warnings against the idea of voting for the wrong candidate.

France used such coercive methods against American voters during the 1796 US elections opposing the democratic republican Thomas Jefferson to the Federalist John Adams. Convinced that a Federalist American government could mean the death of the young Franco-American

[500] See notably ICCPR, *Supra* note 118, Article 1.
[501] Hurst Hannum, "Legal Aspects of Self-Determination" Encyclopedia Pricetoniensis.
[502] "[t]he principle of non-intervention thus protects this aspect of self-determination to wit, the free and authentic expression of will by the people as well as the conditions that enable the people to form authority and will and to make free choices." Tsagourias, *Supra* note 409. Ohlin, *Supra* note 459, at 1595.
[503] See in that sense Damrosch, *Supra* note 431, at 37.

alliance, France undertook to mobilize the American public opinion in support of Jefferson. The French minister to the US, Pierre-Auguste Adet, was tasked to influence the election result.

In reality, France's influence activities in America dated back from the negotiation of the Jay Treaty in 1794 between US and England under the Washington administration. [504] As the Treaty was interpreted as a first step towards a normalization of the relations between the two signatories, France experienced the Jay Treaty as a blow to the Franco-American alliance. Its incapacity to prevent its ratification pushed the French representative Adet to intensify its influence effort. The influence reached its peak during the 1796 elections, eventually turning into a blatant foreign interference in the third American presidential elections. Indeed, in an effort to get Jefferson elected, Adet published several manifestos in a widely-distributed journal titled *Philadelphia Aurora*, calling on French sympathizers to show their support for France and the Republican candidate, occasionally referring to the possibility of a war with France if Adams happened to be elected. He reiterated the threat later in the course of the election, in what constituted a blatant intervention into the American politics.[505] In October 1796, French privateers started attacking American ships at sea, which marked the start of the Quasi-war between the US and the French navy. But regardless of the events that followed the election, the act of pretending that electing the "wrong" candidate would ignite a conflict was a clear demonstration of an early use of coercive methods against a state with the view to impede on a sovereign process and as such, a blatant violation of what, later will be conceptualized as the prohibition of coercive interventions. The conditional threat publicly expressed by Adet was widely circulated a republican journal. US relationship with France gradually deteriorated under Washington, notably due to the signature of the Jay Treaty. This had for effect to anger a substantial amount of pro-France voters, who were easily influenced by Adet's further actions. The overall context of a France already at war against England could have also helped adding credibility to the threat, thereby literally forcing some electors to choose Jefferson in fear of a war. [506] It bears noting that the fact that Adams still got elected does not contradict the assessment here. In attempting to coerce Americans voters, France showed its determination to impose its will on a foreign state.

The 1796 French election meddling in the US differs in many ways from what is today experienced by liberal democracies around the globe. Adet was a well-known political figure.

[504] Alexander DeConde, "Washington's Farewell, the French Alliance, and the Election of 1796." (1957) 43(4) *The Mississippi Valley Historical Review* 641, at 642.
[505] *Ibid.* at 653.
[506] *Ibid.*

He managed to build good relations with American politicians and scientists. Most importantly, Adet's official functions were known to the public, its manifestos were authentic, and its threats were expressed in the open. A credible pressure to get the audience to behave in a certain way, or voting for a particular candidate, would be obvious if the audience is being openly threatened to vote for a specific candidate by a legitimate figure, as it occurred during the French meddling in the 1796 elections. Conversely, underground efforts to spread rumours would hardly reach the coercion threshold in absence of threats. Notwithstanding, according to the traditional meaning of coercion, to be credible, "coercers are still required to demonstrate their resolve"[507], which is impossible outside the context of explicit threats.

The above example highlights the similarities between dictatorial coercion and prohibited threats to use force. In addition to constituting an impermissible interference into the sovereign affairs of the US, it is likely that Adet's communications would today be denounced as a violation of Article 2§4. Unsurprisingly, when contemporary episodes of electoral disinformation did not involve the sending of explicit threats to voters, as during the 2016 US election meddling, "the majority concluded that Russia's actions did not fulfil its conditions in particular that of coercion."[508]

- When disinformation campaigns target a state outside the context of elections

Outside the context of elections, dictatorial coercion will have to be appreciated in the context of interstate relations. A way to establish the credibility of the pressure generated by a disinformation campaign targeting the citizenry of a state involves checking an array of contextual and subjective factors ranging from the characteristics of the target audience to the scale of the communication campaign, its duration, the internal context, and the capacity of the information space of the target to naturally absorb and repel disinformation.[509] Hence, a disinformation campaign that inundates the information space of the target with certain types of narratives such as inflammatory speeches, may be experienced as a form of pressure to act or not to act by the target state, provided that it knows what behaviour is expected from it.

Admittedly, it might be easier to determine the credibility of the pressure when negative consequences have already materialized within the territory of the target, but the challenge in such a situation would be for the victim to prove that current internal disruptions are directly

[507] See Gomez, *Supra* note 473.
[508] Tsagourias, *Supra* note 409, at 3.
[509] Jamnejad & Wood, *Supra* note 374, at 348.

caused by the disinformation campaign in question. On this point, the further away the causation of the alleged requisite effect from the immediate effect of the operation, the less persuasive would be the claim according to which the operation has been designed to force the will of the target state. Unsurprisingly, it is extremely difficult to prove that online disinformation caused a particular coercive outcome.[510]

But even when a sufficient causal nexus is present, the dictatorial reading of coercion requires, by definition, that the target state knows what particular action is expected from it, and it must fear that deleterious measures will be taken against it if it does not comply. In that sense, arguing that disinformation addressed to a population are precisely intended to cause a state to act unwillingly is barely possible. Admittedly, the target state may have an idea of which of its actions has to be reversed to please the disinformant, but even that is uncertain. In fact, as argued above, for coercion to work, the target should not have to speculate on the expected outcome.

In conclusion, dictatorial coercion, *as the act of compelling a state to act or not to act in a certain way*, seldom manifests in online disinformation campaigns.[511] A state exposed to disinformation will be able to perceive what is expected from it only in very rare situations, where the population of a state is in some way manipulated by an external power to carry out its own will, or when the context of the disinformation is sufficiently clear that the target state understands what is expected from it. Insofar as it is barely provable that a disinformation campaign was exercised with the intent to coerce a state to act or refrain from acting a certain way *vis à vis* its reserved domain, the traditional meaning of coercion sets unreachable thresholds for disinformation campaigns, in addition to putting an extremely heavy burden of proof on the target state.

5.7.2 Non-dictatorial coercion and online disinformation campaigns

The above investigation has shown that dictatorial coercion rarely, if ever, occurs in the disinformation context.[512] In fact, some have stated clearly that coercion "is not a tenable standard for restraining the use of cyber means to influence or otherwise restrict sovereign

[510] See Zhao, *Supra* note 117, at 55.
[511] See notably Moulin, *Supra* note 433; Ido Kilovaty, "The Elephant in the Room: Coercion" (2019) 113 *AJIL Unbound* 87; Kilovaty, *Supra* note 81; Lahmann, *Supra* note 92, at 38.
[512] See *Ibid.*

functions"[513] and that most cyber operations do not aim at compelling a state to act—or to refrain from acting—in a given direction. They deprive actors of control and render them powerless instead."[514] This led some commentators and states to articulate less stringent formulations of the concept of coercion.

A review of recent and some less-recent publications touching upon the non-intervention principle reveals that coercion is not always understood as a dictatorial behaviour in the sense of *compelling a state to act or abstain from acting in a certain way*, but that a violation of the principle of non-intervention is also constituted when the target is denied control over a matter upon which it is the sole entity to decide.[515] Schmitt for instance recently noted that "[F]oreign cyber activities that deprive a State of its ability to act vis-à-vis the *domaine réservé* are almost always coercive. They make it objectively impossible or substantially more difficult for the State to pursue a particular policy or activity, [...]"[516] Moynihan similarly interprets coercion as a "pressure on the victim state to deprive the target of its free will in relation to the exercise of its sovereign powers in order to compel conduct or an outcome with respect to a matter reserved to the target state."[517] The Australian Government defined a coercive intervention as one that "effectively deprive another state of the ability to control, decide upon or govern matters of an inherently sovereign nature."[518] More recently, at the occasion of a speech at Chatham House, the UK's Attorney General, Suella Braverman went further and declared; "[W]hile the precise boundaries of coercion are yet to crystallise in international law, we should be ready to consider whether disruptive cyber behaviours are coercive even where it might not be possible to point to a specific course of conduct which a State has been forced into or prevented from taking."[519]

Thus, according to these views, foreign conduct that generates pressure sufficient to deprive the target of the possibility to perform its sovereign functions is coercive, regardless of whether the coercer wants the target to act in a certain way. It is interesting to assess whether this non-dictatorial reading of coercion really changes something in terms of the legality of state-

[513] Kilovaty, *Supra* note 511.

[514] Moulin, *Supra* note 433; "A prohibited intervention is one that interferes by coercive means (in the sense that they effectively deprive another state of the ability to control, decide upon or govern matters of an inherently sovereign nature), either directly or indirectly, in matters that a state is permitted by the principle of state sovereignty to decide freely." See also in that sense the Autralian government, *Supra* note 470, at 5.

[515] See notably Schmitt (2017), *Supra* note 162, at 318; Tsagourias, *Supra* note 409.

[516] Schmitt, *Supra* note 211, at 51; Schmitt, *Supra* note 422.

[517] Moynihan, *Supra* note 470.

[518] "International Law Supplement to Australia's International Cyber Engagement Strategy", *Supra* note 514, at 5.

[519] *S*uella Braverman, the Attorney General, at Chatham House in London (19 May 2022*)*. <https://www.ukpol.co.uk/suella-braverman-2022-speech-at-chatham-house/>

sponsored disinformation campaigns. To that end, three credible hypothesis involving disinformation are analysed below.

- Non-dictatorial coercion in electoral disinformation

According to the "non-dictatorial reading of coercion", a state-sponsored disinformation campaign directed at voters with the aim of disrupting a vote, or preventing an election process from occurring would constitute a prohibited intervention, provided that the disruption caused is serious enough to significantly affect the capacity of the target to hold elections properly.[520] This would notably concern voting suppression *via* disinformation. Indeed, disinformation activities may eventually amount to voting suppression tactics when targeted at voters of an upcoming or ongoing election. This is so, when a disinformation campaign is centred around a core narrative consisting of false claims that the poll is cancelled, or disinformation campaigns purposefully indicating the wrong location of voting parlours, or providing false instructions about the voting procedure at scale, for the purpose of preventing voters from casting their vote.[521] This would also concern, as observed by Germany, the act of deliberately inciting "violent political upheaval, riots and/or civil strife in a foreign country, thereby significantly impeding the orderly conduct of an election and the casting of ballots."[522] Conversely, a disinformation campaign aimed merely at persuading voters to choose one candidate over another, or to lose faith in the electoral system, as this was mostly the case during the famous 2016 US election meddling[523], although striking in the reserved domain of the target state, would hardly amount to coercive intervention because it does not bar the target state from holding elections. Also, the ability to manage its own online information space is not a sovereign prerogative of a state. Arguing otherwise could foster human rights abuses and the fragmentation of the internet.[524] It is useful to note that it is *in principle* impossible for the

[520] *Ibid.* see also Egan, *Supra* note 309, at 174; Federal Government of Germany, *Supra* note 260.
[521] *Ibid.*
[522] Federal Government of Germany, *Supra* note 260. It is worth noting that whether it effectively disrupts sovereign processes, the deliberate incitement of violent political upheaval, riots and civil strife is likely to amount to a prohibited intervention for the simple fact that explicitly inciting violences constitutes an illegal intervention under international law, and even a prohibited use of force in certain circumstances. One may thus wonder why Germany made reference to electoral processes here.
[523] See ODNI, *Supra* note 41.
[524] It should be noted, however, that that the decision to disconnect from internet as a last resort solution to a situation threatening the survival of a state constitutes a sovereign right. In that sense, a disinformation campaign that forced a state to disconnect entirely from the global internet, for example, in order to preserve an ongoing electoral process from foreign interference could also be understood as a coercive act, provided that the severity of the campaign was such that the only available way for the target state to preserve its political integrity is to disconnect from the global information space for the time of the process. See in that sense Schmitt (2017), *Supra* note 162. at 13.

defending state to act on the online information space to mitigate the effects of an alleged malign influence attempts *via* content removal, whatever the internal context, unless the said content amounts to hate propaganda or is of such a nature that it threatens the survival of the said state.

As argued earlier, another way to appreciate coercion in the context of electoral disinformation is to focus on whether the voters themselves are deprived of the ability to decide for which candidate to vote.[525] Tsagourias argues that this happens when an external power attempts to exercise control over the cognitive environment of the voters; "the intervening state controls not only the peoples' cognitive environment within which authority and will are formed but also their choices. It also controls the authority and will of the government that emerges."[526] This view however conflates coercion with manipulation and persuasion. *Coercion occurs at the moment of the formulation of the will, not during its formation*.[527] Furthermore, at an individual scale, disinformation tactics do not require the external force to *directly* control the target. Rather, the influencer merely sets the stage for the taking of particular decisions without resorting to "sticks" or "carrots."[528] In other words, the target is deceived (or persuaded) of the necessity to take a particular action. This relationship is sometimes called "unconscious control", or "reflexive control." Then, if there is indeed a form of control remotely exercised on a subject, the victim does not act under pressure.[529] Of course, one might be tempted to counter that the manipulated individual is in that case deprived of its capacity to exercise control over its own decisions, as the latter are induced by an external influence. Yet, it is hard to see deprivation of control when voters are not *compelled* to take suboptimal decisions. In the end, deceiving or persuading voters into choosing a particular candidate does not deprive the target from its ability to decide upon sovereign matters (in that case, the choice of a political organisation by the individuals vested with decision-making powers). The target ultimately retains its capacity to decide whom to vote for, unless it is explicitly threatened to make a particular choice, or dissuaded from casting its vote.[530]

With that being said, a good case might be made with regard to the fact that the phenomena of information bubbles and ideological compartmentalization on social media drastically

[525] See Schmitt, *Supra* note 422, at 748; Lahmann, *Supra* note 92, at 38.
[526] Tsagourias, *Supra* note 409..
[527] See notably Xuan, *Supra* note 178.
[528] See notably, Kilovaty, *Supra* note 511. In the context of state-sponsored disinformation campaigns, control over the addressee is, at best, exerted in a very distant manner. However, information activities are particular in that they mostly rest on manipulation and persuasion.
[529] See notably Zhao, *Supra* note 117. At 57.
[530] Lahmann, *Supra* note 62, at 202; Xuan, *Supra* note 178 ; Kilovaty, *Supra* note 511.

enhance the level of influence an external agent may exert on an audience today. In contrast to propaganda during the cold war, online disinformation may be more efficient at making the target audience impervious to contradictory streams of information, thus accentuating the degree of influence that an external force will be exercising upon the audience, and making it harder for a population to resist the narrative it is fed.[531] Yet, reaching such a level of control over foreign audiences would certainly require at least, a very long time period and incredibly sophisticated means of influence that a foreign state is unlikely to deploy in the relatively narrow time frame of an election period.

It bears noting that Tsagourias seems to give primacy to the intent of the foreign power rather than the result of its actions in his assessment. Accordingly, the act of targeting foreign voters with disinformation denotes an interventionist intent to control the voters and the outcome of the election process that should fall under the ambit of the non-intervention principle. However, as it does not address the question of *how* the act is experienced by the target, this approach does not appear to focus on coercion *stricto-sensu*.

Alternatively, Schmitt advances the argument that "the covert nature of the troll operation deprived the American electorate of its freedom of choice by creating a situation in which it could not fairly evaluate the information it was being provided."[532] Albeit relevant, such a view does not address the problem of disinformation spread in a transparent way, by real individuals acting in the open, which is equally capable of influencing a vote.[533] For example the US-led investigation regarding the 2016 election meddling identified several Russian media that overtly participated in the campaign alongside internet trolls.[534] Although the covert nature of a disinformation campaign would undoubtably count as an evidence of malicious intent, the idea that the use of false identity makes a difference in the assessment of coercion is unconvincing.[535]

In reality, the above shows that discussions on coercion in the disinformation context tend to deviate from the necessity to assess coercion. Placing the concept of loss of control approach to coercion promulgates the view that, in the particular case of disinformation targeting voters' attitude, there are only two situations where voters would be clearly "coerced" in the sense of a deprivation of control over a choice:

---

[531] On this point it has been noted that an influence that can be resisted is not coercive, see Jamnejad & Wood, *Supra* note 374.

[532] See Schmitt, *Supra* note 211; See also the same idea at Ohlin, *Supra* note 459; Tsagourias, *Supra* note 409.

[533] See in that sense Lahmann, *Supra* note 62, at 205.

[534] See ODNI, *Supra* note 41, at iii.

[535] See notably Zhao, *Supra* note 117, at 56.

- *When a communication is specifically intended to make voters miss the election day, or vote the wrong way* (by recommending to use phone message instead of casting votes at the usual voting parlours)[536]
- *When a substantial number of messages in a communication campaign is formulated as explicit and credible threats to make a particular choice*

Absent conditional threats, a loss of control over the choice of a candidate would occur when a massive disinformation campaign successfully overwhelms the information space of the target state in a way that render other candidates completely invisible to the audience. Take for example a campaign directed at a state that is in the impossibility to repel the narratives because it is in extreme technical disadvantage compared to the disinformant.

Conversely, a disinformation campaign merely intended to manipulate the audience into voting for a specific candidate, or thinking that the election is rigged, or losing faith in the electoral system is *in principle* non-coercive, unless it is proven that the conduct in question was intended to deprive the victim of its ability to making a choice, and that formal and contextual factors reveal that the conduct in question could generate credible pressure to that end.[537] It is unclear how such point could be made with regard to disinformation activities. Persuading voters to change their attitude towards a particular candidate or the election does not result in actual deprivation of control over the choice of a candidate. A case on point can be seen in the "Macron Leaks" in 2017. The French government for example mentioned the "natural defiance" of the French people, and considered that thanks to their conscience, the French did not trust any of the information the disinformant disseminated.[538] Whatever might be the intent of the external force, making the case for coercion when the victim is not *forced* into a course of action is unconvincing.[539]

Therefore, in the particular case of electoral disinformation, focussing on the loss of control aspect of coercion may not make much of a difference compared to dictatorial coercion, at least to the extent that disinformation campaigns targeting voters' attitude towards a candidate in particular, or the election system in general, are concerned. Some might see this as an

[536] See notably in that sense Egan, *Supra* note 250; Paul C. Ney Jr., remarks at U.S. Cyber Command Legal Conference (Mar. 2, 2020); Roy Schöndorf, "Israel's Perspective on Key Legal and Practical Issues Concerning the Application of International Law to Cyber Operations" (2021) 97 *International Law Studies* 395, at 403.

[537] Jamnejad & Wood, *Supra* note 374, at 348.

[538] See notably Jean-Baptiste Jeangène Vilmer, "The "Macron Leaks" Operation: A Post-Mortem" (2019) Atlantic Council Report, <https://www.atlanticcouncil.org/in-depth-research-reports/report/the-macron-leaks-operation-a-post-mortem/>

[539] Xuan, *Supra* note 178; Kilovaty, *Supra* note 70.

unsatisfactory conclusion. Admittedly, election meddling through coordination of inauthentic behaviours for the purpose of deceiving voters into electing a particular candidate or more broadly, lose faith in national institutions, is a particularly daring form of interference when executed by a foreign power. It is also by essence an attack against the exclusive right of peoples to determine their own political system without external interference, and a potential abuse of their right to freedom of thoughts and opinions.[540]

In light of this, moving away from the rigid criterion of coercion could certainly help capturing more disinformation campaigns.[541] For example, because coercion is hard to establish in such a setting, and that "sovereignty" in the legal sense of the term may not suggest the same thing as the will of a nation, Ohlin focusses exclusively on the right to self-determination.[542] Lahmann similarly evacuates coercion from the issue of electoral disinformation and reconciles the enjoyment of internal self-determination with this of sovereignty.[543] As will be stressed in the next chapter, nothing really precludes the operability of an autonomous right to self-determination in that context. In the end, the approach seems to boil down to how persuasive the observer finds the argument of self-determination alone.

- Non-dictatorial coercion in health disinformation

Pursuant to the non-dictatorial reading of coercion, the widespread dissemination of false and misleading information about vaccines could amount to a prohibited intervention if the target is unable to carry out its health policy properly, and that this was the result expected by the disinformant.[544] The recent "Fazze disinformation case," could have constituted an example on point if it had succeeded at triggering widespread vaccine hesitancy abroad. Briefly, on 24th May 2021, some European influencers received suspicious emails from a mysterious public relation agency named Fazze, asking the recipients to publish short videos denigrating the Pfizer vaccine on YouTube, Instagram and other well-known social media in exchange for significant amounts of money, and requiring them not to reveal the arrangement.[545] The campaign did not seem to focus exclusively on a European audience. Influencers from India and Brazil were also

[540] More on that point in the next chapter.
[541] See in that sense Kilovaty, *Supra* note 70; Kilovaty, *Supra* note 511; Moulin, *Supra* note 433.
[542] Ohlin, *Supra* note 142.
[543] Lahmann, *Supra* note 92. at 40-41; Lahmann, *Supra* note 62, at 206.
[544] See generally Schmitt & Milanovic, *Supra* note 326.
[545] Jon Henley "Influencers say Russia-linked PR agency asked them to disparage Pfizer vaccine", (25 Mai 2021), *The Guardian*, <Https://www.theguardian.com/media/2021/may/25/influencers-say-russia-linked-pr-agency-asked-them-to-disparage-pfizer-vaccine>.

approached by the agency. Among the addressees, a French influencer followed at the time of writing by 1,2 million users, was the first to raise the alarm in the face of what turned out to be an attempt at spreading disinformation about the Pfizer vaccine to a young European audience under the cover of the notoriety and the credibility of legitimate content creators. Subsequent investigations firstly revealed several irregularities surrounding the agency in question, which, contrary to what was indicated on its website, did not seem to be based in London. Further, the profiles of the main employees of the company, all showing a connexion to Russia, vanished from social media after the revelations. Later discoveries pointed to the implication of pro-Kremlin personalities in the disinformation campaign, along with troubling redundancy between the data used to disprove the effectiveness and the safety of the Pfizer vaccines that the influencers were asked to communicate to their audience and the information shared on the twitter account of the Sputnik V vaccine. At the time, the French counterintelligence authorities suspected that the Russian Federation had a responsibility in the conduct of the campaign.[546] Of course, two unanswered questions make it difficult to assess this case. First, similar operational pattern and the fact that the operation may indeed profit Russia are, from a legal perspective insufficient in themselves, to betray the involvement of the Russian Federation. Second, even in the eventuality that the conduct was attributed to the Russian Federation, another uncertainty lies in the fact that the intent of the communicator was particularly hard to infer from the communications. It seems indeed that the focus was put on denigrating the Pfizer vaccine especially rather than fostering anti-vaccine sentiment.

Nevertheless, assuming that the conduct is attributable to a state, and that the campaign had for effect to severely disrupt the rollout of COVID19 vaccines to the point that the victim state would not be able to vaccinate a sufficient portion of its population during the pandemic, then, under this more flexible interpretation of coercion, the communication campaign would have constituted a prohibited intervention.

Similarly, a disinformation campaign aimed at fomenting distrust towards the national health authorities at a time of pandemic may constitute a prohibited intervention if sufficient causal nexus between a disinformation campaign and a subsequent rise in uncooperative behaviour can be ascertained. This coincides with the above-cited view of the government of New-Zealand, who singled-out disinformation operations significantly undermining a state's

[546] Matthew Dalton, "France Suspects Russian Role in Campaign to Discredit Pfizer Vaccine*"* (25 May 2021), *The World Street Journal*, <https://www.wsj.com/articles/france-suspects-russian-role-in-campaign-to-discredit-pfizer-vaccine-11621978289>.

public health efforts during a pandemic as a possible violation of the non-intervention principle.[547]

Yet, the same as for appreciating *in concreto* the impact of a disinformation campaign on an election process, it might not always be possible, in practice to ascertain that the paralyzing raise in non-cooperative behaviours regarding the administration of vaccine in a territory is caused by a foreign health disinformation campaign. For all that, the harmful intent of the author of the disinformation campaign might very well be apparent.

- Non-dictatorial coercion in divisive disinformation

The last case has been singled out in the literature and in the first chapter of this study as one area of high difficulty when it comes to assess the legal status of state-sponsored disinformation campaigns.[548] This extremely broad category concerns what has sometimes been defined as "malign influence", that is, disinformation campaigns which do not appear to have been intended to *interfere* with a particular state function, but are damaging nonetheless. It includes "cyber activities involving information or disinformation that does not affect how the election is carried"[549] and extend to any other episodes of disinformation merely aimed at affecting social cohesion.

It bears noting that electoral disinformation, arguably the easier form of disinformation activity to assess, will often generate long term consequences that may be felt long after the election process has passed. Electoral disinformation is capable of enhancing the disruptive potential of later disinformation campaigns constituting mere "malign influence" activities focused on exploiting the distrust, dissatisfaction and the animosity that previous electoral disinformation have eventually triggered.[550] In that sense, electoral disinformation and disinformation targeting society at large by exacerbating social division are intrinsically linked, and their effects may accumulate over time.

Disinformation campaigns having for no other apparent purposes than gradually eroding the social bounds tying a society together are subversive in essence. Yet, they hardly fall in any legal category. Such elusive communication campaigns are common and particularly feared amongst states, especially democracies, that are left with few legal and institutional tools to

[547] The Government of New Zealand, *Supra* note 257.
[548] Schmitt, *Supra* note 422, at 754
[549] *Ibid.*
[550] Delerue, *Supra* note 81, at 256.

repel them.[551] These forms of disinformation campaigns are long-term oriented strategies that are likely to remain undetected as they unfold until an internal situation reached stability-threatening proportions. An example of such type of disinformation is the disinformation deluge that affects Moldova since the Russian invasion of Ukraine in 2022.[552] The campaign, reportedly conducted so as to exacerbate social division, notably with regards Ukrainian refugees was described by the president of the Republic of Moldova as a "coordinated effort."[553] Examples of disinformation campaigns aimed at exacerbating social division can be found in the anti-migrants disinformation that affected Europe in 2016 and 2017[554], or the race-related communication campaign that targeted the US during the 2016 elections period.[555]

Schmitt does not consider such disinformation campaigns as amounting to prohibited interventions. Instead, he notes that they could potentially amount to violations of state sovereignty, notably if they cause substantial death, destruction or injury in the territory of the target state. With regard to non-intervention, while it is true that in absence of a clearly expressed request for compliance to a demand made by the active party, the waging of disinformation campaigns to "drive wedges"[556] in a society hardly constitute a dictatorial interference, could they nevertheless constitute a pressure undermining the capacity of the victim state to control a process falling under its reserved domain?

Non-dictatorial coercion requires that the conduct in question is capable of depriving the target of its ability to decide upon certain matters. In that sense, a particularly tensed internal context may generate a "credible pressure" felt by the state, making it substantially more difficult for it to pass certain policies. In other words, it *can* have a coercive effect. However, that is a long way from saying that the state is effectively deprived of the ability to act upon its reserved domain, and that the act was designed for such a result. In other words, if such disinformation campaigns may effectively generate internal disruptions, and eventually

[551] See generally Wigell, *Supra* note 95; Rid, *Supra* note 94.

[552] Morgan Meaker, "An 'Explosion' of Anti-Ukraine Disinformation Is Hitting Moldova" (8 April 2022), *Wired,* <https://www.wired.com/story/moldova-disinformation-war-ukraine/>.

[553] Liliana Barbăroșie, "Maia Sandu: campania de ură contra refugiaților ucraineni e rodul unui 'efort organizat'" (4 March 2022), *Radio Europa Libera Moldova,* <https://moldova.europalibera.org/a/maia-sandu-campania-de-ur%C4%83-contra-refugia%C8%9Bilor-liliana-barb%C4%83ro%C8%99ie-radio-europa-liber%C4%83/31736056.html>.

[554] See notably "The pro-Kremlin narrative about migrants" (16 May 2017) *EUvsDisinfo*, <https://euvsdisinfo.eu/the-pro-kremlin-narrative-about-migrants/?highlight=rape>; Stefan Meister, "The "Lisa case": Germany as a target of Russian disinformation" (25 July 2016) *NATO Review*, <https://www.nato.int/docu/review/articles/2016/07/25/the-lisa-case-germany-as-a-target-of-russian-disinformation/index.html>.

[555] Report of the select committee on intelligence US senate on Russian active measures campaigns and interference in the 2016 US election Volume 2: Russia's use of social media with additional views, 116th Congress US senate, Report 116-XX

[556] Rid, *Supra* note 94.

fragilize a state in the long term, the resulting disruptions can at best account for contextual factors contributing to the coerciveness of a subsequent conduct. In other words, disinformation activities constituting *mere* "malign influence" would not on themselves amount to prohibited interventions.

Notwithstanding, the assessments become highly speculative in absence of clear evidences that the disinformation campaigns caused social instabilities.[557] Most importantly, it is unclear how the reserved domain requirement would be fulfilled in such a situation. Disinformation campaigns merely aimed at gradually severing the bounds that tie a society together, radicalizing the audience, or fomenting distrust towards the national institutions by planting conspiracies, fake and manipulated stories, libellous hate speeches and rumours may not be pursuing clear objectives at all. Without the possibility to single out specific sovereign competences upon which the target state is specifically deprived the capacity to act, the principle of non-intervention in its current formulation does not seem apply to disinformation campaigns of such kind.[558]

An alternative to the principle of non-intervention in the context of divisive disinformation could be to assess the phenomenon through the lenses of human rights law, particularly, the principle of non-discrimination.

5.7.3 Remarks on the incidence of "coercive" message on the overall legality of a campaign

Assessing the legality of disinformation campaigns against the principle of non-intervention has highlighted yet another issue. As shown earlier in this work, a disinformation campaign is a complex act. It is an aggregate of more or less targeted messages released over time and geared towards achieving the same result. That is not to say, however, that the messages are all similar, nor that the campaign is exclusively covert. In other words, disinformation campaigns are not purely homogeneous acts.[559] A case on point is the 2016 US election meddling, which involved various deception methods in order to mislead the US electorate. In the US election

[557] See Hollis, *Supra* note 334.

[558] Xuan, *Supra* note 178. Most often with non-interventionist disinformation campaigns, the short-term repercussions on the sovereignty of the target will remain undiscernible, which makes it hard to speak of direct infringement on sovereignty. Surely, detrimental effects on sovereignty may still emerge in the long term, but may appear self-inflicted. It transpires that this form of disinformation leans towards influence rather than interference. In this context, online disinformation is more about influencing mass consciousness than micro-targeting decision makers. (See Keir & James & Seaboyer, *Supra* note 528, at 8). A disinformation campaign that does not constitute a clear-cut attempt to interfere with one specific sovereign process should not amount to foreign interference, by incidence, it cannot be interpreted as a prohibited intervention into the reserved domain of the target state.

[559] See Xuan, *Supra* note 178; ODNI report, *Supra* note 41.

case, while some communications were clearly intended to prevent voters from voting, not all were conceived that way.[560] A great portion of the messages was merely aimed at fostering distrust in a particular candidate or in the election system as a whole, or sowing confusion, without explicitly pushing voters towards the wrong voting procedure.[561] The question thus arises, as to how to assess a disinformation campaign that, when approached in its globality, may not constitute a coercive intervention but still contains some messages that straight-up threatens voters from making a particular choice and dissuades voters from casting their votes? The answer seems to be that the "coercive" messages cannot be said to have particular impact on the evaluation of the coercive nature of the whole campaign, unless they compose a substantial part of it. A majority of the literature did not consider the 2016 US election meddling as a coercive intervention into the politics of the US. Instead, as a commentator deplored, the 2016 state-sponsored and state-condoned electoral disinformation left observers "with an overall impression of illegal conduct, but without a clear and unambiguous doctrinal route towards that conclusion."[562] The above assessments seem to confirm this view.

5.8 Alternative views on non-intervention

The introduction of the concept of the denial of control over protected competences is interesting, because it gets rid of the unpractical requirement according to which the coerced must know which particular behaviour is expected by the coercer. This is not much of a stretch from the concept of dictatorial coercion because a loss of control is by definition an effect of coercion.[563]

Unfortunately, it has also been shown that the non-dictatorial reading of coercion makes little difference in practice, when it comes to assess the legality of disinformation activities sponsored by a state and affecting the interests of another. Even though focussing on the more concrete loss of control criterion might be preferable, for that it facilitates the assessments of coercion, the criterion of coercion remains on itself a source of confusion.

There is no agreement in the legal literature as to how to understand coercion in the context of interstates relations. Regarding the application of the non-intervention principle to state-

[560] Nicholas Thompson & Issie Lapowsky, "How Russian Trolls Used Meme Warfare to Divide America" (17 December 2018) *Wired,* <https://www.wired.com/story/russia-ira-propaganda-senate-report/>.

[561] *Ibid*; ODNI report, *Supra* note 41.

[562] Ohlin, *Supra* note 459, at 1593.

[563] For some authors however, such a formulation of the principle of non-intervention is already a departure from the meaning of coercion. See in that sense Moulin, *Supra* note 433.

sponsored disinformation campaigns, an additional hurdle relates to the absence of obvious and sufficiently direct causal nexus between an illegitimate act and a result, or the outright absence of visible effect arising from an unacceptable behaviour. Given that some disinformation campaigns clearly follow interventionist purposes, this state of affair might appear frustrating.

In response, a small portion the literature leans towards an "intensity test" to gauge the coerciveness of non-forceful foreign interferences. For instance, it has been suggested that both "the nature of state interests affected by a cyber operation, the scale of effects the operation produces in the target state, and the reach in terms of number of actors involuntarily affected by the cyber operation in question" would help indicating coercion.[564] Similarly, some commentators have pointed to the fact that a disinformation campaign was not a single action, but a series of acts (disinformation operations) constituting an aggregate that, as such, could be more likely to have coercive effects.[565] By so arguing, these commentators however miss the point of coercion.

Of course, establishing the credibility of the pressure felt by the claimant requires among other, to appreciate some formal and contextual factors surrounding the conduct, such as its scale, but through lending too much credence to formal and contextual factors, there is a risk to disconnect the central question of whether the target has been deprived of its capacity to decide, or straight-up coerced, from the assessment.[566] On this point, a commentator deplored that recent submissions regarding the meaning of coercion emphasized not "whether the victim state was *forced* to decide a matter on which it is generally allowed to decide freely, but whether there was an attempt to affect protected state interests and the *effects* that such an operation produces."[567] More recently, another commentator argued about recent legal propositions that they were "essentially attempts to redefine PNIVIF [The Principle of Non-Intervention and Non-Interference]."[568]

In view of this, one might be tempted to abandon the criterion of coercion for more concrete ones. After all, assessing coercion based on the severity of a conduct essentially equals replacing the "coercion test" by an "intensity test."

The idea that the criterion of coercion should be rejected is not without support. Especially since the advent of cyberspace, the legal literature is filled with legal propositions regarding the inadequacy of the ICJ's interpretation in the context of non-forcible (including cyber-enabled)

[564] Watts, *Supra* note 423.
[565] Rodriguez, *Supra* note 126.
[566] See in that sense Kilovaty, *Supra* note 70; Xuan, *Supra* note 178; Zhao, *Supra* note 117.
[567] Kilovaty, *Supra* note 70, at 171.
[568] Xuan, *Supra* note 178, at 33.

and illegitimate state conduct. For example, Kilovaty called for a re-evaluation of the principle based on the disruption caused by the conduct in question.[569] According to him, a prohibited intervention via cyber means "can be identified by the attack itself as well as the intention behind the attack and its level of invasiveness."[570] A second author has called for a replacement of the coercion criterion by a paradigm of "impermissible manipulation and fraud."[571] Another sees the principle of non-intervention as mere guiding norm from which new specific rules on PYOPS should draw inspirations.[572]

Less recent propositions were also concerned with subtle forms of political interference that may be "as effective as cruder forms of domination", arguing that non-intervention should accommodate human rights values.[573] In the same vein, an author qualified the ICJ's formulation of a prohibited intervention as way too restrictive, arguing "*C'est donc, non pas le critère de la contrainte, mais celui de l'atteinte aux droits souverains qui sert de dénominateur commun à toutes les formes d'actes catalogués d'intervention, ce qui atteste du caractère accessoire du premier critère*."[574] In confusing the coercion criterion with a criterion of "intensity", other authors demonstrated, perhaps unconsciously, the incompatibility of coercion when it comes to assess non-forcible state acts.

Obviously, the scope of the principle of non-intervention is ambiguous. The main reason is that the determination of what amounts to a prohibited intervention in international law is mostly the result of a case-by-case evaluation. Hence, it is difficult to provide an accurate definition of a prohibited intervention, let alone one that is internationally agreed.

With that being said, and taking into account the submissions made by prominent scholars, it is still possible to provide the common features of all prohibited intervention; In general, *a prohibited intervention is a conduct that reflects the determination of its author to impose its will on the target with regard to matters protected by the principle of sovereignty.* What principally distinguishes an illegal interference from an admissible interaction is thus the brazen willingness of the active party, to bring about the conditions leading to a position of strength among the international community, or the conservation thereof, at the detriment of the political independence and/or territorial integrity of one other state or more.

[569] See Kilovaty, *Supra* note 70, at 174.
[570] See Kilovaty, *Supra* note 70.
[571] See Zhao, *Supra* note 117.
[572] See notably Xuan, *Supra* note 178.
[573] Damrosch, *Supra* note 431, at 5.
[574] David, *Supra* note 447, at 353.

Of course, there are some blatantly inadmissible foreign conducts that are relatively easy to qualify. An intervention is primarily a projection of force, a manifestation of hard power. It succeeds when an unconsented transfer of authority from the target to the interfering state occurs. Are thus undoubtfully constitutive of illegal interference those particularly outrageous conducts aimed at getting the target into behaving a certain way through direct or indirect use of force, or the threat thereof in contravention of the UN Charter. The soviet invasion of Czechoslovakia following the Prague Spring in 1968, or the US-backed Bay of Pigs invasion of Cuba in 1961 are clear examples of armed interferences. There are the boldest conducts in term of invasiveness and impact. They both affected the political independence and the territorial integrity of the target states.

Less harmful but no less inadmissible, exhortations assorted with threats of sanctions in case of non-compliance would clearly constitute internationally illegal interventions if the demands relate to the *domain reserve* of the target and if they do not have for purpose to enforce compliance with international law.[575]

Things are not as straight-forwards in the presence of indirect and non-forcible conducts such as PSYOPS. Economic coercion, manipulation tactics aimed at a foreign population in order to undermine social cohesion and other forms of intimidation, blackmail and coordinated harassments thus fall within the vast category of subtler yet inappropriate attempts to get a counterpart into a specific course of action. In this broad category, some conducts may be tailored to influencing decision-making, other may involve targeting a population with divisive information, other may also revolve around the use of economic leverage tactics. Although inappropriate, these conducts are not always internationally wrongful. In other words, these are "grey zone" conduct that defy straightforward legal characterisation. When sustained over a certain period of time, however, they have for effect to cause gradual loss of sovereignty for their target.

Whether the instigator has crossed the line of international legality is therefore established on a case-by-case basis, taking into account the consent of the recipient state, the circumstances of the facts, the characteristics of the act, the inferable intent, and the eventual impact of the act. In some circumstances, a high degree of potential for impairment and a clear dedication to nefariously affects the territorial integrity and/or the political independence of the target could suffice in making the point of a violation of the principle of non-intervention even if coercion *stricto-sensu* does not occur.[576]

[575] See notably David, *Supra* note 447.
[576] See Corn and Helal, supra note 479.

The potential for impairment on the right and interest of the recipient would be assessed based on the characteristics and the *scale* of the act, its level of *invasiveness*, its *duration*, and more generally, the *amount of resources devoted to the realization of its alleged purpose*. The inferable intent, itself, must be interventionist in nature. This means that the act should denote the clear and distinguishable intent of its authors, to undermine the political independence and/or territorial integrity of a state. Conversely, disinformation campaigns that do not threaten in a clear and distinguishable manner a protected interest owned by the target by virtue of the principle of sovereignty cannot credibly be condemned as a prohibited intervention.

