International law and the challenge of disinformation: a patchwork of rights and obligations

Smulders, A.M.

2 Subversive Disinformation

'It is unlawful for Nations to do any act tending to create trouble in another state, to stir up discord, to corrupt its citizens, to alienate its allies.'

– Emmerich de Vattel[1]

2.1 INTRODUCTION

Chapters two through five analyse how existing international legal frameworks govern disinformation. Central to this assessment is the long-standing prohibition on subversive propaganda – foreign communications aimed at destabilising States by influencing their nationals towards insurrection, revolt or strife – dating to the French Revolution.[2] Early bilateral, and later multilateral treaties, stipulated that such influence operations are incompatible with the principle of sovereign equality of States and the prohibition of interference in internal affairs.[3] The government of a State, Whitton and Larson argue in their seminal work on international propaganda, 'is under a legal duty to refrain from spreading subversive propaganda hostile to a foreign country in time of peace.'[4] There are few rules, they continue, that 'are more firmly established' under international law. The scope of these prohibitions has gradually evolved from anti-war propaganda to include broader interference in 'political, economic, social and cultural' affairs.[5]

The Internet, advanced information technology and artificial intelligence, further enabled new forms of interference into sovereign affairs traditionally

---

1 Emmerich de Vattel, *The Law of Nations* (Chitty edn, 1863) 18.

2 Eric De Brabandere, 'Propaganda' (2019) Max Planck Encyclopaedia of Public International Law, para 12; Vernon van Dyke, 'The Responsibility of States for International Propaganda' (1940) 34 American Journal of International Law 1, 58.

3 John B Whitton and Arthur Larson, *Propaganda: Towards Disarmament in the War of Words* (Oceana Publications1963) 95; John Martin, *International Propaganda – Its Legal and Diplomatic Control* (University of Minnesota Press 1958) 90-94; Clark C Havighurst (ed), *International Control of Propaganda* (Oceana Publications1967); Elizabeth A Downey, 'A Historical Survey of the International Regulation of Propaganda' (1984) 5 Michigan Journal for International Law 341, 342-345.

4 Whitton and Larson (n 3) 95.

5 *Case Concerning Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States of America)* Judgement (Merits) [1986] ICJ Rep 14, para 205.

considered beyond the reach of foreign influence.[6] Among these is subversive disinformation. As a subset of subversive propaganda,[7] this type of disinformation encompasses false or misleading information that is created, produced, or disseminated with the intent to disrupt, undermine or obstruct a State's freedom to determine its internal and external affairs. Unlike disinformation or propaganda for war, it is neither directly nor exclusively designed to facilitate or incite armed conflict between States.[8]

The sovereign independence of States forms the cornerstone of international law.[9] The foundational sovereignty of States confers supreme authority over their territory and the conduct of internal and external affairs,[10] embodying equal rights and duties that maintain the stability of international relations, global peace and security.[11] Consequently, disinformation that undermines or disrupts State sovereignty constitutes a direct threat to the international legal order.

This chapter presents the parameters wherein subversive disinformation is incompatible with the obligations deriving from State sovereignty under public international law. It explores the phenomenon of 'subversive disinformation' in more depth, including by charting its historical development and highlighting several contemporary occurrences (2.2). Following, it explores the rights and obligations embodied in the notion of State sovereignty as a substantive rule of international law (2.3), alongside the prohibition of intervention (2.4).[12] While the State sovereignty and non-intervention paradigm overlap in their application to disinformation, this chapter argues in favour of upholding their distinct character:[13] unlawful intervention invariably

6 Lisa J Damon, 'Freedom of Information versus National Sovereignty: The Need for a New Global Forum for the Resolution of Transborder Data Flow Problems' (1986) 10 Fordham International Law Journal 4, 266-267; Joseph Nye, *Cyber Power* (2010) Harvard Kennedy School Belfer Center for Science and International Affairs Essay, 4; Hollis *et al.*, 'Information Operations under International Law' (2022) 55 Vanderbilt Journal of Transnational Law 5, 1217; Kimberly Breedon, 'An International Law Perspective on Political Information Warfare: The Challenges of Combatting the Weaponised Use of Conspiracy Theories and Disinformation to Undermine Democracy' (2022) 15 Saint Thomas Journal of Law and Public Policy 2, 657.

7 Section 1.5.2 'Typology'.

8 Whitton and Larson (n 3) 83.

9 Charter of the United Nations (adopted 24 October 1945) 1 UNTS XVI, Article 2(1); *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 5) para 263.

10 Samantha Besson, 'Sovereignty' (2011) Max Planck Encyclopaedia of Public International Law, paras 56, 69-70.

11 Memorial of the United Kingdom, *Corfu Channel* (United Kingdom v Albania) [1949] ICJ Rep 35, 43 (Separate Opinion of Judge Alvarez).

12 On sovereignty and non-intervention, see James Crawford, *Brownlie's Principles of Public International Law* (2019 Oxford University Press) 431; Harriet Moynihan, 'The Application of International Law to State Cyberattacks' (2019) Chatham House Research Paper.

13 Marko Milanovic, 'Revisiting Coercion as an Element of Prohibited Intervention in International Law' (2023) 117 The American Journal of International Law 4, 607-608.

violates both frameworks, whereas sovereignty violations absent coercive intent to alter State behaviour do not constitute a violation of the non-intervention rule.[14] The analysis engages with the widely voiced critique that both frameworks suffer from legal uncertainty: how do they apply to the online realm and can they adapt to the new sociotechnological reality of contemporary threats?[15] In this 'grey zone' of international law,[16] the chapter identifies conceptual complexities and definitional dilemmas, and proposes a set of indicators to determine the unlawfulness of subversive disinformation.

The legal analysis departs from established frameworks regulating subversive propaganda, whilst engaging with contemporary debates on international law's application to cyberspace. It reveals a prevailing conservative stance on interference and intervention, particularly in the context of influence operations that operate on the cognitive level. While acknowledging *lex lata* limitations, this chapter argues for a more behaviour-centred interpretation of the law to better capture subversive behaviour in the 21st century.

2.2 Propaganda, Subversion and Disinformation

Subversion, by definition, encompasses 'the action of overthrowing a nation, government, [or] ruler.'[17] Such conduct fundamentally contradicts State sovereignty, rendering subversive activities inherently incompatible with obligations to respect and protect it. It is well-recognised that foreign propaganda can threaten the safety and well-being of a State, particularly when information operations or campaigns are *intended* to undermine sovereign independence.[18] Subversive disinformation exerts similar, if not more severe, disruptive effects (2.2.1). While contemporary discourse predominantly approaches disinforma-

14 Wingfield and Wingo conclude that despite the observation that a number of trivial interferences or other minor intrusions do not rise to the level of a violation of sovereignty, 'in any case, sovereignty is implied the lowest of several thresholds […]' in Thomas Wingfield and Harry Wingo, 'International Law for Cyberspace: Competition and Conflict' in Paul Cornish (ed), *The Oxford Handbook of Cyber Security* (2021 Oxford University Press) 579; Mohamed S Helal, 'On Coercion in International Law' (2019) 52 New York University Journal of International Law and Politics 1, 90.

15 'Socio-technological reality' forms the central notion in Kilovaty's argument, in Ido Kilovaty, 'Rethinking Coercion in Cyberspace' in Mitt Regan and Aurel Sari (eds), *Hybrid Threats and Grey Zone Conflict: The Challenge to Liberal Democracies* (Oxford University Press 2024).

16 Michael N Schmitt, 'Grey Zones in the International Law of Cyberspace' (2017) 42 Yale Journal of International Law 1, 46; Gary P Corn, 'Covert Deception, Strategic Fraud, and the Rule of Prohibited Intervention' in Jack Goldsmith (ed), *The United States' Defend Forward Cyber Strategy: A Comprehensive Legal Assessment* (Oxford University Press 2022) 205.

17 Oxford English Dictionary, 'Subversion' (Accessed 28 November 2024).

18 RJ Spjut, 'Defining Subversion' (1979) 6 British Journal of Law and Society 2, 254-261 taking the 1978 definition of subversive activities are those 'threaten the safety or well-being of the state, and which are intended to undermine or overthrow parliamentary democracy by political, industrial or violent mean.'

tion as political subversion through electoral interference, this chapter adopts a broader conceptual framework (2.2.3), illustrating the subversive potential of scientific, economic and cultural disinformation.

2.2.1 From Propaganda to Disinformation

Early discussions on propaganda and the international order implicitly treated propaganda *per se* as a form of subversive communication.[19] Within a legal system that prioritises sovereign States and stable inter-State relations, efforts to influence foreign populations towards 'insurrection, revolt, or civil strife' faced swift and widely supported condemnation. As discussed elsewhere, the emergence of inter-State broadcasting capabilities transformed propaganda's nature by enabling direct and large-scale communication with foreign audiences 'over the heads of their government.'[20] These developments rendered bilateral friendship treaties and unilateral denouncements insufficient to contain propaganda's cross-border effects, leading to the first peacetime regulatory instrument: the 1936 International Convention concerning the Use of Broadcasting in the Cause of Peace.[21] Post-World War II, multiple international legal instruments emerged, as discussed in chapter one. Conventional reference to the falsity of communications was first made in the 1952 Convention on the International Right of Correction, which recommended 'the adoption of measures designed to combat the dissemination of *false or distorted* reports likely to injure friendly relations between States [emphasis added].'[22]

Unlike other categories of disinformation examined in next chapters – defamatory, terrorist and discriminatory – subversive disinformation is not defined by specific thematic narratives or tailored aims. Its defining purpose is disrupting, undermining, or obstructing a State's freedom to determine its internal and external affairs. Malicious actors – States or otherwise – pursue these objectives through virtually unlimited combinations of different types of disinformation. They defame public figures, arouse support for subversive or terrorist organisations and spread discriminatory narratives to fuel ethnic hatred. This overlap, or convergence, between different types of disinformation under the denominator of subversive disinformation is, however, not a new development.

Regarding subversive propaganda, Resolution 2131 on the 'Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of Their

19 Martin (n 3) 5-9.

20 *Ibid*. 2, 7.

21 International Convention Concerning the Use of Broadcasting in the Cause of Peace (adopted 23 September 1936, entered into force 2 April 1938) A/RES/841 [hereafter: Broadcasting Convention].

22 Convention on the International Rights of Correction (adopted 31 March 1953, entered into force 24 August 1961) 191 UNTS 435, preamble.

Independence and Sovereignty' (1965), listed terrorist activities (including propaganda) as prohibited subversive behaviour. Similarly, the 1976 Declaration on Non-Interference, and the 1981 Declaration on the Inadmissibility of Intervention and Interference, identified defamatory propaganda as a form of prohibited subversive communication, including 'campaigns of vilification and intimidation', instances of 'subversion and defamation', and 'any defamatory campaign, vilification or hostile propaganda.'[23] A deeper analysis of these types of disinformation – examining their content, structure, *modus operandi*, and convergence with existing categories of unlawful speech– follows in chapters three, four and five.

Subversive disinformation is, however, more than an accumulation of these specified types of harmful propaganda. Anticipating the legal analysis (sections 2.3 and 2.4.), the present analytical framework upholds that subversive disinformation – as false or misleading information intended to cause harm – is conceptually premised on three distinguishing elements. First, subversive disinformation is characterised by its strategic objectives to destabilise essential State structures or to influence State policy issues.[24] Unlike incidental misinformation that may have far-reaching effects, disinformation comprises a deliberate, coordinated nature and strategic objectives.

Second, achieving these objectives often relies on a dual-track approach targeting two distinct, but interconnected, audiences.[25] At the State level, it targets decision-makers, government leaders and opinion-makers, directly aiming to compromise decision-making processes and provoke envisaged policy responses.[26] Simultaneously, it targets public opinion through mass dissemination strategies across traditional and social media platforms, focusing on shaping public narratives and polarising societal groups. This aspect creates conditions for societal unrest through systematic erosion of institutional trust, aligning with the traditional interpretation of subversive propaganda – characterised by the aim of 'destabilizing State institutions by influencing nationals of another State towards insurrection, revolt, or civil strife'.[27] The

23 UNGA, 'Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of Their Independence and Sovereignty' (adopted 21 December 1965) UNGA Res 2131 (XX); Organisation of African Unity, 'Declaration on Non-Interference in the Internal Affairs of States' (21 July 1976) CM/Res 444 (XXV); UNGA Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States (adopted 9 December 1981) UNGA Res 36/103.

24 Sections 2.3.2.1. 'Territorial Integrity'; 2.3.2.2 'Inherently Sovereign Functions'; 2.4.1 'Internal and External Affairs'.

25 This dual-track strategy, first identified by the U.S. Central Intelligence Agency (CIA) during its analysis of Cold War-era KGB disinformation operations, remains relevant today, in John L Martin, 'Disinformation: An Instrumentality in the Propaganda Arsenal' (1982) 2 Political Communication 1, 47-64, 52.

26 Stephan Lewandowsky *et al.*, 'Misinformation and the Epistemic Integrity of Democracy' (2023) 54 Current Opinion in Psychology 101711, 1-3.

27 De Brabandere (n 2) para 12.

contemporary difference lies in the exploitation of technological tools and online information ecosystems, and the employment of unprecedented persuasion and manipulation techniques to achieve the strategic objectives more gradually yet pervasively – as outlined in chapter one.[28] This context enables foreign States to 'feed information and disinformation into the political debate in order to get the target population or political class to move themselves to a policy position that aligns with the interests of the outside power.'[29]

A third distinguishing feature of subversive disinformation is the employment of attribution manipulation, i.e. falsely attributing the creation or dissemination of false and misleading information to another State, through forging official documents and impersonation of State agencies on seemingly authoritative channels. On one hand, this constitutes a particularly effective strategy to cause internal unrest when it concerns sensitive topics, such as immigration or minorities. On the other, when the falsity of the information is intentionally or unintentionally exposed, but the attribution manipulation upholds, this may effectively erode public trust in government communication.[30] Here lies one of the greatest long-term subversive risks of disinformation; not in tangible outcomes or the falsity of the information, but in creating a permanent state of distrust and scepticism towards State information and communication.[31] Thus while subversive propaganda and disinformation share fundamental characteristics in their aim to undermine State sovereignty, modern subversive disinformation represents a more sophisticated evolution through its dual-track targeting, information manipulation strategies, and exploitation of digital communication systems.

28 Section 1.3 'Societal Context of Disinformation'.

29 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 161-197, 193; this is often referred to as 'reflexive control' ('as a means of conveying to a partner or an opponent specially prepared information to incline him to voluntarily make the predetermined decision desired by the initiator of the action' in Corneliu Bjola and James Pamment, *Countering Online Propaganda and Extremism: The Dark Side of Digital Diplomacy*' (Routledge 2018) 14); Annie Kowaleski, 'Disinformation and Reflexive Control: The New Cold War' (1 February 2017) Georgetown Security Studies Review, Accessed 21 November 2024; See also Sienho Yee, 'Public Opinion' (2013) Planck Encyclopaedia of Public International Law, paras 5, 8, 15-16.

30 Martin (n 25) 47-64, 50.

31 As noted by the WEF, '[g]iven distrust in the government and media as sources of false information, manipulated content may not be needed – merely raising a question as to whether it has been fabricated may be sufficient to achieve relevant objectives' in the World Economic Forum, *Global Risks Report 2024* (10 January 2024) Insight Report, 20.

2.2.2 Beyond Election Interference

While legal discourse on subversive disinformation focuses primarily on electoral interference, the phenomenon spans across multiple State domains, each challenging sovereign authority and stability. Most prominently, the COVID-19 pandemic, and its accompanying 'infodemic' demonstrated how health-related disinformation generates subversive outcomes. Both foreign and domestic actors undermined State pandemic management by eroding trust in public health measures, inciting panic, and ultimately provoking violent resistance against health institutions, politicians, and medical experts.[32] While this involved a plethora of actors, State involvement by Russia, the United States, China, and Iran, is broadly evidenced.[33]

Less recognised, but equally disruptive, is economic disinformation, wherein strategic narratives concerning currency devaluation, economic setbacks, financial instability, insolvency risks, or sector-specific corruption directly impact a State's autonomy in economic policy-making and implementation.[34] Environmental and scientific disinformation present additional challenges; false narratives about climate change undermine both domestic environmental policies and international cooperation by discrediting environmental agreements (including the Paris Agreement) and portraying them as instruments of foreign control.[35] The past decade revealed States increasingly weaponise

32 Maria Mercedes Ferreira Caceres *et al.*, 'The impact of misinformation on the COVID-19 pandemic' (2022) 9 AIMS Public Health 2, 262-277; Luiza Bandeira *et al.*, 'Weaponized: How rumors about COVID-19's origins led to a narrative arms race' (February 2021) Atlantic Council DFRLab Report 15- 46; Sean Quirk, 'Lawfare in the Disinformation Age: Chinese Interference in Taiwan's 2020 Elections' (2021) 62 Harvard International Law Journal 2, 526-567, 556.

33 Luiza Bandeira *et al.*, 'Weaponized: How rumors about COVID-19's origins led to a narrative arms race' (February 2021) Atlantic Council DFRLab Report, 15- 46.

34 Austria expressly qualified these forms of disinformation as prohibited intervention, in Republic of Austria, 'Position Paper of the Republic of Austria: Cyber Activities and International Law' (April 2024) 6 ('[e]xample: State A launches a large-scale campaign against the government of state B that spreads disinformation about the alleged corrupt business practices of state B's government, intended to sow distrust within the population of state B. The campaign eventually causes the government of state B to resign, resulting in a governmental crisis. Such a cyber activity would violate the prohibition of intervention'); Tiziana Assenza *et al.*, 'From Buzz to Bust: How Fake News Shapes the Business Cycle' (March 2024) ECONtribute Discussion Paper No. 287, 1-48 9 ('[…] we show that technology fake news increases macroeconomic uncertainty, exacerbates unemployment, and depresses industrial production').

35 A striking example is the Kremlin-backed dissemination of climate disinformation aimed at derailing climate change mitigation policies and renewable energy investment, by framing global warming as 'a "hoax" and emission-reduction plans as a form of "Western imperialism" engineered to undermine the development of emerging economies', in NATO, 'NATO Climate Change and Security Impact Assessment' (2024) Secretary General's Report, 27-28 Stephan Lewandowsky, 'Climate Change Disinformation and How to Combat It' (2021) 42 Annual Review of Public Health 1-21; Tiffany Hsu and Steven Lee Myers, 'Disinformation

such disinformation to disrupt democratic processes abroad by fostering policy scepticism and governmental distrust.[36] Similarly, social and cultural disinformation exploits existing divisions to weaken internal stability. As examined in chapters three through five, targeted amplification of sensitive issues – immigration, terrorism, or historical revisionism – fuels social unrest and intercommunal tensions.[37] Empirical manifestations of these phenomena include the instrumentalisation of disinformation during the Ukrainian refugee crisis by both Belarus and Poland, and the Hungarian government's strategic deployment of disinformation to attribute increased migration pressure to Romania.[38] While illustrating singular events, these cases represent pervasive patterns.