With regard to state-sponsored PSYOPS, the claimant must bring clear evidences that a disinformation campaign was clearly and distinguishably aimed at undermining one sovereign function or more[577], or at least, and taking into account the disruptive effect of a disinformation campaign, that the instigator could not have ignored that the act would lead to significant infringements on the sovereign functions of the recipient. As a principle, disinformation campaigns having for *verifiable* consequences the disruption of an election process, the disruption of health policies, or the triggering of internal strife would qualify. In order to avoid misattribution and escalation, the causal relationship between the act and the effect should be clear and provable. In very exceptional circumstances, the absence of sufficiently clear and provable causal nexus between the act and subsequent effects in the territory or the governance of the recipient state would not preclude the application of the non-intervention principle, if, and only if, the nature of the interests at stake, the characteristics of the acts and its duration denote a clear intent from the part of the interfering state to detrimentally affect the sovereign functions of the target state.

Obviously, it is hard to imagine disinformation campaigns alone reaching the level of prohibited interventions, even under this slightly relaxed interpretation of the principle.[578] Gauging the contribution of a disinformation campaign in the occurrence of visible disruption, harm, or damage, generally involves a high amount of speculation. But proving clear intent to impair a specific sovereign function is equally difficult in the context of foreign disinformation, unless the disinformation effort is supported by more brazen acts following the same *premeditated design* that is. Apart from this hypothesis, and in absence of explicit threats or calls for civil strife, the disinformation effort would have to be massive in terms of its scale and duration to ever amount to a prohibited intervention.

[577] See in that sense Kilovaty, *Supra* note 70.

[578] Kilovaty, *Supra* note 70. (Noting that even a relaxed interpretation of the principle of non-intervention that lends more credence on the effects and the intentions of a conduct would not apply to disinformation campaigns.)

6. State-sponsored cyber-enabled disinformation campaigns as breaches of state sovereignty

According to some commentators, the principle of sovereignty makes a better legal basis for assessing non-forceful and non-coercive cyber-enabled conduct than the principle of non-intervention.[579] This legal proposition is based on the premise that the principle of sovereignty constitutes a relatively broad rule susceptible of being breached on its own. The question of the legal value of the principle of sovereignty is however contentious. Assessing the legality of state-led disinformation campaigns against the principle of sovereignty thus entails at the same time, that the contentious question of the legal value of the principle of sovereignty is elucidated. This means establishing whether the principle of sovereignty is a distinct part of a regulatory continuum alongside the principle of non-intervention, or whether it is a norm that is only breached when other correlative rules of international law are themselves violated by a state. The latter explanation would mean that disinformation campaigns hardly, if ever, violate the principle of sovereignty. The former explanation would, in contrast, render the principle of sovereignty applicable to disinformation activities.

6.1 The principle of state sovereignty

The origins of the concept of state sovereignty can be traced back from 1576, and the seminal book of the philosopher Jean Bodin; *Les 6 livres de la république*. State sovereignty later materialized in the Westphalia Treaties of 1648. Since then, the principle occupied a central place in international relations and international law. Although its content substantially evolved through the ages, sovereignty remains a pivotal principle of international law.[580]

Sovereignty is an attribute and a precondition to become a subject of international law, and to enjoy international rights and obligations.[581] Pursuant to state sovereignty, a state has a legal right to exercises prerogatives of sovereign authority over a territory, to the extent that it is not at the expense of the authority of another sovereign state.

[579] See notably Watts, *Supra* note 455.

[580] "The Organization is based on the principle of the sovereign equality of all its members." 1945 UN Charter, Article 2§1,

[581] *Corfu Channel* (United Kingdom of Great Britain and Northern Ireland v Albania), (1949) ICJ, judgement. Individual opinion of Judge Alvarez, 43.

The beginning of the 20th century saw the concept taking hold as a definite rule of international law following the oft-cited *obiter dictum* of the then-arbitrator at the Permanent Court of International Justice, Max Huber. The latter held in 1928; "[S]overeignty in the relation between States signifies independence. Independence in regard to a portion of the globe is the right to exercise therein, to the exclusion of any other State, the functions of a State. The development of the national organisation of States during the last few centuries and, as a corollary, the development of international law, have established this principle of the exclusive competence of the State in regard to its own territory."[582] While state sovereignty remains the cornerstone of international relations, its supremacy in both contemporary international law and international relations, the apparition of individuals and international organisation as new subjects of international, and the existence of a cyberspace have fragilized the concept.

6.2 The content of the principle of sovereignty

Sovereignty contains two aspects. The first aspect, the inviolability of the territorial integrity of states, is the most concrete of the two, as it protects a state against unsolicited penetrations into its territory.[583] Similarly, non-physical conduct remotely causing physical destructions, death, or injury into the territory of another state violate its sovereignty. Since the advent of cyberspace, there is also a growing consensus that a cyber operation remotely causing a serious loss of functionality in infrastructures located abroad amounts to a violation of the territorial sovereignty of the target state.[584]

But sovereignty is more than a safeguard against physical trespassing of geographical borders. The territory, as the vessel of sovereign authority, is not the only aspect to be contained in the principle of sovereignty. Sovereignty also means supreme authority, that is, the power that a state owns over a territory, the capacity to exercise sovereign functions over this territory, at the exclusion of other states.[585]

In its first aspect, the principle protects the territorial integrity of a state. In its second aspect, the principle protects the political integrity of a state. Hence, pursuant to the principle of state sovereignty, any non-consented conduct having for effect to prevent the territorial state from properly exercising elements of its authority over the portions of land, air and sea that it is

[582] *Island of Palmas* (United States vs The Netherlands), (1928) Perm. Ct. of Arbitration, at 838.
[583] *Corfu Channel* (United Kingdom of Great Britain and Northern Ireland v Albania) (1949) ICJ, Judgement.
[584] See Schmitt, *Supra* note 162.
[585] *S.S. Lotus* (France vs Turkey), (1927) PCIJ Judgement.

legally entitled to administrate, may constitute an internationally wrongful act. How does this translate in practice, in the context of disinformation activities targeting a state, is studied below.

6.2.1 The uncertainties surrounding the content of the principle of sovereignty

The question of the legal value of the principle of sovereignty in the legal literature is not fully settled. Yet, the view that sovereignty is a rule that can be violated on its own "has taken on a mainstream position within the discipline."[586] Among other authors, a majority of the Tallinn Manual experts are proponents of that view.[587] Drawing upon Max Huber's *obiter dictum*, Schmitt, and a majority of the experts of the Tallinn Manual 2.0 argue that the two aspects of the principle of state sovereignty, territorial integrity and political independence, carry legal value.

Accordingly, alongside the unproblematic cases involving unsolicited physical trespassing into a territory or remote causation of quantifiable damages in the territory of the target state, a foreign conduct may constitute a violation of sovereignty if it undermines the political independence of the target state. On this point, Schmitt and Vihul explain that "[S]uch violations need not be accompanied by any damage or injury, and unlike the prohibition on intervention into a State's *domaine réservé*, no coercive intent or effect is required" […] "the determinative factor is whether the activities interfered with qualify as inherently governmental functions."[588] Therefore, the question of illegality of a foreign conduct is not precluded by the absence of physical manifestation of effects in the territory of the target and the absence of the element of coercion. Whether tangible or not, non-coercive foreign interference into the affairs of a state may violate the sovereignty of the target state if they qualify as usurpation or interference with its inherently governmental functions.[589]

Under this interpretation, the principle of sovereignty would function as a residual, third-level safeguard to the political integrity and the territorial integrity of a state, behind the principle of non-intervention, and the prohibition to use or threaten force.

---

[586] Lahmann, *Supra* note 279, at 66.

[587] See *inter alia*, Schmitt M. (2017), *Supra* note 162; Michael N. Schmitt & Liis Vihul, "Respect for Sovereignty in Cyberspace" (2018) 95 *Texas Law Review* 1639.

[588] Schmitt & Vihul, *Ibid.,* at 1649.

[589] Schmitt & Vihul, *Ibid.,* at 1649. See also Schmitt, *Supra* note 422, at 753.

To be clear, the first function of state sovereignty is not debated here. Although debatable[590], the view according to which the principle of sovereignty functions as a safeguard against a violation of the territorial integrity of a state, which includes cases of unsolicited border trespassing that do not satisfy the conditions set out in the principle of non-intervention and the prohibition of the use of force, is fairly undisputed.[591] In contrast, the position of the literature regarding the existence of a primary rule of international law stemming from the second function of the principle of sovereignty and prohibiting undue interferences with/ or usurpation of/ the inherently governmental functions of a state is less clear.[592] This proves problematic in the present context because disinformation activities are in essence more likely to constitute a violation of the second function of sovereignty.

First of all, commentators tend to disregard the elusive question of the violation of the second aspect of sovereignty in their assessments. Delerue, for example, essentially approached the rule of sovereignty as a safeguard against a violation of territorial integrity, and appreciated the question of a violation of political independence against the principle of non-intervention.[593] Other commentators have similarly focused on the principle of non-intervention when it comes to assess intangible conduct that do not generate physical consequences on the territory of a state.[594] Schmitt himself occasionally expressed doubt about this question.[595]

Such a view is unsurprising. Indeed, the function that a hypothetical rule of sovereignty is said to assume with regard to a states' governmental functions overlaps in many ways with the non-intervention principle. This is all the truer when the non-dictatorial reading of coercion is advanced.[596] Besides, it is difficult to articulate a vision of undue interferences with/ or usurpation of/ the inherently governmental functions of a state that could applied in an objective manner. Notwithstanding, several human rights considerations may stand in the way of such an

[590] The question of the legal consequences arising from the mere violation of territorial integrity of a state is unsolved at this point. It would seem that in case of a violation of territorial integrity via acts that do not in themselves amount to prohibited interventions or unlawful uses of forces, it is left to the territorial state to call, or not, the violation.

[591] Schöndorf, *Supra* note 536; Schmitt & Vihul, *Ibid*; The possibility has notably been contemplated in the context of airspace violation and uncontrolled falls of space debris into foreign territories. See on that point Delerue, *Supra* note 81, at 200-233.

[592] See Wenqing Zhao, *Supra* note 117..

[593] See Delerue, *Supra* note 81, at 200-233.

[594] See for instance, Tsagourias, *Supra* note 409; Kilovaty, *Supra* note 70; Baade, *Supra* note 148; Moulin, *Supra* note 433.

[595] "It is somewhat unsettled as to whether cyber activities that are not directed against the government's systems can violate sovereignty." Michael N. Schmitt "Foreign Cyber Interference in Elections: An International Law Primer, Part II" (16 October 2020) *EJIL:Talk!* <https://www.ejiltalk.org/foreign-cyber-interference-in-elections-an-international-law-primer-part-ii/>.

[596] See Moynihan, *Supra* note 470.

approach to the sovereignty principle in international law, especially when it comes to deal with international communications.

6.2.2 States' views regarding the content of the principle of sovereignty

Some states have questioned the existence of a rule of sovereignty susceptible of being violated on its own. A forerunner in that sense is the UK's Attorney General, Jeremy Wright, who in 2018, rejected the interpretation of sovereignty as a rule capable of being breached on its own in the context of cyber activities; "The UK Government's position is therefore that there is no such rule as a matter of current international law."[597] However, as far as state practice is concerned, a majority of states embrace the view that sovereignty is a standalone rule of international law that applies in the cyber context. This is for example the case of France[598], Brazil[599], Canada[600] and China.[601]

However, few states that have explicitly mentioned the issue of the legality of state-led disinformation in their position papers have done so by reference to the principle of sovereignty. Most of the official positions earlier reported in this work approach state-led disinformation activities from the perspective of the principle of non-intervention.

More generally, it seems that states disagree about the normative value of the second function of state sovereignty. In fact, whereas a majority of states seems to agree that the principle of sovereignty includes the "territorial integrity", not all states that have recognized the existence of a rule of sovereignty in cyberspace have approached the rule in a way that would suggest that non-coercive cyber operations such as online disinformation that do not directly cause material damages nor loss of functionality could breach the principle. Often states will invoke the principle of non-intervention when they express their opinions regarding the legal status of state-sponsored disinformation activities.

---

[597] Jeremy Wright, "Cyber and International Law in the 21st Century" (23 May 2018) *GOV.UK*, <https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century>.

[598] French Ministry of the Armed Forces, *Supra* note 160.

[599] National position contained in *Official compendium, Supra* note 161.

[600] Government of Canada, "International Law applicable in cyberspace" (April 2022) <https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/peace_security-paix_securite/cyberspace_law-cyberespace_droit.aspx?lang=eng>.

[601] Ministry of Foreign Affairs of the People's Republic of China, "China's Positions on International Rules-making in Cyberspace" (2021), <https://www.fmprc.gov.cn/mfa_eng/wjb_663304/zzjg_663340/jks_665232/kjlc_665236/qtwt_665250/202110/t20211020_9594981.html>.

Germany is a notable exception. Germany shares the consensual view regarding physical effect and harm, but interestingly adds "[F]oreign interference in the conduct of elections of a State may under certain circumstances constitute a breach of sovereignty *or*, if pursued by means of coercion, of the prohibition of wrongful intervention."[emphasis added] [602] The wording suggests that Germany also considers the second aspect of the principle of sovereignty, the prohibition of intangible foreign conduct affecting the sovereign functions of a state, to constitute a primary rule of international law.

Excluding Germany, other national view lack clarity when it comes to determine where the scope of a hypothetical principle of sovereignty safeguarding the political integrity of a state begins, and where the scope of the principle of non-intervention stops.

Canada for example reiterates the uncontentious view that the production of effect and/or loss of functionality would result in a violation of territorial sovereignty. It then adds that "[…] an interference with or usurpation of inherently governmental functions of the affected State, would be a violation of territorial sovereignty"[603] Canada however specifies "Cyber activities that have significant harmful effects on the exercise of inherently governmental functions would constitute an internationally wrongful act. For Canada, this would include government activities in areas such as health care services, law enforcement, administration of elections, tax collection, national defence and the conduct of international relations, and the services on which these depend. There can be a violation of territorial sovereignty by way of effects on governmental functions regardless of whether there is physical damage, injury, or loss of functionality. An example would be a cyber activity that interrupts health care delivery by blocking access to patient health records or emergency room services, resulting in risk to the health or life of patients."[604] Yet, it is difficult to imagine that the example provided by the government of Canada does not already constitute a coercive intervention.

France considers what constitute a violation of sovereignty "all attacks against French cyber infrastructure or one causing effects on French territory" but adds that "a cyber-enabled interference in the internal or external affairs of France, i.e. bearing harm or likely to harm the French political, economic, social and cultural system can constitute a violation of the principle of non-intervention."[605] Therefore, in France's view, it is only the principle of non-intervention that safeguards the political integrity of a state.

---

[602] Federal Government of Germany, *Supra* note 260.
[603] Government of Canada, *Supra* note 600.
[604] *Ibid.*
[605] [translated by the author] French Ministry of the armed forces, *Supra* note 160.

The US appears to take a similar position than UK, noting that "it does not appear that there exists a rule that all infringements on sovereignty in cyberspace necessarily involve violations of international law."[606]

The main representatives of the "Westphalian" actors of the international community, China, Russia and Iran, articulate an absolutist vision of state sovereignty, which also provides protection against intangible foreign conduct leveraging the free flow of information online.[607] Adopting an intransigent posture against foreign interferences, they may be willing to attribute full normative character to the two aspects of sovereignty. However, there is the possibility that they simply consider sovereignty as extending over their respective online information spaces. In that sense, unconsented disseminations of information into their online information space by a foreign entity is interpreted as a violation of their "online information space", as a derivative of "air space."[608] Still, this argument may only make sense for the states that have enacted strong regulation on online content, thereby territorializing a portion of the global information space.

In fine, looking more closely at states' treatment of sovereignty in recent position papers does not lead to clear-cut conclusions regarding the relevancy of the principle of sovereignty in the context of cyber-enabled disinformation, specifically with regard to its second function. Notwithstanding, the term of "interference" is often used as a rhetorical device by states rather than a normative concept giving rise to legal consequences.

6.3 State-sponsored disinformation campaigns and the two functions of state sovereignty

As the government of Brazil remarked, "there is neither broad state practice nor sufficient *opinio juris* to generate new customary international norm allowing for the violation of State sovereignty, including by means of ICTs."[609] Then, assuming that a rule of sovereignty does exist, can non-coercive and non-forceful state-sponsored disinformation campaigns still violate the territorial sovereignty or interfere with the inherently governmental functions of a state? In

[606] Hon. Paul C. Ney, Jr.," DOD General Counsel Remarks at U.S. Cyber Command Legal Conference" (2 March, 2020), <https://www.defense.gov/News/Speeches/Speech/Article/2099378/dod-general-counsel-remarks-at-us-cyber-command-legal-conference/>.

[607] See Lahmann, *Supra* note 279.

[608] An interesting case to illustrate that point is the Nuclear Test Order case, in which Australia argued that violated its sovereignty over its territory "[T]he deposit of radio-active fall-out on the territory of Australia and its dispersion in Australia's airspace without Australia's consent" See *Nuclear Tests* (Australia vs France), (1973) ICJ, Order.

[609] See *Official compendium, Supra* note 161.

other words, is a non-coercive interference affecting the sovereign authority of a state essentially appreciated by reference to the principle of non-intervention, or can be appreciated by reference to the principle of sovereignty alone, when the element of coercion is absent? The two functions of the principle of sovereignty, and their relevance to the issue at hand are covered below.

6.3.1 Disinformation campaigns as violations of territorial sovereignty

Commentators mostly agree around the fact that cyber-enabled conduct that tampers with voting machines to modify the vote counts will constitute violations of territorial sovereignty.[610] In the same vein, some have argued that "any unauthorised state sponsored cyber operation constituting an intrusion into a foreign system constitutes a violation of the territorial sovereignty of the foreign State."[611]

Cyber disinformation campaigns are intangible and do not directly generate quantifiable damages. Also, disinformation campaigns are not intrusive in the sense that they do not penetrate a system,[612] and eventual analogies between "online information space" and "air space", as an intangible medium upon which states traditionally exercise sovereignty, do not reflect state practice.

Even though no territorial intrusion *per se* occurs, there is the possibility that disinformation campaigns produce effects in the territory of the target state. Assessing disinformation activities through the lenses of territorial integrity will thus systematically requires a causality-test. Schmitt posits that, in the absence of a causal nexus between the problematic conduct and the materialization of physical effects, no violation of sovereignty can be made out.[613] In such a case, the materialization of physical effect such as death, injury or loss of functionality may also give rise to an uncontentious case of violation of the principle of non-intervention, and eventually constitute an illegal use of force.[614] Again, the question remains as to whether

[610] It bears keeping in mind that directly interfering with the vote count may also constitute a prohibited intervention.

[611] Delerue, *Supra* note 81, at 226.

[612] The present work has used the notion of "online information space" profusely, to define an abstract portion of the global information space where the information is most likely to be consulted by the nationals of a given state. It is a model that helps understanding how disinformation campaign are targeted, but it does not describe a legal reality. Also, even though some parts of the global information space are so tightly controlled that they virtually become territorialized versions of internet, it remains inaccurate to speak of "intrusion" *strictu-sensu* in the context of communications activities leveraging online information.

[613] Schmitt, *Supra* note 422, at 754; Schmitt, *Supra* note 595.

[614] See *Supra* "State-sponsored cyber-disinformation as illegal "force""

minimal physical effects may still amount to a violation of territorial sovereignty without breaching the non-intervention principle. The debate is unsettled at this point. Nevertheless, it is highly improbable that states condemn a disinformation campaign on the sole basis of the territorial aspect of a principle of sovereignty, as this would require for the target to prove the causal nexus between the disinformation campaign in question and the occurrence of harm, destruction or loss of functionality in its territory.[615]

Obviously, it is unlikely that a state ever succeeds at proving that a disinformation campaign substantially contributed to harm, destruction or loss of functionality in its territory. And even when it turns out that disinformation substantially contributed to physical damages, then the conduct might as well cross the thresholds of the principle of non-intervention and the non-use of force.

6.3.2 Disinformation campaigns as interferences with/ or usurpations of/ the inherently governmental functions of a state

It is not clear how the principle of sovereignty functions as a second-level safeguard to the political integrity of state against foreign conduct that do not generate any physical effects in the territory of the target state.[616] With regards to international communications in general, the Tallinn Manual 2.0 explicates that propaganda activities do not in principle amount to a violation of sovereignty. Could state-led online disinformation nonetheless amount to a mere violation of sovereignty for the reason that it interfered with the inherently governmental functions of the target state?

The inherently governmental functions of a state relate to the concept of sovereign authority. The category is mostly fluid and contains competences exclusively owned by a state in its territory. These are sometimes called "official acts." The French equivalent is "*prerogatives de puissance publique,*" which defines the prerogatives that state officials or government employees are entitled to exercise discretionarily for the pursuit of public interest.[617] Hence this

[615] A disinformation campaign would have to reach an extreme level of intensity in order for a sufficiently solid causal nexus between the act in question and quantifiable damages to be established. Surely, the said conduct will be crossing other higher-degree thresholds than the mere violation of sovereignty if it is so intense that it indirectly caused quantifiable damage in the target's territory. In such extremely hypothetical cases, the conduct will be considered and punished as a violation of the highest-tier rule, not solely as a violation of sovereignty.

[616] Schmitt, *Supra* note 422; Zhao, *Supra* note 117.

[617] Nicolas Kada & Mathieu Martial, « Puissance publique » in Nicolas Kada (ed.), *Dictionnaire d'administration publique* (Grenoble ; Presses universitaires de Grenoble, 2014), at 404.

does not apply to matters of private companies or private individuals.[618] Also, a state cannot exercise these prerogatives over a territory that is not its own. Doing so is considered an unconsented exercise of power over the territory of another state, or a usurpation of the governmental functions of a state, and potentially a prohibited intervention.

To a certain extent, the category of official acts overlaps with the *reserved domain* of a state, the main difference between the two concepts being one of perspective. Indeed, while the reserved domain is appreciated from the perspective of international relations and law, the inherent governmental functions are determined in municipal law.

- Direct interference with or usurpation of/ the inherently governmental functions of a state

The capacity to manage an online information space does not constitute an inherently governmental function of a state.[619] Hence, disinformation campaigns cannot be said to interfere *in a direct manner* with the inherently governmental functions of a state for the sole reason that it interfered with the capacity of a state to manage its information ecosystem. On this point, it is necessary to note, as one commentators warned, that "placing greater reliance on sovereignty in the cyber context risks bolstering a position ('cyber sovereignty') in which individuals' freedoms online are undermined, contrary to international human rights law."[620] This would be particularly the case if a state is given the sovereign ability to regulate its online information space against so-called foreign threats in a more intrusive manner than through transparency-increasing measures. Although a state may justify such a position by the need to bolster its capacity to mitigate so-called "foreign malign influence" leveraging online information for political purpose, legally recognising the sovereign competence of a state in doing so may not always profit the public interest, neither may it be done without opening the door to abuses of all kinds.

[618] Moynihan, *Supra* note 470.

[619] Schmitt (2017), *Supra* note 162, at 26, comment 29. Interestingly, it is specified in a note at page 26 of the Tallinn Manual 2.0. that, while "numerous instruments stipulate that States in which other State's broadcast, especially satellite broadcast, are available should have some degree of control over the information that is being transmitted on to their territories. […] the Experts agreed that this premise has not crystallized into a customary international law requirement." This remark has a particular resonance in the present discussion. at some point, this could even be interpreted as to suggest that certain foreign disinformation activities could amount to an interference with or usurpation of the inherent governmental function owned by the recipient state in that it interfered with its sovereign ability to exercise control over its online information space. However, it goes without saying that such an approach to information flows and state sovereignty is untenable in our interconnected world, as this fragilize the principle of free information flow and severely undermine the basis on which cyberspace is built.

[620] Moynihan, *Supra* note 470.

Indeed, there is a great deal of discretion involved in both the identification of inherently governmental functions, and the identification of undue foreign influence. The view that the ability of a state to control its online information space constitutes in itself an inherently governmental function simply cannot be accepted without calling into question the very existence of cyberspace.

- Indirect interference with or usurpation of/ the inherently governmental functions of a state

One likely instance where disinformation campaigns may indirectly interfere with the inherently governmental functions of a state is in the context of electoral disinformation. Recognising that "[i]t is not altogether clear whether all interference with an election is encompassed in the rule",[621] Schmitt nonetheless advances the argument that cyber activities that "indirectly disrupt the election's smooth execution, such as voter suppression activities, would qualify **[**as violation of sovereignty]."[622] Accordingly, "posting incorrect information about how, where, or when to vote reasonably could be characterized as interfering with the State's ability to conduct the election."[623] Such a view makes ample sense. In fact, the holding of presidential or legislative elections constitutes by definition an inherently governmental function exercised by state official, for the public interest. Consequently, foreign attempts to prevent voters from voting *via* disinformation may engage the responsibility of the sponsoring state on the basis of a direct violation of the sovereignty of the target state for the reason that it hampers the proper holding of elections. Also, as noted earlier, suppressing a substantial number of voters may also amount to a coercive intervention in that it deprives individuals vested with decision-making power from the ability to choose a candidate.[624] In this case there seems to be a difference in gravity between "indirect disruption of a sovereign process' smooth execution" amounting to a breach of sovereignty, and a deprivation of control over the vote or the election, or any sovereign function amounting to a prohibited intervention.[625] Yet, where to place the cursor of illegality in that situation is unclear, and it is unlikely that minimal disruptions not breaching the principle of non-intervention are even capable of breaching the sovereignty of the affected state, let alone being denounced as such.[626]

[621] Schmitt, *Supra* note 595.
[622] Schmitt, *Supra* note 422, at 753; Schmitt, sura note 595.
[623] Schmitt*, Supra* note 422.
[624] See also Schmitt, *Supra* note 422; Moynihan *Supra* note 470.
[625] Schmitt, *Supra* note 422.
[626] Moynihan *Supra* note 470.

As regards disinformation campaigns waged to deceive voters into voting for a particular candidate or thinking that the system is rigged, the situation is less clear. Schmitt for example argues that misleading voters into thinking that the information they were exposed to emanated from true nationals would constitute a violation of sovereignty and a coercive intervention.[627] While it can be argued that in absence of a substantial number of messages threatening voters, or providing wrong information as regard the way to vote, foreign-sponsored electoral disinformation campaigns still do not constitute coercive interferences, the possibility remains that they constitute internationally illegal indirect interferences with the inherently governmental function of a state by reasons of their covertness. In fact, opening the door to such possibility does not even pose any issue in terms of potential instrumentalization by unscrupulous actors, as it is anyway required to prove that false identities or bots are used by the disinformants. But there are two downsides to this view. First, it may be too underinclusive. As argued earlier, overt disinformation is equally, if not more, efficient than covert disinformation at deceiving the audience.[628] Second, there are conceptual obstacles to the operability of a principle of sovereignty that functions as a second-tier rule safeguarding the political integrity of a state. (i.e., a rule of non-interference). These will be detailed in a next sub-section.

It is more difficult to place the cursor of illegality outside the context of election. A state-sponsored disinformation campaign involving bots and individuals pretending to be acting on behalf of a state's governmental organs or officials to communicate wrong information about vaccination, or providing the wrong location of a shelter during a natural disaster is likely to amount to a usurpation of a state functions, even if it only caused minor disruptions in the territory of the target state. However, as argued earlier, even in absence of death and destruction, the fact that consequences of the magnitude of a use of force are foreseeable would suffice for the act to breach the non-use of force prohibition, as well as the non-intervention principle.

Disinformation campaigns *directly* causing injury, death or destruction in the territory of the target, would constitute violations of the territorial integrity of the target state, and certainly amount to prohibited interventions and indirect uses of force.[629] How mere disruptions affecting state functions are evaluated in practice is unclear, and relaxing the thresholds to accommodate

[627] Schmitt takes the view that such conduct would certainly amount to a prohibited intervention as well. This could be taken as a proof that there might be no difference, in practice, between an interference with or usurpation of/ the inherently governmental functions of a state and a prohibited intervention. See Schmitt, *Supra* note 422. See Schmitt, *Supra* note 211. at 47 and 51. See also in that sense Tsagourias, *Supra* note 409.

[628] See Lahmann, *Supra* note 62, at 205.

[629] See notably ICJ, *Supra* note 196, at 97, §250.

foreign interference via disinformation is a risky enterprise when the act in question is intangible, and may not generate physical effects as it unfolds.

6.3.3 Conceptual obstacles to the operability of "secondary" non-interference principle stemming from the principle of sovereignty

Lack of clarity surrounds the question of the legal force of the principle of sovereignty and the conditions of its application to non-coercive conduct affecting the political integrity of the target state without resulting in physical damage nor loss of functionality. Alongside the various human rights considerations briefly evoked throughout the previous developments, several conceptual problems stand in the way of a fully operable *rule* of non-interference stemming from state sovereignty;

Firstly, the above investigation has shown that there is no clear indication in the literature as to whether a rule of sovereignty could be violated by a disinformation campaign without simultaneously amounting to a prohibited intervention. Notwithstanding, finding concrete differences between interference with inherent governmental functions and a violation of the principle of non-intervention is particularly difficult. The fact that most commentators have failed to provide examples of hypothetical instances where sovereignty alone is breached by disinformation activities calls into question the relevance of the rule of sovereignty alone in this context. Of course, the focal point ultimately resides in the capacity for the said conduct to generate sufficiently noticeable inconvenience as regards a sovereign function without being coercive. [630] However, opening the possibility for a state to respond to "noticeable inconvenience" affecting its sovereign functions by ways of countermeasures is not conducive to the maintenance of international stability.

Secondly, an obvious tension will arise between the principle of non-intervention and the principle of sovereignty if the former happened to apply to the same type of conduct but in a more straight-forward manner than the latter. In fact, while the experts of the Tallinn Manual failed to provide precise definition of inherently governmental function, the category arguably encompasses the reserved domain of a state.[631] This begs the question of what would be left of the principle of non-intervention if a rule of sovereignty is given similar functions of protection of political integrity in the context of intangible conduct. Unless a mere violation of sovereignty

[630] See Moynihan, *Supra* note 470.
[631] Schmitt (2017), *Supra* note 162, rule 2, Comment 16.

does not give raise to the same consequences as a violation of the principle of non-intervention, there would be practically no reason, in such case, for a prohibition of coercive interventions to even exist.

Thirdly, the ICJ has explicitly acknowledged the role of the principle of non-intervention as a safeguard to states' political integrity.[632] Some commentators even argued that "[A]ccording to one reading, the prohibition on intervention protects the non-territorial, 'metaphysical aspect of sovereignty (a state's political integrity) rather than its physical dimension (a state's territory).'"[633] While the present author does not concur with this view, debates pertaining to the existence of a rule of sovereignty safeguarding the sovereign competences of a state against intangible foreign interferences that do not generate quantifiable damages would have even less reasons to take place from that perspective.

Fourthly, the concept of interference is a very loose one. The principle of sovereignty may rarely be applied in a consistent and objective manner by states. More than the principle of non-intervention, a great deal of discretion may be involved in the application of a rule of sovereignty safeguarding the inherently governmental functions of a state against foreign influence attempts. This has to be put in perspective with the fact that the concept of foreign interference is on itself inherently subjective. It identifies foreign conduct that is felt and disagreed upon by the passive party, and affects a function the performance of which the said state has identified as necessary for the public interest. Often, the determination of such functions will be subject to political and ideological considerations, which in turn renders the principle, as a whole, "politically and ideologically malleable".[634] Not only that, but giving full normative status to a principle of non-interference derived from the principle of sovereignty risks aligning international law with domestic law. This is substantially lowering the requisite thresholds, which could foster conflict escalation, thereby contradicting with the very function of the law as a stabilizer to interstate interactions.

It bears noting that the same remark could be made about the non-intervention principle if an excessively loose interpretation of a prohibited intervention was to be adopted by states.

Fifthly, alongside these conceptual challenges lies the endless ideological debate pertaining to the application of a concept of state sovereignty that act "as a shield against any form of 'outside interference.'"[635] How to reconcile a statist vision of international law with the free

[632] See ICJ, *Supra* note 196. §202 at 96.
[633] See Crootof, *Supra* note at 214. See also in this sense Marcelo Kohen, *Supra* note 432, at 159
[634] Lahmann, *Supra* note 279, at 92.
[635] Lahmann, *Supra* note 279, at 106.

flow of information and human rights in a globalized world is a difficult question that is likely to remain unanswered for the time being.

To conclude, articulating a vision of the principle of sovereignty that also acts as a safeguard to political integrity (i.e., a rule of non-interference) raises several conceptual issues. This calls into question the relevancy of a customary principle of non-intervention protecting the sovereign affairs of a state against coercive behaviours. Although municipal systems may effectively apply broader conceptions of foreign interferences than what the customary principle of non-intervention prohibits, non-interference as a safeguard to the political integrity of states exists in international law only through non-intervention.

Therefore, and to an even greater extent than with non-intervention, the issues encountered while exploring the applicability of a principle of sovereignty to cyber-enabled state sponsored disinformation activities give rise to a noticeable sense of inadequacy. First, the thresholds contained in a rule of non-interference are unknown. Second, and unless non-interference stemming from sovereignty remains a non-binding norm, it is unclear how the rule would cohabit with the prohibition of coercive intervention, which would then assume the same role, but with tremendously less efficacy.

In a nutshell, disinformation activities that do not breach the principle of non-intervention for the reason that they failed the coercive test, or to which the apparition of tangible effects in the territory of a sovereign state cannot be traced back with sufficient certainty, will not, *in principle*, and under current international law, constitute an internationally unlawful act by reason of a violation of sovereignty.

Thus, to conclude on sovereignty, and as a commentator underlined, "[I]n the absence of agreement on how sovereignty applies in the cyber context, the non-intervention principle provides a more objective basis for assessing international wrongfulness than the sovereignty principle."[636]

But even then, the non-intervention principle will find an application to the context of psychological warfare in only very limited circumstances. This is the perfect occasion for states and international lawyers to give serious attention to human rights law. Thus, the next chapter examines whether the boundaries set by international human rights law can be exceeded in the context of psychological warfare.

[636] Moynihan, *Supra* note 470.

Chapter IV.
State-sponsored disinformation campaigns in international human rights law

1. Introduction

Alongside state-centric rules of international law safeguarding the interests of states lie human rights law, which protect the interests of individuals. Because it focusses on the primary target of disinformation, individuals, human rights law allows for a more granular approach to foreign disinformation campaigns. Thus, a human rights approach to disinformation could yield more convincing results in terms of legal qualification.

In a recent video message to the HRC, UN Secretary General Antonio Guterres declared that cyberspace was a "wild west for human rights", referring among other, to the scourge of online disinformation.[637] When they engage in online disinformation campaigns against foreign audiences, state actors may indeed violate a number of human rights treaty provisions.[638] These are the freedom of thoughts, expression, the freedom to hold opinion without interference, the right to self-determination, the right to life and the principle of non-discrimination. These rights are set out in the 1976 International Covenant on Civil and Political Rights (hereinafter ICCPR), which elaborates the rights that were first enshrined in the 1948 Universal Declaration of Human Rights (UDHR). While the latter instrument was mostly declarative and thus non-binding, the Covenant is binding on its state parties. The Covenant is the second binding international instrument after the 1953 European Convention on Human Rights (ECHR) to implement the rights set out in the UDHR, and to expand upon them. The ICCPR currently gathers 173 state parties. In virtue of Article 2, they are duty-bound to protect the rights enshrined in the ICCPR within their respective territories.

[637] "Human rights 'inescapable and powerful': Guterres" (28 February 2022) *UN News*, <https://news.un.org/en/story/2022/02/1112962>.

[638] See in that sense, Jones, *Supra* note 119; Michael Schmitt & Marko Milanovic, "Cyber Attacks and Cyber (Mis)information Operations During a Pandemic" (2020), 11 *Journal of National Security Law & Policy* 247; Hollis, van Benthem and Dias, *Supra* note 61.

Albeit to various extents, many rights contained in the ICCPR relate to disinformation. If the mind is the primary target of the information spreader, individuals are his first victims. Ellul explains that the audience is regarded as a mere asset that can be modelled at will by the propagandist.[639] In other words, PSYOPS are dehumanizing acts. Today, cyberspace is firmly ingrained in people's lives. Technology provides states with extremely efficient means to conduct massive deception work on specific audiences.[640] State actors possess the tools, the capabilities and funds to craft particularly far-reaching and invasive disinformation campaigns that private actors cannot reciprocate.[641]

Disinformation is, by essence, the manipulation of the minds; it exploits the credulity of the target for political or strategical purposes. Therefore, at a time when public opinions have never been so vulnerable to information manipulation tactics, specific attention should be given to human rights.[642]

Alongside the means, the ends of disinformation are no less morally condemnable. Disinformation may be waged by political entities in order to undermine the self-determination of the audience, to sow division in a society, or to erode the trust in science. It results that the issue of online disinformation should not be reduced simply to the preservation of freedoms of thoughts or opinion.

There are rights that appear to be more likely to be breached by disinformation than others, such as the freedom to hold opinion without interference. Other human rights are only breached by disinformation in an indirect manner, or by incidence, and on very rare occasions. For instance, the right to life could be said to have been breached when a disinformation campaign has provoked death, or substantially contributed to a situation of imminent danger for the life of the audience. But human rights law does not only empower individuals to claim and assert their rights *vis à vis* disinformation spreaders. It also sets strict legal limitations to speech regulations. This second aspect of human rights law's role in matters related to disinformation is at least as much important as the first and deserve a distinct focus in this section.

This section provides an analysis on the legality of disinformation campaigns under the international human rights law stemming from international human rights treaties in general, and the ICCPR in particular. The following sections therefore approach human rights law as a

[639] Ellul, *Supra* note 85, at xvi.

[640] Colomina, Sanchez Margalef, Youngs, *Supra* note 57, at 2.

[641] See for instance Andrew Roth, "Russia uses Orwellian propaganda news vans in Mariupol" (26 May 2022) *The Guardian*, <https://www.theguardian.com/world/2022/may/26/russia-uses-orwellian-propaganda-news-vans-in-mariupol>.

[642] See Jones, *Supra* note 119; Irene Khan, *Supra* note 47; Hollis, van Benthem and Dias, *Supra* note 61, at 14.

limitation to the conduct of psychological warfare. From the most to the least obvious, the freedoms of thought, opinion and expression (2), the right to self-determination (3), the principle of non-discrimination (4), the right to life (5) and their correlatives duties are examined in turn. A sixth section briefly recalls the role of human rights law as a limitation to speech regulations (6). A seventh section examine the inevitable question of human right extraterritoriality (7). A conclusion comments about the value of a human-centric approach to state-sponsored disinformation (8).

It is important to keep in mind that the rights examined in this section apply primarily to the relation between governments and their own citizenry. To a certain extent, they may also to the relation between populations and a foreign state. As this work focusses on the regulation of psychological warfare, only the question of the legality of PSYOPS waged by a state against foreign populations will be assessed. Still, it is worth noting that the arguments raised in this chapter are also applicable to the relation between governments and their own citizenry.

2. The freedom of thought and the right to hold opinion without interference

The right to hold opinions without interference, and the freedom to seek, receive and impart information and ideas of all kinds, which also apply in cyberspace,[643] are unquestionably relevant in the present context. They are enshrined in the ICCPR, Articles 18 and 19, which stipulate "[E]veryone shall have the right to freedom of thought, conscience and religion."[644]; "Everyone shall have the right to hold opinions without interference."; "[E]veryone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."[645] These rights are proclaimed in various instruments such as the ECHR, the 1948 American Declaration of the Rights and Duties of Man, the 1969 American Convention on Human Rights, and the 2009 Charter of the Fundamental Rights of the European Union, its Article 3 making reference to the right to "the integrity of the person", which includes its "mental integrity". In the American legal system, the freedoms of thought, opinion and expression are protected by the First Amendment.

[643] See *inter alia* HRC, *The promotion, protection and enjoyment of human rights on the Internet*, (2014), A/HRC/RES/26/13; HRC, The promotion, protection and enjoyment of human rights on the Internet, (2016), A/HRC/RES/32/13. See also notably HRC, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, (2016), A/HRC/32/38.

[644] ICCPR, *Supra* note 118, article 18.

[645] ICCPR, *Supra* note 118, article 19.

There is an extensive international jurisprudence on freedom of thought, opinion and expression, particularly coming from the European Court of Human Rights. Similarly, there is an extremely high number of legal cases associated with the freedom of thought, opinion and expression in domestic law. Usually, these rights are approached as safeguards against unjustified censorship policies. Some commentators however hinted recently that disinformation activities, and the mechanisms on which disinformant rely could threaten these rights in a direct manner.[646]

The freedoms of thought, opinion and expression are deeply interrelated. In fact, there can be no freedom of thought and opinion without a proper access to information and information pluralism. In turn, access to a plurality of ideas is impossible without freedom to hold opinions. Quality information is vital to the proper functioning of modern information societies.[647] Similarly, the freedom to think and to hold opinion without interference is indissociable from the concept of freedom, and a vital component of democracies.[648] The UN Human Rights Committee has notably stressed that "[T]he freedoms of opinion and expression form a basis for the full enjoyment of a wide range of other human rights."[649] However, whilst the freedom of expression and the correlative freedom to manifest religion or belief may be limited on some specific circumstances, the freedom of thought and the right to hold opinions without interference (hereinafter FT&RHO) are unconditionally protected.[650] In its *Ivanova v. Bulgaria* case, the ECHR stated that in virtue of the FT&RHO, "a State cannot dictate what a person believes or take coercive steps to make him change his beliefs."[651] So far, the FT&RHO have mainly been called upon with respect to the maintenance of religious pluralism, in relation to the principle of *laïcité*. It is however submitted that the FT&RHO also extends to the use of manipulative practices by the authorities on the populations.

---

[646] See notably Jones, *Supra* note 119; Simon McCarthy-Jones, "The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century" (2019) 2(19) *Frontiers in Artificial Intelligence* 1; Irene Khan, *Supra* note 47, at 7.

[647] "Without quality information, people cannot understand nor effectively respond to the events that shape their world." "Why information matters A foundation for resilience" (2015) Internews report, Embracing Change: The Critical Role of Information Project, at 5; United Nations Human Rights Committee (HRC), *General comment no. 34, Article 19: Freedoms of opinion and expression*, (2011), CCPR/C/GC/34.

[648] See notably Barush Spinoza, *A Theological-Political Treatise*, (translated by Elwes, R. H. M. (Robert Harvey Monro), 1853) Chapter XX – that in a free state every man may think what he likes, and say what he thinks. (Spinoza speaks of a "free republic").

[649] HRC, *Supra* note 647.

[650] CCPR General Comment No. 22: Article 18 (Freedom of Thought, Conscience or Religion) (30 July 1993) CCPR/C/21/Rev.1/Add.4; Schmitt (2017), *Supra* note 162, at 188, rule 35 "Rights enjoyed by individuals", §5. ("The right to hold an opinion freely is a guarantee so central to the object and purpose of international human rights law that, unlike the freedom of expression, its exercise may not be restricted.)

[651] *Ivanova v. Bulgaria* (2007) ECtHR (fifth section), judgement.

One illustrious proponent of the FT&RHO was the Dutch philosopher Spinoza. In its *Tractatus Theologico-Politicus*, Spinoza firmly defended the idea that the freedom to think was inalienable. According to Spinoza, the freedom to think is a natural right that cannot be abdicated, as doing so would equal giving up on what distinguishes rational beings from beasts. "For this reason government which attempts to control minds is accounted tyrannical, and it is considered an abuse of sovereignty and a usurpation of the rights of subjects, to seek to prescribe what shall be accepted as true, or rejected as false, or what opinions should actuate men in their worship of God", he argues.[652] The idea that no ruler should extend their authority to the minds is central to good governance and legitimacy. The FT&RHO protect the mental autonomy of individuals, and therefore, establish a correlative duty not to subject someone to manipulative or deceptive tactics.[653] In virtue of the FT&RHO, states should ensure that the individuals placed within their respective jurisdiction have access to factual information, and that no authority makes use of invasive and non-transparent public opinion-shaping measures on their populations.

At first glance the FT&RHO appears to set strict limitations to the conduct of psychological warfare. That is not to say, however, that states should plainly give up on the use of PSYOPS to be consistent with the FT&RHO. Indeed, although the exact scopes of the FT&RHO remain unclear, it makes no doubts that they do not cover all forms of influence exercised by ways of communications by an external authority. The opposite would set an impracticable threshold, as peoples are subject to external influence on a daily basis. The question of the legality of state-sponsored communication campaigns under human rights law hinges on the distinction between what can be considered an acceptable influence and an illegitimate manipulation.

Some commentators have attempted to clarify the distinction between legitimate forms of influence and illegal manipulation. In a dissenting opinion to ECHR's landmark case on religious proselytism, *Kokkinakis v. Greece*, judge Valticos noted that "to attempt persistently to combat and alter the religion of others, to influence minds by active and often unreasonable propaganda" would certainly interfere with the freedom of thoughts, conscience and religion.[654] The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan similarly affirmed that "Any effort to coerce the holding or not holding of any opinion is prohibited."[655] The Tallinn Manual experts in turn recalled "State

[652] Spinoza, *Supra* note 648.
[653] See notably in that sense, McCarthy-Jones, *Supra* note 646; Colomina, Sanchez Margalef, Youngs, *Supra* note 57, at 11.
[654] Dissenting Opinion of Judge Valticos to *Kokkinakis v Greece* (1993) ECHR 20, 17 EHRR 397.
[655] Irene Khan, *Supra* note 47, at 7.

conduct that interferes with the freedom of opinion includes online incitement against protected persons, online intimidation, or other forms of harassment conducted on the basis of a person's views, such as political or religious views that are evidenced by membership in a political party or a religious denomination."[656] An international legal scholar determined that illegitimate manipulation would be characterised "as the disabling or circumvention of an individual's ability to rationally appraise information."[657] In the same way, and in relation to online disinformation another commentator recently noted, "[A]s a minimum, freedom of thought entails a right not to have one's opinion unknowingly manipulated or involuntarily influenced, fundamentally linked with the concept of human agency."[658]

In reality, most contributions on the subject point towards the same conclusions; Firstly, that attempting to induce particular behaviours through influence *pure and simple* does not contravene the FT&RHO. Secondly, that illegitimate manipulation is traditionally associated with the use of insidious, invasive, persistent methods of influence without the knowledge of the target, or unconscious manipulation.[659]

Brainwashing and political indoctrination would constitute unlawful methods of unconscious manipulation.[660] An illustration of undoubtably freedom-destroying manipulation methods deployed in the domestic context is North Korea's incredibly invasive state propaganda apparatus. In its 2014 Report, the Commission of Inquiry on Human Rights in the Democratic People's Republic of Korea describes North Korea's propaganda apparatus as an "all-encompassing indoctrination machine that takes root from childhood to propagate an official personality cult and to manufacture absolute obedience to the Supreme Leader."[661] Yet, unconscious manipulation goes way beyond brainwashing, coercion or indoctrination.

Unconscious manipulation precludes human judgement and may, *in fine*, lead to the complete alienation of the subject. Unconscious manipulation, if too invasive and persistent, interferes with the process of the *formation of the will* of the target, undermines its mental autonomy, and as a consequence, undermines the FT&RHO. Cyberspace makes it easier to engage in unconscious manipulation. Online social media platforms are boosting the manipulative power of influence activities to the point that an increasing number of

[656] Schmitt (2017), *Supra* note 162, at 188-189, rule 35.
[657] Mc Carthy-Jones, *Supra* note 646, at 10.
[658] Jones, *Supra* note 119, at 34. Human agency traditionally refers to the capacity of a human to act and make choices intentionally.
[659] McCarthy-Jones, *Supra* note 646, at 11; Irene Khan, *Supra* note 47, at 7.
[660] Irene Khan, *Supra* note 47, at 7.
[661] Report of the commission of inquiry on human rights in the Democratic People's Republic of Korea (7 February 2014) HRC, A /HRC/25/63.

commentators is calling for the creation of a liability regime that would apply to non-state actors, particularly social media companies, which are often accused of negligence for not doing enough to prevent malicious actors from conducting manipulative influence activities on their platforms.[662]

The harmfulness of an influence campaign can be expressed in terms of its manipulative value. Occasional and non-invasive manipulation would have weak manipulative value. They may not generate sufficient harm to conclude on a violation of human rights law. In the same vein, it may be more difficult to make the point of a violation of the mental autonomy of the target when the provenance and authorship of a piece of information is known to the audience, as this would have for effect to weaken the manipulative value of the said communication. Covert influence campaigns, in contrast, are more likely to fall under the scope of the FT&RHO.

Disinformation is by definition a method for unconscious manipulation. The instigator aims at exploiting cognitive bias and inducing behavioural changes. To that end, information spreaders create or distort information and remodel the facts so that they trigger the expected emotions in selected audiences. In the state-to-state context, disinformation is best achieved when the facts are believable, and when the identity of the instigator of the campaign is unknown to the public.[663] For these reasons, state-sponsored disinformation activities are highly harmful to the FT&RHO. It bears noting that the fact that a disinformation campaign was unsuccessful at causing particular change in behaviours does not preclude the possibility that the communications could have undermined the FT&RHO of the audience. In addition, it is interesting to note that the question of the truth-value of a communication is not decisive for the evaluation of the harm in that context.[664]

It is clear that the FT&RHO provide limitations to manipulative activities conducted by an authority. The use of censorship measures to block access to alternative viewpoints combined with intensive communications campaigns deployed by a government against its own population would likely qualify as a violation of the FT&RHO.

As regards psychological warfare, it is submitted that the limits set by the FT&RHO could realistically be exceeded in the context of PSYOPS activities conducted by a state against foreign populations.[665] This is particularly the case when disinformation elements are involved.

---

[662] Jones, *Supra* note 119.

[663] See in general Ohlin, *Supra* note 142.

[664] See Jones, *Supra* note 119, at 37.

[665] Of course, the harm inflicted to the FT&RHO in the context of psychological warfare may be less flagrant than when manipulation activities are conducted by the authorities of a state against their own citizenry. The reason is that the amount of control that domestic authorities have over the national information space is always greater than this of foreign actors.

Consequently, if claims of unconscious manipulation implicating state-sponsored PSYOPS targeting foreign populations were to be made at the occasion of legal proceedings, the harm to FT&RHO should be gauged on the basis of the manipulative value of the act. Future regulations on the question should consider:

- The ubiquity of the communications. - A pervasive communication campaign that literally inundates the online information space of the target would obviously be highly manipulative, because the possibility for users to access alternative information sources would be *de facto* restricted.
- The transparency level in information communications. – Information disseminated covertly, *via* inauthentic account or bots are manipulative by definition.
- The medium used to provide information. – Information dissemination online circulate faster and spread farther than the information conveyed *via* traditional medium. Social media users are particularly vulnerable to modern forms of unconscious manipulation.
- The duration of the effort.
- The education level of the target audience. – Uneducated populations are more vulnerable to manipulation.

In light of the above, the state sponsored disinformation campaign that targeted American voters during the 2016 presidential elections was inherently manipulative in that it involved the use of identity concealment methods by the instigator and took full advantage of social media platform's mechanisms of recommendation, which are in themselves particularly invasive in terms of access to the thoughts of the users.[666] As such, the conduct contravened the FT&RHO of the target audience.

There is a growing voice calling for account to be taken of the role that the FT&RHO could play in setting boundaries to state-sponsored influence. The relevancy of this argument will only grow stronger with the development of virtual-reality spaces such as the Metaverse, and the development of implantable brain–machine interfaces, which is currently spearheaded by the Company Neuralink. Such technological advancements could have dramatic consequences in terms of the manipulative value of PSYOPS in general, and disinformation activities in particular.

[666] Irene Khan, *Supra* note 47, at 7; Vosoughi, Roy and Aral, *Supra* note 51; The legal implications of the use of identity concealment tactics in the context of foreign-led disinformation campaigns go way beyond the potential violation of the FT&RHO Ohlin, *Supra* note 142.

More work is needed to better determine what ought to be considered an illegal form of online manipulation.[667] Lawyers, sociologists and psychologists need both be involved to clarify the contours of the FT&RHO.

3. The right to self-determination

Article 1 of the ICCPR stipulates "[A]ll peoples have the right of self-determination. By virtue of that right they freely determine their political status and freely pursue their economic, social and cultural development."[668] The right to self-determination has an internal and an external aspect. In virtue of external self-determination, an identifiable group of people has a right to access to nationhood, which relates to independence.[669] In virtue of internal self-determination, all peoples have the ability to freely determine their political status and freely pursue their economic, social and cultural development. Although the customary right to self-determination has mainly been of use in the context of the decolonization process, in connection to a right to secession, some commentators have also pointed to the applicability of *internal* self-determination to episodes of foreign electoral interferences targeting established states.[670]

Cassese describes internal self-determination as the ability for a people in a sovereign state "to elect and keep the government of its choice".[671] The right to self-determination is premised on a nation's ability to make properly informed choices. It is not the lack of knowledge about political questions that is problematic here. *Uninformed* voters are no dangers to democracies. However, the possibility that a substantial number of *misinformed* voters is misled by a foreign entity, into choosing a path detrimental to the interests of the collective directly contradicts with the right to self-determination of the voters.[672]

It is the view shared by some commentators, that peoples' right to *internal* self-determination could be breached when a foreign state attempts to interfere with electoral processes, as these are the dedicated ways through which individuals collectively express their

[667] See *Supra* "2. The freedom of thought and the right to hold opinion without interference."
[668] ICCPR, *Supra* note 118, at Article 1.
[669] Hannum, *Supra* note 501.
[670] See Antonio Cassese, "Political Self-Determination – Old Concepts and New Developments" in Antonio Cassese (ed.), *UN Law/Fundamental Rights: Two Topics in International Law* (Florence: Sijthoff & Noordhoff, 1979), at 137; Tsagourias, *Supra* note 409. See also Ohlin, *Supra* note 142. At first in disagreement with these views, Schmitt seems to seriously consider the relevancy of the right to self-determination to the question of foreign interferences. See Schmitt, *Supra* note 739. See also Lahmann, *Supra* note 62.
[671] Cassesse, *Ibid.*
[672] See in that sense and generally, Sandrine Baume, Véronique Boillet, Vincent Martene *Misinformation in Referenda* (London: Routledge, 2020).

preference regarding their political organisation; "[I]n a democratic society, this is the meaning of self-determination [to select their representatives] and it is inextricably linked to the electoral process."[673] The right to internal self-determination, as enshrined in the ICCPR, imposes upon all states a duty to allow and respect the sovereign expression of the will of the people. The right to internal self-determination encompasses several correlative duties, some of which are reconfirmed elsewhere in the ICCPR. Article 25 for instance provides a right to take part in the conduct of public affairs, to vote and be elected at genuine periodic elections. In virtue of the right to internal self-determination, states have a duty to ensure the sincerity of the vote in times of election, to ensure information pluralism, to ensure the free expression of the national will through vote, and to ensure the equal treatment of candidates.

Foreign actors are generally allowed to make public their preferences as regard the direction that a state takes in the process of deciding its political organisation. Foreign leaders are free to praise freshly elected foreign counterparts, to publicly express their discontent when a candidate loses an election, to finance the campaigns of a candidate within the limit of the law of the recipient state, or to criticize the posture of a candidate that is running for election. There are however clear boundaries on what foreign actors can say, or do with regard to matters of self-determination. For example, coercing foreign voters into voting for a particular candidate, or to abstain from voting, would, among other rules, constitute a violation of their right to internal self-determination if conducted or sponsored by state actors. Although no affiliation with the Iranian government could be established, an interesting case is this of two Iranian hackers who have been accused by the US authorities to have send threatening messages to tenth of thousands of American voters during the 2020 presidential election.[674] In their communications, the hackers notably threatened voters with violence if they did not cooperate. In the hypothesis that the hackers acted under the instructions of a state, and that the incriminated conduct was thus legally attributable to the instigator state, the later could bear international responsibility for a violation of the target's right to internal self-determination. Arguably, the instigator state could even be in breach of the principle of non-intervention.

Similarly, the state that attempt to manipulate foreign voters *via* deceptive operations in the hope of disrupting a vote or swaying the election result may violate the right to internal self-determination of the voters. On this point, Ohlin underlines the importance of transparency in

[673] Ohlin, *Supra* note 142, at 9.

[674] "Two Iranian Nationals Charged for Cyber-Enabled Disinformation and Threat Campaign Designed to Influence the 2020 U.S. Presidential Election" (18 November 2021) The US department of justice, <https://www.justice.gov/opa/pr/two-iranian-nationals-charged-cyber-enabled-disinformation-and-threat-campaign-designed>.

the relation between foreign or local authorities and the people. He 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."[675]

When foreign actors engage in covert disinformation activities to deceive voters into thinking that the views expressed by the disinformants are those of authentic nationals, they harm the right to self-determination of the voters, who are manipulated into thinking that the messages come from insiders. As a consequence, they lend value and credibility to the voices of outsiders as if they were coming from fellow citizens that share similar values and interest.[676] Essentially, foreign actors that engage in electoral disinformation campaigns via *astroturfing* tactics, pretending to be insiders, are stealing the right to self-determination of their target. They subvert the sovereign will of the target audience. The degree of harm inflicted to the collective right to self-determination is interpreted based on the quantity of false accounts used in the campaign, and the quantity of messages disseminated in the online information space of the voters. It is possible to imagine other hypothesis of a violation of the right to self-determination resulting from particularly invasive and interventionist disinformation campaigns. *Referenda,* for instance*,* may also relate to the political organisation of a state. Hence, any attempts from foreign states to interfere in the conduct of a referendum abroad *via* disinformation campaigns could constitute a violation of the self-determination of the affected audience, provided that the referendum relates to the political organisation of the state, and that it is clear that the disinformant intended to change the issue of the process, or to foster absenteeism. Unfounded accusations of fraud that cast doubt about the integrity of the vote, when made by foreign state actors, while morally inacceptable, are not constitutive of a violation of internal self-determination. They may however amount to a violation of the principle of non-intervention and the customary prohibition of revolutionary propaganda if clearly aimed at inciting revolt abroad.

Finally, it bears noting that the right to self-determination and the freedoms of thought, opinion and expression are deeply interrelated. This is to say, that there can be no authentic expression of the sovereign will of the people in a state where the freedoms of opinions, thought and expressions are not given due consideration. Consequently, state-sponsored disinformation

[675] Ohlin, *Supra* note 142, at 1.
[676] *Ibid*., at 15.

in times of election would simultaneously constitute a violation of the right to internal self-determination and the freedoms of thought, opinion and expression. [677]

4 The right to equality and non-discrimination

The right to equality and non-discrimination is enshrined in a number of international instruments such as the ICCPR, the ECHR and the International Convention on the Elimination of All Forms of Racial Discrimination (CERD), the later gathering 182 state parties. The right was further elaborated through a series of views, recommendations and judgements by the European Court of human right (ECtHR), the Committee on the Elimination of Racial Discrimination (CmERD) and the Human Rights Committee (HRC). The Declaration on Fundamental Principles concerning the Contribution of the Mass Media to Strengthening Peace and International Understanding, to the Promotion of Human Rights and to Countering Racialism, Apartheid and Incitement to War (or UNESCO Mass Media Declaration) also recalled the importance of human rights, in particular, the Freedom of expression and the right to not be subjected to discrimination in the context of international communications.

The ICCPR prohibits "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence."[678] In virtue of Article 20 ICCPR, states are duty-bound to criminalize certain forms of hate speech, and are prohibited from inciting discrimination, hostility or violence. The right to equality and non-discrimination provides legal standing to victims of defamatory statements and vilification campaign dispersed in the context of incitation to genocide, war propaganda and incitement to discrimination, hostility or violence, in general. These communications are often associated with the broader notion of "hate speech". That is to say, that a state would be liable for not criminalizing incitement, or hate speech that is instrumental in leading to discrimination, hostility or violence, or for not taking the appropriate corrective measures. That is to say, that the victims of a disinformation campaign that contravened article 20 are legally entitled to ask reparation to a state if said state has not taken the appropriate steps to protect the right to egality and non-discrimination of the

[677] The UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression has alternatively appreciated the legality of information communications during electoral process through the lenses of the freedom of opinion of expression. See "Report of Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue" HRC, 26th session, A/HRC/26/30, (2 July 2014).

[678] See also ICCPR, *Supra* note 118. Article 18§3(b)

victim while being in the capacity to do so. The principle of non-discrimination thus set exceptions to the freedom of expression.

Many forms of hate speech simultaneously constitute disinformation campaigns. In fact, as explained in Chapter I, disinformation is mostly used to ignite prejudices or promote hatred. [679] Malicious actors may use disinformation as a mode of production and as a mode of dissemination of hate. The latter relates to the use of bots and fake accounts and other inauthentic behaviours so as to artificially amplify a message. The former relates to the creation of false and misleading content so as to justify intolerance and give substance to a narrative. In the same vein, disinformants may have recourse to hate speech occasionally to catch the attention of a specific audience or to cause disruption without explicitly inciting hatred.[680] The question thus arises, as to whether and under what circumstances disinformation campaigns waged by foreign states and targeting foreign audiences that show aspects of hate speeches contravene the right to equality and non-discrimination?

Over the years, major hate speech cases which also constituted disinformation campaigns have attracted the attention of international lawyers. Some of them constituted blatant violations of international human rights law, clearly motivated by criminal intent, thereby potentially amounting to crimes against humanity. These were the anti-Jews propaganda conducted by the Nazi regime, the propaganda activities of the Bosnian Croats during the Yugoslavian War, the anti-Tutsi hate speeches promoted by the Hutus in the context of the Rwandan genocide, and the more recent discriminatory campaigns aimed at fostering ethnic division within the Burmese society and inciting hatred towards the Rohingyas ethnic group promoted by the government of Myanmar. All these cases involved the propaganda activities of the authorities of a state and the persecution of minorities located in the territory of the propagandist state. In the case of egregious hate speeches such as institutionalized hate speeches, international tribunals or the international criminal Court are competent to adjudicate the matter and to try the instigators.

With regard to transboundary hate speech, the pending hate speech case concerning the conflict between Armenian and Azerbaijan merits a particular intention. Since 2020 serious tensions re-erupted between Armenia and Azerbaijan over a territorial dispute, with armed hostilities taking place regularly around the border near the Nagorno-Karabakh region. The two

[679] Lisa Reppell & Erica Shein, "Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions" (2019) *International Foundation for Electoral Systems* (IFES); Bayer et al. supra note 63.

[680] William J. Aceves, "Virtual Hatred: How Russia Tried to Start a Race War in the United States" (2019) 24 *Michigan Journal of Race and Law* 177.

countries have accused each other of conducting illegal PSYOPS against populations. Azerbaijan notably accused Armenia of conducting and supporting disinformation campaigns in the Armenia-occupied territories to steer hatred against the remaining Azerbaijanis. Azerbaijan recently brought the issue to the attention of the ICJ.[681] In its request for provisional measures to the ICJ, Azerbaijan denounced "cyber disinformation operations to incite and stir ethnic hatred and violence against Azerbaijanis." [682] Precisely, Azerbaijan accuses the Armenian authorities to engage in cyber disinformation operations to incite anti-Azerbaijani hatred, and to turning a blind eye on the activities of Armenian hate groups engaging in similar acts. The claimant bases its claim on the CERD. The case is still pending at the time of writing and the ICJ has yet to pronounce itself on the merits. The impugned communication campaign could potentially contravene to the right to equality and non-discrimination of the aggrieved population.

With that being said, hate speech constituting egregious violations of international human rights law are relatively uncommon. Hate speech may be instrumental in driving political changes abroad, yet, more often than not, disinformation campaigns are articulated around subtle assumptions, stereotypes, inauthentic pieces of opinions or manipulated images rather than explicit incitement to discrimination or violence. That is to say, that neither the instigator or the information spreader is truly motivated by an "intense and irrational emotions of opprobrium, enmity and detestation towards a target group on the basis of a protected characteristic."[683] In the eventuality that a disinformation campaign contains hate speeches, pejorative and discriminatory language are used sporadically and disseminated alongside non-vilifying information to facilitate the spread, catch the attention of a specific audience, and provoke the expected social effects. Does this necessarily mean that disinformation campaigns containing hate speech aspects cannot breach the non-discrimination principles?

First of all, the concept of hate speech is broad. That is to say that not all hate speeches fall into the scope of Article 20 ICCPR. Censorship measures are only admitted in the context of incitement. The UN Human Right Office of the High Commissioner defines hate speech as "any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or

[681] 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.; Application of the international convention on the elimination of all forms of racial discrimination (Azerbaijan v. Armenia) (2021) ICJ, Order - Request for the Indication of Provisional Measures.

[682] Request for provisional measure, *Ibid.,* at 13.

[683] Article 19, "'Hate Speech' Explained A toolkit" (2015), at 9; Aceves, *Supra* note 680, at 236.

discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor."[684] Article 20 however prohibits "incitement to discrimination, hostility or violence", which arguably constitutes a narrower category of conduct than "hate speech" in the sense of the first definition. The question of intent is central to the determination of a breach of the right to equality and non-discrimination.[685] "Incitement", in the meaning of Article 20 indeed discloses a clear intention to discriminate or exclude. The intent to discriminate or exclude is obvious in presence of explicit incitement to discrimination. In Belkacem vs. Belgium, the ECHR was brought to decide whether racist utterances, anti-Christian and anti-atheist messages, insults and incitement to violence justified censorship measures in the name of the non-discrimination principle. The ECHR confirmed the view of the Belgium Court of Cassation that the applicant, had not "simply expressed his views, but had unquestionably incited others to discrimination on the basis of faith and to discrimination, segregation, hatred or violence towards non-Muslims, and had done so knowingly and therefore intentionally."[686] In contrast, the Court held that censorship measures against a Turkish politician that expressed its disagreement about the legal qualification too give to the mass deportation and massacre of Armenians committed by the Ottoman Empire in the 1910's, which is usually considered a Genocide, were ill-proportioned because the applicant merely shared its opinion without the intention to incite hatred.[687] In light of this, disinformation campaigns may contravene the right to equality and non-discrimination, as enshrined in the ICCPR, when the communications contained clear incitements. This substantially reduces the likelihood of a state-sponsored disinformation campaign falling into the scope of Article 20. The ECtHR, the HRC and the CmERD have however shared broader understanding of the category of communications prohibited under the non-discrimination principle.

Although explicit incitement to discrimination, hostility or violence systematically justifies the taking of censorship measures against the advocator, international jurisdictions have occasionally validated interferences to the freedom of expression in absence of explicit

[684] United Nations Strategy and Plan of Action on hate Speech (2019), <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf>.

[685] Toby Mendel, "Hate Speech: Can the International Rules be Reconciled?" (2011) Expert workshops on the prohibition of incitement to national, racial or religious hatred, Workshop for the Americas.

[686] *Belkacem v. Belgium* (2017) ECtHR, Decision on admissibility.

[687] *Perincek v. Switzerland* (2015) ECtHR (Grand Chamber), Judgement.

incitement on the basis that the impugned communications may foreseeably lead to incitement and violences.[688] Such an approach to article 20 is visible for example, in the HRC's communication J. R. T. and the W. G. Party v. Canada. The issue involved the use by the defendant, of dubious phone messages in order to promote his Party. After noting that the messages in question were all constructed so as to warn the listener about the purported "dangers of international finance and international Jewry leading the world into wars, unemployment and inflation and the collapse of world values and principles"[689], the HRC decided that "the opinions which Mr. T. seeks to disseminate through the telephone system clearly constitute the advocacy of racial or religious hatred which Canada has an obligation under article 20 (2) of the Covenant to prohibit."[690] In their individual opinions to the Faurisson v. France case, members of the HRC Elizabeth Evatt and David Kretzmer specified, about the conviction of an individual under Article 20 for an act that does not include the element of incitement, that "there may be circumstances in which the right of a person to be free from incitement to discrimination on grounds of race, religion or national origins cannot be fully protected by a narrow, explicit law on incitement that falls precisely within the boundaries of article 20, paragraph 2. This is the case where, in a particular social and historical context, statements that do not meet the strict legal criteria of incitement can be shown to constitute part of a pattern of incitement against a given racial, religious or national group, or where those interested in spreading hostility and hatred adopt sophisticated forms of speech that are not punishable under the law against racial incitement, even though their effect may be as pernicious as explicit incitement, if not more so."[691]

In some instances, the ECtHR considered the status of the speaker as well as the context of the communication as determining factors justifying the taking of censorship measures even in absence of explicit incitements. In Pastor v. Germany, the ECtHR for instance considered that holocaust denial made at the occasion of a Parliament session by a German Parliament member justified an interference in the freedom of expression of the speaker. Notably, the Court remarked that the fact that the speech had taken place in a democracy that "may be regarded as

[688] See in that sense Mendel, *Supra* note 685. See also *Garandy v. France* (2003) ECtHR Decision on admissibility; *Pastors v. Germany* (2019) ECtHR (fifth section), judgement; *Robert Faurisson v. France* (1996), HRC, communication; *M'Bala M'Bala v. France* (2015) ECtHR, decision.

[689] *J. R. T. and W. G. Party v. Canada* (1983) UNHRC, Admissibility, Communication No 104/1981.

[690] *Ibid.* §8(b).

[691] HRC (1996), *Supra* note 688. For more examples of illegal hate speeches not containing explicit incitement, see also Aceves, *Supra* note 680, at 234.

having a special moral responsibility to distance [themselves] from the mass atrocities perpetrated by the Nazis" had mattered into the decision.[692]

In addition, it has been remarked in the literature that the CERD's provisions on discriminatory speeches were substantially broader in scope than those of the ICCPR and the ECHR.[693] Article 4 of the CERD notably provides that state parties "[S]hall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another colour or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof;"[694] Article 4 CERD thus extends to hate speech simple, regardless of the question of intent or effect.[695] In virtue of Article 4 CERD, disinformation campaigns that include "ideas based on racial superiority or hatred" would breach the CERD, thereby engaging the international responsibility of a state party for the commission or the condoning of the wrongful act.

The 2012 Rabat Plan of Action sponsored by the United Nations High Commissioner for Human Rights attempted to clarify the notion of incitement to hatred.[696] Notably, the experts involved in the writing of the Plan established a six-pronged test for determining whether a communication fell into the scope of Article 20. According to the experts several cumulative parameters should be verified in order to establish responsibility. These are the context; the speaker, the intent, content and form, the extent of the speech act, and the likelihood, including imminence of harm. The Rabat Plan is interesting in two aspects. First, it appears to be incompatible with the CERD, as the later focusses exclusively on the content and the form of the act. Second, it clarifies the term of "incitement" as understood in the ICCPR. Specifically, the experts explained that explicit incitement to hatred are not required for a communication to breach Article 20, what count is that the communications could foreseeably lead to discrimination, hostility or violence and that the instigator clearly sought to steer hatred against individuals on the basis of their membership in a group.

---

[692] ECtHR (2019), *supra note 688.* In the same vein, about discriminatory statements made by a teacher, see Ross v. Canada, (1996) HRC, communication.

[693] Mendel, *Supra* note 685; "Prohibiting incitement to discrimination, hostility or violence" (2012) Article 19 Policy Brief, at 13.

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

[695] Mendel, *Supra* note 685.

[696] *Annual report of the United Nations High Commissioner for Human Rights*, (11 January 2013), A/HRC/22/17/Add.4. See also a shortened version of the report "One-pager on 'incitement to hatred'", <https://www.ohchr.org/sites/default/files/Rabat_threshold_test.pdf>.

It bears noting that incitement is an inchoate offence, which means that the author of the offence incurs responsibility regardless of whether consequences have resulted from the incriminated communications.[697]

In light of this, disinformation campaigns would contravene Article 20 ICCPR in ways that justify the taking of censorship measures against the speaker and engage the international responsibility of the instigating or condoning state if the communications;

- *contains explicit incitements to discrimination, hostility and violence towards a particular group based on the race, gender, religion, or the nationality of its members.*
- *may foreseeably lead to the commission of hate crimes by the members of the audience without containing explicit incitements*. In that case, it is necessary to inquire about the intent of the instigator, and specifically, to pay attention to the context of the spread, the content of the communications and the audience, including its historical background.

Explicit incitements or blatant hate speeches would hardly go undetected and unpunished. Hate speeches made online would undoubtably prompt further investigations to find the real identity of the instigator and his intentions. Therefore, malicious actors who resort to disinformation so as to pit off two ethnics groups against each other, worsen racial tensions within a society, or to merely pollute the online information space of the target state with unnecessary race-related debates are more likely to opt for legitimate-looking pieces of opinions of contentious nature that could in themselves trigger hate feelings among the audience, rather than plain incitements for violence that could become obvious red flags of maliciousness. The second hypothesis might thus constitute a better framework to assess state-sponsored disinformation campaigns targeting foreign populations.

Two cases merit intention here. These are, the migrant-related disinformation that pro-Kremlin outlets have been accused to disseminate in Europe between 2016 and 2017, and Russia's protracted effort to sow division in the American society ahead of the 2016 elections.

As regards the first case, it was indeed reported that following Europe's migrant crisis in 2015, a wave of anti-migrant disinformation communicated by Russian media targeted European audiences to sow division in Europe. Specifically, immigrants were depicted as "cynically abusive of the welfare system, and potential terrorists"[698] while the European

[697] Aceves, *Supra* note 680, at 235. See also *Ibid.*
[698] *EUvsDisinfo*; Meister, *Supra* note 554.

governments were presented as "weak and naïve."[699] The Lisa case is a typical example of anti-European migrant disinformation partially driven by a foreign government. The Lisa case is the story a young German Russian girl who falsely declared after a 30 hours disappearance having been kidnapped and raped by migrants before admitting that she lied. Even after her retractation, the story was abundantly relayed by Russian media to criticize the failure of German authorities to respond effectively to the case and to pinpoint the migration issues in Europe. Specifically, Russian media framed the information in a way that showed a causal relationship between the increased movement of migrant coming to Europe following the war in Syria in 2015 and a rise in insecurity in Europe.[700] Demonstrations were subsequently held in Germany to reclaim justice for the girl. Neo-Nazis groups reportedly joined the demonstrations. The German minister of foreign affairs, Frank-Walter Steinmeier accused the Russian authorities of instrumentalizing the case for political ends.[701]

The second example relates to US's race-related issues. It has been remarked that the foreign-sponsored disinformation campaign which targeted American voters ahead of the 2016 election also included communications clearly intended to sow division via the exploitation of race-related tensions.[702] In a partly declassified report, the US Senate Select committee on intelligence made mention of the acts of internet trolls allegedly affiliated to the IRA.[703] The report notably notes that "Posted to 10 of the IRA's YouTube channels, were 1,063 videos-or roughly 96 percent of the IRA content-dedicated to issues of race and police brutality. The names of the IRA's YouTube channels were consistent with the posted video content and included 'Black Matters,' 'BlackToLive,' 'Cop Block US,' 'Don't Shoot,.' and 'PoliceState.' The content of the videos posted to those channels exploits issues of extraordinary sensitivity inside the African-American community."[704] On this point an author has argued "[N]ot every ad promoted racist views. And not every ad called for violence. In fact, many did not. But every message had a distinct purpose. When the Russian social media campaign is considered *in toto*, it clearly violates the prohibitions against racial discrimination and hate speech."[705]

---

[699] *EUvsDisinfo*; Meister, *Supra* note 554.
[700] Baade, *Supra* note 148, at 7.
[701] "Fall Lisa: Steinmeier kritisiert Spekulationen aus Moskau" (28 January 2016) *Frankfurter*, <https://www.fnp.de/politik/fall-lisa-steinmeier-kritisiert-spekulationen-moskau-10632529.html>.
[702] Aceves, *Supra* note 680.
[703] Report of the select committee on intelligence, *Supra* note 555, at 38.
[704] *Ibid.,* at 58.
[705] Aceves, *Supra* note 680, at 232.

In both instances, the goal of the campaign was to sow division in the target society and diverting public opinion from certain topics. In both instances, the communications could have realistically and foreseeably led to the commission of hate crimes similar to what could have been achieved with explicit incitements to discrimination, persecution of violence. The migrant-related disinformation campaign which targeted Europe in 2016 however lacked the element of intention. It also did not constitute incitement in form. The security danger posed by migrants was certainly a highly implied idea, yet the focus of the campaign was placed on sowing distrust towards the European authorities. The second case is easier to assess in that the campaign clearly included explicit incitements to violence and discrimination.[706] It was also relatively clear that the intent of the instigator was to pit-off two communities against each other by exploiting pre-existing race-related tensions in the US society.[707] For these reasons, it seems more plausible to consider part of the disinformation targeting the US ahead of the 2016 elections as illegal hate speech susceptible of engaging the international responsibility of the instigating or condoning state than the European migrant-related disinformation. Be that as it may, and although hate speech and disinformation are intrinsically tied to each other, not all disinformation campaigns constitute hate speech. In practice, disinformation campaigns rarely contravene Article 20.[708]

5. The right to life

Article 6 of the ICCPR stipulates "No one shall be arbitrarily deprived of his life."[709] This article provides a right to life, in virtue of which any person, including states, is prohibited from engaging in activities that could endanger or end the lives of individuals. The right to life is enshrined in a number of international human rights treaties, including the ECHR, it is also proclaimed in many national constitutions. The right to life sets a correlative duty for states and authorities to ensure that nothing threatens the life of individuals. The same way as the right not to be subjected to torture or to inhuman or degrading treatment or punishment and the FT&RHO, the right to life is absolute. This means that no exceptions to these rights are admitted. The right to life also sets a correlative duty to preserve public health. The protection

[706] Aceves, *Supra* note 680, at 232.
[707] *Ibid.*
[708] Bayer and al., *Supra* note 63, at 36.
[709] ICCPR, *Supra* note 118. Article 6§1

of the rights to life and to health may in some circumstances justify restrictions to some other rights protected by the ICCPR.[710]

Disinformation may endanger lives in many ways. For instance, disinformation during public health emergency or natural disasters may lead to grave violations of the right to life.[711] Take the hypothesis of a malicious actor purposefully indicating unreliable information to a disaster-stricken community, or spreading the lie that a new miraculous cure has been found against a deadly virus in time of pandemic. In these cases, the violation to the right to life and health would be explicit if there is a sufficient causal link between a disinformation campaign and the materialization of harm or casualties. However, as argued in chapter I, it can rarely be affirmed with certainty that disinformation *caused* a particular result. More often than not, disinformation causes a result in an indirect manner, and it is likely that other factors have contributed to a harm. This may lead to absurd situation where disinformation is obviously spread with criminal intent but the speaker remains undisturbed because the communications failed to provoke death and injury, or because so many sources have contributed to the harm that observers fail to establish solid causal nexus.

It has been observed in the literature that purposefully creating the conditions for the loss of life through the dissemination of false or misleading information suffices in establishing a causal nexus between the communications and the harm done to the right to life and to health, regardless of loss of life. Schmitt and Milanovic indeed argue that the "state agents who spread misinformation online that directly affects health or exposes individuals to significantly elevated risks violate their state's obligations to respect and protect the rights to life and health."[712] Accordingly, life-threatening behaviours would engage the right to life of the recipients.[713] Therefore, 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 would be of the nature to violate the right to life of the audience. When conducted by independent private actors, such conduct would engage the responsibility of the territorial state on the basis of its failure to prevent or stop the dissemination. This reasoning would have applied to some episodes of disinformation that contributed to cripple the management of health crisis abroad during the COVID-19 pandemic.[714]

---

[710] See for instance, ICCPR, *Supra* note 118. Articles 12, 18, 19, 21, 22.

[711] See generally Internews report, *Supra* note 647; Hollis, van Benthem and Dias, *Supra* note 61, at 18.

[712] Milanovic & Schmitt, *Supra* note 638, at 268.

[713] Hollis, van Benthem and Dias, *Supra* note 61, at 18.

[714] 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).