The spectrum of potentially subversive disinformation extends to legal and judicial disinformation – undermining trust in legal systems through false narratives about court rulings or judicial integrity,[39] including impartiality. Religious or ideological disinformation that manipulates beliefs or fabricates claims of persecution, may likewise culminate in hostility, sectarian violence and a general deepening of divisions in society.[40] Even corporate and commercial disinformation targeting States, when they operate as commercial actors, can constitute forms of subversion by destabilising economic relationships and market confidence. These illustrations only scratch the surface but already indicate that subversive risks of disinformation extend beyond electoral

---

Is One of Climate Summit's Biggest Challenges' The New York Times (30 November 2023) Accessed 28 November 2024.

36 For a comprehensive introduction, see Naomi Orestes and Erik M Conway, *Merchants of Doubt* (Bloomsbury Publishing 2012); Lewandowsky notes that these forms of disinformation arguably undermine democracy in 'much the same way' as electoral and political disinformation, 'albeit in a more indirect manner', in Stephan Lewandowsky *et al.*, 'Misinformation and the Epistemic Integrity of Democracy' (2023) 54 Current Opinion in Psychology 101711, See also Elena Yi-Ching Ho, 'Climate Disinformation Is Compromising Taiwan's Efforts in Defending Democracy' (31 October 2024) The Diplomatic, Accessed 28 November 2024; Andre Heffernan, 'Countering Fossil-Fuelled Climate Disinformation to Save Democracy' (2024) Centre for International Governance Innovation, Digital Policy Hub – Working Paper, 1-14, Accessed 28 November 2024; Jackson Bellamy, 'Climate Change Disinformation and Polarization in Canadian Society' (18 December 2020) Policy Primer North America and Arctic Defense and Security Network, 1-17.

37 WEF (n 31) 18-19.

38 European Digital Media Observatory, 'Ukrainian Refugees and Disinformation: Situation in Poland, Hungary, Slovakia and Romania' (5 April 2022) European Digital Media Observatory Blog.

39 The US National Center for State Courts have identified common disinformation themes attacking judicial system, including narratives that a 'justice system tolerates, protects and covers up crimes committed by immigrants' or that it 'system tips the electoral map in favour of a particular party', in NCSC, 'Combatting Disinformation: A Playbook Template for State Court' (March 2022) Accessed 28 November 2024.

40 How disinformation on religious matters becomes subversive is most clearly evidenced by the instrumentalisation of disinformation by terrorist organisations as IS and Al-Qaida, in, *inter alia*, Jamil Ammar, 'Disinformation: The Jihadists' New Religion' in Rubén Arcos *et al.*, *Routledge Handbook of Disinformation and National Security* (Routledge 2023) 111-118; see also chapter four.

interference. Their combined deployment presents the greatest information-driven risk to State stability and sovereignty.

2.3 State Sovereignty

Sovereignty as a principle of international law establishes both reciprocal rights and obligations in itself and gives rise to corollaries, such as non-intervention, that both derive from and protect sovereign authority.[41] The first proposition of State sovereignty as a primary rule of international law has, however, endured considerable scepticism. While acknowledging these perpetual concerns, the first part of this section demonstrates the validity of this characterisation (2.3.1). Consequently, this positions subversive disinformation that breaches the territorial integrity of a State or interferes with its inherently sovereign functions, as a violation of State sovereignty and thus an internationally wrongful act (2.3.2). In practice, however, this is not unequivocally accepted. States and academic discourse fear that this recognition leads to overinclusiveness of the rule at the expense of its credibility. To bridge this gap, the section explores existing proposals to include a threshold for breaching sovereignty and introduces 'intent' as an indicator of unlawful interference.

2.3.1 Principle or Primary Rule

The status of sovereignty as a stand-alone substantive rule or as a 'mere' principle of international law has long divided experts. While scepticism persists,[42] its status as a substantive rule, including in the digital domain, has repeatedly been affirmed.[43] The ICJ's predecessor in the Lotus Case established already in 1927 that absent permissive rules, a State 'may not exercise

41 Robert Jennings and Arthur Watts (eds), *Oppenheim's International Law*, Vol. 1: Peace (9th edn, Longman 1996) 428.

42 Gary Corn and Robert Tayler, 'Sovereignty in the Age of Cyber' (2018) 111 American Journal of International Law Unbound 207; Chimène Keitner, 'Foreign Election Interference and International Law' in Duncan B Hollis and Jens D Ohlin (eds), *Defending Democracies* (Oxford University Press 2021) 179, 191.

43 Przemysław Roguski, 'Violations of Territorial Sovereignty in Cyberspace – An Intrusion-Based Approach' in Dennis Broeders and Bibi van den Berg (eds), *Governing Cyberspace: Behavior, Power and Diplomacy* (Rowman & Littlefield 2018) 73; Michael N Schmitt (ed) and NATO Cooperative Cyber Defence Centre of Excellence, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd edn, Cambridge University Press 2017) [hereafter: Tallinn Manual 2.0] Rule 4; see also the consensus report by the UN Group of Governmental in the Field of Information and Telecommunication, 'State sovereignty and international norms and principles that flow from sovereignty apply to the conduct by States of ICT-related activities and to their jurisdiction over ICT infrastructure within their territory' (July 22 2015) UN Doc A/70/174, 15.

its power in any form in the territory of another State', which it interpreted to extend beyond territorial integrity and include political independence.[44] Subsequent international jurisprudence affirmed this position, including the 1928 *Island of Palmas Arbitral Award*, the 1949 *Corfu Channel Case*, the 1974 *Nuclear Test* Case, the 1986 landmark case *Concerning Military and Paramilitary Activities in and Against Nicaragua*, the 1988 *Nicaragua-Honduras Border Case*, the 1990 *Rainbow Warrior Arbitration Case,* and the 2015 Case on *Certain Activities Carried out By Nicaragua.*[45]

The obligation to respect State sovereignty is now firmly established in both conventional and customary international law.[46] While sovereignty is often understood through its derivative prohibitions – notably those against the threat or use of force and intervention – its legal significance extends beyond these corollaries. As a self-standing right, it presents a framework for analysing the lawfulness of influence operations. This approach offers advantages over relying solely on the non-intervention principle, which may be limited in effectiveness due to narrow interpretations of its 'coercion' requirement. It, however, is not only advantageous, but also warranted by the internal cohesion between State sovereignty and the prohibition of intervention. Analysing influence operations exclusively through the lens of the latter prohibition creates a paradox: operations with potentially greater interfering effects than overtly coercive acts may escape legal scrutiny simply because they fall below the coercion threshold.[47]

Explicit rejection or hesitation towards recognising sovereignty as a primary rule often stems from concerns about the unsettled question on the existence,

---

44 *Case of the SS 'Lotus' (France v Turkey)* [1927] PCIJ Rep Series A No 10, 18; in the *Nicaragua Case,* '[t]he ICJ determined that US over flights of Nicaragua which did not constitute uses of force or intervention violated Nicaragua's sovereignty […]' (*Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 5) para 88, 251); *Island of Palmas Case (Netherlands v USA)* (1928) 2 RIAA 829, 838.

45 *Corfu Channel Case (UK v Albania)* (Merits) [1949] ICJ Rep 4, 35-36; *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 5) para 251; *Certain Activities Carried Out by Nicaragua in the Border Area* (Costa Rica v Nicaragua) and Construction of a Road in Costa Rica along the San Juan River (Nicaragua v Costa Rica) [2015] ICJ Rep [229]; *Border and Transborder Armed Actions* (Nicaragua v Honduras) [1988] ICJ Rep 90 [50]-[57]; *Rainbow Warrior* (New Zealand v France) (1990) 20 RIAA 215, 20.

46 UNGA, 'Declaration on Principles of International Law concerning Friendly Relations and Co-operation among States in accordance with the Charter of the United Nations' (24 October 1970) UNGA Res 2625 (XXV); UNGA, 'Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of Their Independence and Sovereignty' (21 December 1965) UNGA Res 2131 (XX) [5]; UNGA Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States (9 December 1981) UNGA Res 36/103 [1]-[2], [2(c)]; Pia Hüsch, 'Non-Intervention Thresholds in Cyberspace – In the Shadow of the Sovereignty Debate?' (2023) 92 Nordic Journal of International Law 3, 371-380.

47 Schmitt (n 16) 67; Barrie Sander, 'Democracy Under The Influence: Paradigms of State Responsibility for Cyber Influence Operations' (2019) 18 Chinese Journal of International Law 1-56, 20.

or demarcation, of a threshold for violating sovereignty. It is a valid concern that acknowledging State sovereignty as a primary rule without a *de minimis* threshold for breaching it, results in overinclusiveness.[48] However, this concern supports developing an appropriate threshold through qualitative and quantitative criteria (2.3.2) rather than rejecting the norm entirely. This position, balancing between *lex lata* and *lex ferenda*, increasingly gains support from experts and States.[49] Alternatively, as convincingly argued by Lahmann, even if sovereignty itself were not recognised as an independent rule, its substantive meaning as a principle would give rise to other specific primary rules – including rights to territorial integrity and political independence.[50] Thus, while the distinction between sovereignty as a principle or primary rule is legally relevant, it is not ultimately decisive.

48 Henning Lahmann, 'On the Politics and Ideologies of the Sovereignty Discourse in Cyberspace' (2021) 32 Duke Journal of Comparative & International Law 61, 90-93.

49 The experts drafting the authoritative Tallinn Manual concluded that 'States shoulder an obligation to respect the sovereignty of other States as a matter of international law', in Tallinn Manual 2.0 (n 43) Rule 4, Commentary 2; Michael N Schmitt, 'Virtual Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law' (2018) 19 Chicago Journal of International Law 30, 40; the notable exception here is the United Kingdom that noted in the United Kingdom Policy Paper on the Application of International Law to States' Conduct in Cyberspace that 'sovereignty, as a general principle, is a fundamental concept in international law. […] The United Kingdom does not consider that the general concept of sovereignty by itself provides a sufficient or clear basis for extrapolating a specific rule or additional prohibition for cyber conduct going beyond that of non-intervention referred to above. […]'; Schmitt and Vihul have analysed the various approaches of international tribunals, States, IOs and academics towards this topic in Michael N Schmitt & Liis Vihul, 'Respect for Sovereignty in Cyberspace' (2017) 95 Texas Law Review 1639; See also Steven Wheatley, 'Election Hacking, The Rule of Sovereignty, and Deductive Reasoning in Customary International Law' (2023) 36 Leiden Journal of International Law, 675-698, 690; Przemysław Roguski, 'Application of International Law to Cyber Operations: A Comparative Analysis of States' Views' (2020) Hague Program for Cyber Norms Policy Brief; Benedict Pirker, 'Territorial Sovereignty and the Challenges of Cyberspace' (2013) in Katharina Ziolkowski (ed), *The Peacetime Regime for State Activities in Cyberspace – International Law, Foreign Affairs and Cyber-Diplomacy* (Tallinn 2013); Sean Watts and T Richard, 'Baseline Territorial Sovereignty and Cyberspace' (2018) 22 Lewis and Clark Law Review 771; Kevin J Keller, 'In Defense of Pure Sovereignty in Cyberspace' (2021) 97 International Law Studies 1432; Henning Lahmann, 'Infecting the Mind: Establishing Responsibility for Transboundary Disinformation' (2022) 33 European Journal of International Law 411; Nicholas Tsagourias, 'The legal status of cyberspace: sovereignty redux?' in Nicholas Tsagourias and Russel Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar 2021) 22; Peter BMJ Pijpers, *Influence Operations in Cyberspace and the Applicability of International Law* (Edward Elgar 2023) 165, 172-175; Ali Strongwater, 'Combatting Disinformation Through International Law' (2023) 55 International Law and Politics 33, 35-38; Björnstern Baade, 'Fake News and International Law' (2018) 29 European Journal of International Law 4; Hitoshi Nasu, 'The Infodemic': Is International Law Ready to Combat Fake News in the Age of Information Disorder?' (2021) 29 Australian Yearbook of International Law 1; Duncan B Hollis, 'Why States Need an International Law for Cyber Operation' (2007) 11 Lewis Clark Law Review 1023.

50 Lahmann (n 48) 90-103.

2.3.2 A Violation of State Sovereignty

The international regulation of subversive information throughout the 20th century was driven by the recognition that it constituted a risk to a State's territorial integrity and/or independence. Its unlawfulness under international law gained widespread and explicit recognition following the 1949 UNGA Resolution 'Essentials of Peace', which urged States to refrain from 'any threat or act, direct or indirect, aimed at impairing the freedom, independence, or integrity of any other State, or at fomenting civil strife and subverting the will of the people in any State'.[51] In the years that followed, the condemnation of subversive propaganda under the banner of upholding and protecting State sovereignty remained the focal point in the discourse on information regulation.[52]

While the emergence of cyberspace and development of new technologies only increased the possibilities of instrumentalising subversive information, the legal implications for State sovereignty have been framed hesitantly. The Tallinn Manual experts, *inter alia*, adopted a sceptical stance, arguing that propaganda – without commenting on its meaning – 'is generally not a violation of sovereignty.'[53] Yet they acknowledged that 'the transmission of propaganda, depending on its nature, might violate other rules of international law' and that '[…] propaganda designed to incite civil unrest in another State would likely violate the prohibition of intervention.'[54] The leading author of the Manual, Michael Schmitt, considered this rejection firmly supported by extensive State practice of States engaging in 'both truthful and untruthful propaganda during foreign elections.'[55] Another author opines that approaching propaganda and disinformation through the lens of State sovereignty is 'misplaced',[56] and that 'its doctrinal requirements are a poor fit' for evaluating specific types of disinformation, including in the context of elections.[57]

Recalling the differentiation between propaganda and disinformation, this section demonstrates how State sovereignty's two components – territorial integrity (2.3.2.1) and inherently sovereign functions (2.3.2.2) have changed over time, establishing non-interference obligations applicable to disinforma-

51 UNGA, 'Essentials of Peace' (1 December 1949) UN Doc A/RES/290, para 3.

52 Section 1.2.1 'Historical Developments of Information Regulation'.

53 They acknowledge that 'numerous instruments stipulate that States in which other States' broadcasts, especially satellite broadcasts, are available should have some degree of control over the information that is being transmitted into their territories,' yet 'subsequently agreed that this premise has not crystallised into a customary international law requirement', in Tallinn Manual 2.0 (n 43) Rule 4 Commentary, para 29, footnote 28.

54 *Ibid.*; Michael Kearney, *The Prohibition of Propaganda for War in International Law* (Oxford University Press 2007) 75-77.

55 Schmitt (n 49) 46.

56 Patrick CR Terry, 'Voting by Proxy – Meddling in Foreign Elections and Public International Law' (2022) 239 Indiana Journal of Global Legal Studies 2, 67-115, 75.

57 Jens D Ohlin, *Election Interference* (Cambridge University Press 2020) 75.

tion. It recognises that conceptual and doctrinal uncertainties as well as persistent interpretative disagreements among States and experts impede a conclusive analysis.[58] To break the impasse and respond to the growing threat of subversive disinformation, section 2.3.2.3 supports proposed thresholds to differentiate between types of disinformation that do and do not violate State sovereignty. Given disinformation's limited direct impact on territorial integrity, the focus lies on interference with inherently sovereign functions, proposing three evaluative factors:

1. The type of disinformation and State function involved
2. The target audience
3. Timing and scale of the operation

The analysis concludes by examining how disinformation's intrinsically malicious intent, while not required to establish a breach of State sovereignty, may nevertheless be indicative of unlawful interference (2.3.2.4).

2.3.2.1 Territorial Integrity

The territorial dimension of State sovereignty entails exclusive authority and power over its territory, and inviolability of and respect for said territory, extending to its territorial sea,[59] airspace and parts of cyberspace.[60] The orthodox understanding equates violations of sovereignty with tangible, physical incursions into a State's territory. The digital revolution and vastly expanding array of cyber tools enabling States to remotely compromise another State's territorial integrity, however, generate consensus that this position is untenable.

Non-digital forms of disinformation nevertheless continue to raise pertinent questions regarding false or misleading information and territorial sovereignty.[61] Analogue disinformation campaigns serve a strategic utility when targeting populations with limited Internet access. They also support hybrid models for disinformation operations, whereby reinforcing online disinformation through offline messaging – platforms, radio broadcasts etc. – increases the narrative's credibility. A clear illustration of the persisting relevance of this analogue dimension is North Korea's accusation that South Korea violated its territorial sovereignty after dropping anti-regime propaganda pamphlets

58 For instance, some experts opine that only physical incursions into another State's territory constitute a violation of its sovereignty, whereas other consider any unauthorised State-attributable interference a violation. Throughout this spectrum, the reoccurring question pertains to the inclusion of a *de minimis* threshold for sovereignty violations.

59 *North Sea Continental Shelf Cases* (Germany v Denmark; Germany v Netherlands) [1969] ICJ Rep 3 [59].

60 Section 1.4.3 '(Dis)information Sovereignty and Jurisdiction'.

61 Martin (n 30) 54 citing Victor Marchetti and John D Marks, *The CIA and the Cult of Intelligence* (Alfred A Knopf New York 1974) 157.

within its territory and installing speakers along the border to broadcast propaganda, which North Korea labelled 'an act of war.'[62]

The territorial implications of online subversive disinformation have remained peripheral to broader debates on the applicability of international law in cyberspace. This discourse does, nevertheless, yield some valuable insights. First, the authoritative Tallin Manual adopts the view that any (cyber)operation against a State or person/entity located in that State while physically present on its territory violates the State's sovereignty.[63] This encompasses, *inter alia*, subversive disinformation campaigns carried out or financed by foreign agents while present on the territory of the target State.[64] Accusations of such violations of State sovereignty have surfaced following involvement of Russian and Chinese diplomatic personnel in disinformation operations that the host State considered incompatible with its national interests,[65] security and other sovereign prerogatives.[66]

Second, most contemporary disinformation operations operate in a deterritorialised manner, executed remotely and/or anonymously. To capture this reality, several interpretative approaches of the existing law emerged. A restrictive approach contends that 'only cyber operations that cause a permanent loss of functionality to cyberinfrastructure, or that results in physical damage or injury are to be considered violations of territorial sovereignty.'[67] While broadly supported by scholars and State practice until roughly five years ago,[68] Chircop rightfully criticises this view as particularly conservative and untenable as it fails to adapt to new forms of interference that have equally – or even more – harmful effects.[69] Conversely, the 'strict inviolability approach' represents the opposite end of the spectrum. Though supported by limited State practice and a growing – yet still limited – segment of aca-

62 Shaimaa Khalil and Thomas Mackintosh, 'South Korea to resume loudspeaker broadcasts over border in balloon row' BBC News (6 June 2024) Accessed 16 October 2024; Hyung-Jin Kim and Kim Tong-Hyung, 'Bizarre psychological warfare using K-pop and trash balloons raises tensions between the 2 Koreas' AP News (11 June 2024) Accessed 16 October 2024; Kim Tong-Hyung, 'South Korea says North Korea is installing its own loudspeakers along the border' AP News (10 June 2024) Accessed 2 December 2024.

63 Tallinn Manual 2.0 (n 43) Rule 4 Commentary 8.

64 Terry (n 56) 106.

65 Africa Centre for Strategic Studies, 'Mapping a Surge of Disinformation in Africa' (13 March 2024) Infographic, Accessed 13 November 2024.

66 Corneliu Bjola and Ilan Manor, 'The Use and Abuse of History by Russian Embassies on Twitter' in Rubén Arcos *et al.* (eds), *Routledge Handbook on Disinformation and National Security* (Routledge 2023) 148-159; Julie Tomiche, 'France Kicks out Russian spies working 'under diplomatic cover' Politico EU (11 April 2022) Accessed 22 October 2024; France24, 'France summons Chinese ambassador over 'unacceptable' tweets' France24 (23 March 2021) Accessed 22 October 2024.

67 Luke Chircop, 'Territorial Sovereignty in Cyberspace After Tallinn Manual 2.0' (2020) 21 Melbourne Journal of International Law 2, 12.

68 Tallinn Manual 2.0 (n 43) Rule 4 Commentary 11-13.

69 Chircop (n 67) 12.

demic discourse,[70] this interpretation posits that all cyber activities exploiting, or conducted within, the infrastructure of another State violate its territorial integrity, even if no tangible harm occurs.[71] Buchan articulates the underlying rationale of this progressive conceptualisation, observing that 'States exert sovereignty over information in cyberspace which belongs to entities and individuals over which they exercise jurisdiction.'[72]

While disinformation often accompanies cyberoperations that breach State sovereignty through this paradigm, chapter one's analyses demonstrates that disinformation – causing indirect, cognitive harm through behavioural change – follows a different logic than cyberoperations.[73] In purely theoretical terms, a different, data-centred conceptualisation of (dis)information could potentially position it as an intrusion into a State's territory because data relies on physical network layers for its dissemination.[74] This perspective would recognise that digital communications require tangible infrastructure to be disseminated, which is located within sovereign territories. Realistically, however, this interpretation is untenable. Global digital infrastructure makes the distinction between disinformation and legitimate digital communications impossible to sustain.[75] Introducing a regulatory dimension that approaches disinformation as a form of unlawful data manipulation, however, may become increasingly plausible with the widespread deployment of AI-generated content. These technologies qualitatively and quantitively degrade data credibility without necessarily compromising confidentiality or availability.[76] Such reconceptualisation, however, follows an infrastructural logic with a focus on technological

70 Russell Buchan, *Cyber Espionage and International Law* (Bloomsbury Press 2018) 51; Wheatley (n 49) 694 also highlight that France and China, and a minority of the Tallinn Manual Expert 'adopt this catch-all position'.

71 Tsagourias (n 49) 22; see also Francois Delerue, *Cyber Operations and International Law* (Cambridge University Press 2020) 226; Sean Watts, 'Low-Intensity Cyber Operations and the Principle of Non-Intervention' (2015) 14 Baltic Yearbook of International Law 137, 142.

72 Russell Buchan, 'Cyber Espionage and International Law' in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing, 2015) 184.

73 Lahmann (n 49) 426; to this end, Hollis observed that 'there is little precedent for treating purely cognitive effects to which [influence operations] ultimately aspire as breaching sovereignty', in Duncan Hollis, 'The Influence of War: The War of Influence' (2018) 32 Temple Journal of International & Comparative Law 36, 42.

74 Chircop (n 67) 12.

75 Roguski (n 43) 79; Pijpers (n 49) 181.

76 Tomlinson and others explore this under the notion of 'data defamation' as 'the deliberate introduction of false or defamatory information into a training set for the purpose of causing that AI to produce false or defamatory outputs', in Bill Tomlinson *et al.*, 'Turning Fake Data into Fake News: The AI Training Set as a Trojan Horse of Misinformation' (2023) 60 San Diego Law Review 741, 664; Positioning disinformation as unlawful data manipulation builds on Tallinn Manual's acknowledgment that 'altering or deleting data stored on cyber infrastructure without causing physical or functional consequences' violates the territorial dimension of State sovereignty; Chircop (n 67) 13; Tallinn Manual 2.0 (n 43) Rule 4, Commentary 14.

information creation instead of the prevailing epistemological approach to disinformation, hence it is not explored further here.

The territorial implications of subversive disinformation become more distinct when combined with unlawful cyberactivities such as unauthorised access to a State's infrastructure (i.e. hacking).[77] Such activities are increasingly employed in the preparation of disinformation and other influence operations, as evidenced during both the 2016 US and 2017 French Presidential elections, and more recently in Romania, Moldova and Georgia in 2024.[78] This operational convergence creates a pattern: the more effective disinformation campaigns and operations become, the greater the incentive for malicious actors and States to resort to cyber activities to enable, facilitate and amplify these forms of subversive interference. While this does not position disinformation itself as a violation of a State's territorial integrity, it indicates a connectedness that warrants structural inclusion of disinformation in frameworks governing cyberspace.

More significantly, disinformation serves as an enabler of cyber activities that unambiguously violate territorial integrity. State-coordinated disinformation campaigns may, for example, be instrumentalised to deceive or manipulate individuals into downloading malware or to disclose sensitive information, indirectly compromising a State's cyber infrastructure and integrity. Disinformation also directly or indirectly triggers physical consequences on a State's territory. As Lahmann illustrates:

> 'if a false or misleading piece of information induced citizens of the target state to ingest a supposedly remedial, but in fact harmful, substance that results in severe illness or even death, the right to territorial inviolability would be breached.'[79]

Disinformation's territorial implications through societal harm may extend to, *inter alia*, gendered disinformation catalysing domestic violence, false information on disposal of waste leading to environmental harm, and disinformation containing propaganda for war precipitating armed conflict.[80]

---

77 Wheatley (n 49) 675-698.

78 Pijpers (n 49) 218; Delerue (n 71) 241-242; Sarah Rainsford, 'Romania hit by major election influence campaign and Russian cyber-attacks' BBC (5 December 2024) Accessed 5 December 2024; Nicholas Chkhaidze, 'Russia emerges as the real winner of Georgia's 'disputed election (12 November 2024) Atlantic Council, Accessed 6 December 2024.

79 Lahmann (n 49) 415.

80 The instrumental use of disinformation as a pretext for the use of force is evidenced in numerous international conflicts, from the US invasion in Iraq following unsubstantiated allegation of Weapons of Mass Destruction (WMD) to Russia's invasion in Ukraine justified by claims of protecting Russian minorities for purported genocide. Another striking illustration can be found in the build-up to the Third Arab-Israeli War, in Jingjin Huang, 'Information Warfare in the Digital Age: Legal Responses to the Spread of False Information under Public International Law' (2024) 28 Journal of Education, Humanities and Social Sciences 1, 179; European External Action Service, 'Study of International Norms for Foreign Informa-

Categorising these territorial effects as disinformation-induced violations of State sovereignty depends on establishing a linkage between the speech and subsequent behaviour. While it is unconvincing to position disinformation as the sole cause of these harms, international law recognises that parallel, complementary or cumulative causal relationship may still lead to a finding of internationally wrongful behaviour.[81]

In sum, while disinformation's enabling and combinatory effects – alongside cyberoperations – raise legally significant concerns for territorial integrity, theoretical constructions of disinformation *per se* as violating State sovereignty's territorial dimension encounter theoretical and practical barriers. The need to prove the nexus between territorial intrusion – or at least *de minimis* effects – requires the kind of factual causation that has proven exceptionally difficult to establish for disinformation.[82]

2.3.2.2 Inherently Sovereign Functions

The a-territorial dimension of sovereignty is encapsulated by a State's inherently governmental or sovereign functions.[83] It represents the exclusive domain of State authority,[84] where usurpation or inference by another State constitutes a breach of sovereignty. Such violations do not necessitate any physical effect to manifest on the State's territory: a violation, Milanovic and Schmitt conclude comprises 'the existence of activities that states alone are entitled to perform.'[85]

While central to State authority, 'inherently sovereign functions' lack precise definition in international law.[86] They are often portrayed as a normative

---

tion Manipulation and Interference (FIMI)' (November 2023) 9, Accessed 16 October 2024.

81 Vladyslav Lanovoy, 'Causation in the Law of State Responsibility' (2022) British Yearbook of International Law 1, 65-69; International Law Commission, 'Report of the International Law Commission on the Work of its 45th Session' (1993) UN Doc A/48/10, 70, Commentary to Former Draft Article 8, para 13.

82 Section 6.2.3 'Causality'.

83 The notion 'inherently *governmental* functions' derives from the domestic context, whereas the *sovereign* functions align better with international law, in Frédéric Mégret, 'Are there "Inherently Sovereign Functions" in International Law' (2021) 115 American Journal of International Law 452, 454; see also Samantha Besson, 'The International Public: A Farewell to Functions in International Law' (2021) 115 American Journal for International Law Unbound 307, 308.

84 UNGA Res 375(IV) 'Draft Declaration on Rights and Duties of States' (6 December 1949) UN Doc A/RES/375, Article 1: 'Every State has the right to independence and hence to exercise freely, without dictation by any other State, all its legal powers, including the choice of its own form of government' and UN Charter (n 9) Article 2(7) ('[n]othing contained in the present Charter shall authorize the United Nations to intervene in matters which are essentially within the domestic jurisdiction of any state').

85 Marko Milanovic and Michael N Schmitt, 'Cyber Attacks and Cyber (Mis)information Operations during a Pandemic' (2020) 11 Journal of National Security Law & Policy 247, 255.

86 Moynihan (n 12) 15.

construct tied to the legal conception of State sovereignty.[87] As Mégret denotes, 'there are no ISFs (Inherent Sovereign Functions) in the absolute'.[88] The content and scope of these functions have been variously associated with the concept of *acta jure imperii*, deriving from the context of State immunity, or domestic classifications, shaped by diverging historical and cultural factors.[89] This notwithstanding, commonly cited functions include the conduct of elections, crisis management, tax collection, diplomacy, and law enforcement.[90] The ICJ's wording in the Nicaragua case, 'the choice of a political, economic, social and cultural system, and the formulation of foreign policy', suggests that inherently sovereign functions extend to 'the activities of the authorities responsible for foreign and military affairs; legislation and the exercise of police power; and the administration of justice.'[91] The implications of this delineation are far-reaching. For instance, experts and international organisations underscore the harm of disinformation for democracy;[92] if democracy as a norm and system of government falls within the scope of inherently sovereign functions, the evidenced undermining and disruptive effect of disinformation on democratic processes qualifies as interference.[93]

Interpretations of 'inherently sovereign' range from pragmatic views that exclude activities performed by non-State entities to more existential questions concerning the legal validity of the term itself.[94] As a construct of international

87 For an international theory on inherently sovereign functions, see Mégret (n 83) 463.

88 *Ibid*.

89 Lahmann (n 49) 426; Mégret (n 83) 461-462; Tallinn Manual 2.0 (n 43) Rule 4, Commentary 16-17, 22; while there is overlap with the concept of *domaine reserve*, neither concept provides a definitive overview of functions from an international legal perspective, in Hazel Fox and Philippa Webb, *The Law of State Immunity* (Oxford University Press 2013) 399.

90 Michael Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention' (2020) 96 International Law Studies 549, 557; Tallinn Manual 2.0 (n 43) Rule 44 Commentary 16.

91 Moynihan (n 12) 15.

92 European Commission, 'Tackling Online Disinformation: a European Approach' (Communication) COM(2018) 236 final; UN Special Rapporteur on the Promotion of the Right of Freedom of Expression, 'Report on Disinformation and Freedom of Expression' (13 April 2021) A/HRC/47/25; IACHR Special Rapporteur on Freedom of Expression, 'Disinformation, Pandemic, and Human Rights' (2023) OEA/Ser.L/V/II/CIDH/RELE/INF/25/23; Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda (3 March 2017) United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information; ASEAN, 'Guideline on Management of Government Information in Combatting Fake News and Disinformation in the Media' (Ministry of Communications and Informatics, March 2024); African Union, 'The Future of Media in Africa (AU Media Fellowship Learning Series, Addis Ababa, 14-15 February 2024).

93 Jean Cohen, 'The Democratic Construction of Inherently Sovereign Function' (2021) 115 American Journal of International Law 312, 312.

94 Tallinn Manual 2.0 (n 43) Rule 4, Commentary 17.

law, can such functions be *inherently* sovereign? Despite ongoing debate, several parameters relevant to disinformation have emerged. First, purely private or commercial actions are excluded, even when performed by the State.[95] Second, privatisation does not alter a function's inherently governmental nature. Similar to the law on State responsibility, delegating functions to non-State actors does not shield States from responsibility. Decisive is the nature of the function rather than the actor performing it.

Even if the functions can be precisely delineated, the question remains when an influence operation *interferes* with or *usurps* said functions.[96] International law does not provide a conclusive definition of either concept. Under its ordinary meaning, usurpation embodies the taking over and performing of governmental functions (i.e. wrongful appropriation),[97] while interference is commonly understood as hampering, frustrating or meddling in a deleterious way', either intentionally or unintentionally.[98] Disinformation as a process of intentional manipulation is unlikely to usurp a sovereign function of another State, but it can undermine the ability of a State to exercise the outlined functions through influencing its population or decision-makers. Compared to the prohibition of intervention as a corollary of sovereignty, the threshold for 'interference' is arguably less deleterious than intervention: latter being interlaced with the qualifier of coercion, thus forming a *de maximus* parameter. Below this 'threshold', however, the grey area remains.

Mere persuasion is not prohibited, but the deliberate employment of false or misleading information, especially on a large scale, may qualify as an 'unauthorized exercise of authority.'[99] A robust differentiation between interference and lawful persuasion necessitates recognition of interference's variable nature across sovereign functions. Each 'inherently sovereign function' operates through distinct mechanisms, involving different State agencies and decision-making processes. Recalling the illustrations from section 2.2, various forms of subversive disinformation may precipitate distinct behavioural responses that constitute subversive interference in these different contexts. The discourse on disinformation, elections, and democracy clearly subscribes to this variability: interference can take place by dissuading voters to cast their ballots, misinforming them on when and where to vote,[100] manipulating their choice towards a particular outcome, or discrediting the outcome of elections.

95 Milanovic and Schmitt (n 85) 255.

96 For a political perspective on the matter, see Kristine Berzina and Etienna Soula, 'Conceptualising Foreign Interference in Europe' (2020) Alliance For Securing Democracy.

97 Schmitt (n 16) 46; Wheatley (n 49) 694.

98 Gardner Dictionary of Legal Usage, 'Meddle; Intermeddle; Interfere; Tamper' (2011) 570.

99 Moynihan (n 12) 29.

100 Marco Roscini, *International Law and the Principle of Non-Intervention: History, Theory, and Interactions with other Principles* (Oxford University Press 2024) 398.

Aside from electoral processes, disinformation targeting public health policy during crises can violate sovereignty, albeit indirectly,[101] depending on the scale, degree of exposure and sophistication of the message. The COVID-19 infodemic contained subversive disinformation, alongside discriminatory and defamatory disinformation.[102] Not because public health policy is necessarily an inherently sovereign function,[103] but because crisis management is.[104] Comparing cyberactivities that cause physical damage to information operations, Milanovic and Schmitt conclude that

> 'it matters not whether the harm to human lives and health is caused by a cyber operation that, say, physically makes COVID-19 testing impossible, or by a misinformation campaign that fatally undermines public confidence in, and willingness to partake of, testing.'[105]

Interference through disinformation during the COVID-19 pandemic comprised altering views and behaviours on different fronts, of which undermining public confidence is but one; encouraging people to disregard social distancing rules, to consume substances making them ill or even resulting in death, inciting violence against medical institutions and staff, or spreading false origin-theories to disrupt inter-State relations.[106] The diversity of narratives targeting different audiences and dimensions of public health, crisis management and external relations reiterates that the multifaceted nature of the term 'interference' requires a matching multilayered response in the context of influence operations.

A third illustration of subversive disinformation amounting to interference with the sovereign conduct of international relations, are campaigns targeting diplomatic or consular staff through extensive defamatory disinformation campaigns (chapter three).[107] These have been instrumentalised to disrupt friendly relations or political stability between third States, as illustrated by China's strategic employment of disinformation regarding Taiwan's alleged

101 Strongwater (n 49) 696; Pijpers (n 49) 182-183; Allison Denton, 'Fake News: The Legality of The Russian 2016 Facebook Influence Campaign' (2019) 37 Boston University International Law Journal; Milanovic and Schmitt (n 85).

102 Chapters three and four.

103 Although some argue that is it, see eg. Priya Urs,'Cross-Border Cyber Operations Targeting Healthcare as Unlawful Intervention in the Affairs of States' (2024) 57 Vanderbilt Journal of Transnational Law, 36; Lahmann (n 49) 9.

104 Milanovic and Schmitt (n 85) 255.

105 *Ibid.* 269.

106 Digdem Soyaltin-Colella and Deniz Sert, 'The Strategic Use of Narratives and Governance of the COVID-19 Pandemic in Major Autocratisers in Europe' (2024) 26 Journal of Balkan and Near Eastern Studies 565; Philipp Sprengholz *et al.*, 'Historical narratives about the COVID-19 Pandemic are Motivationally Biased' (2023) 623 Nature 588.