With regard to the right to the correlative right to health, the Office of the High Commissioner for Human Rights underlines that states have a duty to *facilitate, provide and promote* the right to health of the citizen placed under their respective jurisdiction, notably through the "dissemination of appropriate information relating to healthy lifestyles and nutrition, harmful traditional practices and the availability of services […]"[715] Accordingly, a state-sponsored disinformation campaign promoting dangerous cures to a disease would contravene the right to health as it interferes "directly or indirectly with the enjoyment of the right to health."[716]

To conclude, whether domestic or foreign, state actors who instigate a disinformation campaign that is of the nature to contribute to a situation of imminent danger for the audience violate the right to life of the audience. In addition, individuals are entitled to expect that the authorities of any states ensure that their lives are not threatened. This means, among other things, that all states are duty-bound under the right to life, to censor any communication that is of the nature to contribute to a situation of imminent danger for the lives of their respective citizen, as well as the lives of foreign citizen. This applies with an even greater force to communications made by an authority or a public figure, who by definition enjoy great influence.[717]

6. Freedom of expression and speech regulations

The above investigation has shown that disinformation may violate human rights law. It has also been shown that restrictions to the freedom of expression could be admitted in some circumstances, when the safety of individuals or the authenticity of democratic processes is at stake. Before moving to the conclusion of the chapter, it is important to emphasize the role of the freedom of expression in setting boundaries to speech restriction measures. A brief reminder is all the more needed that recent years have seen an increase on infringements of freedom of expression. (or freedom of speech)

The 1948 International Conference on Freedom of Information sponsored by the Economic and Social Council was the starting point of international debates on instituting the right to

[715] *CESCR General Comment No. 14: The Right to the Highest Attainable Standard of Health,* (Art. 12), (2000), Office of the High commissioner for Human Rights.
[716] *Ibid.*
[717] See for instance *Zeljko Bodrožić v. Serbia and Montenegro* (2005) HRC, views.

freedom of expression as it is known today.[718] The conference devoted attention to information plurality, the free flow of information, the freedom of information, and the importance of truth in international relations. The task of the conference was to "promote friendly relations among nations and to combat the dissemination of false or distorted reports likely to injure the friendly relations between States."[719] In that respect, the Draft Convention on freedom of information annexed to the Final Act of the conference proclaims, in its first article, the fundamental freedom to impart and receive information and opinion.[720] The document notes, however, that "it is the moral obligation of the Press and other agencies of information to seek the truth and report the facts, thereby contributing to the solution of the world's problems through free interchange of information bearing on them, promoting respect for human rights and fundamental freedoms […]"[721] The second article of the Draft Convention on freedom of information enumerates the correlative duties that stem from the freedom of opinion and expression. In doing so, it recognises that penalties and restrictions may be admitted if intended, *inter alia*, to sanction "The systematic diffusion of deliberately false or distorted reports which undermine friendly relations between peoples or States."[722] The United Nations Conference on Freedom of Information declares in the second annexed resolution that "propaganda which is either designed or likely to provoke or encourage any threat to the peace, breach of the peace, or act of aggression, and on the spreading of false and distorted reports; (a) Are contrary to the purposes of the United Nations as defined in the Charter; (b) Constitute a problem of the first importance calling for urgent corrective actions on the national and international planes;"[723] That is not to say however, that state-sponsored disinformation warrant the taking of censorship measures against the concerned entities by the recipient state. Indeed, the second resolution makes mention of fascist propaganda, racism and propaganda for war, which sets the threshold relatively high for censors. Resolution 14 also states that restrictions to the freedom of expression may only be permitted on ground of "public morality."[724] Another resolution mentions the necessity to maintain "national military security", the latter term being interpreted in a relatively restrictive manner.[725] Therefore, it is clear that any attempts to censor information sources for the sole reason of the misleadingness of a report would breach the freedoms of

[718] Downey, *Supra* note 76, at 347.

[719] UN Final act of the United Nations Conference on Freedom of Information, held at Geneva, Switzerland, from 23 March to 21 April, 1948, <https://digitallibrary.un.org/record/3806839>, at 11.

[720] *Ibid.,* at 14; "Draft Convention on Freedom of Information" (1949) A/RES/313 (IV).

[721] *Ibid.,* at 22.

[722] *Ibid.,* at 15.

[723] *Ibid.,* at 23.

[724] *Ibid.,* at 29.

[725] *Ibid.,* at 8 and 28; Downey, *Supra* note 76, at 346; Whitton, *Supra* note 60; Metzl, *Supra* note 180.

expression unless said attempt exclusively aims at ensuring national military security in face of imminent security threat, or preventing discriminatory violence.

The Conference was a successful attempt to centralize the discussions pertaining to freedom of information and its abuses, a necessity a period of aggravating tensions between the Eastern and the Western blocs. Interestingly, the Conference formally installed a relation between truth and the proper enjoyment of the freedom of information. It should however be noted that not all states agreed on the parameters of information regulation in peacetime, the US, as proponent of absolute freedom of expression, firmly opposed the idea of a limitation, whatever the justification.

Developments on the freedom of expression will eventually culminate with the adoption of the UDHR the same year, and later with the ICCPR. The ICCPR proclaims in Article 19 that "[E]veryone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice." In a third paragraph, the ICCPR specifies that restriction on the freedom of expression must be provided by law and are necessary, "(a) For respect of the rights or reputations of others, (b) For the protection of national security or of public order (*ordre public*), or of public health or morals" The freedom of expression is then not absolute. Restrictions are admitted under strict circumstances.[726] The exceptions are reminiscent of the "public morality" and "national military" exceptions mentioned in the 1948 Final act of the United Nations Conference on Freedom of Information. The first paragraph of Article 19 referring principally to discriminatory communications, the second referring to communication that are of the nature to contribute to a situation of imminent danger for the health and security of individuals, including incitement to violence. However, past UN Resolutions and GGE reports have consistently recalled the duty of states to abstain from interfering with individuals' freedoms in the context of information and communication technologies.[727] The UN Human Rights Committee provided examples of legitimate grounds of restriction to freedom of exception. In its 2011 general comment, the Committee clarifies that exceptions to the freedom of expression may be admitted in limited circumstances such as in the context of election, to protect voters from coercion and intimidations that would prevent them from expressing their own will as

[726] See notably HRC, *Supra* note 647.

[727] See *inter alia Developments in the field of information and telecommunications in the context of international security*, (9 January 2014) A/RES/68/243; A/RES/70/237, (30 December 2015); A/RES/75/240, (31 December 2020); *The promotion, protection and enjoyment of human rights on the Internet*, (5 July 2018) A/HRC/RES/38/7; (1 July 2016) A/HRC/RES/32/13.

regards the political organisation of a state.[728] In general, restrictions must meet the following cumulative criteria; be prescribed by law, have a legitimate purpose, be necessary in a democratic society, and be proportional to the interest protected. For example, at the occasion of an allegation of infringement on the freedom of expression of a Turkish journalist by the turkey government, the ECHR considered that harsh criticism of an official" failed the test of "necessary in a democratic society", even though some parts of the impugned communication were "provocative and inelegant and certain expressions could legitimately be classed as offensive".[729] The Court first assessed that the restriction was indeed prescribed by law, and that it pursued the legitimate purpose of protecting the reputation of others. However, the ECHR considered that the restriction was not answering a "pressing social need" to protect the reputation of the concerned official. Notably, the Court considered that the impugned communications were presented in a "satirical way", that they had sufficient factual basis, and that the purpose of the communications was not to wantonly insult the concerned official. In addition, under the right to freedom of expression, national authorities are also duty-bound to dully explain the reasons of a restriction and to notice the speaker of an interference on its freedom of expression.[730]

In spite of the above, recent regulations pertaining to the regulation of disinformation have been a source of concern.[731] For instance, in October 2022 the UN Human Right Office expressed concerns about new amendments made to the Turkish criminal Code that establishes sanctions for "publicly disseminating false information" on digital platforms.[732] In 2021, Singapore passed a controversial 250 pages-long Foreign Interference (Countermeasures) Act (FICA) which introduces countermeasures intended "to prevent, detect and disrupt foreign interference in (…) domestic politics conducted through (i) hostile information campaigns and (ii) local proxies (Politically Significant Persons)." The law was criticized notably for the reason that its provisions are so broadly stated that it would constitute a serious threat to the freedom of expression.[733] Prior to the FICA, Singapore already passed a controversial Bill that aimed to

[728] *Ibid.,* at 7.
[729] Tusalp v. Turkey, (2012) ECHR (second section), judgement.
[730] *Ibid.* See also *Shin v. Republic of Korea* (2004) HRC, views.
[731] Irene Khan, *Supra* note 47.
[732] "Türkiye: Concern mounts over human rights implications of 'disinformation bill'" (14 October 2022) *UN News*, <https://news.un.org/en/story/2022/10/1129567>.
[733] See FICA, *Supra* note 247; Speech of He Ting Ru, a parliament member of the Working Party of Singapore on Foreign Interference (Countermeasures) Bill at the Singaporean parliament on October 4th 2021. Full speech available in YouTube, <https://www.youtube.com/watch?v=LUVnI7xkiMo>; "Singapore: Withdraw Foreign Interference (Countermeasures) Bill" (13 October 2021) *Human Rights Watch*, <https://www.hrw.org/news/2021/10/13/singapore-withdraw-foreign-interference-countermeasures-bill>.

curb online disinformation.[734] More recently Russia passed a law that introduces prison sentences for the act of spreading what the authorities have considered "fake news" about Russia's military involvement in Ukraine.[735]

There is a general trend among states to address the disinformation problem through domestic legal means. Given the lack of clear framework of rules for disinformation, some recent regulations have led to an erosion of the freedom of expression.[736] Some national legislations on disinformation 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. This only underlines the importance of finding neat theoretical and legal boundaries to what may constitute inadmissible yet legal disinformation campaigns, and what constitute illegal disinformation campaigns against which censorship measures are admitted. Simply put, the more obvious the harm or the risk of harm resulting from a disinformation spread, the more plausible the argument that interferences into the freedom of expression of the speaker are needed. On that point, and in view of the previous sub-sections, censorship measures may be admitted against disinformation only under two limited circumstances;

- *When disinformation contributes to a situation of imminent danger for the life of the members of the audience, including acts of incitements to hatred, disinformation during health crisis, natural disaster and wartime.*
- *When a disinformation campaign is of the nature of denying voters their right to internal self-determination, or corrupting democratic processes abroad in ways that impedes the expression of collective will.*

Save as aforesaid, interferences into the freedom of expression of a disinformant are difficult to justify. There is a need to find the appropriate threshold of harm when it comes to

[734] See an updated survey of the actions taken by governments against online misinformation at Funke & Flamini, *Supra* note 377.

[735] 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/>.

[736] "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>; Funke & Flamini, *Supra* note 377; Irene Khan, *Supra* note 47.

assess the impact of disinformation to human rights, and it is obvious that the truth-value of the content alone would not count as a standalone criterion to conclude on the illegality of disinformation.[737] The harm will have to be appreciated alongside an array of cumulative indicators including the context of the spread, the content of the communication and the inferable intent. Of course, there will be situation where the harm or the risk of harm is imminent harm is obvious, when the health and life of individuals are at stake. Conversely, there may be situations where the harm or risk of harm is unprovable, or too elusive. This might be the case with respect to the harm inflicted on the FT&RHO. Indeed, although it has been argued previously that highly manipulative communication practices would undoubtably contravene FT&RHO of individuals, it is difficult to conclude on the illegality of manipulative practices, let alone, to argue that foreign-sponsored manipulative practices should be prohibited, as such an argument could lead to overdetermination.

7. The question of the extraterritorial application of human rights treaties

International human rights treaties are particularly relevant when it comes to set limits to psychological warfare. Human rights treaties, and, at a broader extent human rights law, provide a comprehensive framework to assess state-sponsored online disinformation activities, whether directed abroad or at a domestic audience. It is a well-needed complement to the constellation of relevant state-centric rules which, it has been shown on a previous chapter, seldom apply to the context of psychological warfare.

Yet, the one remaining uncertainty when it comes to using a human-centric framework for assessing the legality of disinformation campaigns relates to the question of extraterritoriality. This issue pertains to the fact that the ICCPR, much like other important human rights treaties, is not of extraterritorial application. Indeed, the ICCPR contains a jurisdiction clause that limits the application of its provisions to a state in the context of its relationship with persons located "within its territory and subject to its jurisdiction."[738] This means that states parties to the treaty are only bound to respect and ensure protection of the rights contained in the text within their respective territories.[739] Although jurisdiction clauses of the sort are not always explicit in

[737] Bayer and al., *Supra* note 63; Irene Khan, *Supra* note 47, at 9.
[738] ICCPR, *Supra* note 118. Article 2.
[739] See among other Schmitt, *Supra* note 422; Milanovic & Schmitt, *Supra* note 638; Ohlin, *Supra* note 459, at 1585.

human rights treaties, it is always understood that human right treaties oblige the parties *vis à vis* persons situated in their territory and placed under their jurisdiction.

First of all, it has been pointed out in many occasions that restrictive interpretations of the scope of human rights treaties could lead to absurd results in terms of human rights protection, in addition to going against the very purpose of human rights treaties. Indeed, if states interact with the populations of other states in a systematic manner, the idea that states are only obliged to respect the rights of the persons located in their territories quickly becomes an untenable paradox. As a result, many scholars and human rights bodies have advocated for human rights extraterritoriality, with the inter-American Court of human right and the EtHR spearheading the movement for extensive interpretation of the extraterritorial application of human rights treaties.[740]

In spite of this, the legal proposition according to which human rights oblige states *vis à vis* all persons including individuals outside their territories remains a contentious one.[741] Specifically, while the idea of human rights extraterritoriality has gained broad support, disagreements have flared up with respect to the proper interpretation to give to extraterritoriality and the underlaying concept of jurisdiction.[742]

On the one hand, proponents of human rights universality consider that states should assume responsibility for the human rights violations arising from the acts or omissions of their agents and organs, whatever the circumstances of the facts. On the other hand, proponents of a more cautious interpretation of human rights extraterritoriality preconize restrictive interpretations of extraterritoriality that hinge on the amount of control state agent or organ really had over the victim at the moment of the commission of the prejudice. They justify their arguments invoking the fact that loose concepts of extraterritorial jurisdiction risk imposing more obligations on states than they initially agreed when taking part in a human rights treaty, which may have a chilling effect on states' future decisions to take part in human rights treaties. The criteria they advance for the establishment of the responsibility of a state for human rights violations

[740] See notably *Al-skeini and others v. The United Kingdom*, (2011) EtHR, judgement.; and Concurring Opinion of Judge Bonello to *Al-skeini and others v. The United Kingdom*, (2011) EtHR, judgement. See also *The environment and human rights*, (2017) Inter-American Court of Human Rights, Advisory Opinion OC-23/17, at 43. (Submitting an extremely broad interpretation of state jurisdiction).

[741] Jones, *Supra* note 119; Samantha Besson, "Due Diligence and Extraterritorial Human Rights Obligations – Mind the Gap!" 9(1) *ESIL Reflections* 1.

[742] Samantha Besson, "The Extraterritoriality of the European Convention on Human Rights: Why Human Rights Depend on Jurisdiction and What Jurisdiction Amounts to" (2012), 25(4) *Leiden Journal of International Law* 857; Jones, *Supra* note 119; Yuval Shany "Taking Universality Seriously: A Functional Approach to Extraterritoriality in International Human Rights Law" (2013) 7(1) *The Law & Ethics of Human Rights* 47, at 68.

occurring in the context of the operations they conducted abroad are various and not all authors seem to agree on the same parameters

To better grasp the source of the disagreement, it is important to understand that the question of human rights extraterritoriality depends on how state jurisdiction is interpreted in the first place; "[J]urisdiction amounts, therefore, at once to a normative threshold and a practical condition for human rights.[743] To give effect to human rights extraterritoriality, the victim of human rights violations arising from transboundary activities must be in capacity to obtain redress from the part of the state of origin. In other words, once an infringement occurs, it must be shown that, in the particular circumstances of each case, the duty-bearing state *owed* human rights protection to the injured individual. For some authors, this is the case if a jurisdictional link exists between the victim (the right-holder) and the state (the duty-bearer).

By ratifying human rights treaties, states accept to be bound by human rights obligations and the corelative duties of diligent prevention *vis à vis* persons under their jurisdiction. State jurisdiction is traditionally understood as the control or authority a state enjoys over the persons, property and activities situated in its territory.[744] A state is presumed to be exercising all elements of jurisdiction locally, over its *own* territory. Therefore, fully constituted states must respect the rights of the individuals residing in their respective territories.

Under the most restrictive interpretation of state jurisdiction, states are not obliged towards persons situated outside their respective territories. However, in human rights circles, a consensus has emerged in support of the view that territory should not be the unique basis for jurisdiction when it comes to the application of human rights treaties.

Proponents of the functional model of extraterritorial jurisdiction for instance consider that state jurisdiction is not limited to its territory, it extends to places where duty-bearer exercises the *functions* of jurisdiction: authority and control. Under the functional model of jurisdiction, jurisdiction is not anymore interpreted *ratione loci*, it is interpreted in terms of the amount of authority and control states exercise over persons, regardless of territorial considerations.[745] As location and jurisdiction are dissociative, state jurisdiction, and so the reach of human rights treaties, extends beyond the territory of the duty-bearing state.[746] However, not all authors agree on the conditions of application of the functional model.

---

[743] Besson (2012), *Ibid*., at 863.

[744] Bernard H. Oxman, Jurisdiction of States, (2007), Max Planck Encyclopedia of Public International Law.

[745] Bonello, *Supra* note 740.

[746] See Janig Philipp, "Extraterritorial Application of Human Rights" in Christina Binder, Manfred Nowak, Jane A Hofbauer and Philipp Janig (eds.), *Elgar Encyclopedia of Human Rights* (2022 Forthcoming); See also *Advisory Opinion Concerning Legal Consequences of the Construction of a Wall in the Occupied Palestinian*

For example, Besson argues that for human right obligations to apply outside the territory of the duty-bearing state, a state must exercise "effective overall control" over a foreign territory, understood as a "the effective overall normative power over people,"[747] as in the case of military occupations. She adds that a state is also obliged towards persons situated outside its own territory, provided that it exercises effective overall control over them.[748] A state exercises effective overall control, extraterritoriality, as it does territoriality, through its organs and agents, including the military, understood as persons vested with the exercise of elements of state authority. The acts of the latter therefore engage the state the same way as they do locally, as long as they exercise effective overall control over persons abroad. Hence, states would incur responsibility for the human rights violations occurring in a territory over which they exercise overall effective control, and more generally, for human rights violations arising from the conduct of their agents as long as they exercise effective overall control over persons abroad.[749]

Judge Bonello, in a concurring opinion to the *Al-skeini and others v. The United Kingdom* case has submitted an alternative model of functional jurisdiction; "a State has jurisdiction for the purposes of Article 1 whenever the observance or the breach of any of these functions is within its authority and control."[750] The interpretation of judge Bonello thus seems to encompass more instances than Besson's.

Arguing that Bonello's interpretation of functional jurisdiction might be too broad and therefore prone to arbitrariness, Shany submits that "an extraterritorial projection of power by a state would qualify as encompassing affected individuals within that state's jurisdiction, if and when, the potential impact of the act or omission in question is direct, significant and foreseeable."[751] Yet, the same way as Besson's, The interpretation of Shany seems to be inapplicable to situations involving the negligence of the territorial state in preventing and abating transboundary human rights infringements committed by private actors situated in its territory. Unfortunately, this specific scenario is the most common in the disinformation context.

As a proponent of effective human rights universality, Milanovic takes a different approach to extraterritorial jurisdiction. Arguing that functionalist models of extraterritorial jurisdiction

Territory, (2004), ICJ. (Where the Court seems to have subscribed to the disjunctive criteria of jurisdiction argument.)

[747] Besson, *Supra* note 742, at 875.

[748] *Ibid*.

[749] *Ibid*.

[750] Bonello, *Supra* note 740.

[751] Shany, *Supra* note 742 at 68.

are prone to collapse, he submits a "mixed model" of jurisdiction. Notably, he argues that negative obligations should be detached from any considerations of territoriality, as long as they involve the acts of state actors. In other words, states should not violate the rights of persons, whatever their locations. Conversely, positive human rights duties, such as the obligations to ensure and enable, only apply locally.[752] His view is particularly compelling and seems to align with the opinion of Judge Bonello even though the latter qualified his interpretation of state human rights jurisdiction "functional.".

In light of the above, it is clear that the question of the extraterritorial application of human rights treaties is far from settled. As jurisdiction is the *sine-qua non* condition for the application of a human rights Treaties, the crux of the matter lies in the meaning to give to "jurisdiction" under human rights law to give raise to effective extraterritoriality. That being said, there is a clear incentive to render human rights extraterritoriality more effective, notably in presence of infringements involving physical harm.

Yet, as unintuitive as they may appear, restrictive views on the extraterritorial human rights jurisdiction of states (and thus on human rights extraterritoriality as a whole) remain relatively common in the literature. They are mostly the result of conciliatory attitudes with regard to state sovereignty and the fear that imposing too much on states could either backlash on human rights law, or end up suggesting a right to intervention in situations when foreign populations are victims of the freedom-destroying policies of a regime. However, conciliatory attitudes regarding states' human rights commitments are not tenable when it comes to customary human rights.

Customary human rights are the most important rights. They are peremptory norms (or *Jus Cogens* norms), from which no derogation is permitted. This concerns, at least, the right to life, the prohibition of torture, the principle of non-retroactivity of criminal offenses, as well as the prohibition of genocide and slavery and the principle of non-discrimination. [753] The ILC and the ICJ also consider the principle of self-determination to being universally applicable.[754] By

[752] Marko Milanovic, *Extraterritorial Application of Human Rights Treaties: Law, Principles, and Policy* (Oxford: Oxford University Press, 2013).

[753] See Malcolm N. Shaw, *International Law* (UK: Cambridge sixth edition, 2008), at 291; Yearbook of the ILC, 1988, vol. II, Part 2, at 64; *Advisory Opinion on the Legal Consequences for States of the Continued Presence of South Africa in Namibia*, (1971), ICJ; *Loizidou v. Turkey* (Preliminary Objections), (1996), ECtHR (Grand Chamber); *Bankovic et al. v. Belgium et al*., (2001) Decision of admissibility, ECtHR (Grand Chamber). Some authors have remarked that the extraterritoriality of human rights is gradually becoming less of a contentious question. See in that sense Hollis, van Benthem and Dias, *Supra* note 61, at 35.

[754] See "Commentary to draft article 50 of the draft articles on the law of treaties", (1966), 2 *Yearbook of the ILC,* chap. II, sect. C, at 247; Yearbook of the ILC, *Supra* note 753; Shaw, *Ibid..*; *Case Concerning East Timor* (Portugal v. Australia), (1995) ICJ judgement, §29: "In the Court's view, Portugal's assertion that the right of

nature consensual, these rights should apply extraterritorially, customary human rights do not need to be grounded on treaties. They exist out of states' collective conviction that they should comply to them. As a result, customary human rights should apply extraterritoriality, regardless of jurisdictional constraints and therefore, should be considered to be owed by states, to every person, whatever their locations.

The case of non-customary human rights is not treated in this work. However, for the sake of completeness, it is acknowledged that some rights may be subject to competing interpretations. This is notably the case of the FT&RHO. [755] Unclear contours tend to render the right prone to arbitrariness in their application and make it difficult to impose clear positive duties of prevention on states. Therefore, a distinction between infringements on customary (moral rights) and infringements on non-customary human rights might be useful in some specific instances, notably involving the conduct of private actors. Due to the fact that states may not all be at the same page when it comes to interpreted these rights, a margin of appreciation in the assessment of whether the point of a violation of non-customary human rights can effectively be made should be granted. That is to say, that criteria such as the scale and duration of the abuse could be applied in a more stringent way than for customary human rights. Most importantly, the traditional considerations of jurisdiction apply to them. That is to say, that it may not be possible to hold states accountable for their negligence in preventing violations of the non-customary human rights of persons located outside their territories resulting from the conduct of private actors in their territories. Basically, non-customary human rights are not always owed by states to everyone, at all time.

This is not to say, however, that states have a positive duty to enforce human rights law in territories they do not themselves administrate.[756] This is why Milanovic's interpretation of human rights extraterritoriality is particularly relevant. States are released from any positive

peoples to self-determination, as it evolved from the Charter and from United Nations practice, has an *erga omnes* character, is irreproachable.".

[755] On this point, it bears noting that the renewed commitments of states and organisations to uphold the rights to freedom of speech and the FT&RHO do not necessarily suggest that the rights in question have become customary law, even though several private associations and human rights watchdogs have persistently claimed that the rights to freedom of expression and FT&RHO were customary human rights. See for instance Richard Carver, "Training Manual on International and Comparative Media and Freedom of Expression Law" (2019) Media Legal Defence Initiative, <https://www.mediadefence.org/resource-hub/resources/media-defence-training-manual-on-international-and-comparative-media-and-freedom-of-expression-law/>; Katie Bresner "Understanding the Right to Freedom of Expression, An International law primer for journalists" (2014) Journalists for Human Rights (JHR), <https://jhr.ca/wp-content/uploads/2019/10/Understanding-Freedom-of-Expression-Primer-ENG-web.pdf>.

[756] See in general Milanovic, *Supra* note 752.

obligations to secure enjoyment of human rights and to prevent their violations in territories they do not administrate. However, positive human rights duties apply internally.

Thus, drawing inspiration from Milanovic's "mixed model" of extraterritorial jurisdiction, states should abstain from violating the customary rights of the audience while conducting domestic and transboundary PSYOPS.[757] This applies to the acts of state organs[758], persons or entities empowered by the law of the state to exercise elements of governmental authority[759], or entities placed at the disposal of the state to exercise elements of governmental authority[760], as their misconduct are directly attributable to the state. As shown below, this will also apply to the acts of private actors upon which the state exercise effective control.

Additionally, states have positive obligations to ensure effective enjoyment of human rights in the territories they administrate. However, when it comes to securing enjoyment of customary human rights law states should not knowingly allow their territories to be used to conduct disinformation campaigns contrary to the customary human rights of populations located outside their territories. As a result, states would not only incur responsibility for violations of customary human rights when the violations were directly committed by state agents or organs, they would also incur responsibility for negligence in preventing infringement on the customary human rights of foreign individuals as long as they are in capacity to act upon the source of the prejudice and that said source is under their jurisdiction. What this means in terms of the corelative positive due diligence duties is explored in the next chapter. For now, it is enough to note that states should fulfil their human rights duties and conduct policies with due diligence.

[757] Hollis, van Benthem and Dias, *Supra* note 61, at 34. See also Milanovic, *Ibid*.
[758] ILC, *Supra* note 81, at 40, Article 4.
[759] ILC, *Supra* note 81, at 42, Article 5.
[760] ILC, *Supra* note 81, at 43, Article 6.

Chapter V.
Attribution, negligence, and due diligence in the context of online disinformation activities

1. Introduction

The previous chapters were focused on the objective element of a breach of international law, that is, the question of the conformity of disinformation campaigns to international law. The previous chapters have demonstrated that cyber disinformation may violate certain human rights. It has also been shown that non-coercive disinformation campaigns could still be considered internationally unlawful when they serve a clear interventionist purpose. The present chapter focusses on the subjective element of a breach of international law, that is, the attribution of the act to a state, in this case, the attribution of disinformation campaigns constituting breaches of international law and/or international human rights law to a state.

Legally attributing an unlawful conduct to a state enables the defendant to invoke the international responsibility of the wrongdoer, which also opens the possibility for the defendant to refrain from performing the international obligations it owns *vis à vis* the wrongdoer. In other words, the circumstances excuse the non-performance by the defendant of its international obligations towards the wrongdoer. The different circumstances precluding wrongfulness (CPW) are spelt out in the fifth chapter of the ARSIWA. As for the main body of rules forming the regime of attribution of a wrongful act to a state, they are to be found in ARSIWA Part 1, Chapter II, in the jurisprudence of the ICJ as well as in jurisprudence of *ad-hoc* international criminal tribunals.

States may incur international responsibility for their actions and omissions.[761] Four possible scenarios may arise in the context of unlawful transboundary disinformation activities. The first possibility is that the organs of a state directly carry out an unlawful conduct. In that case, the process of attribution is straightforward. The second possibility is that non-state actors are carrying out the conduct for the account of a supporting state. In that case it is required to determine whether the case in question satisfies all the conditions to attribute the private

[761] ILC, *Supra* note 81, at 34, rule 2.

conduct to the sponsoring state. The third is that of a state condoning the activities of non-state actors placed within its jurisdiction. In that case, the territorial state might not incur international responsibility for the conduct of these non-state actors. It may instead, be liable for a failure to stop or prevent injurious transboundary and domestic consequences arising from disinformation activities committed by individuals placed under its authority.

In covering these three scenarios, the present chapter explores the different aspects of attribution and state responsibility, and examine the applicability of the rules of attribution and the application of due diligence in the disinformation context. This chapter first seeks to establish whether and under what circumstances the international responsibility of states can be engaged for their direct or indirect involvement in online disinformation campaigns. As both direct and indirect attribution models are seldom applicable to the disinformation context, the second part of the chapter is dedicated to the study of various regimes of state responsibility for negligence in preventing the nefarious transboundary consequences of a private act.

The first section is an introduction to the question of the attribution of unlawful conduct to a state. The question of evidentiary standards and the technical difficulties inherent to the attribution of *cyber* conduct to a state are covered in this first section (2). The second section deals with the attribution of disinformation campaigns to a state, including the attribution of acts acknowledged by a state as its own and the attribution of the acts of non-state actors to a state (3). As an alternative to the perceived unsuitability of classic attribution regimes to the specificities of the present context, the third section looks at state responsibility through the lenses of the regimes of state liability for negligence, and corporate liability (4).

2. Technical aspects of the attribution of cyber-disinformation activities to a state

The process of attribution presents both a technical and a legal dimension. When it comes to attributing cyber-enabled acts to a state, the attribution process is generally a three-steps process. [762] The identification of the machine used to carry out the conduct and the identification of the human behind it together proceed from the technical aspect of attribution, and rely on extensive forensic processes and proof-gathering investigations. The attribution of a conduct to a state proceeds from the legal aspect of attribution, which rely mainly on the frameworks developed in the international jurisprudence and the work of the ILC. This first section is dedicated to the technical aspect of the attribution process of unlawful online disinformation

[762] For a detailed account on the technical aspects of cyber attribution, see Delerue, *Supra* note 81, at 55-85.

campaigns. As such, it covers the peculiarities of the attribution of cyber-enabled disinformation activities (2.1), the evidentiary standards applicable in the context at hand (2.2), and the question of the disclosure of electronic communications (2.3).

2.1. The peculiarities of the attribution of cyber-enabled disinformation campaigns

States incur international responsibility when they commit a breach of international law.[763] A breach of international law may arise when state-centric rules of international law are violated by a state, or when a state intentionally infringes the rights of individuals. The breach may be attributed to a state directly, when its own organs carry out the conduct, or indirectly, when non-state actors such as proxies or other auxiliaries are called upon by the contractor state to carry out its reprehensible deeds.

In general, the attribution of malicious activities to a state is rarely a straight-forward process. However, without any visible movement of troops, personnel, weapons and with barely any quantifiable effects arising from PYOPS, legal attribution of intangible conduct to a state is a particularly arduous task. This is all the truer in cyber-enabled psychological warfare, where instigators can easily cover their tracks, and hide their true identities and locations.[764]

The outsourcing of a conduct to non-state actors is a common practice in psychological warfare. This is to say, that disinformation and propaganda have always been decentralised phenomena. Traditionally, states had recourse to state-controlled media, fake journalists and other undercover agents to plant false news abroad. Since the advent of cyberspace, disinformation has become cyber-enabled, thereby gaining in sophistication.

Today psychological warfare mostly occurs in cyberspace. This adds several layers of complexity to the task of attributing PSYOPS. A first layer of complexity relates to the technical aspect of attribution. Unsurprisingly, high covertness and recourse to non-state actors are recurring features of state-sponsored cyber disinformation campaigns. The involvement of a state in a disinformation campaign is often kept at a minimum, if not made completely invisible, in order to avoid political backlash and ensure plausible deniability for the instigator. Whilst state officials may sometimes engage in overt disinformation activities by themselves, the bulk of a disinformation campaign is always left to non-state actors. In turn, the executor may also

[763] ARSIWA, *Supra* note 81, article 1 & 2, at 32-36.
[764] See *Supra*, chapter I.

use fake or hijacked accounts (bots) and fake websites to hide their tracks. With regard to the alleged Russian interference in the 2016 elections, the ODNI report for instance recorded "a mix of agents of influence, cut-outs, front organizations, and false-flag operations."[765] Consequently, detecting cyber disinformation campaign and tracing back a dissemination to the initiator generally requires long and expansive investigation work that makes timely reaction difficult.[766]

The process of attributing cyber-enabled PSYOPS first requires detecting a variety of indicators ("assets") of coordinated online inauthentic behaviours. This means gathering of information associated with the sources of information such as the registration information of suspicious web pages, the date of creation of the accounts involved in the spread of problematic information, the behaviour patterns of suspicious sources, their IP addresses, and even the characteristics of the content they share.[767] Attribution may also necessitate the identification of the humans or the entity operating behind the suspicious accounts. This is easier said than done, as an author notes, "The best that can be achieved is a degree of confidence that an operation is *probably* run by a particular actor, or that various assets belong to a specific operation, but establishing who is behind the operation is seldom possible with open sources."[768] Finally, the geographical location of the executors is also needed, notably, in regard to state responsibility for negligence and the development of due diligence duties with respect to disinformation activities.

It bears noting that the attribution to a machine is not necessarily a mandatory step in the process of attribution of disinformation to a state. First of all, internet users, national media, or even state organs may sometimes share problematic information overtly. Similarly, it may not always be necessary to attribute a conduct to a person in order to attribute to a state. This is primarily the case when the organs of a state or state-controlled media engage in propaganda activities in their personal capacity, (white propaganda.) But more often than not, states don't risk exposure for spreading disinformation. For this reason, disinformation campaigns are mostly waged covertly, in social media platforms and online messaging apps, *via* proxies using identity concealing methods.

---

[765] ODNI, *Supra* note 41, at 2.

[766] See notably Eric Talbot Jensen & Sean Watts, "A Cyber Duty of Due Diligence: Gentle Civilizer or Crude Destabilizer?" 95 *Texas Law Review* 1555; Lahmann, *Supra* note 152, at 416; Delerue, *Supra* note 81, at 55; Nimmo, *Supra* note 150; Lahmann, *Supra* note 90, at 201-257; Crootof, *Supra* note 214, at 582.

[767] Nimmo, *Supra* note 150.

[768] *Ibid.*

With regard to the attribution to a person, or entity, the universality and opacity of cyberspace, the porosity of online platforms, the dynamics of interactions between the various actors in cyberspace can be sources of complications. A variety of actors interacts in cyberspace in an ongoing basis, regardless of geographical location.[769] With profoundly distinct stakeholders relying on the medium for different reasons, and interoperating at different levels, from different locations, it is all the more difficult to conduct accurate attribution.[770] States may outsource disinformation campaigns to private public relations (PR) agencies or private individuals around the world, thereby forming complex networks capable of reaching large audiences. In turn, executors may leverage the opacity provided by the different cyber tools such as IP-spoofing solutions, fake accounts, and even I.A. to hide their true identity, affiliations and geographical location.[771] Alongside state-sponsored disinformation, private actors, motivated by financial gains or simply patriotism may also spread disinformation for their personal benefit. Whether unwittingly or willingly, private and public actors may even engage in mutually reenforcing information sharing activities. Privately-conducted disinformation can be subsequently repurposed by a state in a disinformation campaign, thereby incidentally turning unaffiliated individuals into *de facto* unwitting agents of the sponsoring state. This creates an environment particularly conducive of mischaracterisations.

The ubiquity of disinformation in social media platform means that major digital companies, such as Twitter, Meta, Sina Weibo or Tencent play a prominent role in detecting and attributing disinformation campaigns. Since 2017, the major platforms Twitter and Facebook take action to better combat and prevent disinformation in their platforms, including updating their user policies, conducting regular investigations, risks assessments, and content removal operations.[772] For instance, Meta established operations centres dedicated to ensure the integrity of national election processes around the world. The fact that social media platforms have access to more tools than any investigators highlights the importance of public-private collaboration in that area.

[769] "Geography is of only limited relevance to the issue of attribution." Schmitt (2017), *Supra* note 162, at 83.
[770] Nimmo, *Supra* note 150.
[771] *Ibid.*
[772] *Ibid*; Francois, *Supra* note 764; Gleicher, *Supra* note 42.

2.2 Evidentiary standards for the attribution of disinformation to a state

Attributing a series of false or misleading communications to a person or entity with the view to legally attribute the conduct to a state requires solid evidences the nature of which international law does not provide clear instructions about. The determination of what are to be considered sufficient evidentiary standards to support an accusation constitute yet another challenge in the context of the attribution of disinformation activities.

With regard to the attribution of a conduct to a state, the general lack of evidentiary standards in international law is a well-known yet under-addressed problem. Lahmann notes that "[…] the International Court of Justice [(hereinafter ICJ)] has adopted the scheme that the more grave an allegation against a state is, the higher the corresponding standard of proof needs to be."[773] The international jurisprudence does not however provide specific prerequisites on this matter. In general, the nature of the violation and the responses it may trigger enlighten on the quality of the proofs to be furnished. Lahmann asserts that for the attribution of a tort amounting to a use of force, the evidences clearly need to be "clear and convincing."[774] In light of this, attribution of particularly unacceptable disinformation campaigns creating a situation of imminent damage or death would require clear and convincing evidences. This could concern those state-sponsored disinformation campaigns amounting to breaches of the prohibition of war propaganda, breaches of the non-use or threat to use force or institutionalized hate speeches, against which responses in the form of proportionate countermeasures could *inter alia* be admitted. Indeed, solid proofs would be all the more needed in such hypotheses that international communications constituting egregious violations of international and human rights laws would open the possibility for the victim state to have recourse to unilateral means of responses against the wrongdoer. Without getting into extensive details, such countermeasure could take the form of website defacements, server overload, targeted injection of factual information in the information ecosystem of the disseminator in order to make the "truth louder"; and radio frequency jamming. As response have to be proportionate, it is difficult to imagine self-defence measures involving military forces be used against disinformation campaign, or any PSYOPS. Responses would be restricted to countermeasures, which are necessarily non-forceful, or other means of lesser intensity.[775] More generally, social

[773] Lahmann, *Supra* note 90, at 73.
[774] *Ibid.,* at 170.
[775] ILC, *Supra* note 81, at 75, article 22 and articles 49-51; Schmitt M. (2017) *Supra* note 162, at 111-133, Rules 20 to 25; Robin Geiss & Henning Lahmann, "Freedom and Security in Cyberspace: Non-Forcible Countermeasures and Collective Threat-Prevention" in Katharina Ziolkowski (ed.), *Peacetime Regime for State*

media platforms would be required by the victim state to take-down problematic content if they did not already, in accordance with their content policy. Meta, Facebook's parent company for instance indicates in its "Transparency Center" that false information is systematically removed from the platform if it is "likely to directly contribute to the risk of imminent physical harm" or "likely to directly contribute to interference with the functioning of political processes." The company adds that "certain highly deceptive manipulated media" may also be removed.[776] Still, if reactions are limited to non-forceful responses, the evidences relied upon to justify a reaction by a state have to be clear and convincing when the allegations relate to a grave breach of international law.

Yet, online disinformation campaigns, as intangible and non-forcible acts *per se*, rarely, if ever, constitute grave violations of international law activating the possibility for the victim to use countermeasures. In fact, in the aftermath of the Russian meddling in the 2016 American elections, the US only responded with measures of retorsion, which, the ILC indicated, are not inconsistent with any international obligation and as such, are to be distinguished from countermeasures and other CPW.[777] Indeed, states are still reluctant to denounce state-sponsored disinformation as violations of international law, and a review of state practice does not bring much as to whether cyber disinformation conducted by a state enables the victim state to adopt countermeasures in order to force the disinformant to fulfil its obligations, should the conduct breach international law. Notwithstanding, the responses themselves will have to comply with human rights standards such as the freedom of speech and opinion. The only imaginable reaction is that social media platforms proceed themselves to the removal of the problematic content when they judge it necessary. With measures of retorsions being the only visible actions undertaken by victim states so far, it could be that the standards of proof are relatively low when it comes to attribute disinformation activities.

Of course, that is not to say that the existence of breach arising from a state-sponsored cyber disinformation operation can be established on the sole basis that a particular message massively circulates online, or that the truth value of a communication is weak. Similarly, allegations should not lend to much credence on the question of who profits the campaign. Caution has to be maintained in the process of attributing disinformation. This implies that a

---

Activities in Cyberspace. International Law, International Relations and Diplomacy, (Tallinn: NATO CCDCOE Publication, 2013), at 621.