107 On diplomatic interference and propaganda generally, see Paul Behrens, *Diplomatic Interference and the Law* (Hart Publishing 2018) 171-191.

deportation of Hong Kong protestors, forming part of its broader 'one country' propaganda designed to undermine the political independence of both States.[108] More violently, Russia's well-calibrated "disinformation playbook" around Niger's Coup hampered peaceful conciliation within its borders, as well as with neighbouring States.[109] The active support of Russia, the UAE and Egypt for disinformation by various proxy forces in Sudan has had a comparable effect.[110] The three scenarios encompass different narratives, actors, timelines and varying degrees of scale, reach and intensity, yet all three demonstrate interference with the internal or external affairs of another State.

The administration of justice presents another critical domain of sovereign functions susceptible to disinformation interference. Systematic and sophisticated disinformation portraying a State's justice system as biased, corrupt, or untrustworthy may undermine compliance, internal stability, and public order.[111] The instrumentalisation of, *inter alia*, inauthentic audio-visual materials (deepfakes) to disrupt judicial investigations particularly threatens a State's exclusive right to conduct criminal justice investigations within its territory – a concern heightened by digital evidence's growing significance in legal proceedings.[112] In addition, the choice to enact legislation to counter disinformation falls squarely within the domain of inherently sovereign functions – albeit subject to State obligations under human rights law.[113] While the operational reality of private entity involvement in blocking, removing or labelling disinformation challenges the underlying authority structure, the *de jure* authority – the legal power to shape the regulatory framework – remains vested in the State. Consequently, when a State deploys aggressive disinformation operations against another, this interferes with the target State's legislative and enforcement prerogatives.

In this context, it is noteworthy that prosecutors, law enforcement and judges enjoy a greater level of protection against speech that damages their reputation and credibility and undermines public trust in them. Under Article 10 of the European Convention on Human Rights (ECHR), the ECtHR permits restricting such speech because the judiciary, '[as] the guarantor of justice'

108 Quirk (n 32) 538, 540.

109 Africa Centre for Strategic Studies (n 65).

110 *Ibid.*

111 Austria, in its 2024 declaration on the applicability of international law to cyberspace, uses this illustration as an example of prohibited intervention: 'State A launches a large-scale campaign against the government of state B that spreads disinformation about the alleged corrupt business practices of state B's government, intended to sow distrust within the population of state B. The campaign eventually causes the government of state B to resign, resulting in a governmental crisis. Such a cyber activity would violate the prohibition of intervention', in Austrian Position on Cyber Activities and International Law (April 2024) 6, Accessed 21 November 2024.

112 On the exercise of law enforcement, evidence gathering and sovereign functions, see Wheatley (n 49) 695.

113 Moynihan (n 12) 14.

must 'enjoy public confidence to carry out its duties', thus protecting it 'against gravely damaging attacks that are essentially unfounded.'[114] While this does not establish unlawful intervention under international law, it signals the judiciary's centrality in the sovereign functioning of a State.

Under the overarching denominator of cultural and discriminatory disinformation, targeted disinformation campaigns against, *inter alia*, diaspora communities reveal patterns of deliberate destabilisation. Russia's operations in the Baltic States as well as China's activities regarding Taiwan and Hong Kong demonstrate how States employ manipulated information tactics to disturb internal stability abroad.[115] Moreover, history is littered with examples of how the fabrication of false narratives about military matters alongside the dissemination of false accusations of imminent violence have provoked international tensions or conflicts, interfering with a State's ability to maintain its national security and conduct foreign relations effectively.[116] Therefore, disinformation operations which intentionally create the perception of an imminent threat of force, may specifically be prohibited under Article 2(4) UN Charter, as argued elsewhere.[117]

Disinformation may also infringe upon State sovereignty through the nexus between inherently sovereign functions and the protection of individual rights and freedoms. An instrumentalist interpretation of sovereignty presupposes that the inherent function of a State is ultimately to serve and protect the rights of its citizens.[118] Given disinformation's widely evidenced harmful impact on fundamental rights,[119] all foreign disinformation could be deemed universally subversive by interfering with the State's function to protect these rights. While theoretically viable, this interpretation relies on a very expansive

114 *Prager and Oberschlick v Austria* App no 15974/90 (ECtHR, 26 April 1995) para 24; *Morice v France* App no 29369/10 (ECtHR GC, 23 April 2015) para 168.

115 Quirk (n 32) 526-567; Breedon (n 6) 636-640.

116 Including the US invasion in Iraq and Russia's invasion in Ukraine, as illustrated in section 2.2; See also Kearney (n 54); Evelyn Aswad, 'Propaganda for War & International Human Rights Standards' (2023) 24 Chicago Journal of Intern International Law 1, 1-30; Andrei Richter, 'International Legal Responses to "Propaganda for War" in Modern Warfare' (2023) 10 Journal of International Media and Entertainment Law 1, 54-80.

117 Section 3.4.1.4 'Defamatory Disinformation as a Threat to Peace'.

118 Malcolm Shaw, *International Law* (Cambridge University Press 2008) 268-269; Juliana Kokott, 'States, Sovereign Equality' (2011) Max Planck Encyclopaedia of Public International Law, para 4.

119 Richard Wingfield, 'A Human Rights-Based Approach to Disinformation' (2019) Global Partners Digital, Accessed 16 October 2024; Kate Jones, 'Online Disinformation and Political Discourse: Applying a Human Rights Framework' (2019) Chatham House Research Paper, Accessed 16 October 2024; Amnesty International, 'A Human rights Approach to Tackle Disinformation' (2022) Submission to the Office of the High Commissioner for Human Rights, Accessed 16 October 2024; UNGA, 'Countering Disinformation For the Promotion and Protection of Human Rights and Fundamental Freedom' (12 August 2022) UN Doc A/77/287.

understanding of inherently sovereign functions, encountering practical obstacles, including overdetermination.

2.3.2.3 Unlawfulness Threshold

Overcoming these ambiguities would benefit from a delineation between permissible and impermissible interference based on standardised criteria. The argument for including such a threshold finds support in theory and practice. Following Oppenheim's view that 'independence is a question of degree,' sovereignty violations similarly become a matter of degree, requiring either a qualified understanding of 'interference' or a defined threshold.[120] Although *lex lata* is inconclusive – with both States and experts divided – continued omission of a threshold risks rendering the rule ineffective and irrelevant through overinclusiveness. While almost all forms of subversive disinformation interfere with a State's sovereign prerogatives to some degree, classifying all such interference an internationally wrongful act can indeed not legitimately be upheld.

States and experts favouring inclusion generally suggest an effects-based approach, evaluating interference as a matter of degree through both qualitative and quantitative factors.[121] Several States, notably France and the Netherlands, seemingly support a higher threshold based on *significant* effects, considering criteria such as severity, scale of societal impact, practical effects on a State's ability to regulate its inherently sovereign functions, and overall gravity.[122] Arguably, however, both standards are flawed because they appear to only retrospectively assess whether interference has taken place and thus

120 Jennings and Watts (n 41) 391.

121 Canada, 'International Law Applicable to Cyberspace' (22 April 2022) para 15 ('cyber activities that rise above a level of negligible or de minimis effects, causing significant harmful effects within the territory of another State without that State's consent, could amount to a violation of the rule of territorial sovereignty with respect to the affected State'); in the United States position in the Official compendium of voluntary national contributions on the subject of how international law applies to the use of information and communications technologies by States, they comment that '[i]n certain circumstances, one State's non-consensual cyber operation in another State's territory, even if it falls below the threshold of a use of force or non-intervention, could also violate international law. However, a State's remote cyber operations involving computers or other networked devices located on another State's territory do not constitute a per se violation of international law. In other words, there is no absolute prohibition on such operations as a matter of international law. This is perhaps most clear where such activities in another State's territory have no effects or de minimis effects. The very design of the Internet may lead to some encroachment on other sovereign jurisdictions', in UNGA, 'Developments in the field of information and telecommunications in the context of international security: Official compendium of voluntary national contributions on the subject of how international law applies to the use of information and communications technologies by States' (13 July 2021) UN Doc A/76/136, 1140.

122 Moynihan (n 12) 22, 26.

failing to enable timely intervention, i.e. to prevent harm.[123] Alternative proposals that would remedy these concerns have gained little traction or are inapplicable to influence operations.

Since interference through disinformation primarily involves manipulating foreign populations through false or misleading information, evaluating its (significant) effects on sovereign functions requires understanding influence, persuasion, and manipulation processes. According to chapter one's findings, disinformation's likelihood of achieving such effects predominantly depends on three factors:

1. *The type of disinformation and the State function(s) it targets*
   Recalling the variable nature of interference, the narrative created and disseminated requires case-by-case assessment relative to the State function it affects. For example, pseudoscientific medical reviews that have been labelled as disinformation by the World Health Organization may interfere with managing a public health crisis as an inherently sovereign function, whereas the same information may have no bearing on other sovereign functions. Thus, whether disinformation constitutes interferences depends on whether the narrative targets essential elements of a specific sovereign function. This assessment varies not only per function, but also per State: culturally and religiously sensitive disinformation may, *inter alia,* disrupt diplomatic relations between certain States while passing unnoticed in others, and false and misleading information on migrants or ethnic minorities may threaten internal stability and public order in different degrees in different States.
2. *The target audience*
   Unlike physical incursions or infrastructure-damaging cyberoperations, influence operations achieve interference through behavioural change, focusing on how altered audience behaviour impacts sovereign functions. When deployed strategically, messages tailored to specific audience profiles increase the likelihood of achieving desired behavioural outcomes. While not decisive, the difference must be recognised between disinformation targeting the public opinion broadly and disinformation which aims to exert influence on decision-makers. The latter more readily constitutes unlawful interference, particularly regarding democratic governance as a sovereign prerogative exercised primarily by government and representatives.[124] Heightened scrutiny should be given to subversive disinformation targeting this group; while not always targeting the State or its officials directly, when it does, it significantly increases the risk of compromising core public services and (the perceived) legitimacy of a State's actions and

123 Roguski (n 43) 74-76.

124 Lukas Willmer, 'Does Digitalization Reshape the Principle of Non-Intervention?' (2023) 24 German Law Journal 508-521, 517.

decisions.[125] While arguably, decision-makers are simultaneously better equipped to identify disinformation and address it, reoccurring patterns of political actors in, *inter alia,* Germany, Italy, the Czech Republic, Serbia and the United States adopting and amplifying foreign subversive disinformation – especially originating in Russia – indicates their continued susceptibility, even within institutional settings that are assumed to be more resilient.[126]

3. *Timing and scale of the campaign or operation*
   It is well recognised that crises periods or sensitive intervals amplify disinformation's impact. State practice subscribes to this view, with many countries tightening regulations on political advertising prior to elections to prevent the spread of mis- and disinformation.[127] Scale considerations become crucial when attempting to influence legislative decision-making or effecting behavioural change by mobilising the general population. Effective interference typically requires prolonged exposure of a substantial audience to multiple national and foreign disinformation narratives, as strikingly illustrated by various COVID-19 disinformation campaigns. Different narratives aiming to reduce trust in institutions and undermine scientific credibility whilst simultaneously attributing legitimacy to alternative sources and promoting substitute narratives, are most "successful". Drawing from la Coeur's introduced differentiation between disinformation stories, campaigns and operations, these patterns imply that disinformation stories – as singular false (news) narratives – opposite disinformation campaigns or operations – characterised by long-term strategic

125 Indra Spiecker gennant Dohmann, 'Public Officials' Lies and the Proliferation of Disinformation' in András Koltay, Charles Garden and Ronald Krotoszynski (eds), *Disinformation, Misinformation and Democracy* (Cambridge University Press 2025) 77-114; Eduardo Bertoni, 'Public Official' Lies and Their Effects on the Proliferation of Disinformation' András Koltay, Charles Garden and Ronald Krotoszynski (eds), *Disinformation, Misinformation and Democracy* (Cambridge University Press 2025) 115-130; Talita de Souza Dias and Antonio Coco, *Cyber Due Diligence in International Law* (March 2021) Final Project Report Oxford Institute for Ethics, Law and Armed Conflict, 74; it is evidenced that decisionmakers rely on mass media which inevitable shapes their beliefs and orientations with regard to the policy they make, in Michael Kunczik, *Images of Nations and International Public Relations* (Routledge 1990) 20, 58, 86; Einar Östgaard, 'Factors Influencing the Flow of News' (1965) 2 Journal for Peace Resolutions 39, 54.

126 Elina Treyger *et al.*, 'The Denazify Lie: Russia's Use of Extremist Narratives Against Ukraine' (RAND Corporation 2025); Sebastian Geisler and Roman Eichinger, 'Die Entzauberung der AfD' BILD (5 May 2024) Accessed 8 April 2025; Pavel Havlíćek, 'Czechs Deluged by Russian Disinformation After Spy Brawl' (Center for European Policy Analysis, 7 May 2021) Accessed 8 April 2025; Hafsa Khalil, 'Zelensky says Trump living in Russian "disinformation space"' BBC News (19 February 2025) Accessed 8 April 2-25; Alina Polyakova and others, 'The Kremlin's Trojan Horses 2.0: Russian Influence in Greece, Italy, and Spain' (Atlantic Council, November 2017) Accessed 8 April 2025.

127 Marko Milanovic and Philippa Webb, 'False Speech' in Amal Clooney and David Neuberger (eds), *Freedom of Speech in International Law* (Oxford University Press 2024) 224-236.

coordination,[128] should categorically fall below a potential threshold for impermissible interference under international law.

In sum, while theoretical and practical arguments support the inclusion of a threshold for sovereignty violations – through disinformation or otherwise – international legal doctrine and State practice do not unambiguously recognise its existence. Not incorporating a threshold, however, will become increasingly untenable considering growing cyber and influence capabilities. Interference tools and strategies will only continue to expand, inevitably leading to inflation of accusations of 'sovereignty violations' without the creation of graduated interference thresholds – a trend already visible in the context of disinformation. For disinformation and other indirect, influence operations, whether these will rise to the level of impermissible interference depends on their anticipated impact, which in turn requires an understanding of the multifaceted nature of influence operations in general (chapter one) alongside case-by-case assessments of specific campaigns. For the latter, this analysis should consider at least the three outlined factors: the typology and targeting of disinformation vis-à-vis specific sovereign functions, the strategic identification of target audiences (particularly decision-makers), and the temporal and scalar dimensions of influence campaigns.

2.3.2.4 Intent as Indicator

Subversive disinformation encompasses false or misleading information characterised by the intent to disrupt a State's exercise of its sovereign functions. While intent is not doctrinally required for breaching State sovereignty, academic discourse increasingly considers it relevant in determining legality of State conduct in the digital realm broadly.[129] In the context of information operations, the presence of malicious intent distinguishes disinformation from other forms of information manipulation that are generally considered less grave and malicious. Although unintentional disinformation (misinformation) may cause harmful effects, sovereignty-undermining behavioural change can hardly be achieved without an element of strategy and coordination.[130] The intentional nature of subversive disinformation therefore is a significant factor in assessing interference with inherently sovereign functions, in contrast to propaganda or misinformation.

When foreign States reveal their intent to undermine electoral processes, disrupt public health strategies during crises, or cause internal strife, it is

128 Section 1.5.1.2 'Form, Presentation and Content'.

129 Roguski (n 43) 76.

130 This characteristic has led experts favouring inclusion of disinformation in the category of cyber acts that carry a risk of causing significant harm, in Dias and Coco (n 125) 146.

counterintuitive to ignore this.[131] Building on the general 'Indicators of intent to cause harm' in chapter one, the structure of subversive disinformation campaigns often provides significant contextual evidence that can cumulatively substantiate a claim of subversive intent. Does the campaign align with the strategic interests or objectives of the foreign State? Do the narratives mirror (official) State policy? Are there recurring patterns of disinformation – national and international – operated across various platforms? Does the campaign appear – or is amplified – during a time of crisis or unrest? Is it executed alongside other diplomatic or military influence attempts, or does it occur within a broader pattern of cyberoperations? And finally, are there clear and persistent efforts to disguise the involvement of a foreign State and/or the source of information? In conjunction, these questions can indicate the level of strategy and coordination that cannot occur unintentionally and is thus indicative of malicious intent and, in the context of influence operations, of impermissible subversive disinformation.

2.3.3 Interim Conclusion

The analysis of subversive disinformation under international law demonstrates that State sovereignty, as a primary rule of law, offers a potentially valuable frame for addressing modern disinformation threats, but that its effectiveness is limited by conservative interpretations and the unclear existence of a threshold for violating sovereignty. Traditional interpretations of sovereignty violations focus on physical incursions and tangible harm; hence disinformation presents a unique challenge through its indirect, cognitive effects. Rather than dismissing its impact due to a lack of physical damage, the chapter argued in favour of including a threshold for sovereignty violations – one that particularly evaluates disinformation's interference with sovereignty case-by-case based on three key factors: the type and target of State function affected, the strategic targeting of specific audiences (especially decision-makers), and the timing and scale of operations. These factors, combined with consideration of intent as an indicator of unlawful interference, provides a more nuanced and practical method for distinguishing between permissible influence and sovereignty violations.

While this framework for evaluating disinformation's interference with sovereignty represents *lex ferenda* rather than *lex lata*, it addresses a critical

131 This resonates with, *inter alia*, the U.S. framework against foreign interference, which centralises intent. According to the Cybersecurity & Infrastructure Security Agency, certain acts are considered 'interference' if they are 'designed to sow discord, manipulate public discourse, discredit the electoral system, bias the development of policy, or disrupt markets for the purpose of undermining the interests of the United States and its allies' in CISA, 'Foreign Interference Taxonomy' (July 2018) Department of Homeland Security, Accessed 9 February 2023; Corn (n 16) 226.

legal gap. The current open-ended conception of sovereignty and lack of objective criteria for identifying violations create two significant risks: inconsistent domestic interpretations that may escalate international tensions,[132] and States' reluctance to label information operations as sovereignty violations could be misinterpreted as implied consent or *opinio juris* accepting such activities.[133] This legal uncertainty regarding the evolving nature of influence operations, underscores the need for delineation criteria to prevent both over-expansive claims and dangerous under-enforcement of sovereignty violations.

2.4 The Principle of Non-Intervention

The duty to refrain from intervention in the exclusive jurisdiction of other States is an essential corollary of State sovereignty, and a cardinal obligation for all States.[134] The prohibition lacks universal codification, but is reflected in many regional treaties,[135] international judgements and is considered 'part and parcel' of customary international law.[136] Its seminal importance has been repeatedly emphasised in UNGA resolutions,[137] and has taken a central role in the debate on the legality of State conduct in cyber space.[138] While attempts to consolidate a definition are plenty,[139] its exact scope and meaning remain contested,[140] prompting scholars to characterise it as the 'most elusive

132 Moynihan (n 12) 20; Berzina and Soula (n 96).
133 Breedon (n 6) 677-679.
134 Crawford (n 12) 431.
135 Convention on the Rights and Duties of States (adopted 26 December 1933, entered into force 26 December 1934) 165 LNTS 19 (Montevideo Convention) Article 8; Charter of the Organization of American States (adopted 30 April 1948, entered into force 13 December 1951) 119 UNTS 3, Articles 9 and 13; Vienna Convention on Diplomatic Relations (adopted 18 April 1961, entered into force 24 April 1964) 500 UNTS 95; Constitutive Act of the African Union (adopted 11 July 2000, entered into force 26 May 2001) 2158 UNTS 3 Article 41; Treaty of Amity and Cooperation in Southeast Asia (ASEAN) (adopted 24 February 1976).
136 *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 7) para 202.
137 An overview is presented in Urs (n 103) 9-12.
138 Pijpers (n 49) 148-162.
139 Among others, the negotiations preceding the 1969 Vienna Convention on the Law of Treaties, the 1970 Declaration on Principle of International Law concerning Friendly Relations and Cooperation among States (1970) and the 1996 Draft Code of Crimes against the Peace and Security of Mankind (Draft Code on Crimes Against the Peace) reveal the various obstacles and considerations in the debate on defining intervention; Helal (n 14) 55-56 citing Stuart S Malawer, 'A New Concept of Consent and World Public Order: "Coerced Treaties" and the Convention on the Law of Treaties' (1970) 4 Vanderbilt Journal of Transnational Law 1, 16-17; Robert Rosenstock, 'The Declaration of Principles of International Law Concerning Friendly Relations: A Survey' (2017) 65 American Journal of International Law 5,726-729; John Linarelli, 'An Examination of the Proposed Crime of Intervention in the Draft Code of Crimes Against the Peace and Security of Mankind' (1995) 18 Suffolk Transnational Law Review 1.
140 Roscini provides a comprehensive overview of the discussion, in Roscini (n 100) 145-195.

of all international principles' and 'the Rubik's cube of international law.'[141] Nevertheless, where definitional uncertainty regarding State sovereignty generates scepticism about its status as a substantive rule of international law, unlawful intervention attributable to a State uncontroversially constitutes an internationally wrongful act.

The non-intervention principle as a binding norm was authoritatively clarified by the ICJ in the *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States of America)*,[142] presenting a definition that reflects its customary international law status:[143]

> 'A prohibited intervention must … be one bearing on matters in which each State is permitted, by the principle of State sovereignty, to decide freely. One of these is the choice of a political, economic, social and cultural system, and the formulation of foreign policy. Intervention is wrongful when it uses methods of coercion in regard to such choices, which must remain free ones. The element of coercion, which defines, and indeed forms the very essence of, prohibited intervention, is particularly obvious in the case of an intervention which uses force, either in the direct form of military action, or in the indirect form of support for subversive or terrorist armed activities within another State.'[144]

From this judgement, in conjunction with subsequent case law, over thirty-five General Assembly resolutions and regional instruments, the consensus emerged that unlawful intervention consists of two cumulative elements: interference into the domestic jurisdiction of another State – often indicated by the notion of *domaine réservé* – which is coercive in nature.[145] As discussed elsewhere,

141 Mohammed S Helal, 'Intervention, Force & Coercion – A Historical Inquiry on the Evolution of the Prohibition on Intervention' (2024) 103 International Law Studies – U.S. *Naval War College, Stockton Center for International Law*, 3; Vaughan Lowe, *International Law* (Oxford University Press 2007) 104; John Norton Moore, 'Legal Standards for Intervention in Internal Conflicts' (1983) 13 Georgie Journal of International and Comparative Law 191, 191.

142 The principle has also been codified in various treaties largely coinciding with its meaning in the Nicaragua case, see e.g. Charter of the Organisation of Islamic Cooperation (adopted 14 March 2008) Article 2(4)-(5); Charter of the Association of Southeast Asian Nations (adopted 20 November 2007) 2624 UNTS 223 Article 2(2)(e)-(f); Constitutive Act of the African Union (adopted 11 July 2000) 2158 UNTS 3 Article 4(g); Charter of the Organization of African Unity (adopted 25 May 1963) 479 UNTS 39 Article 3(2); Charter of the Organization of American States (adopted 30 April 1948) 119 UNTS 3 Articles 19-20.

143 Arguably, the prohibition is part of *jus cogens*, in Separate Opinion of Judge Sette-Camara, *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 7); Philip Kunig, 'Prohibition of Intervention' (2008) Max Planck Encyclopaedia of Public International Law, though the attribution of this status is not unchallenged, in Maziar Jamnejad and Michael Wood, 'The Principle of Non-Intervention' (2009) 22 Leiden International Journal of International Law 345-381, 358.

144 *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 7) para 205.

145 Roscini (n 100) 145-147; Jamnejad and Wood (n 143) 348.

it is widely accepted that certain forms of subversive propaganda fall *prima facie* within the scope of the prohibition.[146]

It is not challenged that the prohibition extends into the digital realm,[147] though new declarations of 'information sovereignty' and the development of 'manipulative coercion' do question the traditional understanding of both components.[148] While the meaning and scope of the prohibition have always been contested, these developments sparked criticism that perpetual (over)reliance on the ICJ's wording from 1986 risks excluding new forms of subversive behaviour.[149] Cyberspace and the emergence of modern influence operations have divided experts over whether – and if so, how – the norm needs redefining,[150] including by potentially moving away from the notion of coercion.[151] While laudable as efforts in the pursuit of clarity, the outcome of these debates has not necessarily been fruitful in relation to disinformation.[152] Persuasion, manipulation and coercion remain arbitrarily or inconsistently delineated, the growing role of non-State actors is insufficiently addressed and renewed interpretations rely excessively on disinformation as electoral

146 UNGA Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of Their Independence and Sovereignty, UNGA Res 2131 (XX) (21 December 1965) UN Doc A/RES/2131; UNGA Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States, UNGA Res 36/103 (9 December 1981) UN Doc A/RES/36/103.

147 Roscini (n 100) 377-380 ('[s]tates from all geographical regions and political affiliations […] have expressly affirmed that the principle of non-intervention applies to cyberspace').

148 Kilovaty, *inter alia*, challenges the relevance of the principle in today's world, labelling coercion as 'an outdated standard', in Kilovaty (n 15) 130; Chesterman equally notes that the principle of non-intervention 'fail[s] to keep pace with technological advancements that render territorial limits irrelevant' in Simon Chesterman, 'Secret Intelligence' (2019) Max Planck Encyclopaedia of Public International Law, para 23.

149 Corn (n 16) 212.

150 Katja S Ziegler, 'Domaine Réservé' (2013) Max Planck Encyclopaedia of Public International Law, Section E 'The End of Domain Reserve; Ido Kilovaty, 'The International Law of Cyber Intervention' in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing 2021); Ido Kilovaty, 'The Elephant in the Room: Coercion' (2019) Symposium on Dan Efrony & Yuval Shany, a Rule Book on the Shelf? Tallinn Manual 2.0 on Cyberoperations and Subsequent State Practice, AJIL Unbound 113; 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* (2020 Rowman & Littlefield); Thibault Moulin, 'Reviving the Principle of Non-Intervention in Cyberspace: The Path Forward' (2020) 25 Journal for Conflict and Security Law 423.

151 Kilovaty (n 15) 140-147; Ohlin (n 57) 88; Roscini (n 100) 397; Xuan W Tay, 'Reconstructing The Principle of Non-Intervention and Non-Interference – Electoral Disinformation, Nicaragua and the Quilt-work Approach' (2022) 40 Berkeley Journal of International Law 1.

152 As Willmer rightly notes, the renewed interest in the principle of non-intervention concerning digital matter has affirmed its applicability and relevance, but not clarified the application, in Willmer (n 124) 511; Helal similarly notes that 'there is no support in the history of the evolution of non-intervention for the definitions of coercion that some states and scholars have recently espoused', in Helal (n 141) 7.

intervention, or interference with public health in the context of the COVID-19 pandemic.

Although State practice indicates a growing acceptance towards disinformation as a potential non-intervention violation, this acknowledgement remains limited and lacks specificity. Several States, nevertheless, attempted greater clarity. Iran, for instance, has provided one of the clearest condemnations in this regard, stating that '[m]easures like cyber manipulation of elections or engineering the public opinions on the eve of the elections [are] examples of gross intervention.'[153] Other States take a more nuanced and broader approach: Germany opines that disinformation operations are coercive if they 'deliberately incite violent political upheaval […] significantly impeding the orderly conduct of an election' because then they 'may be comparable in scale and effect to the support of insurgents',[154] and France considers forms of digital interference unlawful 'which causes or may cause harm to France's political, economic, social and cultural system […],'[155] echoing the ICJ's language from the *Nicaragua case*. Poland's position on international law in cyberspace qualifies 'a wide-scale and targeted disinformation, in particular when it results in civil unrest' as prohibited intervention;[156] Austria condemns '[l]arge-scale cyber activities, including disinformation campaigns, conducted by or attributable to a state may also constitute, if undertaken to compel another state to involuntarily change its behaviour' as a violation of the prohibition of intervention;[157] and Costa Rica considers 'subversive or hostile propaganda', 'dissemination of false news', and 'electoral disinformation campaigns' prohibited coercive behaviour.[158] Other States that omit explicit references to disinformation, nevertheless have referred to 'unduly influencing public opinion' or consider any behaviour with the 'potential for compelling the target state to engage in an action that it would otherwise not take' to be unlawful intervention.[159]

153 Willmer (n 124) 513 citing Press Release, 'General Staff of Iranian Armed Forces Warns of Tough Action to Any Cyber Threat' (2020).

154 German Government, 'German Government Notes Sanctions Against Nordstream 2 and Turkstream with Regret' (Press Release 432, 21 December 2019) Accessed 22 October 2023.

155 French Ministry of Defense (*Ministère des Armées*), 'International Law Applied to Operations in Cyberspace' [*Droit International Appliqué aux Opérations dans le Cyberspace*] (9 September 2019) 7.

156 Republic of Poland, 'The Republic of Poland's Position on the Application of International Law in Cyberspace' (29 December 2022) 4.

157 Republic of Austria (n 34) 5.

158 Costa Rican Ministry of Foreign Affairs, 'Costa Rica's Position on the Application of International Law in Cyberspace' (2023) 1, 8-10

159 UNGA, 'Official Compendium of Voluntary National Contributions on the Subject of How International Law Applies to the Use of Information and Communications Technologies by States' (UNODA, August 2021) UN Doc A/76/136, 68-69. Danish Government, 'Denmark's Position Paper on the Application of International Law in Cyberspace' (2023) 450.

Instrumentalising disinformation can constitute unlawful intervention, as argued in the following section. To substantiate this argument, section 2.4.1 examines the prohibition's protective scope in both offline and online domains, demonstrating disinformation's interference with protected affairs and exploring how the information sovereignty discourse has positioned disinformation regulation within a State's *domaine réservé.* Subsequently, section 2.4.2 addresses how disinformation pertains to the threshold of 'coercion' as 'the very essence of prohibited intervention'.[160] It explores the relevance of disinformation's intrinsic malicious intent and substantiates the premise that its manipulative influence undermines or circumvents the audience's autonomous decision-making process, making it a form of coercion.

2.4.1 Internal and External Affairs

Intervention embodies interference by a State in the internal or external affairs of another State,[161] making it prohibited when 'bearing on matters in which each State is permitted, by the principle of State sovereignty, to decide freely.'[162] While encompassing diverse jurisdictional aspects, the prohibition is not unlimited. Drawing from earlier UNGA resolutions,[163] the ICJ has articulated that its scope covers in any case 'the choice of a political, economic, social and cultural system, and the formulation of foreign policy.'[164] More concretely, the Montevideo Convention articulates these prerogatives as a State's right to 'organise itself as it sees fit, to legislate upon its interests, administer its services and to define the jurisdiction and competences of its courts.'[165] These competencies are commonly captured by the notion of *domaine réservé*, though some authors prefer 'domestic jurisdiction.'[166] Despite

160 *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 7) para 205.
161 Kunig (n 143) para 1; Roscini (n 100) 309-311.
162 *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 7) para 205.
163 *Inter alia,* the 1970 Declaration on Principle of International Law concerning Friendly Relation among States in accordance with the Charter of the United Nations: '[n]o State or group of States has the right to intervene, directly or indirectly, for any reason whatsoever, in the internal of external affairs of another State. Consequently, 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, are in violation of international law,' in UNGA, 'Declaration on Principles of International Law Concerning Friendly Relations and Co-Operations among States in Accordance with the Charter of the United Nations' (adopted 24 October 1970) UN Doc A/RES/2625 (XXV).
164 *Ibid.*
165 Montevideo Convention on the Rights and Duties of States (signed 26 December 1933, entered into force 26 December 1934) 165 LNTS 19, Article 3.
166 Roscini criticises the use of *domaine reserve* for not being as accurate of 'domestic jurisdiction', in Roscini (n 100) 381; Ossoff also argues that 'the notion fails to capture to entirety of 'matters in which each State is permitted, by the principle of State sovereignty, to decide freely' in William Ossoff, 'Hacking the Domaine Reserve: The Rule of Non-Intervention

apparent overlap with 'inherently sovereign functions' (2.3.2.2),[167] Roscini distinguishes that 'not all matters under a State's domestic jurisdiction [i.e. *domaine réservé*] are 'inherently' sovereign or 'inherently governmental': what identifies them is rather the fact that they are not regulated by any international law applicable to the concerned state.'[168]

Roscini also identifies a fundamental controversy in delineating the notion of *domaine réservé*: traditionally circumscribed as encompassing matters in which a State is free of international obligations,[169] it functions as a residual category (2.4.1.1).[170] Globalisation, digitalisation, and the growing role of non-State actors, however, pressure the binary distinction between internationally regulated and unregulated State affairs (2.4.1.2). Therefore, while preserving the conceptual significance of *domaine réservé*, the chapter proposes a paradigmatic reorientation: from the reserved *domain* to the *object* the prohibition aims to protect, i.e. the ability of States to freely decide on their sovereign course of action (2.4.1.3) In identifying this object, valuable insights derive from the doctrine on self-determination, with an expanding corpus of scholarly opinion suggesting the doctrine's role in substantiating the non-intervention principle.

2.4.1.1 Domaine Réservé

Domaine réservé ostensibly suggests a clear demarcation between domestic and international spheres of competence, yet its scope and contemporary utility remain subject to academic debate.[171] Thus, the '*domaine réservé* describes areas where States are free from international obligations and regulation',[172]

---

and Political Interference in Cyberspace' (2021) 62 Harvard International Law Journal 1, 306.

167 The exact differentiation with the notion of inherently sovereign functions remains unsettled yet they coincidence in covering 'political alliances, diplomatic positions adopted at the international level, and foreign policy in general must be decided on an independent basis without external interference' in Marcelo Kohen, 'The Principle of Non-Intervention 25 Years After the Nicaragua Judgement' (2012) 25 Leiden Journal for International Law 157, 159; See also Marco Roscini (n 100) 381.

168 Against the backdrop of the previous section on sovereignty and jurisdiction, it is important to note that the *domaine reserve* is broader than the domestic jurisdiction of States, in Helal (n 14) 65; Roscini (n 100) 381.

169 Roscini (n 100) 381.

170 Helal (n 14) 66; Cecilia Yue Wu, 'Challenging Paternalistic Interference: The Case of Non-Intervention in a Globalized World' (2023) 65 Harvard International Law Journal 1, 262.

171 Ohlin even states that '[b]ut the mere use of a foreign phrase – and italics, no less – adds nothing more than the illusion of precision', in Ohlin (n 57) 72.

172 Ziegler (n 150) para 1; Kilovaty (n 15) 132; Schmitt (n 49) 45; Jens D Ohlin, 'Did Russian Cyber Interference in the 2016 Election Violate International Law?' (2017) 95 Texas Law Review 1579, 1588.

which explains the uncertainty regarding its scope.[173] Dependent on international obligations and regulations, a State's *domaine réservé* is subject to constant change through developments in international law.[174] As already recognised by the PCIJ in 1923, it is a fluid and relative concept, which 'depends on the development of international relations'.[175] As a direct consequence, the scope of *domaine réservé* may also differ per State.[176] Intensified international cooperation has limited the affairs falling *exclusively* within the domestic sphere; that which traditionally was considered the *domaine réservé* is becoming increasingly internationalised, including 'core matters' such as the regulation of nationality, the unfettered use of State territory and the structuring of its political system.[177] In other words, there are hardly any affairs that are entirely unregulated by international law.[178]

It is, however, counterintuitive to interpret such voluntary international regulatory commitments – which serve individual State's interests and the broader international community – as diminishing the relevance of the principle of non-intervention through disappearance of the *domaine réservé.* While academic discourse subscribes to this position,[179] it is unclear when or by which standards this line will be drawn. Attempting to demystify the matter, Milanovic positions the *domaine* within the bounds of international law, arguing that 'internal and external affairs belong to the reserved domain if a State *has any measure of discretion*' within such bounds.[180]

While the needed granularity remains absent, Milanovic's approach aligns with the purpose of the doctrine of non-intervention, which, as argued by Higgins, is to 'provide an acceptable balance between the sovereign equality and independence of States on the one hand and the reality of an interdependent world and the international law commitment to human dignity on the other.'[181] These considerations apply no differently online than in the analogue realm,[182] although the 'characteristics of the online realm have

173 Already in his 1974 work, RJ Vincent pointed towards these difficulties, in Robert J Vincent, *Nonintervention and international order* (Princeton University Press 1974) 299.

174 In *Nationality Decrees Issued in Tunis and Morocco (French Zone)* (Advisory Opinion) PCIJ Rep Series B No 4 (7 February 1923) Article 24, the PCIJ noted that 'solely within the domestic jurisdiction of a state' refers to matters that 'are not, in principle regulated by international law'; Roscini (n 100) 381; Ziegler (n 150) paras 1-2; Tsagourias (n 150) 48; Ossoff (n 166) 305.

175 *Nationality Decrees Issued in Tunis and Morocco (French Zone)* (n 174) Article 24; Milanovic (n 13) 612; Helal (n 14) 67-68; Helal (n 141).