[776] The two first categories of information to be removed by the platform may potentially constitute violations of international law if attributed to a state, and spread in a sufficiently intense manner. See Meta, Transparency Center, Policy details on misinformation, <https://transparency.fb.com/en-gb/policies/community-standards/misinformation/>.

[777] ILC, *Supra* note 81, at 128.

substantial number of elements must point to the direct or indirect involvement of a state. A certain level of coordination among various identified actors, an obvious targeting, a strong political connotation and manifest geopolitical stakes surrounding the dissemination of problematic information are as many elements that should be proven in the course of an attribution process. *In fine*, attributing disinformation operations to an entity remains, unless fallacious and misleading messages are overtly shared by significant individuals in a repetitive manner, or unless the instigator publicly admits its wrongdoing, the result of a body of convincing evidences dully gathered over a sufficient time period.

In practice, states are reluctant to provide much information about the evidences at their disposal. Revealing too much about an investigation would enable malicious actors to develop bypassing capabilities which would further harden the attribution.[778] How could it then be assured that the aforementioned evidentiary standards are dully followed by a state accusing another of breaching its international obligations? The Tallinn Manual 2.0 only briefly addresses the issue of public release of evidences with respect to the attribution of cyber operation. Taking note of the absence of clear state practice in that regard, the experts do not provide much insights. It seems however reasonable, as outlined in the Manual, to expect a minimum level of transparency, even when the accuser attributes an international tort, be it a relatively "minor" violation. In any case, the question of evidentiary standards in the context of the attribution of unlawful disinformation activities to a state should not be overlooked.

2.3 The question of the disclosure of electronic communications

Two noticeable difficulties may be encountered by investigators during the evidence gathering process. These are worth mentioning in a dedicated sub-section because of the numerous implications they may have with respect to accountability efforts in the context of disinformation campaigns.

The gathering of a sufficiently large number of communications, including images, videos and written statements is essential to the attribution process of disinformation campaigns. However, online content, including image, video and recording, whether public or private, are easily removed by their authors. In addition, social media platforms conduct audit and may eventually proceed to content removal before the attribution process is initiated. Also, it is a fairly established proposition that private, or encrypted communications, which may equally be

[778] Schmitt, *Supra* note 211, at 21.

useful for accountability efforts, are protected by international law.[779] Many applications use end-to-end encryption to ensure the privacy of the communications. Some are notoriously used by malicious actors to disseminate disinformation, such as Telegram.[780] These may however be covered by the principle of confidentiality, or do they?

First of all, the right to privacy and the correlative principle of confidentiality do not extend to public information. In principle, investigators are free to access content published by private persons as long as the content was meant to be available to the public.[781] This concerns websites, blog and public publications made on social media platforms. It is worth noting, however, that publications made on social media or applications of communications may sometimes be meant to limited audiences even though they do not appear private in the stricter sense of the term. On this point, the Tallinn Manual experts note that there is "no clear threshold at which the right to privacy is implicated on the basis of the accessibility of communications."[782] Be that as it may, it can be assumed that the larger the intended audience, the weaker the protection afforded by the right to privacy.

With regard to content subsequently withdrawn by the communicator or deleted by the platform, the emerging tendency is to consider social media platforms duty-bound by an obligation of disclosure for the information they stored when said information can serve as evidence of an offence of international character. As once stated by Facebook's Oversight Board, "Facebook has a responsibility to collect, preserve and, where appropriate, share information to assist in the investigation and potential prosecution of grave violations of international criminal, human rights and humanitarian law by competent authorities and accountability mechanisms."[783]

The 2021 *The Gambia v Facebook* decision covered the question of the disclosure of electronic communications made on an American social media platform for the purpose of supporting a charge of genocide committed by Myanmar against the Rohingyas minority. The case featured a request made by The Gambia to the US District Court, to order the disclosure of deleted communications constituting anti-Rohingyas disinformation. The Court concluded that the content accessible to the public prior to the deletion by Facebook fell within the ambit of the statutory exception to the non-disclosure rule of the Stored Communications Act

[779] Schmitt M. (2017) *Supra* note 162, at 189-191, rule 35 "Rights enjoyed by individuals", §6-11. ("the right to privacy encompasses the confidentiality of communications.")

[780] Bermet Talan, "How journalists can address misinformation on Telegram?" (2021) Journalist Fellowship Paper, Thomson Reuters Foundation.

[781] *Ibid.,* at 191, §10.

[782] *Ibid.*

[783] Facebook Oversight Board, Case decision 2021-001-FB-FBR, at 37.

(SCA).[784] Facebook was thus ordered to disclose communications previously deleted from the platform that were now part of the platform's "backup storage." The Court based its decision on the fact that the communications in question could not enjoy the protection afforded by the right of privacy anyway, because they amounted to "coordinated inauthentic behaviour" (i.e. disinformation campaign) violating the platform's terms of service. Most importantly, the Court acknowledged "that the right to privacy in this case must be balanced against the need to uncover the cause of the Rohingya genocide."[785] The non-disclosure obligation remained applicable to private messages exchanged on the platform.

It transpires from these recent developments, that the international right to privacy may not apply to deleted online communications when they are components to actual or suspected egregious violation of international law. By the same logic, it is safe to assume that covert communications are not protected by human rights law if addressed to a relatively large audience. Again, strictly private online communications between a restricted number of individuals remain protected, whatever the nature of the exchanged content.

3. The attribution of unlawful disinformation campaigns to a state

A state incurs responsibility for its wrongful conduct on the basis of the regime of direct responsibility when organs of a state or persons possessing governmental authority (i.e. the capacity to exercise public powers, or *prerogatives de puissance publique*) engage in said unlawful conduct (3.1). A case of direct attribution to a state may also be made when the organs of a state or persons vested with governmental authority adopt the conduct of non-state actors as their own (3.2). A state may in also incurs responsibility for the acts of private actors on the basis of indirect state responsibility when it assists or supports the acts of other without directly participating in the misdeeds (3.3).[786] It is however often pointed out that current attribution regimes are mostly ill-fitted to provide effective attribution mechanisms in the context of cyber-enabled conduct, which concerns cyber disinformation, too (3.4).

[784] *The Gambia v. Facebook*, (2021) US District Court, (First Instance Court), Decision; Michael A Becker, "The Gambia v Facebook: Obtaining Evidence for Use at the International Court of Justice (Part II)" (6 October 2021) *EJIL:Talk!,* <https://www.ejiltalk.org/the-gambia-v-facebook-obtaining-evidence-for-use-at-the-international-court-of-justice-part-ii/>.

[785] *Ibid.,* at 18.

[786] The last situation being the most common in the context of interstate disinformation activities, the present work intentionally refers to state-sponsored activities rather than state-conducted activities.

It is worth recalling that for the various attribution regimes to be applicable, the behaviour of a state or private actors must first contravene international law. In absence of specific treaty provision specifically regulating the dissemination of false and misleading statement by a state with the intention to undermine the political integrity of another party, or to cause harm abroad, references should be made to general international law. This means that the victim state will have to prove that a state engaged in its personal capacity, or knowingly endorsed content originating from private actors for the specific purposes of intervening in the politics of foreign states processes abroad, and/or to cause harm or put the lives of foreign individuals in danger, which, as argued earlier, are the only hypotheses where claims of violation of international law can be made with respect of transboundary disinformation.

3.1 Disinformation activities conducted by a state directly

Disinformation campaigns constituting internationally unlawful acts may be attributed directly to a state if conducted by the organs of said state. Article 4 ARSIWA provides "1. The conduct of any State organ shall be considered an act of that State under international law, whether the organ exercises legislative, executive, judicial or any other functions, whatever position it holds in the organization of the State, and whatever its character as an organ of the central Government or of a territorial unit of the State. 2. An organ includes any person or entity which has that status in accordance with the internal law of the State."[787] Article 5 adds "The conduct of a person or entity which is not an organ of the State under article 4 but which is empowered by the law of that State to exercise elements of the governmental authority shall be considered an act of the State under international law, provided the person or entity is acting in that capacity in the particular instance."[788]

Sate, as abstract constructs, incur international responsibility for the conduct of their own organs, which include state officials, if the latter are purposefully engaging in unlawful conduct. For instance, the US would incur international responsibility for the actions of the CIA in case of a violation of international law by the latter. The CIA is an official organ of the US, officially instituted through the adoption the 1947 National Security Act, making it a *de jure* organ of the US in the sense of Article 4 ARSIWA.[789] Similarly, the Main Intelligence Directorate of the Russian Army (hereinafter GRU), is an organ of the Russian Ministry of Defence, is a *de jure*

[787] "The conduct of any organ of a State must be regarded as an act of that State"; ILC, *Supra* note 81, at 40.
[788] *Ibid.,* at 42.
[789] Schmitt, *Supra* note 211, at 4.

organ of the state, as opposed to a *de facto* organ that would not owe its status from the law of the national state. All its actions thus have to be interpreted as those of the Russian state itself. Finally, France's COMCYBER, which commands the French cyber forces, was officially instituted by decree in 2017 and operates under the authority of the Chief of Staff of the French Army. The framework of action of the French COMCYBER extends to transboundary online influence operations.[790] The French cyber forces exercise official functions of defence and influence in cyberspace. Violations of international law committed by the French cyber forces would be attributable to France on the basis of direct attribution for the conduct of organs of a state.

In contrast, the acts of non-state private persons or entities, such as the infamous Internet Research Agency (hereinafter IRA), a Russian Company blamed for conducting disinformation campaigns abroad, notably during the 2016 US election, in principle, would not engage the international responsibility of the national state, at least to the extent that direct attribution is concerned.[791] With regard to the IRA, despite alleged connexions between the IRA's founder and the Russian president, the organisation does not have the status of an organ under the law of the Russian Federation. In the same vein, atrocities and disinformation campaigns committed by paramilitary groups are generally not attributable to the national state of origin on the basis of direct attribution for the conduct of organs of the state, because such groups are not by definition organs of a state.

With that being said, some accusations were levelled against the GRU, based on findings pointing to its involvement. These notably concerned the hacking of the emails of the US democratic party and their subsequent release to the public.[792] These charges, if substantiated, would give rise to a case of direct attribution to the Russian Federation. The same way, the "Macron leaks", a series of information operations that targeted the team of Emmanuel Macron during the 2017 presidential election generated claims of Russian involvement. In spite of "circumstantial and contextual elements pointing to Russia"[793] no legal conclusion could be drawn with certainty at the time of the elections. However, two years later, the investigations conducted by the *FireEye* group and *Google* finally reached the conclusion, that the GRU, an organ of the Russian Federation, may have been involved in the "Macron Leaks."[794]

---

[790] L2I, *Supra* note 33.
[791] *Ibid.,* at 4; ODNI, *Supra* note 41, at 4.
[792] See in general ODNI, *Supra* note 41.
[793] Jeangène Vilmer, *Supra* note 538.
[794] Martin Untersinger, "Les preuves de l'ingérence russe dans la campagne de Macron en 2017", (6 December 2019), *Le Monde,* <https://www.lemonde.fr/pixels/article/2019/12/06/macronleaks-des-hackeurs-d-etat-russes-ont-bien-vise-la-campagne-presidentielle-de-2017_6021987_4408996.html>.

The acts of state-owned media, who often play a leading role in spreading disinformation abroad for the account of the owner state are similarly difficult to attribute to a state. Obviously, state owned media are not an official organ of the state.

Direct responsibility may also arise when the entity is formally empowered by the law of a state to exercise "elements of governmental authority." The concept of "elements of governmental authority" is not well defined in international law.[795] The determination of what constitute elements of governmental authority is mostly a matter of appreciation. On this point, the commentaries under Article 6 ARSIWA provide that "[B]eyond a certain limit, what is regarded as 'governmental' depends on the particular society, its history and traditions. Of particular importance will be not just the content of the powers, but the way they are conferred on an entity, the purposes for which they are to be exercised and the extent to which the entity is accountable to government for their exercise. These are essentially questions of the application of a general standard to varied circumstances."[796]

Yet, in the eventuality that state-owned media and paramilitary groups could be said to exercise elements of governmental authority, respectively, to conduct cyber influence operations[797] or to conduct military operation for the account of a state, Article 5 clearly require that the entities in question are "empowered by the law of a state" to exercise such functions, which implies a formal arrangement.[798] As a result, proving that a formal delegation of governmental functions from the state of origin to a media it owns or a paramilitary group occurred with respect to certain acts committed by the said entities is difficult. Realistically, states would not take the risk of making such arrangements. Therefore, in the context of the relationship between the state and the media it owns, or between a paramilitary group and the state of origin, the nexus between the two entities is often too tenuous to make the point of a formal delegation of governmental functions.

3.2 Attribution of a conduct acknowledged and adopted by the state as its own

There is a rare possibility that state organs or agents openly and purposefully relay disinformation emanating from private sources, or straight-up encourage or reward non-state

[795] James Crawford, *State Responsibility: The General Part* (Cambridge: Cambridge University Press, 2013), at 129.

[796] ILC, *Supra* note 81, at 43, Article 5, §6.

[797] Even though see Lahmann, *Supra* note 152, at 416. (Arguing that it is questionable that the dissemination of information via news reports or other media activity can be described as exercising state authority).

[798] ILC, *Supra* note 81, at 43, Article 5; Crawford, *Supra* note 795, at 132.

actors for conducting disinformation campaigns against foreign populations. Encouragements or incitation proffered by state organs and agents *after* the private actors initiated the campaign, or *during* the campaign, may amount to *a posteriori* acknowledgement and adoption of a conduct by a state as its own, thereby falling under the scope of Article 11 ARSIWA.[799]

Acknowledgement and adoption differ from incitement *stricto-sensus* in that public authorities do not order or encourage private actors to engage in a behaviour *before* it has started, but instead endorse the actions of private actors as they unfold.[800]

In general, the cumulative acknowledgement and adoption criterion have to be inferred from the conduct. It is only when the intention of the agent or organ of a state is clearly to support and intensify private conduct that the state's own actions can be considered an acknowledgement and adoption of a conduct by a state as its own. The ARSIWA provides that "[…] what is required is something more than a general acknowledgement of a factual situation, but rather that the State identifies the conduct in question and makes it its own."[801] Citing the *United States Diplomatic and Consular Staff in Tehran* case concerning Iranian students seizing the US embassy and consulates in Iran and holding hostage their personnel[802], the ILC adds "[…] as a general matter, conduct will not be attributable to a State under Article 11 where a State merely acknowledges the factual existence of conduct or expresses its verbal approval of it."[803] The ILC makes it rather clear that in order to activate Article 11, "[…] what is required is something more than a general acknowledgement of a factual situation […]"[804] such as an active behaviour from the part of the state.

Hence, publicly refereeing to the communications campaign in question, or the thesis advanced by the campaigners, even in a favourable way, is not sufficient in itself to engage the responsibility of the authorities of the state on the basis that it has acknowledged and adopted the conduct as its own. "The language of "adoption", on the other hand, carries the idea that the conduct is acknowledged by the State as, in effect, its own conduct" insists the ILC.[805] As a matter of fact, the Court considered in its *United States Diplomatic and Consular Staff in Tehran* case that the decree by which the Ayatollah Khomeini officially ordered the Iranian militants to *only* release the black people and the women who could prove that they were not

[799] ILC, *Supra* note 81, at 52, Article 11.

[800] It bears noting that incitement for the commission of unlawful acts is not expressively regulated by the ARSIWA. The inciting state may however bear international responsibility on the basis that it has exercised direction and control over the private entities it ordered. See Roscini, *Supra* note 161, at 39.

[801] ILC, *Supra* note 81, at 52, Article 11, §6.

[802] *United States Diplomatic and Consular Staff in Tehran*, (1980) ICJ Judgment.

[803] ILC, *Supra* note 81, at 52, Article 11.

[804] *Ibid.*

[805] *Ibid.*

spying for the account of the United States, could not be seen as a mere expression of complacency. Although not mentioned *verbatim* by the Court on the merits, the "adoption" criterion was nonetheless inferable from the declarations and policies adopted by the Iranian authorities regarding the violation of the American diplomatic premisses by Iranian militants.

First of all, applying Article 11 ARSIWA to private disinformation entails some risks for human rights law, notably, the freedom of thought, opinion and expression as enshrined in Articles 18 and 19 of the ICCPR.[806] Although it is clear that a certain level of diligence is expected from a responsible authority while communicating information, even public authorities have a right to make mistake. The point here is not to sanction poorly documented information reshared by an authority, nor is it to capture isolated reshare of false allegations by an authority. Mere criticisms made by a representative of a state, or inconstant value-based judgements, be they justified or not, ought not to constitute acknowledgements and adoptions of a pending privately-conducted disinformation campaign that seems to go along the same lines.

In the context of disinformation activities by private actors, the cumulative acknowledgement and adoption criterion could translate in explicit encouragements made by the authorities of a state, affixed to problematic content and addressed to their authors online, or in the act of relaying publicly problematic content shared by private sources in a repeated manner. The "adoption" criterion could also be considered fulfilled when the authorities repetitively and in a systematic manner, praise a national media outlet over its biased, false, or misleading coverage of a specific matter.[807]

Concretely, the situation of organs or representatives of a state repetitively resharing private messages online would not constitute an acknowledgement and adoption of a private disinformation campaign in the sense of Article 11 ARSIWA until it is made clear, from the content of the communications, the occurrence of mutually reinforcing declarations made by the authorities, and the absence of public rectification regarding the content thus endorsed, that the state was willing to encourage and enhance the manipulative value and the disruptive potential of a campaign. Therefore, the threshold for determining that the conduct of a state *vis à vis* private disinformation amount to an acknowledgement and adoption of the said conduct is unlikely to be met. In that sense, Baade notes that the awarding of the "Order of Service to the Fatherland" medal to the editor-in-chief of the state-owned media outlet Russia Today (RT) by Russian president Putin for their contestable coverage of the Crimean conflict in 2014, while

[806] See ICCPR, *Supra* note 118, Articles 18 and 19.
[807] See Baade, *Supra* note 148.

clearly amounting to endorsement and approval, "cannot in good faith be understood as the Russian Government adopting international responsibility for that news coverage."[808]

It is worth noting that the refusal by the territorial state, to put an end to wrongful acts perpetuated by its citizens does not fall under Article 11 regime of application. At best could passivity account for a failure to exercise due diligence duties of prevention and mitigation of harmful conducts perpetuated by nationals, but the territorial state would not endorse the responsibility for the commission of the unlawful conduct in question.[809]

3.3 Attribution of unlawful private conduct to a state

Disinformation campaigns rarely, if ever, are the product of organs of a state, or persons or entity exercising governmental functions. While public actors may occasionally engage in disinformation activities in an overt manner, the recourse to *proxies* is a recurrent feature of cyber operation waged by states.[810] Most disinformation campaigns are outsourced to private actors such as troll farms and other shady "public relation agencies." The regime of state responsibility for the conduct of private individuals deals with such a scenario.

In principle, a state is not responsible for the wrongdoing of private entities. International law nonetheless provides for the possibility of holding a state accountable for the wrongdoing of private actors that are "neither agents nor organs of a State, nor members of its apparatus."[811] However, the criteria adopted by the international jurisprudence are quite stringent, to the point that many, if not all, cyber-enabled behaviours may remain unattributable under the current regime.[812] Nevertheless, since its *Nicaragua* case, the ICJ quite consistently applied the strict attribution criteria it expounded. The process entails a two-pronged approach to attribution of private conduct to a state; Firstly, through conducting a strict test that consists of verifying whether the private individuals acted as a "de facto" organ of the accused state, thereby acting as a mere extension of the latter's will. Secondly, through conducting a more relaxed "effective control" test, that consists of verifying whether the accused state personally enforced the acts

[808] *Ibid.,* at 1362.
[809] ILC, *Supra* note 81. See also section 3 and no-harm, infra.
[810] *See among other* ODNI, *Supra* note 41.
[811] ICJ, *Supra* note 196, Separate Opinion of *Judge* Ago, at 146, at 178, §16; ILC, *Supra* note 81, article 8, at 47.
[812] See generally Lahmann, *Supra* note 90.

committed by the private actors.[813] Worth emphasizing, the severity of the burden of proof that the accuser will have to assume will vary depending on which of the following situations it is referring to while attributing a conduct. The first implies that a large amount of evidence is brought in order to prove the extensive involvement of the state in the whole campaigns conducted by the private actors. The second implies an atomistic approach, that looks at specific acts individually, with the need for fewer yet seldom obtainable evidences to be handed.

This section looks at the de facto agency test (3.3.1), and the effective control test (3.3.2), and shares some reflexions about their applicability to privately-conducted disinformation campaigns. A third sub-section explores potential developments in the regime attribution of private disinformation campaigns to a state and their potential implications to private disinformation (3.3.3).

3.3.1 The *de facto* agency test, or "strict control test"

The methods of attribution of privately-conducted acts to a state were first expounded by the ICJ in its Nicaragua case. The case featured a revolutionary private group, the *contras*, which had conducted violent activities in the territory of Nicaragua with the aim of overthrowing the then-Sandinistan government. By the time, it became apparent that the United States was providing support to the *contras*.[814] Nicaragua accused the United States, *inter alia*, of being involved in the revolutionary activities of the *contras*, arguing that the defendant was "effectively in control of the contras, that it devised their strategy and directed their tactics, and that the purpose of that Government was, from the beginning, to overthrow the Government of Nicaragua."[815]

The Court first tasked itself with finding out whether the *contras*, although not an official organ of the US government nor an entity vested with governmental functions, could nevertheless be considered a *de facto* organ of the United States. A positive answer would have meant that the acts committed by the contras are attributable to the United States "en bloc."[816] To elucidate this question, the Court looked for various indicators such as the provision of funds prior to a subsequent subversive operation committed by the group, the nature of the entity that

[813] *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v Serbia and Montenegro) (*hereinafter Bosnian genocide test*)*, (2007), Judgment, ICJ Rep 43, at 171, §407.

[814] ICJ, *Supra* note 196, at 11, §20.

[815] *Ibid.*

[816] Crawford, *Supra* note 795*,* at 148.

created the group, and the nature of the US's overall involvement in the activities of the *contras*. Essentially, the Court had to find whether "the relationship of the contras to the United States Government was so much one of dependence on the one side and control on the other that it would be right to equate the contras, for legal purposes, with an organ of the United States Government, or as acting on behalf of that government."[817]

Despite strong evidence that the contras were so dependent on the US that they could not have conducted some of their operations without US support, the Court was unable to confirm that the support provided by the US extended over at least the "great majority" of the contras' operations.[818] Basically, the idea was that the private entity should not have been able to pursue *any* activity without the financial and operational support of the instigator. Hence, the Court established that the contra force could not be equated to a *de facto* organ of the US. The crux of the matter was for the Court to ascertain that the US exercised a "degree of control in all fields as to justify treating the contras as acting on its behalf."[819] The fact that the group has been created by the state, or that the state retains some form of influence over it does not suffice to attribute its actions to the state. Indeed, "to equate persons or entities with State organs when they do not have that status under internal law must be exceptional, for it requires proof of a particularly great degree of State control."[820]

The cumulative criteria as applied by the Court related to factual and temporal aspects, with the required relationship of complete dependence materializing only when the state exercises a full and uninterrupted control over all the activities of the private entity. This is the "strict control test."[821]

Gathering the required proofs to establish such a relation of dependence and control necessitates expansive investigations taking place at least over all the period when the alleged violations occurred. For that reason, the strict control test would rarely prove conclusive in practice. It is common sense that a state will not take the risk to exercise high amount of control over private entities when they rely upon them to engage in reprehensible behaviour. Overall, the strict control test may be of little use when it comes to assess attribution of privately-conducted disinformation campaigns.

---

[817] ICJ, *Supra* note 196; ICJ, *Supra* note 813, at 52, §109 and followings.

[818] "(A)dequate direct proof that all or the great majority of contra activities during that period received this support has not been, and indeed probably could not be, advanced in every respect." ICJ, *Supra* note 196, at 53, §111.

[819] *Ibid.*

[820] ICJ*, Supra* note 813, at 166, §393; ILC, *Supra* note 81, at 47, Article 8.

[821] Delerue, *Supra* note 81, at 111; Stefan Talmon, "The Responsibility of Outside Powers for Acts of Secessionist Entities" (2009) 58(3) *The International and Comparative Law Quarterly* 493.

3.3.2 Responsibility of a state for specific acts perpetrated by private entities and individuals, or "effective control test"

In its Nicaragua case, the ICJ tried to assess whether, even if the *contras* were not to be considered a *de facto* organ of the US, some specific acts committed by the group could still be attributed to the US. In doing so, the Court laid down the "effective control test." This test is based on the rationale that, whereas the relationship between the private actors and the accused state does not amount to complete dependence of the first upon the latter, a state could still be held accountable for the violations committed by non-state actors if it contributed in a material manner to the alleged violations.

Unlike the *de facto* agency test, the effective control test does not focus on the overall nature of the relationship between a state and non-state actors. Instead, the effective control test serves to establish the existence of a causal nexus between the commission of specific unlawful acts and the upstream issuance of instructions by the state. Hence, a state that did not act by itself can still be held responsible for the unlawful conduct of private entities if it contributed directly to the conduct by *directing, controlling, or providing instructions* to the private actors.[822]

It bears noting that there might be some overlap between the agency test and the effective control test. In fact, there seems to be quite a few instances in practice, where the satisfaction of the effective control test requirements does not reveal, by the same token, a relation of complete dependence of the private entity under scrutiny on the accused state that concomitantly satisfy the strict test. This observation should however be qualified. At least on paper, dependency on the basis of the blind obedience to orders does differ from a situation of complete dependence that also entails the granting of subsidies by the state with a view to finance and organize the operations in their entirety. A relationship of complete dependence might include the issuance of orders, yet instructions and a level of direction exercised by a state do not suffice with the *de facto* agency test.

Both the ICJ and the International Criminal Tribunal for the former Yugoslavia adds that mere financing and equipping of the agents does not suffice in proving that all acts committed by the private entity are to be attributed to the state.[823] In other words, the state must have "[…] directed or enforced the perpetration of the acts contrary to human rights and humanitarian law

[822] *See* ICJ, *Supra* note 196; ICJ, *Supra* note 813; Schmitt, (2017), *Supra* note 162, at 94, Rule 17.

[823] See ICJ, *Supra* note 196, at 64-65, §114-116, & at 128, §277; *The Prosecutor v. Dusko Tadić,* (1995), international criminal tribunal for Yugoslavia (hereinafter ICTY), (Appeals Chamber), IT-94-1-AR72, Decision, §131.

alleged by the applicant State."[824] The presence of some forms of incentive from the part of the accused state to get some private actors to act in a certain way, or the fact that the accused state provided substantial assistance to the private entity are not sufficient, as it must be proven that some instructions have been given specifically to commit the violations.[825]

Interestingly, in its 2007 *Bosnian Genocide* case, the ICJ further contended that the fact that the alleged violation is composed of separate specific acts iterated over a certain period in time does not mean that the effective control should be interpreted "[…] in relation to the whole body of operations carried out by the direct perpetrators of the genocide."[826] Adding that "[G]enocide will be considered as attributable to a State if and to the extent that the physical acts constitutive of genocide that have been committed by organs or persons other than the State's own agents were carried out, wholly or in part, on the instructions or directions of the State, or under its effective control."[827] What the Court appears to suggest here, is that an effective control test should be carried out for each operations that composed the alleged violation when the violation arises from an aggregate of acts.[828] However, in pointing to the ARSIWA, Crawford notes that the Court did not intend to require evidences of the issuance of specific orders to engage in each specific act. Instead, he argues that in the context of a violation comprised in a multitude of operations, "general instruction which leaves it open as a method of fulfilling the directive (or perhaps implies a preference for it) suffice[?]"[829]

Applying the above to the context of online disinformation activities conducted by non-state actors however raises questions about the prospects of success of the effective control test. Instructions may be issued through written contracts concluded between the state and the private actors, or issued directly in the field.[830] With respect to cyber disinformation, proofs of instructions or commands to conduct disinformation campaigns may remain extremely difficult to gather for the accuser. In fact, it is even difficult to visualize how such order could look like in practice. Orders to "exacerbate social discontent or division", to "deflect criticism", "destabilize a society" or "tarnish the reputation of someone" do not constitute instructions to commit unlawful acts *per se*. That is to say, that as far as disinformation campaigns are concerned, instructions would have to be quite specific to ever pass the effective control test. On this point, the ILC provides that it may sometimes be necessary to establish whether "the

[824] ICJ, *Supra* note *196*, at 52, §110.
[825] Crawford, *Supra* note 795, at 144.
[826] ICJ, *Supra note 813*, at 169, §401.
[827] *Ibid.,* at 170.
[828] Crawford, *Supra* note 795, at 145.
[829] *Ibid*.
[830] *Ibid*.

unlawful act was incidental to the mission or clearly went beyond it."[831] While this may *in fine* broaden the category of instructions susceptible of passing the effective control test, the burden of proof assumed by the accuser remains significant. The perceived stringency of the effective control test eventually led the ICTY and the ECtHR, to develop their own standards of attribution.[832] Even the ILC's interpretation seems to depart from the initial interpretation.

3.3.3 Subsequent developments regarding the effective control test

Setting the details of the International Tribunals and the European Court's own attribution tests aside, the ILC and the ICJ itself seem to have departed from the initial interpretation of the Nicaragua case. For the application of the effective control test, the notions of control, direction and enforcement were mingled in one single assessment in the *Nicaragua case*. Although they relied on the same notions of control, direction and enforcement, the ICJ in the *Bosnian case*, and the ILC in the ARSIWA subscribed to slightly different approaches. Yet, although a loosening in the stringency of the requirements of the effective control test originally put forth in the *Nicaragua case* was attempted both the ICJ and the ILC have remained vague regarding the responsibility of a state for specific acts perpetrated by private entities and individuals.

Article 8 of the ILC's ARSIWA deals with two scenarios. The first involves "[…] private persons acting on the instructions of the State in carrying out the wrongful conduct. The second deals with a more general situation where private persons act under the State's direction or control."[833] These provisions are interesting in that they seem to distinguish between the responsibility resulting from the *effective* control test, defined by the ICJ as the issuance of specific orders, and a responsibility resulting from more malleable concepts of "direction" and "control."[834] The commentaries under the article suggest that the ILC may have attempted an alternative approach to state responsibility for the conduct of private entities and individuals with involves treating the criteria of direction, instruction and control as disjunctive criteria.

Regarding the responsibility of a state for specific acts perpetrated by private entities and individuals, the second paragraph of commentary under Article 8 of the ARS submits "[T]he

[831] ILC, *Supra* note 81, article 8, §8.
[832] See notably *Prosecutor v. Tadić*, (1999), ICTY; Appeal Chamber; Judgment Case No IT-94-1; *Loizidou v.* ECHR, *Supra* note 753; Talmon, *Supra* note 821.
[833] ILC, *Supra* note 81, at 47, article 8, §1.
[834] Delerue, *Supra* note 70. at 139.

attribution to the State of conduct in fact authorized by it is widely accepted in international jurisprudence. In such cases it does not matter that the person or persons involved are private individuals nor whether their conduct involves 'governmental activity.' […] These include, for example, individuals or groups of private individuals who, though not specifically commissioned by the State and not forming part of its police or armed forces, are employed as auxiliaries or are sent as 'volunteers' to neighbouring countries, or who are instructed to carry out particular missions abroad."[835] This way stated, a state may assume responsibility for the acts perpetrated by private entities and individuals if it employed and send them in an operational field, or instructed them to carry out specific operations. This echoes the effective control test as it was originally expounded by the ICJ in its *Nicaragua case,* and does not seem to depart from it. The approaches of the ILC, the ICJ, and the ICTY are relatively similar when it comes to this specific aspect of attribution.[836]

With regard to the more contentious criteria of direction and control, the ILC first admits that "[M]ore complex issues arise in determining whether conduct was carried out 'under the direction or control' of a State. Such conduct will be attributable to the State only if it directed or controlled the specific operation and the conduct complained of was an integral part of that operation."[837] This *prima facie* seems to relate to the effective control as developed in the *Nicaragua case*. The ILC however underlines that the criteria of direction, control, and instruction are disjunctive.[838] Therefore, for the responsibility of a state to be engaged on the basis of *specific* acts perpetrated by private actors, the state must have enforced the act, or directed *or* controlled the operations that led to the violation. Thus, the ILC adds more granularity to the regime of responsibility for *specific* acts perpetrated by private entities and individuals without changing the *de facto organ* attribution regime. In doing so, the ILC appears to have loosened the attribution regime of private conduct to states. As an author noted, "[T]he direction and control criteria go beyond the ICJ's approach but are not as broad as the ICTY's overall control test."[839]

A clear loosening in the effective control test could have been achieved, if the perceived schism between the question of the enforcement of actions and the control criterion was formally sanctioned by the ICJ. In that regard, the approach developed by the ICJ in its 2007 *Bosnian* case is underwhelming. The Court, while dealing with the question of responsibility

[835] ILC, *Supra* note 81, at 48, article 8, §2.
[836] Delerue, *Supra* note 70. at 143.
[837] ILC, *Supra* note 81, at 48, article 8, §3.
[838] *Ibid.,* §5.
[839] Delerue, *Supra* note 70. at 140.

of the state for the killings perpetrated by the private entity, distinguished between private entities acting under the direction or the instruction of the state, and privately-conducted act committed under the effective control of that state; "[…] it has to be proved that they acted in accordance with that State's instructions or under its "effective control."[840] The ICJ appears to treat the "direction" and "instruction" criteria together. However, the ICJ does not provide much details regarding the meaning and relationship between the three criteria. Through articulating a reasoning similar to the previous *Nicaragua case*, the ICJ left the reader with no clue whatsoever, of what ought to be understood as an effective control that differs from direction and instruction.[841] This lack of indication as regards "control" without instruction is all the more perplexing that 8 years prior to the Bosnian case, the ICTY introduced an "overall control test" that precisely filled the gap of effective control without instructions.[842]

The ICJ remained hostile to the idea of applying the "overall control test" developed by the ICTY in its Tadic test, for the reasons that the overall control test served other goals than establishing state responsibility, and that adopting the reasoning of the ICTY would have excessively broadened state responsibility.[843] Yet, the "overall control test" would have proved a way better regime of attribution than the strict control or effective control tests in the disinformation context.[844] Indeed, a sponsoring-state relying on private entities may not bother giving specific instructions other than a general guidance. In general, the agency conducts its activities based on topics of interests chosen on a daily basis depending on the information landscape, the "hot" topics of the day. The role of the sponsoring state in that case may be limited to partial oversight, and the issuance of broad operational guidelines.

Applying the above to the relationship between a state and private companies such as the IRA or other PR companies, or between a state and the media it owns, and in the eventuality that it is impossible to prove that the accused state exercised complete control over the individual so as to consider the individuals completely dependent upon that state, or that it is impossible to prove that specific instructions were issued to the agents involved in the information operation to conduct disinformation campaigns against foreign audiences, the applicant would in theory still have the possibility to resort to the effective control criterion.

[840] ICJ, *Supra* note 813, at 169, §400.

[841] Antonio Cassesse, "The Nicaragua and Tadic Tests Revisited in Light of the ICJ Judgment on Genocide in Bosnia", (2007), 18(4) *The European Journal of International Law* 649; See also Schmitt, *Supra* note 211.

[842] Cassesse, *Ibid*., at 657. See also ICTY, *Supra* note 832, §120.

[843] ICJ, *Supra* note 813, at 169- 171.

[844] In order to attribute the acts of private actors to a state, the ICTY distinguished between the situations of private individuals carrying out the conduct on behalf of a state, for which the classic effective control test applies, and the situation of "organized and hierarchically structured groups", for which an overall control test that did not require issuance of specific orders suffices.

Halas, the ICJ does not provide clear indications as to the requirements to fulfil to prove effective control without having to prove direction or instruction.

Legal scholars have tried to make sense of effective control while avoiding referring to the ICTY's overall control test. Schmitt for example suggested interpreting the degree of effective control irrespective of the issuance of instructions in terms of the power exercised by an authority on the private group. Accordingly, "the relationship between the State and the group must be such that the State can, if it wishes, compel the non-State group to desist in the operation or alter the conduct thereof."[845] In the same vein the ILC states that effective control materializes when a state uses "[…] its ownership interest in or control of a corporation specifically in order to achieve a particular result."[846] In contrast, the mere facts that a private entity is created by a state, or that a state owns a company or a portion of it does not create effective control.[847]

In fine, effective control is first and foremost interpreted as a matter of degree. Perhaps the best conclusion to give with respect to the attribution of private acts to a state is a quote from the ARSIWA; "[I]n any event it is a matter for appreciation in each case whether particular conduct was or was not carried out under the control of a State, to such an extent that the conduct controlled should be attributed to it."[848]

It is however difficult to make sense of an effective control test in the disinformation context going beyond the most concrete criteria of instructions without drifting too far into wishful thinking. Not only the law is unsettled at this point, but observers do not always have access to all the necessary evidences to conduct legal assessments. As argued earlier, states may be reluctant to reveal all the evidences they have at their disposal. In the end, the question of the attribution of privately-conducted cyber-enabled acts falls into the grey areas of the law.[849]

Many authors have argued that the criteria applied by the ICJ were impracticable, and could not satisfactorily apply to new trends on the utilization of private actors by states.[850] Some also referred to the jurisprudences of the ICTY and the ECtHR to suggest that a lower threshold of attribution should be adopted by the ICJ to avoid further fragmenting international law.[851] Other authors have argued that a more important issue is the identification of the perpetrator in cyberspace and that the standards of attribution is a secondary problem that does not necessarily

[845] Schmitt, *Supra* note 211, at 24.
[846] ILC, *Supra* note 81, at 48, article 8, §6.
[847] *Ibid*.
[848] *Ibid*., §5.
[849] See in that sense Schmitt, *Supra* note 211, at 25.
[850] Cassesse, *Supra* note 841; Jensen & Watts, *Supra* note 766.
[851] Talmon, *Supra* note 821; Cassesse, *Supra* note 841.

need complete reappraisal.[852]

Be that as it may, and regardless of the recommended approach, most scholars seem to agree that legally attributing cyber-enabled conduct is often too much of a lengthy process to allow the injured state to take legal countermeasures against the offending state without risking breaching international law for their own belated response.[853] As a result, timely reaction is often precluded, leaving victim states with the only possibility to take lows-intensity measures of retorsions.[854] Finally, it is important to note that private disinformation campaigns are often unattributable to a state, not because the classic attribution regime is defective, but because no state is involved in the problematic act at all. That is to say, a state may simply condone private disinformation without interacting with the executor.

There exist alternatives to the classic attribution regime that may be used, both to bypass the lengthy investigation process required to attribute cyber-enabled act to a state, and to address the situation of a state harbouring and/or condoning disinformation activities on its soil. Lahmann for example suggests drawing inspiration from the Convention Relating to Intervention on the High Seas in Case of Oil Pollution Casualties to create a "cyber emergency regime."[855] Instead of relying on a lengthy and difficult attribution process, he argues that a distressed state should be legally entitled, under the cyber emergency regime, to take unilateral responses to avert a peril affecting its most important infrastructures. Unlike classic attribution, cyber emergency does not require the victim state to legally attribute the conduct to a state at the moment of the breach. Thus, if an intervention on an electronic asset located in the territory of another state is needed to abate the peril, the distressed state should first notify the origin state of a malicious use of its own cyber assets before intervening. Yet, realistically, states victim of foreign disinformation campaigns would rarely find themselves under a situation of distress necessitating such immediate defensive reaction.

Arguably, the regime of state liability for negligence may offer a more adequate legal framework for dealing with foreign disinformation campaigns in general. The regime applies to situations when states fail to exercise due diligence in preventing the nefarious extraterritorial consequences arising from transboundary acts initiated by private actors from their territories.

---

[852] Geiss & Lahmann, *Supra* note, 775.

[853] "An injured State may only take countermeasures against a State which is responsible for an internationally wrongful act in order to induce that State to comply with its obligations under Part Two. "[…] "Countermeasures are limited to the non-performance for the time being of international obligations of the State taking the measures towards the responsible State." ILC, *Supra* note 81, at 129, Article 49 "object and limits of countermeasures."