176 Helal (n 141) 4; Kilovaty (n 150) 100; Niki Aloupi, 'The Right to Non-Intervention and Non-Interference' (2015) 4 Cambridge Journal of International Law 566, 574.

177 Ziegler (n 150) para 5.

178 Milanovic (n 13) 610.

179 Urs (n 103) 20.

180 Milanovic (n 13) 610.

181 Pijpers (n 49) 152 citing Rosalyn Higgins, 'Intervention and International Law' in *Themes and Theories* (2009) 273.

182 Roscini (n 100) 381.

increased the possibilities to access the reserved domains.'[183] The online information ecosystem in which disinformation operates provides a 'facilitative environment' for access to information and engagement with foreign audiences on domestic aspects of State policies,[184] consequently making the *domaine réservé* even less distinctive.[185]

The interaction between disinformation and the *domaine réservé* is twofold: regulation of disinformation as a part of the *domaine réservé* of a State and the dissemination of disinformation as an interference in this domaine. First, several States maintain that their online information environment falls within their *domaine réservé*, characteriding foreign disinformation – which 'pollutes' their information ecosystem or disrupts State-controlled information dissemination – as prohibited intervention. The Tallinn Manual affirms that content regulation online, subject to international human rights law constraints, resides within a State's *domaine réservé*,[186] a position some scholars extend to online communications generally.[187] This jurisdictional assertion correlates closely with State's broader approaches to sovereignty and jurisdictional authority in the digital realm. China and Russia, notably, advocate expansive control over their online information environments through 'cyber sovereignty' claims, explicitly invoking disinformation suppression to buttress this position.[188] They invoke authoritative international sources, including the 1981 Declaration on the Admissibility of Nonintervention and Noninterference, stipulating

> '[t]he right of States and peoples to have free access to information and to develop fully, without interference, their system of information and mass media and to use their information media in order to promote their political, social, economic and cultural interests and aspirations […].'[189]

183 Wheatley (n 29) 167; Pijpers (n 49) 159.
184 *Ibid.*; Tsagourias (n 150) 48.
185 Kilovaty (n 150) 100.
186 Tallinn Manual 2.0 (n 43) Rule 66, Commentary 11.
187 Katharina Ziolkowski, 'Peacetime Cyber Espionage – New Tendencies in Public International Law' (2013) 425 Peacetime Regime for State Activities in Cyberspace: International Law International Relations and Diplomacy, 434; Kilovaty (n 15) 133.
188 Ossoff (n 166) 307 citing Adam Segal, 'Year in Review: Chinese Cyber Sovereignty in Action, *Council on Foreign Relations* (January 8 2018); Lucas Kello, 'Digital Diplomacy and Cyber Defence' in Corneliu Bjola and Ilan Manor (eds), *The Oxford Handbook of Digital Diplomacy* (Oxford University Press 2024) 121-137; Eugene EG Tan and Benjamin Ang, 'ASEAN Ambiguity on International Law and Norms for Cyberspace' (2022) 20 Baltic Yearbook of International Law Online 1, 133-162; Iran's position paper stated in 2020 stated that 'Every state enjoys the inherent right to the full development of information system and mass media and their employment, without intervention, to advance their own political, social, economic, and cultural interests and aspirations' in Declaration of General Staff of the Armed Forces of the Islamic Republic of Iran Regarding International Law Applicable to the Cyberspace (August 2020) Article III, Accessed 21 November 2024.
189 UNGA 'Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States' (9 December 1981) UN Doc A/RES/36/103, Article 1(c).

Recalling earlier analysis,[190] this extended territorial control argument equally encompasses infrastructure-based control over Internet platforms within a State's territory.[191] Nevertheless, both from a content and infrastructure perspective, regulatory competence over disinformation does not reside solely with States: there is range of conventional and customary law prohibitions for engaging with unfriendly, harmful, defamatory, inciting or discriminatory information, including the Convention Concerning the Use of Broadcasting in the Cause of Peace and the Convention on the International Right of Correction.[192] Thus while regulating disinformation falls within a State's *domaine réservé* , they do not enjoy sole discretion over such online communications.

Secondly, States and experts consider specific instances of disinformation as interference with internal and external affairs – particularly electoral processes and public health policy, both firmly established within the *domaine réservé*.[193] In their argument on cyber-attacks and cyber (mis)information operations during a pandemic, Milanovic and Schmitt, *inter alia*, state that:

> '[i]t is unquestionably within the *domaine réservé* of a state to determine how it will handle a health crisis, as is the actual handling of that crisis. The scope of this authority is not limited to action carried out by government agencies but instead deals with activities by both government and private health care providers, and any other relevant public health entities. Therefore, if a cyber operation by or attributable to one State obstructs the execution of another state's plan for responding to the pandemic, the former will have engaged in prohibited intervention.'[194]

Tailoring this to content-based activities in the digital realm, they conclude that

190 Section 1.4.4 '(Dis)information Sovereignty and Jurisdiction'.

191 While sovereignty and *domaine réservé* are not identical, when it comes to control over territory, there is 'no distinction between the reach of State sovereignty and the concept of *domaine réservé*', in Kunig (n 143) para 1.

192 Broadcasting Convention (n 21); Convention on the International Rights of Correction (n 22).

193 UNGA Res 44/147, '[…] any extraneous activities that attempt, directly or indirectly, to interfere with the free developments of national electoral processes, in particular in developing countries, or that intent to sway the results of such processes, violate the spirit and letter of the principles established in the Charter and in the Declaration on Principles of International Law concerning Friendly Relation and Co-operation among States in accordance with the Charter of the United Nations', in UNGA Res 44/147 (15 December 1989) UN Doc A/RES/44/147, para 2 (adopted 113 to 23, 11 abstentions); UNGA Res 60/164 'Respect for the principles of national sovereignty and diversity of democratic systems in electoral processes as an important element for the promotion and protection of human rights' (adopted 16 December 2005) UN Doc A/RES/60/164; Schmitt (n 49) 48; Kilovaty (n 150) 100; Lahmann (n 49) 4; Nasu (n 49) 65-77.

194 Milanovic and Schmitt (n 85) 257.

> '[i]f misinformation directly causes part of the target state's crisis management plan to fail and was designed to do so, as in falsely announcing that a particular hospital was no longer receiving COVID-19 patient of that testing at a certain location has ended, our view is that the coerciveness elements if required. […] Such actions would be analogous to undisputed examples of intervention, like manipulating election machinery or altering a vote count. They block a state's ability to execute a plan with respect to its *domain reserve*.'[195]

In the context of electoral interference, Schmitt states elsewhere that

> 'as a general matter, 'the process by which a State selects its officials is left to the determination of [a] State and is broadly unregulated by international law. Accordingly, cyber activities by foreign States that affect either the process by which elections are conducted, or their outcome, qualify as prohibited intervention, so long as the second prong of the intervention test, coercion, is satisfied.'[196]

Electoral processes and public health policies exemplify core sovereign prerogatives. This sphere of States prerogatives, however, extends to other domains: admission of foreign nationals, immigration control, refugee management, national security imperatives, and environmental threats.[197] As demonstrated earlier in this chapter, these traditionally recognised components of the *domaine réservé* are challenged by mis- and disinformation campaigns. The scope of potential interference also extends to targeted xenophobic disinformation against specific nationalities or minorities, as well as disinformation weaponising narratives around national security – another undisputed sovereign prerogative.[198] Analysis of these approaches indicates the limited utility of the *domaine réservé* concept in delineating non-intervention's applicability to disinformation.[199] The risk of over-inclusiveness threatens to expand protected domains beyond reasonable bounds, potentially encompassing virtually all policy areas affected by disinformation. Conversely, a restrictive interpretation fails to acknowledge international law's evolving role in this sphere.

---

195 *Ibid*. 269.

196 Schmitt (n 49) 49.

197 Ziegler (n 150) para 5.

198 European Commission, 'Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Region, "Tackling Online Disinformation: A European Approach"' (2018) COM(2018) 236 Final, 6; Unver and Arhan provide a comprehensive overview of countries that have 'contextualised disinformation as a national security threat', including the United States, Russia, China, the UK, France, Italy, South Africa, Turkey and Kenya, in Akin Unver and Ertan Arhan, 'The Strategic Logic of Digital Disinformation' in Rubén Arcos *et al.* (eds), *Routledge Handbook of Disinformation and National Security* (Routledge 2023) 194.

199 Ziegler (n 150) paras 30-32; Moynihan (n 12) 29 citing Lisa Damrosch, 'Politics Across Borders: Non-intervention and Non-forcible Influence over Domestic Affairs' (1989) 83 The American Journal of International Law 1, 34.

2.4.1.2 Inter-State Norms and Non-State Actors

While chapter six concentrates on formal attribution and non-State actors' responsibility, the inter-State character of the prohibition poses significant obstacles in applying the *domaine réservé* to the online environment.[200] Most significantly is the power asymmetry and the growing power of private entities in the disinformation landscape, including social media platforms, but also data analytics companies, political consulting firms and hacking groups.[201]

At the core of this tension lies the predominant private ownership and maintenance of online information and communications infrastructure.[202] Some scholars argue that, '[s]ince private actors, such as Facebook and Twitter, fulfil no sovereign function, it seems antithetical to consider any activity taking place on these platforms as '*domaine réservé*'.'[203] While this position is descriptively accurate and normatively appealing, whether or not the prohibition of intervention has been violated depends on whether the activities affect government functions, not whether they take place on public platforms. These emerging private powers, nevertheless, do simultaneously represent an evolution towards 'privatization of the *domaine réservé*' that constitutes 'a new type of sovereignty – digital rather than Westphalian'.[204] Ultimately, the private digital transformation fundamentally reconfigures democratic processes and the exercise of governmental functions within the *domaine réservé*. When private platforms control information flows, political discourse, and individuals' data, they exercise quasi-sovereign powers without public accountability. This growing public-private asymmetry challenges the traditional boundaries upon which the scope of the *domaine réservé* rests, making the argument that this digital transformation foundationally reshapes its contours increasingly compelling.

Some instances of non-State actor interference – including the 2012 cyber- and disinformation operations executed by the 'Armenian Cyber Army' and the perpetual disinformation operations by the Russia-controlled Internet Research Agency –[205] can be brought within the prohibition's reach by attributing their conduct to a State in accordance with the law on State responsibility. This equally applies to information operations carried out by

200 Watts (n 71) 253.

201 *Ibid.*; Kilovaty (n 15) 135, 139; Anupam Chander and Haochen Sun, 'Sovereignty 2.0' (2021) Georgetown Law Faculty Publications and Other Works 2024, 22-23; Luciano Floridi, 'The Fight for the Digital Sovereignty: What It Is, and Why It Matters' (2020) 33 Philosophy and Technology 369, 371; Agata Kleczowska, 'Trapped in the Grey Zone – International Law Applicable to Non-State Actors' in Mitt Regan and Aurel Sari (eds), *Hybrid Threats and Grey Zone Conflict: The Challenge to Liberal Democracies* (Oxford University Press 2024) 425-445.

202 Kilovaty (n 15) 136; Joseph Nye, 'Power Shifts' Time (9 May 2011) Accessed 22 November 2024.

203 Kilovaty (n 15) 103; Wingfield and Wingo (n 14) 585.

204 Kilovaty (n 15) 136; Milanovic and Schmitt (n 85) 257.

205 Roscini (n 100) 400.

private military companies, including the Wagner group and the Israeli "Team Jorge",[206] on the instructions of States.[207] Absent a degree of control or formal involvement of a State – or the inability to identify this, as is often the case in the growing digital influence industry –[208] a grey zone exists. The evolving landscape of non-State conduct presents challenges that transcend mere attribution issues: the undelegated *de facto* governance of private entities over the information environment enables States and non-State actors to effectively interfere with domestic political processes, public health, and other *domaine réservé* matters without directly targeting State institutions or sovereign prerogatives.[209] By 'leveraging the reach and scale of social media platforms, microtargeting methods, and manipulation techniques', they can achieve the same goal.[210] Thus, while in theory the non-intervention rule could effectively counter non-State actor interference, its potential is critically hampered by the prevailing State-centric normativity, and – as outlined in chapter six – an outdated reliance on State-control as the doctrinal tool to bridge the gap.[211]

2.4.1.3 The Object Protected

Alternatively, shifting focus from the domain protected from intervention (a State's sovereign prerogatives) to the object of intervention (the ability to make free choices on these matters),[212] aligns better with these changing circumstances, without fundamentally altering the scope of the rule. From the perspective of disinformation's operational reality and cognitive impact, the protected object embodies the process through which States form authoritative decisions and exercise their own political will. This process underlies both internal sovereign functions (e.g. democratic deliberation and policy formation) and external prerogatives (e.g. diplomatic positioning and international co-operation).[213] As Pijpers articulates in his comprehensive work on influence operations in cyberspace, 'the object of an intervention in the reserved domain

206 Africa Centre for Strategic Studies (n 65).
207 Council of the European Union General Secretariat, 'The Business of War – Growing risks from Private Military Companies' (31 August 2023) Analysis and Research Team Research Paper, Accessed 20 November 2024.
208 Emma L Briant, 'Researching Influence Operations: 'Dark Arts' Mercenaries and The Digital Influence Industry' in Corneliu Bjola and Ilan Manor, *The Oxford Handbook of Digital Diplomacy* (Oxford University Press 2023) 80.
209 Kilovaty expressly references the of use of deep-fake technology – which generally falls within the concept of disinformation, in Kilovaty (n 150) 104; See also section 1.3.1 'Technological Development and Digital Infrastructure'.
210 Kilovaty (n 15) 139
211 *Ibid.;* section 6.2.2 'Attribution'.
212 Tsagourias (n 150) 51.
213 *Ibid*.

is related to undermining the State's ability to make free choices in the political system and organisation.'[214]

Many commentators, either implicitly or explicitly, support this approach. Notably, Ossoff argues that the primacy of the notion of *domaine réservé*, sparked by the ICJ's wording in the *Nicaragua case* (i.e. 'the choice of a political, economic, social and cultural system, and the formulation of a foreign policy') never sufficiently captured the essence of the prohibition. In support of this argument he refers to the Friendly Relations Declaration, which formulates that '[e]very State has an inalienable right to choose its political, economic, social and cultural systems.'[215] This approach prioritises States' capacity 'to decide freely'[216] and maintain 'independent authority to make choices among various lawful courses of action on a subject regulated by international law.'[217] While this theoretical reorientation is convincing in the context of influence operations, it does raise the question: should Ossoff's conception be narrowly construed as centralised governmental decision-making, or broadly understood through the lens of non-intervention and the free will of the State as a whole, including public opinion?[218] The latter posits that 'a government's authority and will remain free only when its sourcing is also free,'[219] introducing the dimension of self-determination of the people of a State into the scope of the non-intervention doctrine.

Self-determination

The internal dimension of the *domaine réservé indeed* includes the right of a State's population to self-determination,[220] encompassing the collective right of people to express their independent sovereign will by freely choosing their

214 Pijpers (n 49) 153; Tsagourias (n 150) 48.

215 Ossoff (n 166) 307.

216 This is also the position adopted by the Tallinn Manual: '[i]ntervention into the *domaine réservé* of a State need not be directed at State infrastructure or involve State activities. Rather, the key to satisfaction of this first element of intervention is that the act in question must be designed to undermine the State's authority over the *domaine réservé*', in Tallinn Manual 2.0 (n 43) Rule 66, Commentary 11.

217 Moynihan (n 12) 34.

218 Self-determination in this context 'refers to the right of peoples to determine freely and without external interference their political status and to pursue freely their economic, social, and cultural development' in Tsagourias (n 150) 51 citing International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 Article 1(1) [hereafter: ICCPR]; UNGA A/RES/2625 (n 170).

219 Tsagourias (n 150) 51; Jens Bartelson, 'Sovereignty and the Personality of the State' in Robert Schuett and Peter M R Stirk (eds), *The Concept of the State in International Relations* (2015 Edinburgh University Press) 90, '[…] the personality of the state today more often is believer to derive from the unity of the represented […]'. This idea, however, precedes the present debate and traced back to the writings of Pufendorf and Vattel, in Samuel Pufendorf, *Of the Law of Nature and Nations* (Basil Kennet tr, Oxford University Press 1729) 635.

220 Pijpers (n 49) 151; Tsagourias (n 150) 51-52; Roscini (n 100) 247-309.

form of governance.[221] In connection with the non-intervention rule, Tsagourias explains, this right '[…] does not cease once a State has been created, but thereafter self-determination refers to the 'right to authentic self-government, that is, the right of a people really and freely to choose its own political and economic regime'.[222] The prohibition thus protects against external interference in the expression of authority and will by the people, and protects the conditions that enable the people to form authority and will freely, and make free choices.[223] The main implication of this normative alignment between the two concepts, Tsagourias continues, is that:

> 'the domain and object of intervention shift from the government to the actual power holder, the people, and to the process of forming authority and will through which the goal of free choice is also attained. Whereas the government as the depository of such authority and will is protected by the principle of non-intervention, it is not the primary object of protection as the traditional reading holds, but a derivative one; the primary object of protection are the people and the process of authority and will formation.'[224]

Other scholars echo this interpretation. Efrony and Shany notably state that '[the coercive aspect of] intervention is not confined to operations curbing governmental powers and policies, but also to the freedom of choice exercised by the people of a state.'[225] Moynihan advances this position, arguing that a State and its population in this area are inseparable – an attempt to manipulate the will of people in another State thus amounts to an attempt to undermine that State.[226] Notwithstanding that this interpretation remains contested, Kohl's observation that it is 'useful to remind oneself that sovereignty, and implicitly jurisdiction, seek to protect the self-determination of a people and legal diversity across the globe', remains pertinent.[227]

Free Will of the State and the People

Chapter one's analysis on the right to hold and form opinions without interference alongside emerging concepts of cognitive liberty, argued that disin-

221 ICCPR (n 218) Article 25.

222 Tsagourias (n 150) 51; Antonio Cassese, *Self-Determination of Peoples: A Legal Reappraisal* (Cambridge University Press 1995) 137.

223 Tsagourias (n 150) 51 stating that '[a]ccording to the Universal Declaration on Human Rights, Article 21(3): "[t]he will of the people shall be the basis of the authority of the government", citing UNGA Res 217A (III) (10 December 1948) UN Doc A/810.

224 Tsagourias (n 150) 52; Roscini (n 100) 399-400.

225 Dan Efrony and Yuval Shany, 'A Rule Book on the Shelf? Tallinn Manual 2.0 on Cyberoperations and Subsequent State Practice' (2018) 112 American Journal of International Law 583, 641.

226 Moynihan (n 12) 42.

227 Uta Kohl, 'Jurisdiction in network society' in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar 2021) 77.

formation distorts individuals' perception of reality and influences the process of opinion formation. The collective expression of genuine, unimpeded opinions of a State's population constitutes a population's free choice.[228] If disinformation indeed inherently undermines this free choice and strategically targets those who represent the population,[229] a direct risk to a State's capacity for sovereign decision-making emerges.

Both aspects are represented in the doctrine on self-determination. Particularly in the context of electoral processes, foreign influence converts self-determination into 'other-determination', making election outcomes – partly – a reflection of the will of 'outside entities rather than the entity that is holding the election.'[230] Ohlin, suggesting that the self-determination angle is most suitable to address electoral interference, emphasises that deliberation and debate are key, enabling processes that are contingent upon legally protected freedom of thought, freedom of speech, freedom of the press and unrestricted access to information.[231] Access to fraudulent or corrupted information compromises informed deliberation and decision-making processes.[232] This distortion stands in stark contrast with self-determination: it impedes independent opinion formation and subsequent autonomous decision-making,[233] particularly when its credibility is amplified through synthetic content and the concealment or falsification of the actors behind it.[234]

The use of deceptive technologies indicate prohibited intervention when they, as Tsagourias notes, are 'designed and executed in such a way as to manipulate the cognitive process where authority and will are formed and take control over peoples' choices of government.'[235] When disinformation, he continues, is combined with identity falsification, this further 'distorts, undermines, or inverses this process and nullifies the genuine expression of authority and will by the people'.[236] Misrepresentation or concealment of this identity often indicates that the communicator 'engages in subversion that

228 For the shifting conception of the idea of "sovereign will" within the literature on civil strife, see Tay (n 151) 72-75; Breedon (n 6) 685-685.

229 C Athony Pfaff, 'Coercing Well' in Mitt Regan and Aurel Sari (eds), *Hybrid Threats and Grey Zone Conflict: The Challenge to Liberal Democracies* (Oxford University Press 2024) 369.

230 Ohlin (n 57) 13.

231 *Ibid.* 98-100.

232 *Ibid.* 100-102 citing William Lewis, 'War, Manipulation of Consent, and Deliberative Democracy' (2008) 22 The Journal of Speculative Philosophy 4, 271; See also Pfaff (n 229) 369.

233 The Human Rights Committee has repeatedly touched upon the importance of information during elections, in e.g. Human Rights Committee, 'General Comment No 25: The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service (Article 25)' (12 July 1996) UN Doc CCPR/C/21/Rev.1/Add.7, para 12.

234 Kilovaty (n 150) 105; Unver and Arhan (n 198) 194; Bobby Chesney and Daniella Citron, 'Deep fakes: A looming challenge for privacy, democracy and national security' (2019) 107 California Law Review 1754, part II.B.2.

235 Tsagourias (n 150) 54.

236 *Ibid.* 51; Jens D Ohlin, 'Election Interference: The Real Harm and the Only Solution' (2018) 18-50 Cornell Law School Research Paper, 1-26.

undermines and sabotages the political process, an act that should be proscribed by the prohibition on intervention'.[237] In contrast, is it presumed that when the identity of the actor behind the information operation is known, both the State and the people are able to question the genuineness of such information.[238]

This reorientation of the *domaine réservé* and alignment between the object of the rule and new forms of interferences diverts from the traditional non-intervention paradigm. As Kilovaty persuasively argues, this shift in focus from 'State authority and control to the people who collectively shape the decision-making of their respective government' is necessary to 'respond to the challenging nature of technology and the digital information space'.[239] Notwithstanding the importance of highlightingthis evolution to preserve the principle's relevance, the non-intervention rule's utility in responding to modern disinformation threats ultimately depends on whether disinformation amounts to 'coercion' and under what conditions.

2.4.2 The Meaning of Coercion

Coercion differentiates between unfriendly acts and minor interferences, and intervention in contradiction with international law. 'Interference pure and simple', Oppenheim notes, 'is not intervention.'[240] Though the requirement is generally undisputed,[241] international law remains inconclusive regarding what constitutes coercion or coercive behaviour, with States upholding considerably diverging interpretations.[242] The fact that coercion is not by defini-

237 Helal (n 14) 116; Pijpers (n 49) 159-169; Kilovaty (n 150) 90-91; as a step in the right direction, David Sloss suggests that social media users must declare their nationality and that State agents must be banned completely from social media, in part because they rely on fictious identities, in David L Sloss, *Tyrants on Twitter: Protecting Democracies from Chinese and Russian Information Warfare* (Stanford University Press 2022) 159, 168-169.

238 Helal (n 14) 116.

239 Kilovaty (n 150) 105.

240 Jennings and Watts (n 41) 432.

241 There are some scholars that, whilst not denying it to be an element, substantially attribute little to no consideration or value to it, in Antonios Tzanakopoulos, 'The Right to Be Free from Economic Coercion' (2015) 4 Cambridge Journal of International and Comparative Law 616, 623; Evan Criddle, 'Humanitarian Financial Intervention' (2013) 24 European Journal of International Law 583, 591.

242 Jens Ohlin (n 172) 1581; Where Helal considers 'persuasion' to antithesis of coercion, Buchan states 'influence' as opposite coercion, in Helal (n 14) 72; Russell Buchan, 'Cyber Attacks: Unlawful Uses of Force or Prohibited Intervention' (2012) 17 Journal of Conflict and Security Law 2; Pontus Winther, *International Humanitarian Law and Influence Operations: The Protection of Civilians from Unlawful Communication Influence Activities during Armed Conflict* (2019) Dissertation Uppsala University, 223-225; Mohamed Helal, 'Coercion in Cyberspace – Taking Stock of the Debate' (Ohio State Legal Studies Research Paper No 877, 2024) 1-41.

tion unlawful further exacerbates this inconclusiveness;[243] coercive measures can play an important role in the enforcement of international law and occasionally in restoring international peace and security.[244] Consequently, 'coercion' in itself has no normative significance.[245] To some authors, the articulation of the threshold is therefore even 'so vague as to be almost useless',[246] making it underinclusive and thus failing to protect the essential interests of States.[247] In response, States and experts have favoured either a reinterpretation of the coercion threshold whilst relying on the traditional doctrinal approach, or suggest a more radical departure from the existing definition of non-intervention towards a new norm.[248]

Initially understood as forcible or dictatorial action,[249] 'coercion' was often equated with the use of force.[250] Gradually, this interpretation has become less restrictive: the rules on the use force are now but one specific application of the principle.[251] In the 20th century, States and commentators introduced the idea of propaganda as a potential form of coercion, viewing it as a threat to both territorial integrity and political independence.[252] This understanding has expanded further in recent decades, with growing support for applying the concept of coercion to influence and information operations.[253] A spectrum of 'pathways of coercion' emerged, encompassing both direct forms

243 Michael W Reisman, 'Coercion and Self-Determination: Construing Charter Article 2(4)' (1984) 78 The American Journal of International Law 3, 642-645; Tom J Farer, 'Political and Economic Coercion in Contemporary International Law' (1985) 79 The American Journal of International Law 2, 405-413; Jeremy Rabkin and John Yoo, 'A Return to Coercion: International Law and New Weapon Technologies' (2014) 42 Hofstra Law Review 1187-1226; Omri Sender, 'Coercion' (2021) Max Planck Encyclopaedia of Public International Law.

244 Sender (n 243) para 18.

245 Farer (n 243) 406.

246 Derek Bowett, 'International Law and Economic Coercion' (1976) 16 Virginia Journal of International Law 2, 245, 248.

247 Efrony and Shany (n 225) 641; Corn (n 16) 212.

248 Helal (n 141) 82-83; Pia Hüsch (n 46) 384; Kilovaty (n 150) 88; Wu (n 170) 276-280; Steven J Barela and Samuli Haataja, 'Rethinking the International Law of Interference in the Digital Age' in Mitt Regan and Aurel Sari (eds), *Hybrid Threats and Grey Zone Conflict: The Challenge to Liberal Democracies* (Oxford University Press 2024) 398.

249 Kunig (n 143) para 5.

250 *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 7) para 205; Kunig (n 143) para 6; some scholars still lean towards such an understanding, in Natalino Ronzitti, 'Respect for Sovereignty, Use of Force and The Principle of Non-Intervention in the Internal Affairs of Other States' (European Leadership Network 2015) 3.

251 Moynihan (n 12) 29; Milanovic (n 13) 613; Corn (n 16) 215-216; Helal (n 141) 103; Jamnejad and Wood (n 143) 348-349; Wu (n 170) 263-263.

252 Roscini (n 100) 149, 178; Vincent (n 173) 312, 316; *Inter alia,* Australia, China, Israel. Kenya, Rwanda, Saudia Arabia, Thailand and Upper Volta (now Burkina Faso) argued during the drafting of the 1965 Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States that intervention could be undertaken 'through the means of propaganda, psychological warfare, radio broadcasts and other mass media' in Helal (n 141) 64.

253 Moynihan (n 12) 29; Milanovic (n 13) 613; Helal (n 14) 75-76; Terry (n 56) 88; Watts (n 71) 11-12; Kunig (n 140) para 6; Damrosch (n 199) 5.

– such as forcible means and dictatorial interference – and more subtle, indirect approaches.[254] Coercion now manifests across multiple dimensions, through forcible and non-forcible means,[255] resulting in context-dependent interpretations that establish it as a dynamic rather than static concept.[256]

Applying 'coercion' to influence operations while upholding the ICJ's Nicaragua Judgment characterisation of coercion as 'the very essence' of the non-intervention rule,[257] centralises Oppenheim's authoritative understanding of coercion as constituting control.[258] In light of the doctrine's overarching purpose – preventing subversion –[259] the essence of *coercive* influence operations is thus that they deprive the target State of control.[260] This consequentially limits the scope of the non-intervention rule to subversive disinformation that is designed to deprive another State of its freedom of choice.[261]

Coercion-as-control

Applying the coercion-as-control interpretation to disinformation depends on whether the deception and manipulation of individuals' views, and ultimately

254 Tsagourias (n 150) 53; Wheatley (n 29) 168 citing Joel Feinberg, *Harm to Self* (Oxford University Press 1986) 189; the recent EU position also expressly underlining that the rule 'prohibits a State from directly or indirectly intervening in the affairs of another State', in Council of the European Union, 'Declaration on a Common Understanding of International Law in Cyberspace' (18 November 2024) 15833/24 ANNEX, para 2.

255 Sender (n 243) paras 5-13, 22.

256 *Ibid.* para 4; Milanovic (n 13) 614-615.

257 Some experts argue that this characterisation is unfounded, in Corn (n 16) 215-217; Helal (n 141) 82.

258 Kilovaty (n 15) 134; Wu (n 170) 263.

259 Corn (n 16) 217; See also the position of, *inter alia*, the European Union, the Netherlands, Switzerland, Finland, Israel, Estonia, Norway, Romania, New Zealand, Australia and Germany, in Council of the European Union, 'Declaration on a Common Understanding of International Law in Cyberspace' (18 November 2024) 15833/24 ANNEX, para 33; Netherlands Minister of Foreign Affairs, 'Letter of 5 July 2019 to the Parliament on the International Legal Order in Cyberspace, Appendix: International Law in Cyberspace' (Government of the Netherlands, 5 July 2019) 3; Ministry of Foreign Affairs of Finland, 'International Law and Cyberspace: Finland's National Positions' (15 October 2020) 3; New Zealand Foreign Affairs and Trade, 'The Application of International Law to State Activity in Cyberspace' (1 December 2020) para 9; Commonwealth of Australia, Department of Foreign Affairs and Trade, 'Australia's International Cyber and Critical Tech Engagement Strategy' (2021) 9; Federal Government of Germany, 'On the Application of International Law in Cyberspace, Position Paper' (March 2021) 4-5; Switzerland Federal Department of Foreign Affairs, 'Switzerland's Position Paper on the Application of International Law in Cyberspace' (2021); 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, 403.

260 Ido Kilovaty (n 150)105 citing Jennings and Watts (n 41) 430-439.

261 This embodies the first part of the Tallinn Manuel's definition of coercion as 'an affirmative act designed to deprive another State of its freedom of choice, that is, to force that State to act in an involuntary manner or involuntarily refrain from acting in a particular way', in Tallinn Manual 2.0 (n 43) Rule 66 Commentary 18.

behaviour, constitutes control. Critics of classifying disinformation as a form of unlawful coercion opine that merely providing false or misleading information 'to influence someone's decision-making process does not bend the will of the target and […] its will is also not replaced' and thus falls short of prohibited intervention.[262] They maintain that coercion 'implies compulsion with some degree of *forcible* conduct' and that therefore 'deceptive manipulation by way of a disinformation campaign cannot be conceived as coercion.'[263]

These objections refer to the Tallinn Manual 2.0 which expressly distinguishes coercion 'from persuasion, criticism, public diplomacy, propaganda … retribution, mere maliciousness, and the like in that, unlike coercion, such activities merely involve either influencing (as distinct from factually compelling) the voluntary actions of the target State.'[264] Notably, however, this represents a narrower conception of coercion than the Manual's first edition, which identified 'the manipulation by cyber means of elections or of public opinion on the eve of elections, as when online news services are altered in favour of a particular party, false news is spread, or the online services of one party are shut off' as prohibited forms of intervention.[265]

In contrast, most authors reject a categorical exclusion of influence operations from the scope of 'coercion', instead emphasising the need to differentiate between lawfully persuasive and unlawfully coercive information operations online. Political scientist Joseph Nye articulates this perspective by focussing on the degree of manipulation, rather than dichotomously separating it from coercion. If, he argues, 'the degree of manipulation is so deceptive that it destroys voluntarism, the act becomes coercive' – a position increasingly reflected in States' positions.[266] This emphasis on voluntarism, commonly equated with 'free will' or 'freedom of choice' reflects the outlined protected

262 Roscini (n 100) 397.

263 Lahmann explores these objections, in Lahmann (n 49) 201.

264 Interestingly, though the Experts were not undivided, from the outset this is somewhat contradictory with its position stated in the context of State sovereignty, where they concluded that the transmission of propaganda generally does not violate sovereignty, but that 'the transmission of propaganda, depending on its nature, might violate other rules of international law. For instance, propaganda designed to incite civil unrest in another State would likely violate the prohibition of intervention (Rule 66)' in Tallinn Manual 2.0 (n 43) Rule 4 Commentary 29, Rule 66, Commentary 21.

265 Michael Schmitt (ed), *Tallinn Manual on the International Law Applicable to Cyber Warfare* (Cambridge University Press 2013) 45; Sander (n 47) 24.

266 Joseph Nye, 'Protecting Democracy in an Era of Cyber Information Was' (13 November2018) Governance in an Emerging New World; Canadian Government, 'International Law Applicable in Cyberspace' ('designed to deprive the affected State of its freedom of choice'); New Zealand Foreign Affairs and Trade (n 259) para 9(b) ('an intention to deprive the target state of control over matters falling within the scope of its [IHS']); Czech Ministry of Foreign Affairs, 'Position Paper on the Application of International Law in Cyberspace' (27 February 2024) para 9 ('[c]oercion is an activity intended to deprive, directly or indirectly, the State of its ability to exercise control or govern matters within its internal and external affairs […] coercive, i.e. intentionally aims to influence the State's free will and choice').

interests of the *domaine réservé*. In this line of reasoning, 'a disinformation campaign, cyber or otherwise, can severely impact on the victim state's ability to control its internal affairs, namely its political system.'[267]

Coercion-as-extortion

Beyond coercion-as-control, coercion retains its traditional meaning of dictatorial force, encompassing pressure that amounts to extortion through threats. This represents overt manipulative behaviour seeking 'an advantage of some kind by depriving the target state of its free will over the exercise of its sovereign powers.'[268] In this regard, coercion encompasses actions 'aimed at coercing a State to do *or abstain from doing* something it is entitled to do under international law [emphasis added]' as well as those 'actions *merely restricting a state's choice* with respect to a course of action [emphasis added].'[269] In either case, the essence of coercion lies in inter-State pressure designed to vitiate the target State's autonomous exercise of sovereign rights, thereby compelling particular outcomes (coercion-as-extortion) or resulting in a 'denial of the ability to pursue the choices that it wanted to pursue (coercion-as-control).[270]

Disinformation can manifest as both forms of coercion: as a credible threat perceived as extortion or as manipulative control. Evaluating whether it meets either threshold requires assessing its foreseeable consequences, employed means, and actor(s) intent.[271] While some scholars have suggested these factors in isolation, no single element provides a framework that can successfully identify 'coercive disinformation'. The analysis requires cumulative consideration of:

1. Scale, intensity and magnitude of the disinformation operation or campaign, which indicates the likelihood of successful intervention (its coercive effect) (2.4.2.1).

267 Milanovic (n 13) 648; Breedon (n 6) 680-681 ('the question is whether cyber-based information warfare using conspiracy theories and disinformation to manipulate democratic opinion-making in the target State qualifies as coercion under the non-intervention principle. The general consensus appears to be yes').

268 Moynihan (n 12) 31; Milanovic (n 13) 617.

269 Moynihan (n 12) 31 citing Watts (n 71) 256; Other authors similarly recognise coercion as a form of control, as well as extortion. Tsagourias, *inter alia*, in referencing the Friendly Relations Declaration, describes coercion as 'to imply compulsion whereby one state compels or attempts to compel another state to take a particular course of action *against its will* […] [emphasis added]', in Tsagourias (n 150) 48; Schmitt speaks of coercion as subordination of the sovereign will of the target State, in Schmitt (n 15) 51; Terry imposes an equally high threshold, opining that 'interference turns coercive when the targets State cannot terminate the outside meddling at its pleasure, in Terry (n 56) 70.

270 Milanovic (n 13) 617; Moynihan (n 12) 32.

271 Several States argue 'it suffices that a State intends to coerce another State, employs coercive methods, *or* eventually causes coercive effects in another State', in e.g. Costa Rica (n 158) para 24.

2. Tools and methods of creation and dissemination, which suggest both potential coercive effect and intent, though not decisively (2.4.2.2).
3. Intent as a requisite element of intervention, applicable to both coercion-as-extortion and coercion-as-control, with heightened significance in the latter (2.4.2.3).

Finally, the argument that knowledge on behalf of the victim State that it is being coerced is required, is rejected (2.4.2.4). Such knowledge serves 'merely' as probative evidence for claims of coercive intervention. The analysis substantiates the argument that disinformation becomes coercive when these elements collectively demonstrate an intentional campaign of sufficient reach, scale and intensity to undermine the free will of the state.