[854] Jensen & Watts, *Supra* note 766.

[855] See notably Lahmann, *Supra* note 90, at 201-257.

The content of state liability for negligence will be analysed in a last section.

4. International liability for negligence in the disinformation context

The previous sub-section has shown that attributing unlawful disinformation campaigns to a state is rarely a straight-forward process. The uncertainties affecting the effective control test and the apparent stringency of its requirements make the existing framework relatively ill-adapted to the attribution of unlawful cyber disinformation campaigns to a state. It has for instance been noted that states may be harbouring or condoning disinformation "industries" without interacting with the executors at all. In absence of positive actions undertaken by a state, when the state of origin does not play an *active* role into the illegitimate conduct of non-state entities, the application of the aforementioned attribution regime is precluded. The regime of state responsibility for negligence addresses this issue. For this reason, authors are increasingly turning to this regime for dealing with non-attributable cyber-enabled conduct.[856]

Hence, this last section looks at disinformation through the lenses of the different regimes of state responsibility for negligence. In doing so, it shows whether the responsibility of a state of emission could be triggered for its negligence in preventing a violation or mitigating a risk arising from private disinformation activities conducted from its soil.

Negligence, in international law, is the failure to exercise due diligence in the process of preventing or abating harm.[857] Due diligence is a fairly innovative legal proposition that is worth considering seriously, notably for its flexibility. It is the missing piece of state responsibility, in that it allows to take into account the errors and grave omissions of authorities in the process of preventing harm and international law violations for the purpose of establishing state responsibility.

---

[856] See notably Crootof, *Supra* note 214; Jensen & Watts, *Supra* note 766.

[857] See notably Timo Koivurova & Krittika Singh, "Due diligence" (2022) Max Planck Encyclopedias of International Law, Encyclopedia entries. It bears noting that some authors confuse due diligence with no-harm. These are however two different things. Due diligence defines a standard of conduct and attaches to a rule whilst no-harm is a rule on itself. No-harm incorporates due diligence. Therefore, failure to exercise due diligence may trigger state responsibility under the no-harm principle. Other legal frameworks may however incorporate due diligence obligations. See in that sense and notably Heike Krieger & Anne Peters, "Due diligence and Structural Change in the International Legal Order" in Heike Krieger, Anne Peters & Leonhard Kreuzer (eds.), *Due diligence in the international legal order*, (New York: Oxford University Press, 2020). The World Court for proclaimed in its Corfu Channel case that every state is under the obligation "not to allow knowingly its territory to be used for acts contrary to the rights of other States." See ICJ, *Supra* note 583, at 22. See also in that sense Hollis, van Benthem and Dias, *Supra* note 61, at 47; Brunnée, *Supra* note 858; Antonio Coco & Talita de Souza Dias, '"Cyber Due Diligence': A Patchwork of Protective Obligations in International Law" (2021) 32(3) *European Journal of International Law* 771.

Briefly, a state might incur liability for negligence in preventing or mitigating infringements on the rights of persons under its jurisdiction, on the rights of other states, and, to a certain extent, on the human rights of *persons* located outside its territory if it was entitled and in capacity to act upon the source of prejudice in the first place.

In analysing the regimes of responsibility for negligence, this section explores the extent of the due diligence duties of states when it comes to prevent nefarious transboundary consequences in the psychological warfare context.

As it is with respect to the no-harm principle that negligence is devoted the most attention, a first sub-section introduces the no-harm principle and explains its content (4.1). A second sub-section focusses on the applicability of the no-harm principle to online disinformation campaigns (4.2). Moving on from no-harm, the third section examines the other regimes of state liability for negligence with respect to the diligent prevention of infringements arising from private activities with transboundary effects. Notably, it is argued that states should assume due diligence duties to prevent and mitigate infringements on the customary human rights of persons situated outside their territory arising from the conduct of private persons located in their jurisdiction (4.3). A last sub-section covers the recent concept of corporate due diligence, and underlines its relevancy to the case at hand (4.4).

4.1 The no-harm principle, or state responsibility for negligence in preventing transboundary harm arising from the physical consequences of activities not prohibited by international law

The no-harm principle, or the duty to prevent and to redress transboundary harm is a remnant of the old maxim *sic utere tuo ut alienum non laedas*, which can be translated as "use your own property in such a way that you do not injure other people's."[858] It is sometimes alternatively referred to as the principle of "good neighbourliness."

Under the no-harm principle, states are liable for transboundary harm arising from activities placed under their control, irrespective of the question of intent, if they demonstrated negligence in the process of preventing, stopping, or redressing transboundary harm they were in the

[858] Jutta Brunnée, "Sic utere tuo ut alienum non laedas" (2022) Max Planck Encyclopedias of International Law, Encyclopedia entries.

capacity to prevent or stop.[859] Failure to redress or compensate the harm will trigger their international responsibility.

The contemporary duty to prevent transboundary harm finds its origins in environmental law. It is consecrated by the Trail Smelter arbitral tribunal, which determined that Canada had to pay an indemnity to the United States for the environmental damages caused by the toxic fumes emanating from one of its Trail Smelter in the nearby territories of the US. In its arbitral sentence, the tribunal famously declared that "no State has the right to use or permit the use of its territory in such a manner as to cause injury by fumes in or to the territory of another or the properties or persons therein, when the case is of serious consequence and the injury is established by clear and convincing evidence."[860]

The no-harm principle has since made several appearances in international law, primarily in international treaties pertaining to environmental protection.[861] For example, the no-harm principle is enshrined in the 1992 *Rio Declaration on Environment and Development*, which proclaims as a principle that states "have the responsibility to ensure that activities within their jurisdiction or control do not damage the environment outside their jurisdiction."[862] Similarly the 1972 Declaration on the Human Environment (The Stockholm Declaration*)* makes references to necessity to prevent environmental harm.[863] The 1972 *Convention on the International Liability for Damage Caused by Space Objects* (The Liability Convention) installs a liability regime in the context of space objects. The 1992 Convention on the Protection and Use of Transboundary Watercourses and International Lakes (Water Convention) in turn calls on the parties to take all appropriate measures to prevent, control and reduce any "significant adverse effect on the environment resulting from a change in the conditions of transboundary waters caused by a human activity […]."[864] The UNCLOS also provides in its

[859] "Under State responsibility, the breach of primary rules and obligations results in the application of secondary rules in State responsibility. On the other hand, under international liability, a breach will generate secondary obligations that must be fulfilled under the law of State responsibility." Sucharitkul, *Supra* note 859, at 833. See also Hollis, van Benthem and Dias, *Supra* note 61, at 48; Sompong Sucharitkul, "State Responsibility And International Liability Under International Law" (1996) 18 *Loyola of Los Angeles Int'l. & Comp. L.J*. 821, at 828. See also ILC, *International Liability for Injurious Consequences arising out of acts not prohibited by international law* (liability in case of loss from transboundary harm arising out of hazardous activities) (2004), A/CN.4/543 86, at 96, §21-22.

[860] *Trail smelter* (United States, Canada), (1938-1941), 3 UN *Report of International Arbitral awards* 1905, at 1965.

[861] For a more exhaustive list of international instruments where the no harm principle is to be found, see Brunnée, *Supra* note 858; ILC, *Supra* note 859.

[862] *Rio Declaration*, (1992), Report of the United Nations Conference on Environment and Development, A/CONF.151/26 (Vol. I).

[863] *Declaration on the Human Environment*, (1972), United Nations Conference on the Human Environment, 2994/XXVII, 2995/UVII and 2996/XXII, at §3.

[864] *Convention on the Protection and Use of Transboundary Watercourses and International Lakes*, (1996), (as amended in 2013), at 6.

Article 110 a liability clause with regard to the compensation of any loss and damages sustained during a boarding of foreign merchant ships based on unfounded suspicions. Finally, the work of the ILC with regard to the prevention of transboundary harm greatly contributed to the elaboration of the no-harm principle.[865] The ICJ for example relied on the ILC's *Draft articles on Prevention of Transboundary Harm from Hazardous Activities* in its Pulp Mills on the River Uruguay case.[866] Therefore, the no-harm principle has the status of customary law.

The no-harm principle applies to conduct not prohibited by international law. As such, it acts as a sort of stopgap to compensate the lengthy development of international law. Through linking state responsibility to the question of the failure to prevent, stop or redress transboundary harm arising from non-prohibited activities, the no-harm principle circumvents the difficulties inherent to the legal qualification of conduct not prohibited *per se* under international law, due to a lack of knowledge about the harmful character of an activity at the time of its initiation, for example. Therefore, no-harm applies to conduct, irrespective of future qualification as unlawful conduct, when the risk posed by said conduct becomes more clearly known.[867]

There are three saliant points with regard to the no-harm principle. The first relates to the absence of "classic" attribution process, the second relates to the definition of "harm" falling within the scope of the principle, the third relates to the way states should discharge their obligation of prevention with respect to no-harm.

The no-harm principle does not require that the hazardous activities are attributed to the state of origin. Instead, it suffices to prove that the state of origin has been negligent in the process of preventing and mitigating transboundary harm.[868] In other words, it is negligence that is attributed to a state, not the harmful activity.[869] The first prerequisite to establishing state liability is then to trace back the harm to its source, in the territory of said state. The second perquisite is that the state of origin must have been entitled and in capacity to act upon the risk of transboundary harm.[870] This means that the discharge of due diligence duties should not be

[865] ILC, "Draft articles on Prevention of Transboundary Harm from Hazardous Activities" *(2001) 2 Part Two, Yearbook of the International Law Commission 148*.

[866] *Pulp Mills on the River Uruguay* (Argentina v. Uruguay), (2010), ICJ, Judgment. (Specifically, the Court emphasized the importance of conducting environmental impact assessments to mitigate hazardous consequences.)

[867] *Ibid*. *at 150, Article 1, §6.*

[868] Sucharitkul, *Supra* note 859, at 835.

[869] ILC, *Supra* note 865, at 150, Article 1, (8).

[870] ICJ, *Supra* note 813, at 182, §430: "Various parameters operate when assessing whether a State has duly discharged the obligation concerned. The first, which varies greatly from one State to another, is clearly the capacity to influence effectively the action of persons likely to commit, or already committing, genocide. This capacity itself depends, among other things, on the geographical distance of the State concerned from the scene of the events, and on the strength of the political links, as well as links of all other kinds, between the authorities of

done at the expensive of the sovereignties of other states, that the state had constructive knowledge of the risk of harm, and that it was in capacity to act upon the source of harm. These parameters will be developed later.

Regarding the notion of "harm," the no-harm principle was primarily associated with the mitigation of transboundary damages done to the environment.[871] Authors have occasionally pointed to the lack of international instruments containing references to no-harm outside the context of environmental harm to underline the absence of state practice suggesting a broader no-harm principle applicable outside the context of environmental harm.[872] At the same time, the legal literature is increasingly considering no-harm when risks of transboundary harm to persons and material destruction are involved.

There are no reasons that no-harm should not extend to other type of transboundary harm than environmental harm. The ILC itself has clarified in its *Draft articles on Prevention of Transboundary Harm from Hazardous Activities* that "harm" "means harm caused to persons, property or the environment;"[873] At least, the conduct or activities must present a "risk of significant transboundary harm" resulting from its physical consequences and leading to "real detrimental effect on matters such as, for example, human health, industry, property, environment or agriculture in other States" to fall under the scope of the rule.[874]

A couple of authors have argued that "transboundary harm" also encompasses non-physical harm arising from activities that are not themselves prohibited by international law such as reputational damage or moral and psychological harm arising, for instance, from libel or defamation.[875] It may however be difficult with such an interpretation to avoid overreach. The ILC itself rejected the view that non-physical harm should fall into the scope of no-harm.[876] Article 1 provides that harm must arise from the "physical consequences" of a conduct.[877]

---

that State and the main actors in the events. The State's capacity to influence must also be assessed by legal criteria, since it is clear that every State may only act within the limits permitted by international law […]" See also generally Tony Cabus, *Due Diligence and the High Seas*, (London: Routdlege, 2021), at 16 & 46. These requirements are further developed below.

[871] Brunnée, *Supra* note 858.

[872] See Brunnée, *Supra* note 858, about the statement of principle in the Corfu Channel case; "It is less clear, however, whether these broadly applicable notions can be traced back to the *sic utere tuo* maxim or simply to the mutual limitation of sovereign rights."

[873] ILC, *Supra* note 865, at 152, Article 2, (b).

[874] *Ibid*., at 151, Article 1, (13). And at 152, Article 2, (4).

[875] ILC, *Supra* note 865, at 154, Article 3, (11)

[876] "[…] the significant transboundary harm must have been caused by the "physical consequences" of such activities. It was agreed by the Commission that in order to bring this topic within a manageable scope, it should exclude transboundary harm which may be caused by State policies in monetary, socio-economic or similar fields. The Commission feels that the most effective way of limiting the scope of these articles is by requiring that these activities should have transboundary physical consequences which, in turn, result in significant harm." ILC, *Supra* note 865, at 151, Article 1, (16).

[877] *Ibid*., at 151, Article 1, (16).

Whether the transboundary activity is physical or immaterial is inconsequential. What counts is that it can have physical consequences.

It is important to keep in mind that, while the no-harm principle is triggered once transboundary harm occurs, the due diligence obligations contained in the no-harm principle do not have temporal scopes. A state should be diligent from the moment it grants permission for a potentially harmful activity, and even before, as, the state should not grant its authorization to highly hazardous activities anyway.

The state of origin of the harm only incurs liability for said harm if it did not exercise due diligence in the process of preventing and/or mitigating the harm. To discharge its obligation of prevention with respect to no-harm, a state of origin must exercise due diligence, understood as the obligation to take "all appropriate measures" in the process of preventing, minimizing or abating transboundary harm.[878] On this point, the ILC submits "[T]he standard of due diligence against which the conduct of the State of origin should be examined is that which is generally considered to be appropriate and proportional to the degree of risk of transboundary harm in the particular instance."[879] Therefore, the level due diligence expected from the state of origin with respect to transboundary harm prevention will be modulated according to various factors, including the nature of the harmful activity, the amount of available knowledge surrounding the potential for harm of said activity, and the technical and scientific means available to the state of origin before and during the execution of the harmful activity.

When transboundary harm does occur, the state of origin must show that it acted diligently to prevent, minimize and abate the risk of transboundary harm. If it is found that the state of origin did not act diligently, then it must abide to its obligation of liability, and compensate for the damages. Failure to compensate the damages will engage its international responsibility.[880]

4.2 Limitations to the enforcement of the no-harm principle with respect to disinformation activities

When transboundary disinformation campaigns are unattributable to a state, a likely scenario is that the states harbouring disinformation industries turn a blind eye on them. In some cases, it is by lack of means that territorial states do not act. In other cases, states demonstrate

[878] *Ibid*., at 153, Article 3.
[879] ILC, *Supra* note 865, at 154, Article 3, (11).
[880] Timo Koivurova & Krittika Singh, *Supra* note 857.

blatant negligence, or knowingly allow disinformation campaigns to be disseminated from their territories. The no-harm principle might have an application in both cases.

Disinformation activities do not have to be prohibited by international law to fall under the scope of the no-harm principle. Coincidently, as matter of *lex lata*, online disinformation campaigns are not unlawful *per se*. However, the activities must come with a risk of causing "significant transboundary harm" resulting from their physical consequences to fall under the scope of the no-harm principle.

The spread of disinformation to a foreign audience would entail a clear risk of significant transboundary physical harm when said spread is of the nature to put the recipients in situations of imminent danger for their health or lives. This could concern communications campaigns promoting violence or health disinformation, for which there exists a foreseeable risk of transboundary harm. For instance, the state that remains passive in face of disinformation campaigns encouraging foreign populations to take poisonous solutions to "cure" a dangerous disease instead of taking a vaccine, or coordinated incitation to discrimination and hatred targeting foreign individuals, could be considered to have failed in its duties of due diligence in preventing and mitigating transboundary harm, provided that it was entitled and in capacity to act upon the risk of harm in the first place.[881]

But there is a problem with the application of the no-harm principle to the disinformation context. Although the negligence of the state of emission is sometimes clear in the disinformation context, for instance, when the state of origin is reluctant to collaborate with the authorities of other states to investigate egregious private disinformation cases, the no-harm principle may simply never be triggered. The issue relates to the requirements in terms of chain of causation. [882]

It is almost impossible to establish a clear causal link between the physical transboundary harm in question and the communications. That is to say, that while disinformation may effectively have contributed to serious internal disturbances including episodes of physical violences, deaths, or an increase in the number of hospitalizations, chances are that other factors have also contributed to the transboundary harm. As a result, high levels of certainty regarding the real impact of a disinformation event are rarely obtainable. In Hollis's words; "[a]bsent some futuristic mind control technique, it will be difficult to demonstrate that an influence

[881] See in that sense Hollis, van Benthem and Dias, *Supra* note 61, at 50; Schmitt & Milanovic, *Supra* note 326.

[882] See an in-depth treatment of the question at Lahmann, *Supra* note 152.

operation "caused" any particular outcome even if it is clear that the IO had some visible role in the processes that led up to that outcome."[883]

It is unclear whether relaxed models of causation should be admitted under the no-harm regime.[884] Lowering the causality requirements risks opening the no-harm principle to abuses. Also, there is somewhat of an introspective inquiry to be expected from the part of states. Societal, economic or political instabilities are not necessarily the product of foreign-led disinformation. At the same time, it would not make much sense to require *absolute* causality with respect to dangerous disinformation efforts such as health disinformation or disinformation during natural disasters, as in these situations the lives of persons are at stake. On this point, perhaps moving from clear causality to "imminence" of harm based notably on the scale and the context of the disinformation could resolve part of the issues associated with the speech causality problem, but even that is unsure.

Of course, this does not mean that states are exempted from exercising due diligence in preventing and mitigating the risk of transboundary harm arising from the physical consequences of the disinformation activities conducted by their nationals. For due diligence obligation to arise under the no-harm principle, the mere possibility that disinformation campaigns cause physical harm or destruction suffices. Although uncertain and hardly quantifiable, it is at least plausible that health disinformation, disinformation during natural disaster or even climate disinformation causes physical transboundary harm abroad. States should thus exercise particular care not to allow disinformation networks promoting the aforementioned disinformation types to strive in their information space, especially, in the social media platforms they own. But without a clear insight into the chain of causation leading to the occurrence of transboundary harm, the no-harm principle may not be enforceable to the disinformation context. On itself, the no-harm principle may not provide enough incentive for states to discharge their due diligence duties with respect to certain disinformation types, even those that could contribute to transboundary physical harm.

There are other avenues for engaging the responsibility of a state for its negligence in the context of psychological warfare. The no-harm principle applied to conduct not prohibited by international law and for which there exist a risk of significant transboundary harm arising from the physical consequences of an activity. Due diligence however extends beyond the no-harm context and attaches to negative obligations safeguarding legally protected interests such as the territorial integrity and the political independence of states as well as the rights and freedoms

[883] Hollis, *Supra* note 334; Schmitt, *Supra* note 211; Zhao, *Supra* note 117.
[884] *Ibid*.

of individuals. The next sub-sections explore the question of state responsibility for the failure to exercise due diligence in the prevention of infringements on the rights of other states and on the human rights of individuals.

4.3 State liability for negligence in preventing and mitigating infringement on the rights of other states and of persons located outside the territory of the state of emission

The obligation to act diligently in preventing nefarious transboundary consequences is not confined to the no-harm context.[885] In fact, the legal proposition that states are liable for failure to exercise due diligence in preventing infringements on the rights of other states arising from activities conducted by private actors in their respective territories is a well-established principle of customary law. Unlike the no-harm principle, this regime encompasses activities that are not producing any physical consequences (4.3.1). In the same vein, state responsibility for negligence also applies in the context of the prevention and mitigation of infringements on human rights (4.3.2).

Before getting into the hearth of the subject, it is important to clarify the distinction made between the rights of states and the rights individuals. As argued earlier, the "rights of states" are the rights owned by the collective in virtue of the principle of sovereignty. These include, the right to territorial integrity, and the right to political independence, which are safeguarded by a variety of negative obligations such as the non-intervention principle. Human rights, in contrast, are rights owned by each private individuals in virtue of their human condition.

Sometimes, a violation of human rights law can give rise to a violation of the rights of the state of nationality of the victim. Take the example of an execution conducted by a foreign state into the territory of another. Conversely, a violation of the right of a state may also involve infringements on the rights of individuals. Take the example of a foreign intervention aimed at provoking regime change against the collective will, or an armed attack.[886]

However, in many instances, the point of a violation of the rights of individuals can be made without it being possible for the state of jurisdiction to make the claim of a violation of its own rights. This applies, obviously, to situations involving a state's own human rights violations committed on the territory it administrates. This also applies to situations involving human

[885] See, for the sake of illustration, ICJ, *Supra* note 813, at 181, §429.
[886] See notably Damrosch, *Supra* note 431; Tsagourias, *Supra* note 409.

rights violations that do not meet the eligibility requirements set out by the negative rules safeguarding the "rights of states" interpreted *stricto-sensu*.

4.3.1 Due diligence and the prevention and abatement of transboundary disinformation activities infringing on the rights of other states

As a general principle, states should abstain from infringing the political integrity or the political independence of other states while conducting transboundary PSYOPS.[887] This applies to the acts of state organs[888], persons or entities empowered by the law of the state to exercise elements of governmental authority[889], or entities placed at the disposal of the state to exercise elements of governmental authority[890], as their misconduct are directly attributable to the state. As shown earlier, this also applies to private entities under the effective control of a state.

Alongside this general principle to abstain from violating the rights of other states, states have positive duties to prevent such violations from occurring. In its 1949 Corfu Channel case, the ICJ submitted that every state is under the obligation "not to allow knowingly its territory to be used for acts contrary to the rights of other States."[891] Specifically, the Court had to determine, among other things, whether Albania was responsible for the damages inflicted by water mines explosions to English warships and military personnel then effectuating a safe passage in an international strait in Albanian waters. After noting that Albania knew about the existence of the minefield but still did not notify the UK of their existence as they approached, the Court held Albania responsible for the damages; "nothing was attempted by the Albanian authorities to prevent the disaster. These grave omissions involve the international responsibility of Albania."[892] Since then, the principle proclaimed in the Corfu Channel case has acquired the status of customary law.

The principle of the Corfu channel is referenced as the "due diligence principle" by the authors of the Tallinn Manual 2.0. Article 6 of the Manual provides; "[A] State must exercise due diligence in not allowing its territory, or territory or cyber infrastructure under its governmental control, to be used for cyber operations that affect the rights, and produce serious

[887] Hollis, van Benthem and Dias, *Supra* note 61, at 34.
[888] ILC, *Supra* note 81, at 40, Article 4.
[889] ILC, *Supra* note 81, at 42, Article 5.
[890] ILC, *Supra* note 81, at 43, Article 6.
[891] ICJ, *Supra* note 583, at 22.
[892] *Ibid*., at 23.

adverse consequences for, other States."[893] Alternatively, the Corfu Channel case has been presented as an application of the no-harm principle.[894] Some authors have however remarked that the principle proclaimed in the Corfu Channel case, the "Corfu principle"[895] is slightly different than no-harm. Indeed, the "Corfu principle" also appears to be applicable to activities producing non-physical adverse effects, as opposed to the no-harm principle, which applies to non-prohibited activities producing severe adverse physical effects in foreign territories.[896] The distinction lies in the use of the expression "rights of states", which seems to encompass conducts undermining the political integrity of a state without producing physical consequences on its territory.[897]

With that being said, the distinction between the Corfu-principle and the no-harm principle is in many regards superfluous. There are many scenarios where the two "principles" will be applicable as one. Arguably, transboundary harm affecting the territory, infrastructures, and persons located abroad could also be interpreted as affecting the right to territorial integrity the recipient state. Thus, the no-harm principle and the Corfu principle are inherently the same legal construct.

Based on the Corfu principle, states are bound by a general duty to prevent, diligently, foreseeable nefarious consequences on the rights of others states arising from private disinformation conducted from their territory or under their jurisdiction. [898] In fact, the principle even predates the Corfu case.

Various bilateral and multilateral treaties regulating private transboundary communications and private subversive activities in general seem to support the view that states have due diligence obligations to prevent and mitigate infringements arising from private hostile propaganda emanating from within their territories regardless of whether they themselves participate in the misdeeds.[899] For instance, the Broadcasting Convention sets an obligation for the territorial state to prevent and cease harmful communications from emanating from its own territory. Article 3 provides for instance that "The High Contracting Parties mutually undertake to prohibit and, if occasion arises, to stop without delay within their respective territories any

[893] Schmitt (2017), *Supra* note 162, Rule 6, at 30.
[894] See for instance Crootof, *Supra* note 214.
[895] Expression notably employed in Hollis, van Benthem and Dias, *Supra* note 61.
[896] See in that sense Hollis, van Benthem and Dias, *Supra* note 61; Coco & Dias, *Supra* note 857, at 786-787.
[897] Coco & Dias, *Supra* note 857, at 787.
[898] Hollis, van Benthem and Dias, *Supra* note 61, at 47.
[899] Wright, *Supra* note 338, at 84.

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."[900]

Yet, upon examination of the doctrine, it appears that the legal proposition that states are bound by a duty of due diligence to prevent private individuals and organizations from spreading hostile propaganda from their territory is subject to contention.[901] Indeed, the idea that states are responsible for unacceptable communication activities of their nationals is often presented as the exception rather than the rule. With regard to the type of communications concerned, scholars were generally more inclined to admit the existence of such a duty for incitation to revolt or propaganda for war made by private actors, or defamation of diplomatic agents, than for private libellous and defamatory speeches.[902] The exception only to apply to the most brazen forms of private communication activities.[903] Thus, it seems that, in absence of treaty provisions establishing a responsibility for all private communications deemed hostile by the recipient state, states would not incur responsibility for the transboundary communications of private persons under their jurisdiction unless the conduct in question are blatantly violative of the sovereignty of a third state.

The regime of state responsibility for negligence in preventing and abating hostile transboundary communication campaigns is fully applicable to the disinformation context. Once private communications campaigns incorporate overt calls for revolts or overt threats or raise to the level of prohibited interventions, then the state of emission is under an obligation to abate the resulting infringement on the rights of the recipient state. The state of emission should also exercise continuous due diligence to ensure that its territory does not harbour disinformation industries

To incur responsibility for the acts of private actors operating from within its soil or under its jurisdiction, the state of emission must be entitled and in capacity to act to prevent and abate the infringement.[904] This means that the state of origin must not discharge its due diligence duties in ways that contravene the sovereignties of other states. The state of emission must also exercise effective overall control of the authors of the disinformation campaigns.

In the situation of an interstate dispute, a state owes respect to the rights of other states. A state is also presumed to be in full control of the territory it administrates and thus, in the

[900] Broadcasting Convention, *Supra* note 280.
[901] For an exhaustive analyse of state practice in that sense, see Dyke, *Supra* note 5. See also Whitton & Larson, *Supra* note 60, at 133-166.
[902] *Ibid*., at 71.
[903] Whitton & Larson, *Supra* note 60, at 166.
[904] See notably Cabus, *Supra* note 870.

position to "influence effectively the action of persons likely to commit, or already committing […]"[905] illicit activities.

As for the capacity requirement, the state of origin must have constructive knowledge of the risk of international law violations arising from the acts of private individuals operating on its soil. Constructive knowledge means that the state of origin is presumed to have knowledge of a risk of infringement when it is clear, given the circumstance, that it should have been aware of the risk.

Due diligence should not be understood as imposing excessive obligations to states. This is to say, that states do not have to screen their whole online information space in search of problematic communications to discharge their due diligence obligations.[906] However, the knowledge requirement may be considered fulfilled once the state of emission is made aware, by the imperilled state, of potentially unlawful activities conducted by private actors under its jurisdiction. Knowledge should also be assumed once the disinformation activities in question are massive.

More importantly, due diligence should not be abused to unduly restrict the free flow of information and freedom of speech. Preventive and excessive speech restrictions allegedly taken to prevent infringements on the rights of other states would not meet the expected diligence level. They could even constitute violations of free speech. In the same vein, the recipient states should not abuse its rights to demand undue restrictions on the freedom of speech of individuals located abroad.

As a matter of the prevention of transboundary infringements on the rights of other states, all states should have adequate speech regulation in place, and exercise due diligence while enforcing them.

With regard to less pernicious communications, however, the "Corfu principle" is of limited use. The bulk of disinformation campaigns would simply not exceed the frameworks set by the non-intervention and non-use of force principles. When health disinformation endangers the health and lives of individuals without effectively causing them to lose their lives, or when electoral disinformation campaigns *corrupt* a vote without overtly coercing large numbers of voters, observers should consider human rights law instead of state-centric rules.[907]

The question thus arises, as to whether a principle similar to this of the Corfu Channel case exists with respect to the protection of human rights? The answer is resoundingly positive. A

[905] ICJ, *Supra* note 813.
[906] See notably Jensen & Watts, *Supra* note 766.
[907] See *Supra*, Chapter IV.

brief overview of human rights instruments, the jurisprudence of human rights bodies, and the legal literature clearly confirms that states assume positive due diligence duties with respect to the prevention and cessation of human rights abuses occurring within their jurisdictions.[908]

4.3.2 Due diligence and the prevention of human rights impairments arising from transboundary disinformation activities

In human rights law, duties to prevent human rights violation are attached to most human rights.[909] The importance of prevention in the human rights context has been stressed multiple times.[910] The prevention of human rights violations translates in practice in a set of positive duties of conduct (due diligence) that states should follow to meet their human rights commitments.[911] Same as previously, a state incurs liability for human rights violation if it has failed to discharge its obligation of due diligence in preventing said violations, and will engage its international responsibility if it fails to redress the harm.[912]

As argued in the previous chapter, certain disinformation campaigns could constitute human rights violations on their own, for their scale and content, irrespective of the effective materialization of physical harm. This is the case for the right to life, the right to internal self-determination, the right to FT&RHO and the principle of non-discrimination. In that sense, the human-centric framework palliates the problem of causality encountered in the application of the no-harm principle.

It is a fairly consensual legal proposition that states incur liability for the failure to exercise diligence in the prevention of human rights violations occurring in the territory they administrate.[913] Thus, a first section indicates how a recipient state could fulfil its human rights due diligence duties *vis à vis* the protection of the human rights of its own population in the

[908] See, among other Timo Koivurova & Krittika Singh, *Supra* note 857; Bjornstjern Baade, "Due diligence and the Duty to Protect Human Rights" in in Heike Krieger, Anne Peters & Leonhard Kreuzer (eds.), *Due diligence in the international legal order*, (New York: Oxford University Press, 2020), See for an exhaustive list of human rights instrument explicitly reference to an obligation to prevention human rights violation at *The Role of Prevention in the Promotion and Protection of Human Rights*, (2015) A/HRC/30/20, at 4, §5.

[909] Besson, *Supra* note 741.

[910] HRC, *Ibid*; *The contribution of the Human Rights Council to the prevention of human rights violations*, (2018), A/HRC/RES/38/18; *Report of the Office of the United Nations High Commissioner for Human Rights*, (2014), A/HRC/28/30; *The role of prevention in the promotion and protection of human rights*, (2013) A/HRC/RES/24/16; *The contribution of the Human Rights Council to the prevention of human rights violations*, (2020), A/HRC/RES/45/31. *Human Rights Council*, (2006), A/RES/60/251, §f.

[911] Besson, *Supra* note 741, at 4.

[912] See Baade, *Supra* note 908, at 92; HRC, *The Nature of the General Legal Obligation Imposed on States Parties to the Covenant,* (2004) General Comment No. 31.

[913] Besson, *Supra* note 743.

disinformation context (4.3.2.1). The second sub-section inquiries about the possibility to hold a state liable for its negligence in the context of the prevention of human rights violations affecting persons located outside the effective overall control of states but arising from activities initiated in their territories (4.3.2.2). A last sub-section looks at corporate due diligence and the responsibility of social media companies and online news media in making the online information space less prone to manipulation (4.3.2.3).

4.3.2.1 Due diligence and the prevention and abatement by states of transboundary conduct infringing on the rights of persons located within their territory or jurisdiction

This section looks at human right due diligence in the disinformation context from the perspective of the recipient state.

In principle, a state has jurisdiction over the persons, property and activities located in its territory.[914] When taking part to human rights treaties, states accept to be bound by several obligations of due diligence when it comes to ensure and protect the human rights of persons under their jurisdiction. This includes persons situated in their territory, and persons over which states exercise effective overall control located abroad.

By definition, the territorial state is entitled to act to prevent the damages arising from conduct affecting its own territory or population, or the population over which it exercises effective overall control, regardless of the location of the source of the damage.[915] Similarly, the territorial state is presumed to be "in capacity" to *prevent* human rights violation arising from domestic and foreign activities generating freedom-destroying consequences in the territory they administrate. They will however only be in capacity to *act* upon the source of the harm when he exercises jurisdiction over the source of activity. On this point, the territorial state should be presumed to be in full capacity to act upon the source of harm when it owns the domestic news providers and digital companies that have enabled the prejudice. Indeed, the territorial state assumes obligations of prevention proportional to the levels of control they exercise over domestic news providers and digital companies.[916]

State liability has a particular use when it comes to mitigate harm arising from PSYOPS that put the lives of its nationals in a situation of immediate danger, such as disinformation

[914] Oxman, *Supra* note 744.
[915] Besson, *Supra* note 742; See also *case of Oneryildiz v Turkey* (2004) EtHR, judgement.
[916] Whitton & Larson, *Supra* note 60, at 164.

campaigns recommending false cures to serious diseases (health disinformation), hate speeches (divisive disinformation), or disinformation campaigns aimed at meddling with the collective will of the state in times of elections (electoral disinformation).[917]

As already recalled by a number of scholars, due diligence should not impose an excessive burden upon states. [918] With regard to the knowledge requirement, there is no obligation for the state of origin to continuously monitor its own online information space in search of problematic communication campaigns in order to discharge its due diligence duties *vis à vis* person under its jurisdiction. Not only would this require extraordinary efforts, but such measure would obviously clash with other human rights such as the right to privacy. However, knowledge of a risk would be presumed if the state of emission has directly made the recipient state aware of a serious and foreseeable risk arising from the misconduct of its private actors, if social media platforms themselves inform the recipient state of suspicious coordinated activities online targeting its nationals, or eventually, once the campaign reached such a proportion that the recipient state could not have ignored the risk.

With respect to the means available to a state to prevent or abate human rights violations arising from private communication activities, the general principle is that the more hazardous an activity, the higher the expected standard of care in designing policies and degree of vigour in enforcing them.[919] Censorship measures including content removal should be reserved for the most problematic transboundary communications, those that pose clear and imminent danger to the recipient. The recipient state should also diligently prosecute the offenders when the infringement is substantial. Alternatively, the recipient state may issue corrections, or diligently conduct official communication campaigns in its territory to "make the truth louder" when the risk of harmful consequences arising from private disinformation campaigns is foreseeable, as is the case of disinformation during election or *referenda*. Due diligence, in that case, would impose that the communication efforts of a state does not, in turn, undermine the rights of the domestic audiences, particularly their FT&RHO.

The recipient state should remain cautious while deploying measures of human rights-harm prevention, especially in the disinformation context. Indeed, due diligence imposes, by definition, that the risk of harm is managed appropriately, in a reasonable manner. As Lahmann recalls, "a positive legal duty to act cannot extend further than the state's capacity to act which

[917] Hollis, van Benthem and Dias, *Supra* note 61, at 50; Schmitt & Milanovic, *Supra* note 326.
[918] See Schmitt (2017), *Supra* note 162, Rule 16, at 40-43. See also in that sense Jensen & Watts, *Supra* note 766, at 1566.
[919] ILC, *Supra* note 865, at 154, Article 3, (11).

comprises not only factual capabilities but also legal constraints, such as applicable human rights of all individuals within its jurisdiction."[920] On this point, the framework developed by Sunstein in order to help democracies valuing free speech designing regulations on falsehoods is helpful. Taking into account the many variables involved in disinformation, including the state of mind of the speaker, the magnitude of harm, the likelihood of harm and the timing of harm, Sunstein submits that, as a principle, "[F]alse statements are constitutionally protected unless the government can show that they threaten to cause harm that cannot be avoided through a more speech-protective route."[921]

Excessive speeches restrictions (i.e., censorship measures inflicted to non-harmful speeches or speeches threatening to cause harm that can be managed and prevented *through* other measures than censorship) could also constitute a failure to exercise due diligence in preventing transboundary harm arising from domestic or foreign disinformation, in addition to represent potential violations on the freedom of speech of foreign and domestic individuals.[922]

When it comes to taking protective measures for the human rights of their users, multinational companies must cooperate with the recipient state in the fulfilment of the latter's due diligence duties.[923]

Due to jurisdiction issues, the recipient state may not be able to enforce the law against the instigator. This applies to disinformation campaigns conducted from abroad. However, harm prevention in the disinformation context relies mainly on the prevention aspect of due diligence. The incapacity to *act* upon the source of harm does not exonerate the recipient state from its due diligence obligations of harm prevention.

As a matter of prevention, the recipient state must have adequate speech regulations in place. In addition, the recipient state must ensure the authenticity of democratic participation in times of national elections, *referenda,* and all other instances when the population exercises its right to participate in the political affairs of the state in a direct manner. To that effect, adequate protections must be put in place so as to protect against foreign-led attempts to corrupt the expression of the collective will and by extension, the rights of internal self-determination of its population and eventually the right to political participation. Measures of protection could

[920] Lahmann, *Supra* note 152, at 420; See also in that sense Bjornstjern Baade, "Due diligence and the Duty to Protect Human Rights" in in Heike Krieger, Anne Peters & Leonhard Kreuzer (eds.), *Due diligence in the international legal order*, (New York: Oxford University Press, 2020), at 92.

[921] Sunstein, *Supra* note 63, at 4. See also Uyiosa Omoregie, "The 'Harm Principle' and Information Disorder Online" (2021), 1999 *Academia Letters*.

[922] Lahmann, *Supra* note 152, at 420.

[923] See notably Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (DSA).

take the form of national agencies entrusted with preserving the transparency of the information accessible to the voters, as is for instance the case in France, where The Vigilance and Protection Service against Foreign Digital Interference (VIGINUM) and the audio-visual regulator (ARCOM) are ensuring the preservation of the authenticity of democratic processes.

Additionally, an efficient way for recipient states to prevent and mitigate the nefarious impact of foreign led transboundary PSYOPS on their populations resides on awareness and information literacy. Individuals should be taught how to form independent thoughts by their states of residence. They should also be taught early in their life how to interact with internet.[924] Plurality of information and contradictory debates play a crucial role in the mitigation of the risk arising from certain information misuses. States should therefore ensure that sufficient guaranties are in place to maintain a "marketplace of ideas" in their respective online information space.

4.3.2.2 The problematic question of the prevention and abatement by states, of human rights violations affecting persons located outside their effective overall control but arising from activities initiated in their territory

States rarely engage in disinformation campaigns against foreign audiences in a direct manner. More often than not, PSYOPS are outsourced to private entities. Yet, the previous section has shown that the criteria for the indirect attribution of the acts of private actors to states were rarely met in practice. There is also the possibility that the states of emission are not involved in privately-led disinformation campaigns, but finds themselves unwilling to act to address a risk. In that case, the states of emission could be in violation of their due diligence duties of human rights violations prevention. The question thus arises, as to whether states are legally bound to ensure, diligently, that their respective territories are not used for acts contrary to the human rights of foreign individuals?

First of all, and as shown in the previous chapter, state agents and organs are responsible for human rights violations they directly committed. Likewise, human rights treaties give rise to an obligation for states parties to prevent and abate infringement on the rights of persons under their jurisdiction.

[924] See notably Bronner et al., *Supra* note 54.

Although it would make no sense to engage the responsibility of a state for its negligence in preventing an unlawful conduct it itself committed, it is worth noting that, as a principle, states should demonstrate professionalism when communicating with foreign audiences. This entails, notably, that incorrect statements made by its own organs should not be left unaddressed by a state, especially if they touch upon sensitive topics. As for independent private actors operating in the territory of the state of emission, however, the situation is not so clear.

The same way as with the no-harm and the Corfu principle, the prerequisites to engaging the responsibility of a state for its negligence in preventing or mitigating human rights violations arising from the extraterritorial consequences of the activities of persons situated under jurisdiction, are that the territorial state must be entitled and in *capacity* to act upon the source of the risk of harm.

As shown in the previous chapter, some authors consider that states should have jurisdictional competence over the right-holder for any negative and positive human rights duties to arise.[925] Accordingly, a jurisdictional link is created between a state and the right-holder only when foreign individuals are directly affected by the conduct of state agent or organs, or when the state in question administrates the territory of residence of the right-holder. In other words, states do not owe human right protection to individuals abroad at all time, under all circumstances. As a result, states would not incur responsibility for their negligence in preventing and abating the freedom-destroying effects of transboundary activities conducted by private actors from within the territory they administrate, because they are not obliged to do so in the first place.

First of all, it is important to keep in mind that the problem of jurisdictional competences in the context of freedom-destroying transboundary private activities does not always create difficulties. For instance, the due diligence measures of prevention deployed by a state may simply suffice to alleviate the risk of human rights violations arising from the conduct of other states in its own territory. Similarly, there would be no issues of jurisdiction when states willingly commit to the prevention of transboundary harm or when the states of origin of the harm cooperate with the victim and redress the harm, as if the victims were under their jurisdictions. In the same vein, there may be situations in which the state of origin takes the decision to extradite the perpetrators of human rights violations to the state where the violations have been committed. A last unproblematic yet unlikely scenario is when the state of emission and the recipient state are mutually bound by a bilateral treaty to ensure that private propaganda

[925] Besson, *Supra* note 743.

hostile to the human rights of the population of the other party are kept in check in their respective territories.

Still, the possibility remains that the state accused of negligence in preventing transboundary human rights violations refuses to cooperate with the victims. Under restrictive views on human rights extraterritoriality, the victims are then left with no avenues for enforcement purposes, even if it is apparent that the state of origin was negligent in preventing infringements on the rights of persons situated outside their territory and jurisdiction. Another consequence is that the recipient state itself is left with a heavy burden to assume in terms of its own duty of harm prevention.

Some organisations and human rights bodies have drawn inspiration from the no-harm principle to further expand human right extraterritoriality and facilitate human rights effectivity so that it palliates the problem of jurisdiction when the state of origin and the victim are not linked by a direct relationship of authority/control. In doing so, they have introduced the theory of "jurisdiction *qua* control over the source of the prejudice." Thus, according to some, a jurisdictional link is created between a state and an individual when an activity harmed the right-holder, whatever his physical location, provided that the state of origin was exercising control over the harmful activity and was in capacity to prevent or stop the harm. Basically, "the control that states have over activities that might adversely affect the enjoyment of human rights of an individual abroad is considered as jurisdiction over that individual itself."[926] This opens the possibility for engaging the responsibility of the state of origin for its negligence in preventing human rights violations arising from the transboundary effects of private activities initiated from its territory.

The Inter-American Court of Human Rights appears to have subscribed to this interpretation of extraterritorial jurisdiction. In a 2017 advisory opinion, the Court stated; "for the purposes of the American Convention, when transboundary damage occurs that affects treaty-based rights, it is understood that the persons whose rights have been violated are under the jurisdiction of the State of origin, if there is a causal link between the act that originated in its territory and the infringement of the human rights of persons outside its territory."[927] Thus, in order to establish extraterritorial jurisdiction, and thus to ground due diligence duties, a relation of causality must be established between, on the one hand, the failure of a state to address a risk of transboundary harm upon which it was entitled and in capacity to prevent, and the effective

[926] Philipp, *Supra* note 746., (36).
[927] The Inter-American Court of Human Rights, *Supra* note 740, at 43.

occurrence of transboundary harm. Basically, "territorial jurisdiction over the activities is extended extraterritoriality through causality."[928]

The authors of the Maastricht Principles, a document that aims at reenforcing human rights extraterritoriality, similarly submitted that a state has obligations to respect, protect and fulfil economic, social and cultural in "a) situations over which it exercises authority or effective control, whether or not such control is exercised in accordance with international law; b) situations over which State acts or omissions bring about foreseeable effects on the enjoyment of economic, social and cultural rights, whether within or outside its territory; c) situations in which the State, acting separately or jointly, whether through its executive, legislative or judicial branches, is in a position to exercise decisive influence or to take measures to realize economic, social and cultural rights extraterritorially, in accordance with international law."[929]

Some authors have criticized this approach to human rights jurisdiction. Besson notably argues that this risks "diluting the relational specificity of human rights and, by extension, undermining the coherence of international human rights law as a whole."[930] She explains that this interpretation of extraterritorial jurisdiction conflates the requirements for due diligence duties of prevention to arise (the authority of a state over the source of harm), and the conditions for human rights jurisdiction itself (effective overall control over a territory or the injured person).[931] In her words, "this simply puts the cart, due diligence, before the horse, the obligation of due diligence."[932] Accordingly, in absence of jurisdictional link between a state through its agents and organs and foreign individuals, a state of emission simply cannot owe human right protection to the victims of the transboundary freedom-destroying activities of private persons situated under their jurisdiction.

Restrictive views on human rights extraterritoriality are unintuitive in many aspects. The unprecedented level of interconnectivity of the modern world has led to a multiplication of points of contacts between foreign and domestic actors. Meanwhile, private actors have a growing impact on societies all over the world. In such a context, the possibility for interferences and frictions between the foreign and the domestic dimensions is increased.[933]

---

[928] Frederica Violi, "The function of the Triad 'Territory', 'jurisdiction', and 'Control' in Due diligence Obligations" in in Heike Krieger, Anne Peters & Leonhard Kreuzer (eds.), *Due diligence in the international legal order*, (New York: Oxford University Press, 2020) 75, at 82.

[929] *Maastricht Principles on Extra-Territorial Obligations of States*. (2011), at 6.

[930] Besson, *Supra* note 741, at 2.

[931] *Ibid*.

[932] Samantha Besson, "Due diligence in international law" (2022) Audiovisual Library of International Law.

[933] Oxman, *Supra* note 744.

Restrictive models of jurisdiction may result in absurd situations where the state of origin was clearly in the position to act upon the source of harm but willingly chose to turn a blind eye on it, thereby enabling an environment for human rights violations by private actors it harbours. In other words, the negligence of the state of origin is left unaddressed. This perpetuates a vicious circle of impunity where states can easily hide behind the veil of plausible deniability even if they are blatantly negligent in the management of their territories. *In fine*, the risk is that some states become safe haven for disinformation industries.

As argued in the previous chapter, *customary* human rights should be considered owed to everyone, at every time by states. As a result, to the extent that the state of origin was entitled and in capacity to act upon the source of prejudice, the jurisdictional link between the victim of customary human rights violations resulting from transboundary private activities and the state of origin of the conduct should be presumed. This means, in practice, that all states should ensure that their respective territories are not used for acts contrary to the self-determination, rights to life, and right to non-discrimination of foreign audiences, at all time.

Hence the same considerations as previously regarding the fulfilment of human rights due diligence duties would apply to the prevention of abatement of private disinformation campaigns that are of the nature to corrupt foreign election processes and *referenda* or putting the health of foreign audiences in danger.

Few clarifications are in order. Indeed, it has sometimes been pointed that human rights extraterritoriality would put an untenable burden on states. On that point, it is important to keep in mind that the discharge of due diligence obligations should not be done at the detriment of the rights of other states. In other words, states do not have a duty to enforce customary human rights on other states. Second, the stringency of the due diligence duties assumed by the state of origin should always be modulated according to the capacities of each state, and the level of control it holds over the source. Moreover, in the disinformation context, the preventive aspect of due diligence is relatively light in terms of the obligations assumed by the state of origin. Considering the risk of human rights abuses, measures of risk abatement should only be taken in presence of particularly nefarious disinformation activities, those that may put the lives of foreign audiences at risk. This should only concern ongoing health disinformation and egregious hate speeches when they reached a certain scale. Outside these extreme scenarios, due diligence should not serve as a justification for interferences into the freedom of expression of disinformants. Measures of mitigation of the risks arising from information misuses should remain reasonable, and be limited to educational measures and eventually the communication of correction to the aggrieved audience. State liability would arise if the state of emission

demonstrated repeated negligence in dealing with private disinformants involved in transboundary disinformation campaigns, or in cooperating with the authorities of the recipient states to investigate egregious transboundary disinformation cases, provided that said disinformation efforts were constitutive of human rights law violations, or could foreseeably lead to human rights harm.

Non-customary human rights such as the freedom of speech and the FT&RHO remain subject to the requirements of effective overall control upon the victim. However, this is not to say that states are exempted from due diligence obligations of prevention when it comes to the international protection of non-customary rights. States are still expected use their territories in good faith. However, failure to address harmful activities conducted from their territories or from persons situated under their jurisdiction will not trigger an obligation of liability if non-customary human rights are infringed upon. Greater levels of accountability could still be implemented in a gradual manner, once a sufficient number of states demonstrates the same readiness to assume more responsibilities with respect to the extraterritorial protection of the FT&RHO of individuals, as these rights are arguably the most vulnerable rights in today's "world-life-web" where attention economy is dominant.

With that being said, it should be relatively straightforward for a state to discharge its due diligence duties with respect to the prevention of international law breaches arising from information misuses by private actors, whether the disinformation affect domestic or foreign audiences. One should be careful not to expect too much from states in that regard. The main reason is that harsh due diligence obligations would inevitably clash with the freedom of expression. Additionally, there is reason to doubt as to whether states can realistically discharge their few relevant due diligence duties in an appropriate manner. Whether through improving information literacy domestically, or through issuing public corrections, it is likely that a certain degree of political bias and conflicts of interests is involved. This is why liability for negligence in the prevention of excessive risks arising from information misuse should be shared with social media platforms and news providers.

4.3.2.3 The need for a regime of corporate liability of large social media platforms and online news providers for human rights impairments arising from online information misuses

States are not the only entities to be subject to due diligence duties in the disinformation context. A way to release the burden of states with respect to the prevention of human rights violations arising from transboundary disinformation activities, and to reduce the influence of the politics into the management of information misuses could be to encourage social media platforms and news providers to implement due diligence duties in their own activities without relying on states.[934] In the context of online information transmission, this should concern in priority the largest social media platforms and online news providers.

Social media platforms and online news providers are the main enabler of information distribution and by incidence, psychological warfare. Thanks to cyberspace, they have gained considerable influence on individuals around the world. For example, it is estimated that around 60% of world population uses at least one social media platform and that an internet user spends an average of 2,5 hours per day on a social media platform.

Obviously, owners of social media platforms and news providers hold a responsibility in the mitigation of human rights violation arising from cyber-enabled PSYOPS.[935] For that reason, they should themselves assume human rights due diligence duties, and be held accountable in case of grave negligence leading to human rights violations arising from information misuses in their services. Thus, alongside states, human rights responsibilities should be expected from social media companies and news providers operating online directly. Not only would this approach be fairer to states, but it would also allow to circumvent the theoretical difficulties associated with the question of human rights extraterritoriality.

In recent years, the UN, the EU, and several human rights associations have pushed for the establishment of a general a regime of corporate due diligence in the context of human rights protection. In 2011, the UNHRC took concrete steps in this direction. In its "Guiding Principles on Business and Human rights."[936] (also known as the Ruggie Principles), the HRC proclaimed as a foundational principle that "Business enterprises should respect human rights. This means that they should avoid infringing on the human rights of others and should address adverse

[934] See in that sense, Jones, *Supra* note 119; Besson, *Supra* note 741; Rebeca Crootof, "The Internet of Torts: Expanding civil liability standards to address corporate remote interference" (2019) 69 *Duke Law Journal* 583.
[935] Jones, *Supra* note 119.
[936] HRC, "Guiding Principles on Business and Human rights" (2011).

human rights impacts with which they are involved."[937] Article 13 of the Guiding principles makes a direct reference to due diligence. The Article provides that corporations should "[S]eek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts."[938] An interpretative guide published in 2012 provided further guidelines to states and private corporations on how to meet their responsibility with respect to human rights protection. Similarly, in 2019, France acted the *loi de viligance* (or law on due diligence duties) which sought to improve corporation accountability in the domain of human rights protection. In early 2022, the European Union followed suit with a Proposal on Corporate Sustainability Due Diligence.[939]

More recently in October 2022, the European Parliament and Council adopted the Digital Services Act (DSA).[940] The DSA lays the foundations for a regime of corporate liability for negligence in the mitigation of human rights-related risks arising from online information misuses including online disinformation, in social media platforms and online search engines. The DSA is applicable to all online platforms and search engines including foreign companies that have a business in the European Union.

As a European regulation, the DSA is directly applicable in the EU. Although states do not have to transpose the instrument for it to apply domestically, unlike for European directives, the DSA still partially relies on the EU member states in order to take effect. Notably, under the DSA, each member state is granted with some enforcement powers and is required to designate an independent "Digital Service Coordinator" to ensure compliance with the regulation and coordination within the EU. From this perspective, albeit focussed on the regulation of the activities of digital corporations, the DSA is also setting new due diligence duties for member states to fulfil. The core of the DSA is however composed of due diligence obligations directly applicable to online "service providers" that have a certain part of their activity in one European state or more. The European Commission is granted full supervisory and enforcement powers over the largest platforms. The companies Meta (Facebook, Instagram), Google (Google search, Youtube), Twitter and ByteDance (Tiktok) are among the biggest players who are directly concerned by the regulation.

---

[937] *Ibid*., at 13.

[938] *Ibid*., at 14.

[939] European Commission, "Proposal for a Directive on corporate sustainability due diligence" (2022) Directorate-General for Justice and Consumers, <https://commission.europa.eu/publications/proposal-directive-corporate-sustainability-due-diligence-and-annex_en>.

[940] See DSA, *Supra* note 923.

Under the DSA, “very large online platforms (VLOPs)”, those that reach more than 10% of the European population, assume the most stringent obligations while smaller platforms, hosting services and intermediary services will assume less stringent and fewer due diligence duties. Due diligence duties common across all entities include, among other, the duty to conduct transparency reporting, to cooperate with European authorities, following orders, and to report criminal offences. The due diligence duties only applicable to VLOPs include among other, an obligation to share data with authorities and researchers, to comply with external & independent auditing, and interestingly, the obligation to give users the choice not to have recommendations based on profiling. Most importantly, all online platforms are bound by a due diligence duty of transparency, notably regarding the algorithms used for content recommendation. On this point, the DSA notes “When recipients of the service are presented with advertisements based on targeting techniques optimised to match their interests and potentially appeal to their vulnerabilities, this can have particularly serious negative effects. In certain cases, manipulative techniques can negatively impact entire groups and amplify societal harms, for example by contributing to disinformation campaigns or by discriminating against certain groups. Online platforms are particularly sensitive environments for such practices and they present a higher societal risk. Consequently, providers of online platforms should not present advertisements based on profiling”[941]

The DSA is interesting in that it rightfully focusses on the responsibilities of the enabler rather than the individuals that shared information. This approach to online disinformation has the merit to avoid heavy-handed approaches to the issue.

The DSA is built around the concept of “systemic risk”, and recognizes the dissemination of "illegal content" online as a primary source of systemic risk. The category of illegal content notably includes incitements to hatred, pornographic content and terrorist propaganda. The category also includes information that "relates to an illegal practice" such as the advertising of counterfeit products and online harassment. Other risks are considered by the DSA. For example, the corruption of democratic processes, health-related risks and privacy violations are considered long-term risks that disinformation, like other manipulative practices, may exacerbate.

Under the DSA, large platforms must exercise due diligence in assessing and mitigating systemic risks. This means that content moderation must be proportional to the intensity and immediacy of the risk incurred. Consequently, platforms will be required to remove manifestly

[941] *Ibid*., (69).

illegal content. In contrast, content removal is proscribed when it comes to the other contributors to the aforementioned risks. That is to say, that mere disinformation does not justify interferences into the freedom of speech of the speaker. Rather, the DSA requires platforms to exercise due diligence in the way online services providers design their websites and platforms so that they are less prone to manipulative behaviours. Thus, sections 25, 26 and 27 of the DSA prohibit, respectively, addictive interfaces, non-transparent advertising practices, profiling through the collection of certain personal information, and opaque recommendation systems. The DSA also urges major platforms to comply with the Code of Practice on Disinformation. Drafted earlier in 2022, the code notably imposes on service providers the removal of financial incentives for disinformation and the systematization of fact checking. These measures are essential, as the rapporteur for the promotion of freedom of expression and opinion at the United Nations underlined: "[B]y designing their products with highly personalized content to encourage addictive engagement, companies further promote a system that significantly undermines people's agency and choice in relation to their information diet."[942]

Part of the DSA entered into force in November 2022. Following their designation after February 2023, the concerned online entities will have four months to comply with the regulation according to their designation. Failure to comply with their due diligence obligations would result in fines up to 6% of their annual global revenues. Repeated failures to comply with the regulation could eventually lead to a temporary ban of the concerned platforms from the European market.

Although not exclusively focussed on online disinformation, the European DSA is an important tool in the fight against the phenomenon. *In fine*, the European regulators expect to harmonize the regulations of the European members with respect to the regulation of problematic content, including disinformation and hate speech. Additionally, it is expected that the DSA enables greater levels of protections to the political integrities of the members states and the rights of the European citizen, notably their mental autonomy, through rendering political manipulation more difficult to conduct online.

It is still too early to tell whether the European initiative will be successful. Moving forward, it will be interesting to see how the concerned actors will implement the regulation. After all, outrage brings engagements, engagements bring advertisers, and advertisers bring money to platforms and news providers.[943]

[942] Irene Khan, supra note 47, at 14.
[943] Munn, supra note 89.

The DSA is likely to inspire regulatory frameworks in other states. Similar regulations are already in preparation in the UK, Germany and Korea. The problem in these cases however, is that the incentive comes from states rather than an international organisation or an independent organism. As argued earlier, this does not guaranty unbiased approach to the issue. One can only hope that national authorities do not take the opportunity to impose politically-stained and abusive regulations on online entities for the alleged need to "secure the online information space."

Concluding remarks

PSYOPS, particularly when conducted through cyberspace, pose a challenge for legal practitioners and international law as a whole. However, this is not an insurmountable challenge. Through a general review of the relevant Law, this work has shown that adequate limits to cyber-enabled psychological warfare can already be found in international law. These limits are set by human rights law and the international law safeguarding the territorial integrity and the political independence of states. It is possible to summarize the findings of this work as follows;

- The applicable laws

The customary principles of non-intervention and non-use or threat to use force provide strict limitations to the conduct of interstate psychological warfare. Under these principles, PSYOPS in general, and state-sponsored disinformation campaigns in particular, should unquestionably be considered internationally unlawful if they are of sufficient scale AND display the following characteristics:

- *PSYOPS made of open and repeated calls for revolt*
- *PSYOPS made of open and repeated calls for war*
- *PSYOPS made of repeated threats to use violence against members of the audience*

Thus, outside the very unlikely scenarios of a PSYOPS containing explicit and believable threats to use force addressed to the decision makers of a state, or a PSYOPS composed of explicit calls for revolt distributed to the population of a state, there are few plausible instances where disinformation campaigns contravene state-centric rules of international law.

The inadequacy of the state-centric framework governing the use PYOPS by states against foreign populations should not come at a surprise. State-sponsored disinformation mostly have indirect effects on a state. The non-use of force, as well as the principle of non-intervention and the prohibition of undue physical intrusions on the territories of states are best suited to the regulation of activities generating direct and quantifiable effects. As a consequence, it is difficult to apply these frameworks to state-sponsored PSYOPS without speculating excessively about the impact of an information flow at state scale. Similarly, the alternative approach consisting in according more prominence to the intent of the author of the act may simply allow too much place for arbitrariness. As for the sovereignty principle, this work has

explained that the existence of a rule of sovereignty protecting states from interferences into their political integrities was not substantiated and would anyway directly conflict with the non-intervention principle.

Many states and international scholars have devoted their resources and time towards deciphering and rethinking state-centric rules in light of new technological advancements in the information & communication domain. But even so, recent submissions regarding the non-intervention principle are too far apart conceptually to allow for hope of a satisfactory and consensual definition in the near future.

Of course, this is not to say that there are no avenues for a better legal coverage. Yet, one would be hard pressed to find clear, compelling standards for states that apply to interventionist state-sponsored disinformation campaigns without, by the same token, making the principle even more vulnerable to political manipulation than it already is. Be that as it may, lawyers should leave open the possibility for a future adjustment of the non-intervention principle, one that better protects a population's sovereignty against protracted foreign interferences resulting in the creeping loss of political independence of the audience.

But if it is clear that state-centric rules of international law, under their current interpretations, are mostly ill-fitted to face contemporary forms of psychological warfare, their inadequacy alone, should not be used as a justification for broadening the scopes of the rule that protect the sovereignty of states. Indeed, not only has it never been the point to regulate every aspect of psychological warfare, but any inconsiderate broadening of the scope of application of state-centric rules could make the law extremely prone to arbitrary applications, foster escalatory trends between states, and risk severely downplaying the importance of human right in this area because all states may not be willing to reconcile non-intervention with the protection of human rights.

State-centric rules are not the only available tools to combat state-sponsored disinformation. The individuals are directly affected by disinformation. Thus, the legality of cyber-enabled interstate psychological warfare should be assessed preferably through the lenses of international human rights law. After all, although it is ultimately to harm the interests of a state and realize strategic objectives that state-sponsored disinformation campaigns are waged in cyberspace, reasoning in terms of the damages done to the interests of a state instead of the damages done to individuals' cognitive abilities and rights misses the most critical point of psychological warfare; the fact that populations, as the immediate targets, are the first to suffer from state-led deceptive campaigns.

Of course, human rights law is not an enclosed area. State-sponsored disinformation campaigns may simultaneously violate human rights and state-centric customary rules of general international law. The dyadic relationship between the state-centric and the individual-centric rules of international law is easily proven. For example, foreign-led disinformation campaigns that led to substantial damages to life abroad would also engage the international responsibility of the influencing state on the basis of the non-use of force and the non-intervention principles.[944] In the same way, electoral disinformation may breach the right to internal self-determination of the audience and, if massive enough, may also amount to a prohibited intervention.[945] However, human rights are more easily breached by disinformation campaigns than state-centric rules. A human rights approach to disinformation also allows practitioners to take into account domestic disinformation cases, when a government targets its own citizen.

As argued earlier, states should abstain from infringing the human rights of foreign audiences while conducting transboundary PSYOPS through their organs and agents.[946] Likewise, harbouring states should incur liability for the freedom-destroying disinformation campaigns conducted by private entities located on their soil or under their jurisdiction, unless they show that appropriate measures were taken to prevent transboundary harm. This applies, at least, to the extent that violations of customary human rights arising from private disinformation campaigns are concerned.

Both the customary principles of self-determination, non-discrimination, the right to life, as well as their correlative duties of prevention find an application in the disinformation context. Given their customary nature, these rights should apply extraterritorially. Thus, to the extent that customary human rights are concerned, foreign states should not use the exception of jurisdiction in order to evade responsibility for infringement on the human rights of foreign audiences arising from PSYOPS they conducted, nor to evade responsibility for infringements on the human rights of foreign and domestic audiences arising from information manipulation activities conducted by domestic actors they were entitled and in capacity to prevent.

Of course, this is not to say that disinformation campaigns systematically breach the above human rights. Instead, it is submitted that the rights explored in this work should be interpreted as setting limitations on what is considered an acceptable use of PSYOPS by a state.

[944] Milanovic & Schmitt, *Supra* note 638.
[945] See in that sense Tsagourias, *Supra* note 409.
[946] Hollis, van Benthem and Dias, *Supra* note 61, at 34. See also Milanovic, *Supra* note 752.

Yet so far, attempts at addressing the question of disinformation mostly failed to acknowledge how important the human factor are in that context.[947] Not only a more human-centric approach to disinformation enables more granular approaches to disinformation without the downsides associated with the reliance of state-centric rules, but it also provides individuals with the legal tools to oppose abusive regulations taken by their own governments supposedly to defend against foreign PSYOPS.

Human rights law could also inspire future law creation if states finally agree on the necessity of a clear operational framework for the conduct of psychological warfare. A treaty dealing with state-sponsored PSYOPS should thus focus on prohibiting:

- *PSYOPS contributing, in context, to a situation of imminent danger for the life of the members of the audience, including acts of incitements to hatred, health disinformation, and disinformation about basic safety measures during natural disaster and wartime*,
- *PSYOPS that are of the nature to denying voters their right to internal self-determination, or corrupting democratic processes abroad in ways that impedes on the expression of collective will*.

It bears noting that the above principles should be modulated based on the circumstances of each case, taking into account the scale of the communication campaigns, their duration, their truth-value the general narrative conveyed, and the vulnerability of the target society, including the average education and digitalisation levels. It is interesting to note that, although what distinguishes disinformation from other PSYOPS, deceptiveness and high manipulative value, may very well constitute aggravating factors in an international offence, these are never the determining factors in the choice of a legal framework, and should not be.

The same way as espionage, PSYOPS and other information manipulation tactics that do not display the aforementioned criteria, or that remain too modest in scale should be considered "not illegal yet illegitimate" conduct.

- States should exercise self-restraint

The long-term effects of disinformation operations are still poorly known. This generally results in a high degree of overreaction in the way states apprehend information misuses, be

[947] See Hollis, van Benthem and Dias, *Supra* note 61. (Arguing that human rights law is not given enough attention in the disinformation context, and that scholars tend to give primacy of the question of the protection of the rights of states instead.)

they from domestic or extraterritorial origin. Yet, especially when it comes to democracies, abusive crackdown on information would have severe repercussions on the brand image of the concerned states. This, in turn, could fuel online disinformation. States should thus always be aware of the long-term consequences of their actions when they attempt to respond to disinformation and exercise self-restraint in responding to foreign disinformation campaigns.[948]

To ensure the perennity of internet as we know it, states should always exercise due diligence while dealing with disinformation, be it foreign state-sponsored or not.

When there is no doubt as to the authorship of a problematic PSYOPS, legal measures of retorsions should always be prioritized over heavy-handed responses. Measures of retorsions could take the form, *inter alia*, of official protests and diplomatic notices addressed to the enabler or sponsor state. Retaliations *in kind* should be avoided at all cost.

When doubts remain about the authorship of a PSYOPS, recipient states should notice the state of emission of the behaviour. Social media platforms should similarly be required to notice inauthentic coordinated behaviours to the state of emission and the recipient state once detected. Authorities of the recipient state should not intrude into the decision-making processes of their populations to render them impervious to foreign influence at large.[949] States should leave online users drawing their own conclusions regarding the quality of online content, which notably entails investing in educational tools for the improvement of digital literacy and critical thinking.[950] Additionally, states and social media platforms should ensure that the highest level of transparency is maintained so as to allow the individuals to decide whether a given piece of information is legitimate and authentic. Non-obtrusive communication campaigns strictly aimed at re-establishing the facts should be admitted. In the same vein, it should always be possible for a foreign state to, diligently, issue corrections to foreign audiences, online, or through other ways, as long as the communication remains non-obtrusive, non-invasive and transparent.

Temporary website defacements and content removal measures should be reserved to highly disruptive disinformation operations waged by a foreign state and posing serious and immediate threat to the political independence or the territorial integrity of the targeted state such as disinformation operations *clearly* aimed at arousing insurgent groups or internal violence, terrorist propaganda, and health disinformation.

[948] See notably Bronner *Supra* note 54.
[949] *Ibid*.
[950] *Ibid.*

- The ways forward

Online disinformation is an ever-evolving phenomenon. It is very likely that disinformation gains in efficiency and sophistication in the future.[951] The exponential development of A.I technologies including machine-learning and Big Data directly play in the disfavour of those who seek to curb online disinformation.[952] For instance, new methods of autonomation could greatly facilitate the practice of *Astroturfing*. A.I technology could automatize false accounts and identities generation, and even automatize the whole process of content generation and content publication for performing cheap, reactive, highly targeted and massive disinformation operations.[953] A.I is already contributing to the explosion of Deepfake online, featuring fake videos and voice records that are almost indistinguishable from real ones. These are already used in disinformation campaigns and will likely to be even more used in the future.

In spite of this, states and regulators should resist the temptation of promoting too loose interpretations of the available Laws, as doing so entails a high risk of capturing benign and acceptable conduct, in addition to putting some human rights at great risk.

There are ways to improve the resilience of societies in the context of over-evolving digital threats without stretching existing rules to their breaking point. It would be through focussing on human rights due diligence. To give full effect to human rights due diligence in the disinformation context would require states and legal practitioners to further flesh-out the content of due diligence duties they are bound to follow when dealing with online information flows. On this point, this work has shown that the discharge of due diligence duties should not be done at the expense of the sovereignties of other states, nor should it lead to unjustified interference in the freedom of expression.

However, due diligence in international law is still at its infancy. Clear procedural guidelines regarding how states should comply with their due diligence duties with respect to the prevention of human rights infringements arising from information misuses are needed. Corporate human rights due diligence is a promising development, but it should not come at the expense of states' own human rights due diligence duties. Basic due diligence duties that states should follow at any time should include;

- *To work on improving digital literacy*
- *To support encourage independent fact-checking*

[951] Kim Hartmann & Keir Giles, "The Next Generation of Cyber-Enabled Information Warfare" (2020) *2020 12th International Conference on Cyber Conflict* 233.
[952] *Ibid*., at 235.
[953] *Ibid*., at 241-246.

- *To conduct regular audits of their own PSYOPS capabilities*
- *To avoid putting excessive regulations on the freedom of speech*
- *To notify the state of emission once potentially illegal disinformation campaigns are successfully traced back to a territory*
- *To collaborate with the recipient state and to investigate high-scale domestic information misuses once they are detected*

Future developments should also focus on elaborating some human rights in light of the characteristics of our digital world. Without clear delineation, human rights risk being interpreted narrowly, which may curtail their use in many instances. For example, there is a need to uphold the freedom of expression and FT&RHO and to clarify their contours. Particularly, there should be insistence on the right to *equal* access to information. Indeed, the free and unabated flow of information between societies is an enabler of peace. States should not take advantage of the disinformation "trend" to render their populations impervious to so-called "external influence." Hence, an international treaty on media pluralism and the diversity of information sources setting a duty to guarantee that individuals have unabated access to information online and in real life would be a welcome development.

At any rate, any oversimplification in the treatment of the issue of disinformation cannot be seen as a step in the right direction in terms of human rights protection. Robust responses to disinformation should only be considered in the most extreme cases and no regulation should ever be enacted without a careful evaluation of the risk/benefit balance. This warning applies, notably, to the ongoing doctrinal debate regarding the principle of non-intervention.

"The force of propaganda is a direct attack against man."[954] said Ellul. Elusive security threats such as PSYOPS should be seen as an opportunity to elaborate the law and to innovate in ways that give primacy to the protection of fundamental freedoms rather than the sovereignties of states.

The blame for the raise of online disinformation must be shared between the most powerful instigators, states and the enablers, digital service providers, especially social media platforms. Thus, a safe way forward is to implement a liability regime to hold states accountable when they fail to act diligently in following their relevant due diligence duties. Additionally, a regime of corporate liability for negligence in dealing with high-scale information misuses taking advantage of their services should be imposed on the digital corporations that own the largest platforms, as they have a more direct control over the online information space. The emphasis

[954] Ellul, supra note 77, at xvi.

should be put on recommendation systems and platforms designs. *In fine*, it is believed that the implementation of due diligence duties in the context of online disinformation could pave the way for a new tort of negligent enablement of problematic disinformation campaigns that is applicable to states and online service providers. In that sense, instruments like the European DSA are steps on the right direction, but the way to a true World-Wide-Web that is less prone to manipulative practices faces many challenges.

Above everything, internet must remain a tool for emancipation, not manipulation. The future of the global information space is at stake.

Select bibliography

Books

Badouard, Romain. *Le désenchantement de l'Internet : désinformation, rumeur et propagande.* (Limoges: FYP Editions, 2017).

Baume, Sandrine, Véronique Boillet, and Vincent Martene. *Misinformation in Referenda* (London: Routledge, 2020).

Brownlie, Ian. *International Law and the Use of Force by States.* (UK: Oxford University Press 1963).

Cabus, Tony. *Due Diligence and the High Seas*. (London: Routledge, 2021).

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

Crawford, James. *State Responsibility: The General Part.* (Cambridge: Cambridge University Press, 2013).

Delerue, Francois. *Cyber Operations and International Law*. (Cambridge: Cambridge University Press, 2020).

Deschaux-Dutard, Delphine. *Introduction à la sécurité internationale.* (Grenoble: PUG, 2018).

Ellul, Jacques. *Propaganda : The Formation of Men's Attitudes.* (New York: Vintage, 1965).

Ertzscheid, Olivier. *Qu'est-ce que l'identité numérique ? Enjeux, outils, méthodologies.* (Marseille: OpenEdition Press, 2014).

Jowett, Garth S. & O'Donnell, Victoria J. *Propaganda and Persuasion*. (Thousand Oaks: SAGE publication, 2006 (4th edition)).

Lahmann, Henning. *Unilateral Remedies to Cyber Operations Self-Defence, Countermeasures, Necessity, and the Question of Attribution.* (Cambridge: Cambridge University Press, 2020).

Lippman, Walter. *Public Opinion.* (Lanham: Start Publishing LLC, 2015 (reissue)).

Milanovic, Marko. *Extraterritorial Application of Human Rights Treaties: Law, Principles, and Policy.* (Oxford: Oxford University Press, 2013).

Noël, Jacques. *Le principe de non-intervention : théorie et pratique dans les relations interaméricaines.* (Bruxelles : Édition de l'Université de Bruxelles - Bruylant, 1981).

Roscini, Marco. *Cyber Operations and the Use of Force in International Law*. (Oxford: Oxford University Press, 2014).

Shaw, Malcolm N. *International Law.* (UK: Cambridge sixth edition, 2008).

Spinoza, Barush. *A Theological-Political Treatise*. (1670) (translated by Elwes, R. H. M. (Robert Harvey Monro), 1853).

Schmitt, Michael N. (ed.) *Tallinn Manual on the International Law applicable to Cyber Warfare*. (Cambridge University Press: Cambridge, 2013).

Schmitt, Michael N. (ed.) *Tallinn manual 2.0 on the international law applicable to cyber operations*. (Cambridge University Press: Cambridge, 2017).

Sunstein, Cass R. *Liars Falsehoods and Free Speech in an Age of Deception.* (New York: Oxford University Press, 2020).

Vincent, R.J. *Non-Intervention and International Order*. (Princeton: Princeton University Press, 1974).

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

Articles

Aceves, William J., "Virtual Hatred: How Russia Tried to Start a Race War in the United States" (2019) 24 *Michigan Journal of Race and Law* 177.

Anderson, Scott, "Coercion", (2011), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.).

Baade, Björnstjern, "Fake News and International Law" (2018) 29(4) *European Journal of International Law* 1357.

Baade, Bjornstjern, "Due diligence and the Duty to Protect Human Rights" in in Heike Krieger, Anne Peters & Leonhard Kreuzer (eds.), Due diligence in the international legal order, (New York: Oxford University Press, 2020), at 92.

Balkin, Jack M., "The Information Society from an Antihumanist Perspective" in Ramesh Subramanian and Eddan Katz (eds.) *The Global Flow of Information: Legal, Social, and Cultural Perspectives*, (New York: New York University Press, 2011).

Balkin, Jack M., "The First Amendment Is an Information Policy" (2012) 41(1) *Hofstra Law Review* 13.

Besson, Samantha, "Due Diligence and Extraterritorial Human Rights Obligations – Mind the Gap!" 9(1) *ESIL Reflections* 1,

Besson, Samantha, "The Extraterritoriality of the European Convention on Human Rights: Why Human Rights Depend on Jurisdiction and What Jurisdiction Amounts to" (2012), 25(4) *Leiden Journal of International Law* 857.

Blay, Samuel K. N., "Territorial Integrity and Political Independence" (2010) *Max Planck Encyclopedias of International Law* [MPIL] Encyclopedia entries.

Boghardt, Thomas, "Soviet Bloc Intelligence and Its AIDS Disinformation Campaign" (2009) 53(4) *Studies in Intelligence* 1.

Buchan, Russel, "Cyber Attacks: Unlawful Uses of Force or Prohibited Interventions?" (2012) 17(2) *Journal of Conflict and Security Law* 212.

Cassese, Antonio, "Political Self-Determination – Old Concepts and New Developments" in Antonio Cassese (ed.), *UN Law/Fundamental Rights: Two Topics in International Law* (Sijthoff & Noordhoff, 1979).

Cassesse, Antonio, "The Nicaragua and Tadic Tests Revisited in Light of the ICJ Judgment on Genocide in Bosnia", (2007), 18(4) *The European Journal of International Law* 649.

Chauvancy, Francois, "Cyberespace, operations d'information, influence" in Olivier Kempf, Stéphane Dossé et Christian Malis (dir.) *Le cyberspace - Nouveau domaine de la pensée stratégique* (Paris: Economica, 2013).

Chen, Cheng, Wu, Kui, Srinivasan, Venkatesh & Zhang, Xudong, "Battling the internet water army: Detection of hidden paid posters" (2013) IEEE/ACM *International Conference on Advances in Social Networks Analysis and Mining* 116.

Coco, Antonio, Dias, Talita de Souza, "'Cyber Due Diligence': A Patchwork of Protective Obligations in International Law" (2021) 32(3) *European Journal of International Law* 771.

Conforti, Benedetto, "Le principe de non-intervention" in Bedjaoui, Mohammed (ed.) *Droit international*: *bilan et perspectives*, (Paris: Pedone, 1991), 489.

Cormac, Rory, "Techniques of covert propaganda: the British approach in the mid-1960s" (2019) 34(7) *Intelligence and National Security* 1064.

Corn, Gary, "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>.

Corn, Gary P. "Covert Deception, Strategic Fraud, and the Rule of Prohibited Intervention" (2020), 2005, Aegis Series Paper.

Crootof, Rebecca, “International Cybertorts: Expanding State Accountability in Cyberspace” (2018) 103 *Cornell Law Review* 565.

Damrosch, Lori Fisler, “Politics across Borders: Non-Intervention and Nonforcible Influence over Domestic Affairs” (1989) 83 *American Journal of International Law* 1, at 12.

David, Eric “Portée et limite du principe de non-intervention” (1990), 23 *Revue Belge de Droit International* 350.

DeConde, Alexander, “Washington’s Farewell, the French Alliance, and the Election of 1796.” (1957) 43(4) *The Mississippi Valley Historical Review* 641.

De Brabandere, Eric, “Propaganda” (2019) *Max Planck Encyclopedias of International Law*, Encyclopedia entries.

Dillon, Sara, “The Propaganda Conundrum: How to Control This Scourge on Democracy” (2022) 23 *Oregon Review of International Law* 123.

Distefano, Giovanni, “Fait continu, fait composé et fait complexe dans le droit de la responsabilité” (2006) 52 *Annuaire français de droit international* 1.

Downey, Elizabeth A., “A Historical Survey of the International Regulation of Propaganda” (1984) 5(1) *Michigan Journal of International Law* 341.

van Dyke, Vernon, “The Responsibility of States for International Propaganda” (1940) 34(1) *The American Journal of International Law* 58.

Egan, Brian J., “International Law and Stability in Cyberspace” (2017) 35 *Berkeley Journal of International Law* 169.

Fisher, Louis, “Destruction of the Maine (1898)” (2009) *The Law Library of Congress*.

Foltz, Andrew C., “Stuxnet, Schmitt Analysis, and the Cyber “Use-of-Force” Debate” (2012) 67(4) *Joint Force Quartely* 40.

Geiss, Robin & Lahmann, Henning, “Freedom and Security in Cyberspace: Non-Forcible Countermeasures and Collective Threat-Prevention” in Katharina Ziolkowski (ed.), Peacetime Regime for State Activities in Cyberspace. International Law, International Relations and Diplomacy, (Tallinn: NATO CCDCOE Publication, 2013), at 621.

Gerrits, André W.M., "Disinformation in International Relations: How Important Is It?" (2018) 29(1-4) *Security and Human Rights* 3.

Gomez, Miguel Alberto, “Coercion and Cyberspace” (11 September 2018), Real Instituto Elcano, <https://www.realinstitutoelcano.org/en/analyses/coercion-and-cyberspace/>

Von Glahn, Gerhard, “The Case for Legal Control of “Liberation” Propaganda” (1966) 31 *Law and Contemporary Problems* 553.

Hall, Todd H., “On Provocation: Outrage, International Relations, and the Franco–Prussian War” (2016) 26(1) *Security Studies* 1.

Hartmann, Kim & Giles, Keir, “The Next Generation of Cyber-Enabled Information Warfare” (2020) *2020 12th International Conference on Cyber Conflict* 233.

Hayden, Michael V., “The Future of Things Cyber” (2011) 5(1) *Strategic Studies Quarterly* 3.

Higdon, Nolan, “What is Fake News? A Foundational Question for Developing Effective Critical News Literacy Education” (2020) 29(1) *Democratic Communiqué*

Helal, Mohamed S, “On Coercion In International Law”, (2019), 52 International Law and Politics 1.

Hoffman, Frank G., “Examining Complex Forms of Conflict: Gray Zone and Hybrid Challenges” (2018) 7(4) *PRISM* 31.

Hollis, Duncan B., “The Influence of War; The War for Influence” (2018) 32(1) *Temple International & Comparative Law Journal* 31.

Hollis, Duncan B., Tsvetelina J van Benthem and Talita Dias, “Information Operations under International Law” (2022) 55(597) *Vanderbilt Journal of Transnational Law* 1.

Ido, Kilovaty, “The Elephant in the Room: Coercion” (2019) 113 *AJIL Unbound* 87.

Janig, Philipp, “Extraterritorial Application of Human Rights” (2021) in Christina Binder, Manfred Nowak, Jane A Hofbauer and Philipp Janig (eds.), *Elgar Encyclopedia of Human Rights* (2022 Forthcoming).

Jensen, Eric Talbot & Watts, Sean, “A Cyber Duty of Due Diligence: Gentle Civilizer or Crude Destabilizer?” 95 *Texas Law Review* 1555.

Jincheng, Wei, “Information war: a new form of people’s war” (1996) *Military Forum column, Liberation Army Daily*, Chinese intelligence agency-related documents.

Talmon, Stefan, “The Responsibility of Outside Powers for Acts of Secessionist Entities” (2009) 58(3) The International and Comparative Law Quarterly 493.

Tsagourias, Nicholas, “Electoral Cyber Interference, Self-Determination and the Principle of Non-Intervention in Cyberspace” in Dennis Broeders and Bibi van den Berg (eds.) Governing Cyberspace: Behaviour, Power and Diplomacy (Rowman & Littlefield, 2020).

Kada, Nicolas & Martial, Mathieu, « Puissance publique » in Nicolas Kada (ed.), *Dictionnaire d'administration publique* (Grenoble : Presses universitaires de Grenoble, 2014).

Keohane, Robert O., & Nye, Joseph S., “Power and Interdependence in the Information Age.” (1998) 77(5) *Foreign Affairs, Council on Foreign Relations* 81.

Kiessling, Elizabeth K., “Gray Zone Tactics and the Principle of Non-Intervention: Can “One of the Vaguest Branches of International Law” Solve the Gray Zone Problem?” (2021) 12 *National Security Journal* 116.

Kilovaty, Ido, “Doxfare: Politically Motivated Leaks and the Future of the Norm on Non-Intervention in the Era of Weaponized Information” (2017) 9 *Harvard National Security Journal* 148.

Kinslow , Christopher H., “Game of Code: The Use of Force Against Political Independence in the Cyber Age” (2019) July-Aug. 2018 *Army Lawyer* 28.

Kohen, Marcelo “The Principle of Non-Intervention 25 Years after the Nicaragua Judgment” (2012) 25(1) *Leiden Journal of International Law* 157.

Koivurova, Timo & Singh, Krittika, “Due diligence” (2022) Max Planck Encyclopedias of International Law, Encyclopedia entries

Krieger, Heike & Peters, Anne, “Due diligence and Structural Change in the International Legal Order” in Krieger, Heike, Peters, Anne & Kreuzer, Leonhard (eds.), *Due diligence in the international legal order*, (New York: Oxford University Press, 2020).

Kuehl, Dan, “Defining Information Power” (1997) 115 *Strategic Forum* 1.

Lachowski, Zdzislaw, “Confidence-Building Measures” (2006) *Max Planck Encyclopedias of International Law*, Encyclopedia entries.

Lahmann, Henning, “Infecting the Mind: Establishing Responsibility for Transboundary Disinformation” (2022) 33(2) *European Journal of International Law* 411.

Lahmann, Henning, “Information Operations and the Question of Illegitimate Interference under International Law” (2020) 53(2) *Israel Law Review* 189.

Lahmann, Henning, “On the Politics and Ideologies of the Sovereignty Discourse in Cyberspace” (2021) 32 *Duke Journal of Comparative & International Law*, 61.

Lahmann, Henning, “Protecting the global information space in times of armed conflict” (2020) 102 (915) *International Review of the Red Cross* 1227.

Larson, Arthur, “The Present Status of Propaganda in International Law” (1966) 31(3) *Law and Contemporary Problems* 439.

Lazer, David M. J. et al., “The science of fake news Addressing fake news requires a multidisciplinary effort” (2018) 359(6380) *Science* 1094.

Lee, Sangkuk, “China’s ‘Three Warfares’: Origins, Applications, and Organizations” (2014) 37(2) *Journal of Strategic Studies* 198.

Leiser, Mark "AstroTurfing, ‘CyberTurfing” and other online persuasion campaigns" (2016) 7(1) *European Journal of Law and Technology* 1.

Leiser, Mark, “Regulating Computational Propaganda: Lessons from International Law” (2019) 8(2) *Cambridge International Law Journal* 218.

LeRoy, Howard S., “Treaty Regulation of International Radio and Short-Wave Broadcasting” (1938) 32(4) *The American Journal of International Law* 719.

Libicki, Martin, “What is information warfare?” (1995) 28 *Strategic Forum,* Center for Advanced Concepts and Technology, Institute for National Strategic Studies, National Defense University.

Lin, Herbert, “The existential threat from cyber-enabled information warfare” (2019) 75(4) *Bulletin of the Atomic Scientists* 187.

Lin, Herbert & Kerr, Jaclyn, “On Cyber-Enabled Information Warfare and Information Operations” in Paul Cornish (ed.) *The Oxford Handbook of Cyber Security* (Oxford: Oxford University Press, 2021).

Little, Richard “Revisiting Intervention: A Survey of Recent Developments” (1987) 13(1) *Review of International Studies* 49.

Mačák, Kubo, “On the Shelf, But Close at Hand: The Contribution of Non-State Initiatives to International Cyber Law” 113 *AJIL Unbound* 81.

Mačák, Kubo, “Is the International Law of Cyber Security in Crisis?” in N. Pissanidis, H.Rõigas, M.Veenendaal (Eds.) 2016 *8th International Conference on Cyber Conflict: Cyber Power* (Tallinn: NATO CCDCOE Publications.

Mazarr, Michael J., “Mastering the Gray Zone: Understanding a Changing Era of Conflict” (2015) *US Army War College*, *Monographs, Books, and Publications* 428.

McCarthy-Jones, Simon, “The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century” (2019) 2(19) *Frontiers in Artificial Intelligence* 1.

Mendel, Toby, “Hate Speech: Can the International Rules be Reconciled?” (2011) Expert workshops on the prohibition of incitement to national, racial or religious hatred, Workshop for the Americas.

Meister, Stefan, “The "Lisa case": Germany as a target of Russian disinformation” (25 July 2016) NATO Review, <https://www.nato.int/docu/review/articles/2016/07/25/the-lisa-case-germany-as-a-target-of-russian-disinformation/index.html>.

Metzl, Jamie Frederic, “Rwandan Genocide and the International Law of Radio Jamming.” (1997) *91(4) The American Journal of International Law* 628.

Moulin, Thibault, “Reviving the Principle of Non-Intervention in Cyberspace: The Path Forward” (2020) 25(3) J*ournal of Conflict and Security Law* 423.

Moynihan, Harriet, “The vital role of international law in the framework for responsible state behaviour in cyberspace” (2021) 6(3) *Journal of Cyber Policy* 394.

Munn, Luke, “Angry by design: toxic communication and technical architectures” (2020) 7 *Humanities and Social Sciences Communications* 1.

Nicolas, Ashley C., “Taming the Trolls: The Need for an International Legal Framework to Regulate State Use of Disinformation on Social Media” (2018) 107 *Georgetown Law Journal* 36.

Nye, Joseph S., “Public Diplomacy and Soft Power” (2008) 616 *The Annals of the American Academy of Political and Social Science* 94.

Nye, Joseph S., "Soft Power and the Public Diplomacy Revisited" (2019) 14 *The Hague Journal of Diplomacy* 1.

Nye, Joseph S., “Soft power: the evolution of a concept” (2021) 14(1) *Journal of Political Power* 196.

Nye, Joseph S., “The Information Revolution and Soft Power” (2014) 113(759) *Current History* 19.

Ohlin, Jens David, “Did Russian Cyber Interference in the 2016 Election Violate International Law?” (2017) 95 *Texas Law Review* 1579.

Ohlin, Jens David, “Election Interference: The Real Harm and The Only Solution” (2018) *18(50) Cornell Legal Studies Research Paper* 1.

Omoregie, Uyiosa, “The ‘Harm Principle’ and Information Disorder Online” (2021), 1999 *Academia Letters*.

Pajon, Céline, “Japan’s Coast Guard and Maritime Self-Defense Force in the East China Sea: Can a Black-and-White System Adapt to a Grey-Zone Reality?” (2017) 23 *Asia Policy* 111.

Pielemier, Jason, “Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?” (2020) 2020(4) *Utah Law Review* 917.

Pomson, Ori, “The Prohibition on Intervention Under International Law and Cyber Operations” (2022) 99 *International Law Studies* 180.

Pronay, Nicholas & Taylor, Philip M., “’An Improper Use of Broadcasting…’” The British Government and Clandestine Radio Propaganda Operations against Germany during the Munich Crisis and after” (1984) *19(3) Journal of Contemporary History* 357.

Qiao, Shitong, “Whither China’s Non-Interference Principle?” (2011) *ESIL 2011 4th Research Forum*.

Raynova, Denitsa D. “Towards a Common Understanding of the Non-Intervention Principle” (2017) European Leadership Network, Post-Workshop Report.

Rid, Thomas, “Cyber War Will Not Take Place” (2012) 35(1) *Journal of Strategic Studies* 5.

Rodriguez, Manuel, “Disinformation Operations Aimed at (Democratic) Elections in the Context of Public International Law: The Conduct of the Internet Research Agency During the 2016 US Presidential Election” (2019) 47(3) *International Journal of Legal Information* 149.

Schmitt, Michael N., “Foreign Cyber Interference in Elections” (2021) 97 *International Law Studies* 739.

Schmitt, Michael N., “The Use of Cyber Force and International Law” (2015) in. Marc Weller (ed) *The Oxford Handbook of the Use of Force in International Law* (UK, Oxford University Press 2015).

Schmitt, Michael N., “’Virtual’ Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law” (2018) 19(1) *Chicago Journal of International Law* 30.

Schmitt, Michael N., “Foreign Cyber Interference in Elections: An International Law Primer, Part II” (16 October 2020) *EJIL:Talk!,* <https://www.ejiltalk.org/foreign-cyber-interference-in-elections-an-international-law-primer-part-ii/>.

Schmitt, Michael N., “Grey Zones in the International Law of Cyberspace” (2017) 42(2) *Yale Journal of International Law Online 1*.

Schmitt, Michael N. & Vihul, Liis, “Respect for Sovereignty in Cyberspace” (2017) 95(7) *Texas Law Review* 1739.

Schmitt, Michael N. & Milanovic, Marko, “Cyber Attacks and Cyber (Mis)information Operations During a Pandemic” (2020), 11 *Journal of National Security Law & Policy* 247.

Sucharitkul, Sompong, “State Responsibility And International Liability Under International Law” (1996) 18 *Loyola of Los Angeles Int'l. & Comp. L.J*. 821.

Falk, Richard A., “On Regulating International Propaganda: A Plea for Moderate Aims” (1966) 31(3) *Law and contemporary Problems* 622.

Schöndorf, Roy, “Israel’s Perspective on Key Legal and Practical Issues Concerning the Application of International Law to Cyber Operations” (2021) 97 *International Law Studies* 395.

Simon, Herbert A., “Designing organizations for an information-rich world” in Martin Greenberger (Ed.), *Computers, Communications, and the Public Interest* (Baltimore: Johns Hopkins Press, 1971).

Smyczek, Peter J., "Regulating the battlefield of the future: the legal limitations on the conduct of psychological operations (PSYOP) under public international law" (2005) 57 *Air Force Law Review* 209.

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

Stephens, Dale, "Influence Operations and International Law" (2020) 19(4) *Journal of Information Warfare* 1.

Sucharitkul, Sompong, "State Responsibility And International Liability Under International Law" (1996) 18 Loyola of Los Angeles Int'l. & Comp. L.J. 821.

Tay, Xuan W., "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.

Thomas, Timothy L., "Russia's Reflexive Control Theory and the Military" (2004) 17 *Journal of Slavic Military Studies* 237.

Tladi, Dire, "The Duty Not to Intervene in Matters within Domestic Jurisdiction" In Jorge E. Viñuales (Ed.), *The UN Friendly Relations Declaration at 50: An Assessment of the Fundamental Principles of International Law* (Cambridge, Cambridge University Press, 2020).

Tworek, Heidi, "Disinformation: It's History" (14 July 2021), Centre for International Governance Innovation.

Violi, Frederica, "The function of the Triad 'Territory', 'jurisdiction', and 'Control' in Due diligence Obligations" in in Heike Krieger, Anne Peters & Leonhard Kreuzer (eds.), *Due diligence in the international legal order*, (New York: Oxford Univhellalersity Press, 2020) 75.

Vosoughi, Soroush, Roy, Deb and Aral, Sinan, "The Spread of True and False News Online" (2018) 359(6380) *Science* 1146.

Watts, Sean, "Low-Intensity Cyber Operations and the Principle of Non-Intervention" in Jens David Ohlin, Kevin Govern, and Claire Finkelstein (eds), *Cyber War: Law and Ethics for Virtual Conflicts* (Oxford: Oxford University Press, 2015), 249.

Wenqing, Zhao, "Cyber Disinformation Operations (CDOS) And A New Paradigm of Non-intervention" (2020) 27(1) U.C. *Davis Journal of International Law & Policy* 35.

Whitton, John B., "Cold War Propaganda" (1951) 45(1) *American Journal of International Law* 151.

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

Wigell, Mikael, "Hybrid interference as a wedge strategy: a theory of external interference in liberal democracy" (2019) 95(2) *International Affairs* 255.

Wright, Quincy, "Subversive Intervention." (1960) 54(3) *The American Journal of International Law*, 521.

Wright, Quincy, "The Crime of "War-Mongering." (1948) 42(1) *The American Journal of International Law* 128.

Wright, Quincy, "The Prevention of Aggression*"* (1956) 50(3) *The American Journal of International Law* 514.

Watts, Sean, "International Law and Proposed U.S. Responses to the D.N.C. Hack" (14 October 2016) *Just Security*, <https://www.justsecurity.org/33558/international-law-proposed-u-s-responses-d-n-c-hack/>

Zeng, Jinghan, Stevens, Tim, Chen, Yaru, "China's Solution to Global Cyber Governance: Unpacking the Domestic Discourse of 'Internet Sovereignty'" (2017) 45(3) *Politics & Policy* 432.

Ziolkowski, Katharina, "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, Estonia: NATO CCD COE Publication, 2013), at 165.

Studies and reports

Bayer, Judith et al., "Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States" (2019), European Parliament.

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

Bradshaw, Samantha, Bailey, Hannah, & Howard, Philip N., "Industrialized Disinformation 2020 Global Inventory of Organized Social Media Manipulation" (2021) Oxford UK: Computational Propaganda Project Working Paper.

Bradshaw, Samantha & Howard, Philip N., "The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation" (2019) Oxford, UK: Computational Propaganda Project Working Paper.

Charon, Paul & Vilmer, Jean-Baptiste Jeangène, « Les Opérations d'Influence Chinoise Un moment Machiavélien » (2021) études de l'IRSEM.

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

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

Geiss, Robin & Lahmann, Henning, "Protecting Societies: Anchoring A New Protection Dimension In International Law In Times Of Increased Cyber Threats" Working Paper Geneva Academy. Université de Genève.

Grossman, Shelby, Buchatskiy, Catarina, B.-B., Benjamin, D., Kate, DiResta, Renee, H., Christina & Steinberg, Julia, "Full-Spectrum Pro-Kremlin Online Propaganda about Ukraine" (22 February 2022) Freeman Spogli Institute for International Studies.

Internews, "Why information matters A foundation for resilience" (2015) report, Embracing Change: The Critical Role of Information Project.

Jack, Caroline, "Lexicon of Lies" (2017) Data & Society Research Institute, <https://datasociety.net/library/lexicon-of-lies/>.

Jamnejad, Maziar & Wood, Michael, "The Principle of Non-intervention" (2009) 22 *Leiden Journal of International Law* 345.

Johnson, Christian & Marcellino, William, "Bad Actors in News Reporting Tracking News Manipulation by State Actors" (2021) Santa Monica, CA: RAND Corporation.

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

Kelly, Sanja et al., "Manipulating Social Media to Undermine Democracy" (2017) Freedom House, Freedom on the Net.

Kleis, Rasmus Nielsen & F., MacKenzie, Common, "Submission to UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression report on disinformation" (2021) Reuters Institute.

Larson, Eric V., et al., "Foundations of Effective Influence Operations A Framework for Enhancing Army Capabilities" (2009) Santa Monica, CA: RAND corporation.

Legatum Institute, "Information at War: From China's Three Warfares to NATO's Narratives" (2015), Beyond Propaganda series.

Libicki, Martin, "Cyberdeterrence and Cyberwar" (2009), Santa Monica, CA: RAND corporation.

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

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

Mazarr, Michael J., Bauer, Ryan, Casey, Abigail, Heintz, Sarah, & Matthews, Luke J., "The Emerging Risk of Virtual Societal Warfare: Social Manipulation in a Changing Information Environment" (2019) Santa Monica, CA: RAND Corporation.

Morris, Lyle J. et al., "Gaining Competitive Advantage in the Gray Zone Response Options for Coercive Aggression Below the Threshold of Major War" (2019) Santa Monica, CA: RAND Corporation

Moynihan, Harriet, "The Application of International Law to State Cyberattacks" (2019) Chatham House Research Paper.

Nimmo, Ben, "The case for information defence, a pre-emptive strategy for dealing with the new disinformation wars" in "Information at War: From China's Three Warfares to Nato's Narratives" (2015) Legatum Institute.

Noel, Jean-Christophe, "What is Digital Power?" (November 2019), IFRI Study Series.

Parton, Charles, "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.

Reppell, Lisa & Shein, Erica, "Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions" (2019) International Foundation for Electoral Systems (IFES) Brief.

Simon, Felix, Howard, Philip N., Nielsen, Rasmus Kleis, "Types, sources, and claims of COVID-19 misinformation", (2020), the Oxford Internet Institute, <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>.

Talan, Bermet, "How journalists can address misinformation on Telegram?" (2021) Journalist Fellowship Paper, Thomson Reuters Foundation.

Thompson, Jared, Doxsee, Catrina & Bermudez, Joseph S. Jr., "Tracking the Arrival of Russia's Wagner Group in Mali" (2 February 2022), *Center for Strategic and International Studies,* <https://www.csis.org/analysis/tracking-arrival-russias-wagner-group-mali>.

Thompson, Nicholas & Lapowsky, Issie, "How Russian Trolls Used Meme Warfare to Divide America" (17 December 2018), *Wired,* <https://www.wired.com/story/russia-ira-propaganda-senate-report/>.

Vilmer, Jean-Baptiste Jeangène, "The "Macron Leaks" Operation: A Post-Mortem" (2019) Atlantic Council Report, <https://www.atlanticcouncil.org/in-depth-research-reports/report/the-macron-leaks-operation-a-post-mortem/>

Ward, Megan, Pierson, Shannon & Beyer, Jessica, "Formative Battles: Cold War Disinformation Campaigns and Mitigation Strategies" (2019) Wilson Center Science Technology Innovation Program.

Wineburg, Sam and McGrew, Sarah, Breakstone, Joel and Ortega, Teresa, "Evaluating Information: The Cornerstone of Civic Online Reasoning" (2016) Stanford Digital Repository.

Center for Strategic and International Studies, (CSIS), "Competing in the Gray Zone Countering Competition in the Space between War and Peace" International Security Program, <https://www.csis.org/features/competing-gray-zone>.

Graphika, "Unheard Voice-Evaluating five years of pro-Western covert influence operations" (2022), Stanford internet observatory, Cyber publications.

UN documents

Additional Protocols to the Geneva Conventions *of 12 August 1949* relating to the Protection of Victims of International Armed Conflicts (*1977)* 1125 UNTS 3

Advancing responsible State behaviour in cyberspace in the context of international security, (2020), A/RES/75/32.

Annual report of the United Nations High Commissioner for Human Rights, (2013), UNGA, A/HRC/22/17/Add.4.

Article 19, Freedoms of opinion and expression, (2011), CCPR/C/GC/34.

CESCR General Comment No. 14: The Right to the Highest Attainable Standard of Health, (Art. 12), (2000), Office of the High commissioner for Human Rights.

Condemnation of Propaganda Against Peace, (1951), A/RES/381(V).

Consideration of principles of international law concerning friendly relations and co-operation among States in accordance with the Charter of the United Nations, (1962), A/RES/17/1815.

Convention on the International Right of Correction, (1952) A/RES/630(VII).

Declaration by Migual Rodriguez, Representative of Cuba, at the final session of GGE developments in the field of information and telecommunication in the field of information and

telecommunication in the context of international security, (23 June 2017), <https://www.justsecurity.org/wp-content/uploads/2017/06/Cuban-Expert-Declaration.pdf>.

Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States, (1981), A/RES/36/103.

Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States, (1976), A/RES/31/91.

Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of Their Independence and Sovereignty, (1965), A/RES/20/2131 (XX).

Declaration on the Promotion Among Youth of the Ideals of Peace, Mutual Respect, and Understandings between Peoples, (1965), A/Res/2037 (XX).

Definition of Aggression, (1974), A/RES/29/3314.

Developments in the field of information and telecommunications in the context of international security, (2020), A/RES/75/240.

Developments in the field of information and telecommunications in the context of international security, (2018), A/RES/73/27.

Developments in the field of information and telecommunications in the context of international security, (2015), A/RES/70/237.

Developments in the field of information and telecommunications in the context of international security, (2014), A/RES/68/243.

Disinformation and freedom of opinion and expression, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression Irene Khan, (2021), A/HRC/47/25.

Draft Declaration on Rights and Duties of States (with commentaries), (1949), adopted by the UNGA in Draft Declaration on Rights and Duties of States A/RES/375 (6 December 1949).

Final act of the United Nations Conference on Freedom of Information, held at Geneva, Switzerland, from 23 March to 21 April, 1948.

Final Substantive Report of Open-ended working group on developments in the field of information and telecommunications in the context of international security, (2021), A/AC.290/2021/CRP.2.

Friendly Relations Declaration, (1970), A/RES/2625 (XXV).

General Comment No. 11: Prohibition of propaganda for war and inciting national, racial or religious hatred (Art. 20): 29/07/1983.CCPR General Comment No. 11. (General Comments) (1983), Office of the United Nations High Commissioner for Human Rights.

General Comment No. 22: Article 18 (Freedom of Thought, Conscience or Religion), (1993) CCPR/C/21/Rev.1/Add.4.

Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, (2017), Report of the Secretary-General, A/RES/72/327.

Human Rights Council, (2006), A/RES/60/251.

International Convention concerning the Use of Broadcasting in the Cause of Peace (1936), (17 December 1954), A/RES/841.

International Covenant on Civil and Political Rights, 999 UNTS 171.

International Covenant on Civil and Political Rights, UNHRC General comment No. 34, Article 19: Freedoms of opinion and expression, (2011), CCPR/C/GC/34.

International Law Commission, Articles on State Responsibility, (2001), 2 *Yearbook of the International Law Commission* (Part II) 31.

International Law Commission, "Draft articles on Prevention of Transboundary Harm from Hazardous Activities" *(2001) 2 Part Two, Yearbook of the International Law Commission* 148.

International Law Commission, "International Liability for Injurious Consequences arising out of acts not prohibited by international law" (liability in case of loss from transboundary harm arising out of hazardous activities) (2004), A/CN.4/543 86.

International Law Commission, "Ways and Means of Making the Evidence of Customary International Law More Readily Available" (1950) 2 *Yearbook of the International Law Commission* 367, A/1316.

International Law Commission, "Commentary to draft article 50 of the draft articles on the law of treaties", (1966), 2 *Yearbook of the ILC,* chap. II, sect. C.

Letter dated 9 January 2015 from the Permanent Representatives of China, Kazakhstan, Kyrgyzstan, the Russian Federation, Tajikistan and Uzbekistan to the United Nations addressed to the Secretary-General, (2015), A/69/723.

Measures to be taken against propaganda and the inciters of a new war, (1947), A/RES/2/110.

Official compendium of voluntary national contributions on the subject of how international law applies to the use of information and communications technologies by States, (2021), A/76/136.

Prohibition of propaganda for war and inciting national, racial or religious hatred (Art. 20): 29/07/1983. CCPR General Comment No. 11. (1983), Office of the United Nations High Commissioner for Human Rights.

Proposals submitted to, and decisions of, various United Nations organs relating to the question of State responsibility: supplement prepared by the Secretariat to document, (1969), A/CN.4/165.

Report of the commission of inquiry on human rights in the Democratic People's Republic of Korea (7 February 2014) HRC, A /HRC/25/63.

Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, (2021), A/RES/76/135.

Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, (2015), A/RES/70/174.

Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, (2013), A/RES/68/98.

Report of the Office of the United Nations High Commissioner for Human Rights, (2014), A/HRC/28/30.

Report of Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue, 26th session, (2014), A/HRC/26/30.

Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, (2016), A/HRC/32/38.

Report of the 1966 Special Committee on Principles of International Law Concerning Friendly Relations and Cooperation Among States, (1966), A/6230.

Respect for the principles of national sovereignty and non-interference in the internal affairs of States in their electoral processes, (1989), A/RES/44/147.

Role of States in countering the negative impact of disinformation on the enjoyment and realization of human rights, (2022) A/HRC/RES/49/21.

Special Committee on principles of international law concerning friendly relations and co-operation among states: summary record of the 29th meeting, (1964), A/AC.119.SR-2.

Strengthening of peace through the removal of barriers to free exchange of information and ideas, (1954), A/RES/819 (IX).

Strengthening the role of the United Nations in enhancing periodic and genuine elections and the promotion of democratization, (2017), A/RES/72/164.

Strengthening the Role of the United Nations in Enhancing Periodic and Genuine Elections and the Promotion of Democratization, (2019), A/RES/74/158.

The contribution of the Human Rights Council to the prevention of human rights violations, (2020), A/HRC/RES/45/31.

The contribution of the Human Rights Council to the prevention of human rights violations, (2018), A/HRC/RES/38/18.

The Nature of the General Legal Obligation Imposed on States Parties to the Covenant, (2004) General Comment No. 31.

The promotion, protection and enjoyment of human rights on the Internet, (2016), A/HRC/RES/32/13.

The promotion, protection and enjoyment of human rights on the Internet, (2014), A/HRC/RES/26/13.

The promotion, protection and enjoyment of human rights on the Internet, (2018) A/HRC/RES/38/7.

The promotion, protection and enjoyment of human rights on the Internet, (2016) A/HRC/RES/32/13.

The Role of Prevention in the Promotion and Protection of Human Rights, (2015) A/HRC/30/20

The role of prevention in the promotion and protection of human rights, (2013) A/HRC/RES/24/16

Trail smelter (United States, Canada), (1938-1941), 3 UN *Report of International Arbitral awards* 1905.

United Nations Strategy and Plan of Action on hate Speech, (2019), <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf>.

Other resources

Article 19, "'Hate Speech' Explained A toolkit" (2015).

Article 19, "Prohibiting incitement to discrimination, hostility or violence" (2012) Policy Brief.

Australian government, "Foreign Influence Transparency Scheme Act" (FITSA), (2018).

Australian government, "National Security Legislation Amendment (Espionage and Foreign Interference) Act" (2018).

Australian Government, Department of Foreign Affairs and Trade, "Supplement to Australia's Position on the Application of International Law to State Conduct in Cyberspace", (2019).

Australian Government, Department of Home Affairs, "Countering Foreign Interference" (2020).

Besson, Samantha, "Due diligence in international law" (2022) Audiovisual Library of International Law.

Braverman, *S*uella, the Attorney General, at Chatham House in London (19 May 2022*),* <https://www.ukpol.co.uk/suella-braverman-2022-speech-at-chatham-house/>.

Bronner, Gérard 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>.

Burkhard, Thierry, "Vision stratégique du chef des état-major des armées" French Ministry of the Armies, October 2021.

Constitution and Convention of the International Telecommunication Union (1992)

Declaration of General Staff of the Armed Forces of the Islamic Republic of Iran Regarding International Law Applicable to the Cyberspace (August 2020).

"Declaration of rights and duties of journalists" (Munich Charter), (1971).

ERGA, "ERGA's united response to foreign disinformation and information manipulation: European media regulators strengthen their cooperation" (10 February 2022), Press Release.

European Federation of Journalists (EFJ), "Fighting disinformation with censorship is a mistake" (2022), <https://europeanjournalists.org/blog/2022/03/01/fighting-disinformation-with-censorship-is-a-mistake/>.

European Parliament, "Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States" (2019) LIBE Committee Study.

"Exchange of Communications Between the President of the United States and Maxim M. Litvinov People's Commissar for Foreign Affairs of the Union of Soviet Socialist Republics" (1934) 28(1) *The American Journal of International Law* 2.

Federal Government of Germany, "On the Application of International Law in Cyberspace", Position Paper (2021).

Federal Department of Foreign Affairs "Switzerland's position paper on the application of international law in cyberspace" (May 2021).

French Ministry of the Armies, "Doctrine militaire de lute informatique d'influence (L2I)" (October 2021)

French Ministry of the Armed Forces, "Droit International Appliqué Aux Opérations dans le Cyberespace" (2019).

Funke, Daniel & Flamini, Daniela, "A guide to anti-misinformation actions around the world" *Poynter*, <https://www.poynter.org/ifcn/anti-misinformation-actions/>.

German Federal Government, "On the Application of International Law in Cyberspace" (2021).

Government of Canada, "International Law applicable in cyberspace" (2022).

Government of New Zealand, “The Application of International Law to State Activity in Cyberspace, (2020).

“How is the Term ‘Armed Conflict’ Defined in International Humanitarian Law?” (2008) International Committee of the Red Cross (ICRC) Opinion Paper.

International federation of journalists, “Global Charter of Ethics for Journalists” (12 June 2019)

International federation of journalists, “Declaration of Principles on the Conduct of Journalists” (Bordeaux Declaration), (1954),

Italian Ministry for Foreign Affairs and International Cooperation, "International law and cyberspace", Italian position paper, <https://www.esteri.it/wp-content/uploads/2021/11/Italian-Position-Paper-on-International-Law-and-Cyberspace.pdf>.

Kingdom of the Netherlands, “Appendix: International law in cyberspace”, (26 September 2019) Government of the.

Koh, Harold Hongju, “International Law in Cyberspace” Remarks as Prepared for Delivery by to the USCYBERCOM Inter-Agency Legal Conference Ft. Meade, MD, (18 September 2012), <https://2009-2017.state.gov/s/l/releases/remarks/197924.htm>.

Ministry of Foreign Affairs of Japan, “Basic Position of the Government of Japan on International Law Applicable to Cyber Operations” (2021).

Ministry of Foreign Affairs of the People’s Republic of China “China’s Positions on International Rules-making in Cyberspace”.

NATO, “Allied Joint Doctrine for Psychological Operations”, AJP-3.10.1(A).

Netherland, “the application of international law to cyber operations” (2019).

Organisation pour la sécurité et la coopération en Europe, “Joint Declaration on Freedom of Expression and ‘Fake news’, disinformation and propaganda” (3 March 2017), The representative on the freedom of the media, FOM.GAL/3/17.

OSCE, “International Law and Policy on Disinformation in the Context of Freedom of the Media”, (2021), Brief Paper for the Expert Meeting organized by the Office of the OSCE Representative on Freedom of the Media, <https://www.osce.org/files/f/documents/8/a/485606.pdf>.

Oxford Institute for Ethics, Law and Armed Conflict, “The Oxford Process on International Law Protections in Cyberspace.”

Oxford Institute for Ethics, Law and Armed Conflict “The Oxford Statement on International Law Protections in Cyberspace: The Regulation of Information Operations and Activities.”

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (DSA).

"Résolutions Adoptées Par l'institut de Droit International à La Session d"Aix-En-Provence 22 Avril-1er Mai 1954" (1953-55), 52 *Die Friedens-Warte,* Berliner *Wissenschafts-Verlag.,*

Report of the select committee on intelligence US senate on Russian active measures campaigns and interference in the 2016 US election Volume 2: Russia's use of social media with additional views, 116th Congress US senate, Report 116-XX.

Rid, Thomas, "Disinformation: A primer in Russian active measures and influence campaigns" (2017), Hearings before the select committee on intelligence, United States Senate, 115th congress, first session.

Russian Federation, "Conceptual Views on the Activity of the Russian Federation Armed Forces in Information Space" (2011), Section 1, Fundamental Terms and Definitions.

Singapore's Foreign Interference (Countermeasures) Act, (FICA) (2021).

Swiss Federal Department of Foreign Affairs, "Switzerland's position paper on the application of international law in cyberspace" (2021).

The Ministry of Foreign Affairs of the Russian Federation, "Doctrine of Information Security of the Russian Federation*",* Fundamental documents, (5 December 2016).

The Russian Ministry of Defence, "Conceptual Views Regarding the Activities of the Armed Forces of the Russian Federation in the Information Space" (2011), <https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/Russia_Cyber_Security_Strategy.pdf>.

United Kingdom Department for Digital, Culture, Media & Spor "Online Harms White Paper" (2019)

United Kingdom National Cyber Security Centre, "Reckless campaign of cyber-attacks by Russian military intelligence service exposed" (3 October 2018) <https://www.ncsc.gov.uk/news/reckless-campaign-cyber-attacks-russian-military-intelligence-service-exposed>.

US Air Force (USAF), "Cornerstones of Information Warfare" (Washington, D.C.: Department of the Air Force, 1995).

US Joint Chief of Staff, "Psychological operations" (7 January 2010), Joint Publication 3-13.2;

US Office of the Director of National Intelligence, (ODNI), "Background to 'Assessing Russian Activities and Intentions in Recent US Elections': The Analytic Process and Cyber Incident Attribution" (2017), Intelligence Community Assessment.

US Intelligence Community Assessment, "Background to "Assessing Russian Activities and Intentions in Recent US Elections": The Analytic Process and Cyber Incident Attribution", (6 January 2017), <https://www.dni.gov/files/documents/ICA_2017_01.pdf>.

Watts, Clint, "Disinformation: A Primer in Russian Active Measures and Influence Campaigns", Statement Prepared for the U.S. Senate Select Committee on Intelligence hearing, (2017).

Wright, Jeremy, "Cyber and International Law in the 21st Century" (23 May 2018) GOV.UK Speech, <https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century>.

World Press Freedom Index, (2022 edition), <https://rsf.org/en/rsfs-2022-world-press-freedom-index-new-era-polarisation>.

Encyclopaedia entries

Brunnée, Jutta, "Sic utere tuo ut alienum non laedas" (2022) Max Planck Encyclopedias of International Law, Encyclopedia entries.

Hannum, Hurst, "Legal Aspects of Self-Determination" Encyclopedia Pricetoniensis.

Kearney, Michael G., "Propaganda for War, Prohibition of" (2009) Max Planck Encyclopedias of International Law, encyclopedia entry.

Zemanek, Karl, "Armed attack" (2013) Max Planck Encyclopedias of International Law, Encyclopedia entries.

Ziegler, Katja S., "Domaine Reserve" (2013), *Max Planck Encyclopedias of International Law*. Encyclopedia entries.

Kunig, Philip, "Intervention, Prohibition of" (2008) Max Planck Encyclopedia of Public International Law, Encyclopedia entries.

International Convention on the Elimination of All Forms of Racial Discrimination (1965).

Table of cases

Permanent Court of Arbitration

Island of Palmas (United States vs The Netherlands), (1928), PCA, 2 U.N. Rep. International Arbitration Awards 829.

Permanent Court of International Justice

Phosphates in Morocco, (1938) PCIJ, Judgement.

S.S. *Lotus* (France vs Turkey), (1927) PCIJ Judgement.

Nationality Decrees Issued in Tunis and Morocco, (1923) PCIJ. Advisory Opinion.

European Court of Human Rights

case of Oneryildiz v Turkey (2004) EtHR, judgement.

Belkacem v. Belgium (2017) ECtHR, Decision on admissibility.

Pedersen & Baadsgaard v. Denmark, (2003) ECtHR, judgement.

Ivanova v. Bulgaria (2007) ECtHR (fifth section), judgement.

Kokkinakis v Greece (1993) ECtHR, judgement.

Perincek v. Switzerland (2015) ECtHR (Grand Chamber), Judgement.

Garandy v. France (2003) ECtHR Decision on admissibility

Pastors v. Germany (2019) ECtHR (fifth section), judgement

Loizidou v. Turkey, (1996), ECtHR, Judgment.

M'Bala M'Bala v. France (2015) ECtHR, decision.

Tusalp v. Turkey, (2012) ECtHR (second section), judgement.

Al-skeini and others v. The United Kingdom, (2011) EtHR, judgement.

International ad hoc tribunals

The Nuremberg Trial, (1946) International Military Tribunal (Nuremberg), judgement.

Prosecutor v. Tadić, (1999), ICTY, (Appeal Chamber), Judgment.

The Prosecutor v. Dusko Tadić, (1995), ICTY, (Appeals Chamber), Decision.

The Prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza, Hassan Ngeze (2003) ICTR judgement.

Prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza, Hassan Ngeze (2007) ICTR Appeal Judgment.

International Court of Justice

Advisory Opinion Concerning Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory, (2004). ICJ.

Advisory Opinion on the Legal Consequences for States of the Continued Presence of South Africa in Namibia, (1971), ICJ, opinion.

Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v Serbia and Montenegro), (2007), Judgment, ICJ Rep 43.

Application of the international convention on the elimination of all forms of racial discrimination (Azerbaijan v. Armenia) (2021) ICJ, Order - Request for the Indication of Provisional Measures *Case Concerning Armed Activities on the Territory of the Congo* (Democratic Republic of the Congo v. Rwanda), (2006) ICJ.

Case Concerning Armed Activities on the Territory of the Congo (Democratic Republic of the Congo v. Uganda), (2005), ICJ, Judgement.

Case Concerning East Timor (Portugal v. Australia), (1995) ICJ judgement.

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

Case Concerning Oil Platforms (Islamic Republic of Iran v. United States of America, (2003), ICJ, Judgement.

Corfu Channel (United Kingdom of Great Britain and Northern Ireland v Albania), (1949) ICJ, judgement.

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.

North Sea Continental Cases (Germany v Denmark and the Netherlands) (1969) ICJ, Reports.

Nuclear Tests (Australia vs France), (1973) ICJ, Order.

Right of Passage over Indian Territory (Portugal v. India), (1960) ICJ, Judgment.

United States Diplomatic and Consular Staff in Tehran, (1980), ICJ, Judgment.

Legality of the Threat or Use of nuclear weapons, (hereinafter Nuclear Weapon advisory), (1996), ICJ, Advisory Opinion.

Obligations Concerning Negotiations Relating to Cessation of the Nuclear Arms Race and to Nuclear Disarmament (Marshal Islands vs India), (2016) ICJ, Judgement.

UN Human Right Committee

JRT and WG Party v. Canada (1983) HRC, communication.

Robert Faurisson v. France (1996), HRC, communication.

Ross v. Canada, (1996) HRC, communication.

Shin v. Republic of Korea (2004) HRC, views.

Zeljko Bodrožić v. Serbia and Montenegro (2005) HRC, views.

National courts

The Gambia v. Facebook, (2021) US District Court, (First Instance Court), Decision.