2.4.2.1 Consequence and Intensity

For disinformation to constitute coercion, it must (be able to) deprive or substantially impair the target State's independence in governance.[272] This requirement raises three fundamental questions: must an intervention succeed to breach the non-intervention rule? What linkage must exist between the act and its coercive effect if success is not required? And lastly, how can such linkage or nexus be established and evidenced – which is outlined in chapter six.[273]

Some experts contend that prohibited intervention only occurs if 'the internal or external process is successfully disrupted by a foreign power […] and the victims suffers severe domestic or international consequences.'[274] The prevailing view, however, maintains that intervention does need not to succeed to constitute a breach of the rule.[275] The ICJ's language in the *Nicaragua* case – referring to 'methods of coercion' and 'with a view to coerce' – implies that success is indeed not decisive.[276] As Moynihan aptly analogises, '[j]ust as an act involving use of force that misses its target would still constitute a use of force.'[277]

Establishing a causal link between disinformation campaigns or operations and coercive effect is thus unnecessary, and could paradoxically even encour-

272 Corn (n 16) 218.
273 Section 6.2.3 'Causality'; see also section 1.5.1.4 '(Likelihood of) Harm'.
274 Ido Kilovaty, 'Doxfare: Politically Motivated Leaks and the Future of the Norm on Non-Intervention in the Era of Weaponized Information' (2019) 9 Havard National Security Journal 146-179, 173.
275 Wingfield and Wingo (n 14) 585; Corn (n 16) 218; Wheatley (n 29) 168; Helal (n 141) 112; Tallinn Manual 2.0 (n 43) Rule 66 Commentary 29 ('[t]he international group of experts agreed that the fact that a coercive cyber operations fails to produce the desired outcome has no bearing on whether this Rule has been breached').
276 Milanovic (n 13) 646; Wheatley (n 29) 183 citing *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 7) paras 205, 241.
277 Moynihan (n 12) 32; Corn (n 16) 218; Roscini (n 100) 382-383.

age more coercive attempts, as failed interventions would escape legal consequences.[278] Requiring causality would constrain the rule's application to disinformation, challenging the framework's broader credibility. For example, military intervention exemplifies coercion with clear causal links and reversible effects – the coercive military interference can be reversed if the forces withdraw. The coercive effect(s) of subversive disinformation, however, often prove (partly) irreversible. Interdisciplinary research demonstrates the continued influence of disinformation, its resistance to correction and tendency to become even more strongly entrenched when correction is attempted.[279] While potential irreversibility does not determine doctrinal boundaries *per se*, it would be normatively appealing to differentiate between reversable – or remediable – forms of coercion and those with potentially permanent effects.

Still, establishing a minimum threshold of foreseeability or likelihood between the act (disinformation) and effect (undermining of the State's free will) is imperative to prevent overinclusiveness.[280] The coercive effect the non-intervention rule seeks to prevent is inherent to disinformation's nature: it aims to alter individuals' views and subsequent behaviour in ways harmful to other individuals or society by spreading false or misleading information.[281] Subversive disinformation in particular operates by creating confusion about facts, and 'undermining confidence in the capacity of the democratic system to deliver the best policy outcomes.'[282] Introducing false facts into the decision-making process may thus be perceived as coercive in the sense of exercising control. As Baade argues, 'since the projection of a different set of facts constrains one's freedom to act by making certain options and conclusions no longer seem viable or making others seem mandatory.'[283] This position finds support in well-established general principles from domestic jurisdictions where measures of deception 'are commonly recognised […] as cognisable harm precisely because they are means of undermining the exercise of free will.'[284]

This reveals disinformation's *prima facie* coercive nature through dual mechanisms: first, it actively manipulates public debate and sovereign decision-making through intentional falsehoods and second, it simultaneously drowns out other voices. Beyond spreading falsehoods, disinformation systematically

278 Schmitt concludes that 'it remains unresolved whether coercion requires a direct causal nexus between the act in question and the coercive effect', in Schmitt (n 15) 51; Terry (n 56) 91.

279 Jonas de Kaarsmaecker and Arne Roets, ''Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions' (2017) 65 Intelligence 107-110.

280 The Tallinn Manual captured this requirement as 'indirect causation of coercive effect', in Schmitt (n 15) 51.

281 Section 1.3.2 'Persuasion and Manipulation Techniques'; Sender (n 243) para 13.

282 Wheatley (n 29) 193.

283 Watts (n 71) 261.

284 Corn (n 16) 220.

denies access to accurate and trustworthy information, manipulates cognition through bias, creates information isolation and covertly distorts perceptions of reality.[285] Bhagevatula Satyanarayana Murty, in this seminal work on *The International Law of Propaganda* from 1968, identified this dual harm, arguing that both amount to psychological pressure constituting coercion by subjecting people to 'a high degree of constraint in the choice of alternatives in shaping his conduct.'[286]

Disinformation and coercion-as-extortion

International law recognises that unmaterialised harm can qualify as coercion, evidenced by Article 2(4) of the UN Charter's prohibition on threats of use of force. As examined in 3.2.2, disinformation can manifest as such threats – an unequivocally recognised act of coercion. As a cornerstone of international law, Article 2(4) UN Charter set out that '[a]ll Members [of the United Nations] 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.'[287] While the parameters defining a 'threat of use of force' remain contested,[288] two elements of 'threat' are recognised: it constitutes a communicative act that is 'intended to be conveyed to the target State' either explicitly or in an implied manner.[289] Additionally, experts propose that for a threat of force to be unlawful, it must demonstrate credibility.[290]

If, for instance, State A threatens State B with an imminent nuclear attack unless territory is surrendered, and State B perceives this as credible, such communication risks constituting an unlawful threat of force and thus a

285 Wheatley (n 29) 189; sections 1.3.1 'Technological Development and Digital Infrastructure'; 1.3.2 'Persuasion and Manipulation Techniques'.

286 Bhagevatula Satyanarayana Murty, *The International Law of Propaganda: The Ideological Instrument and World Public Order* (Yale University Press 1989) 28.

287 UN Charter (n 9) Article 2(4).

288 Marco Roscini, *Threats or Armed Force and Contemporary International Law* (2007) 54 Netherlands International Law Review 229, 231; Nicholas Tsagourias, 'The Prohibition of Threats of Force' in Nigel White and Christian Henderson (eds), *Research Handbook on International Conflict and Security Law* (2013) 67; Duncan B Hollis and Tsvetelina van Benthem, 'Threatening Force in Cyberspace' in Laura A Dickison and Edward W Berg (eds), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold* (Oxford University Press 2023) 55-89; Erin Pobjie, 'Prohibited Force: The Meaning 'Use of Force' in International Law' (Cambridge University Press 2024) 132-158.

289 Helal (n 14) 73; Tallinn Manual 2.0 (n 43) Rule 70.

290 Hollis and van Benthem argue that credibility must be assessed by looking at 'among others, the author's identity and capacities, the relationships between States, imminence, and the nature of the threat and what is threatened', in Duncan B Hollis and Tsvetelina van Benthem, 'Threatening Force in Cyberspace' in Laura A Dickison and Edward W Berg (eds), *Big Data and Armed Conflict: Legal Issues Above and Below the Armed Conflict Threshold* (Oxford University Press 2023) 59.

coercive intervention.[291] This risk intensifies when the threat incorporates strategically generated narratives using fabricated documents, images, videos, and artificially created footage of military operations and practice sites – all synthetically fabricated to create a credible perception a State A's capability and intent to follow through on its threat. Significantly, even absent an actual threat, a convincingly constructed perception thereof may destroy a State's voluntarism and thus qualify as coercion.[292]

The threat-based analysis of disinformation captures, however, only one form of coercive disinformation. The coercion-as-control logic better captures disinformation's harm, although most disinformation operations will ultimately fail to meet the invasiveness threshold required for intervention. Only disinformation which likely cause *significant* effects –[293] opposite *any kind* of negative effect – can rise to the level of coercion-as-control. Although international law remains inconclusive on whether this threshold is indeed doctrinally required, the diversity in disinformation campaigns and operations – in terms of scale and level of manipulation – entails that disinformation cannot be categorically in- or excluded from the notion of coercion.

In light of the established inchoate nature of the non-intervention rule, this most accurate way to determine which disinformation does, or does not, amount to coercion, is though observing the actual impact, but instead to rely on the weighted assessment of an information operation's scale, reach, and intensity.[294] A disinformation campaign, Lahmann illustrates, 'that aims at disrupting the target state's public health efforts during a pandemic that no one notices or takes up will entirely fail to produce any coercive effect and thus would not constitute a violation of the principle of non-intervention.'[295] However, how this assessment should take form is not subject to any consensus, making its existence and parameters *lex ferenda* at best.

Substantiating the assessment of significant effects can rely on established theories and models. McDougal and Feliciano's proposed 'consequentiality' test remains authoritative for assessing behaviour's coercive effect.[296] While

---

291 Justin Ling, 'Russia Is Ramping Up Nuclear War Propaganda' (WIRED, 4 November 2022) Accessed 22 April 2025.

292 Urs (n 103) 26.

293 *Ibid.* 31.

294 Watts (n 71) 257; Jamnejad and Wood (n 143) 348 ('[o]nly acts of a *certain magnitude* are likely to qualify as 'coercive'); Tsagourias (n 150) 50 ('because of the particular features of cyberspace, such as its interconnectedness and anonymity, the pathways or coercion can diversify whereas the scalability, reach and effects of intervention enhance'); Tay (n 151) 47 ('coercion, on its plain meaning, only contemplates interpositions of sufficient magnitude'); Steven J Barela, 'Cross-Border Ops to Erode Legitimacy: An Act of Coercion' (2017) Just Security ('the significance and expanse, both in scale and reach).

295 Lahmann (n 49) 424.

296 Myres S McDougal and Feliciano P Florentino, 'International Coercion and World Public Order: The General Principles of the Law of War (1958) 67 Yale Law Journal 771, 782-783;

they depart from actual, rather than potential harm, their assessment contains valuable, transferable indicators: first, 'the importance and number of values affected'; second, 'the extent to which such values are affected'; and third, 'the number of participants whose values are so affected'.[297] Corn, in this work on covert deception, aligns this test with the prophylactic nature of the non-intervention rule, concluding that these factors facilitate an 'assessment of the inverse relationship between the relative value of the targeted interests and the anticipated extent to which the interest will be affected.'[298] Applied to disinformation, this implies, first, that there are necessarily qualitative differences 'among the bundle or rights falling within the *domaine réservé* .' For heavily weighed interests such as elections, 'there should be lower tolerance for interventions aimed at undermining their independence.'[299] Hence the threshold for establishing significant effects – and, consequently, for the qualification of coercion – is lower. The rationale behind this differentiation lies in the varying susceptibility of interests to manipulation. Interests that are associated with higher degrees of emotional decision-making – such as elections or immigration policy, or that require decision-makers to process large amounts of nuanced information, are more easily 'significantly' influenced.

Second, the weighting of these three considerations varies by State, depending on its vulnerability to disinformation, population susceptibility, and dependence on the coercing State.[300] The importance of including this variance into a legal assessment is evident when, for example, comparing intervention through disinformation by Russia in the 2017 European Elections and Russia's perpetual undermining of the electoral processes in various African States that are in the precarious process of democratic transition.[301]

Many developed States possess substantial resources – financial, technological, and human capacity – to detect, discredit, deter, and prosecute intervention. They also benefit from established democratic governance with supervisory and corrective mechanisms, alongside preventive and responsive measures (including digital literacy and education). In other words, while the essence of coercion-as-control is uniform, its application to inter-State relations is not. Hence, a more standardised approach requires that these contextual circumstances weight into the assessment of (potential) consequences, Together with the subsequently elaborated tools and methods, elements of covertness, falsity of information and strategic coordination, this presents a model for scenarios where a strong presumption of coercion exists in disinformation.

Kilovaty (n 15) 142; Corn (n 16) 224-225; Watts (n 71) 257; Barela and Haataja (n 248) 418-419.

297 McDougal and Florentino (n 296) 782-783.

298 Corn (n 16) 225.

299 *Ibid.*

300 Wu (n 170) 278.

301 Africa Centre for Strategic Studies (n 65).

2.4.2.2 Tools, Features and Methods

Commentators including Wue, Helal and Ohlin have argued that 'the *means* of intervention must involve 'methods of coercion',[302] and that 'to constitute coercion, the coercing state must use *unlawful instruments* to compel the coerced state to comply with its demands [emphasis added].'[303] Technological advancement and growing non-State actor involvement in inter-State communications, however, increasingly enable States to 'unduly and substantially influence international or external affairs of another State,'[304] without relying on predetermined coercive tools or methods.[305] While the means and methods of interference may indeed indicate coercion, restricting the determination of coercion to such a narrow interpretation, 'ignores the ability of other uncoercive methods – such as manipulation, deception, disruption, and disinformation – to trigger the involuntary actions of the victim state.'[306] This is particularly true for disinformation: while neither inherently coercive nor unlawful,[307] it can become coercive when instrumentalised to restrict the audiences' choices through 'sophisticated techniques to undermine the authority and agency of individuals by exploiting their human vulnerabilities.'[308] Especially disinformation and other information operations that are not *per se* unlawful, are 'ideal means of coercion', because they are generally low in cost and difficult to attribute.[309]

In differentiating between coercive and non-coercive disinformation, four features indicate coercion. Beyond foreseeable harm and malicious intent, subversive disinformation generally instrumentalises: (1) false information,

---

302 Wu (n 170) 262.

303 Helal (n 14) 73; Terry (n 56) 83; Ohlin (n 57) 83 ('the legal requirement of coercion does not track the size of the operation or the result; it tracks the methods used A small and insignificant operation could satisfy the requirement of coercion, while at the same time a massive and devastating operation might fail to satisfy the coercion prong').

304 Kilovaty (n 15) 135; Duncan Hollis, 'Russia and the DNC Hack: What Future for a Duty of Non-Intervention' Opinio Juris (25 July 2016), Accessed 20 November 2024.

305 Kilovaty (n 150) 88; Common Position of the African Union on the Application of International Law to the Use of Information and Communication Technologies in Cyberspace (2024) para 29, Accessed 21 November 2024 ('[t]he prohibition on intervention applies to the use of any instrument, including armed, political, economic, or any other means, and instruments of information, that may be used by a State for the purposes of intervening in the internal or external affairs of a foreign State. The prohibition on intervention is especially pertinent in the context of cyberspace given the increasing connectivity between States and societies which provides greater opportunities for malicious actors, including States and non-State actors the acts of which are attributable to States, to misuse ICTs for the purpose of intervening in the internal and external affairs of States').

306 Kilovaty (n 15) 141; Kilovaty (n 150) 88.

307 Sender notes unconditionally that 'propaganda across national frontiers' amounts to a 'political means of coercion', in Sender (n 243) para 12.

308 Kilovaty (n 15) 143-144.

309 Pfaff (n 223) 364; section 6.2.2 'Attribution'.

often artificially generated; (2) which is disseminated in a coordinated and calculated manner; (3) by anonymous actors; (4) that use false, artificially created, identities.

False Information

That the falsity of information serves as a key indicator of coercion in subversive information campaigns, is widely recognised in State practice and academic discourse.[310] This understanding dates back to the 1981 General Assembly Declaration on the Inadmissibility of Intervention, which obliged States 'to combat […] the dissemination of false or distorted news which can be interpreted as interference in the internal affairs of other States […].'[311] Analysing the lawfulness of foreign broadcasts, Jamnejad and Wood opine that when content

> 'is deliberately false and intended to produce dissent or encourage insurgents, the non-intervention principle is likely to be breached. If factual and natural, it is doubtful that the broadcast will constitute intervention, regardless of the effect it may in fact have.'[312]

The consensus that 'just providing facts' and 'honest opinions' are excluded from the scope of prohibited intervention thus exists alongside the condemnation of deliberately spreading false or fake news as indicative of 'illegal coercive interference'.[313]

The emergence of information technologies for generating false narratives – including videos and audio fragments – presents a further hallmark of coercion in disinformation. Using deepfake technology and instrumentalising synthetic content to create a distorted version of reality carries the presumption of malicious intent. How could, *inter alia,* the identified use of deepfake

310 Roscini (n 100) 399 ft. 230 (Costa Rica noted that coercion can occur if a State 'deprives another State of the capacity to make free *and informed* choices pertaining to its internal or external affair [including through] the dissemination of false news').

311 UNGA Res 36/103 (9 December 1981) UN Doc A/RES/36/103, para 2(III)(d); However, because the Declaration was adopted by 120 votes to 22, following opposition by Western states, it is not regarded as reflecting customary international law.

312 Jamnejad and Wood (n 143) 374; Breedon (n 6) 667 ('if an operation involves disinformation covertly promoted by a foreign State to affect the voting behavior of another State's polity, the 'the attempts to manipulate the will of the people could constitute intervention because it undermine[s] the target State's sovereign will over its choice of political system''); Strongwater (n 49) 38 ('[t]here is a strong argument for coercion in cases involving the deliberate publication of false information to interfere in the domestic affairs of another state').

313 Wheatley (n 29) 187-188; Harold H Koh, 'The Trump Administration and International Law' (2017) 56 Washburn Law Journal 413-469, 450; 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' (2018) 47 International Journal of Legal Information 3,170.

imagery of Ukrainian president Zelensky calling upon the Ukrainian army to lay down their weapons and to surrender to Russia,[314] be interpreted as anything else but a wilful attempt to undermine Ukrainian sovereign authority?

Coordinated Dissemination

Second, in influence and information operations, coercion-as-control (opposite coercion-as-extortion) constitutes a process rather than an isolated act. Effectively undermining a State's control over its own affairs through manipulation and deception requires both sufficient scale and reach, as well as a degree of strategy and coordination, to establish control.[315] Strategic coordination manifests in the creation and dissemination of subversive disinformation. Past instances of disinformation in violation of the non-intervention rule revealed patterns in exposure to particular groups, narrative framing, cross-platform strategies, and gradual build-up of long-term narratives and specific campaigns. In its creation, new technologies – including demographic and micro-targeting alongside psychographic profiling –[316] strategise vast amounts of data about individuals, groups and even State behaviour to precisely target audiences and tailor messages to most effectively compromise informed deliberation and decision-making.[317] This calculated and coordinated nature serves as a crucial indicator for distinguishing coercion from mere persuasion.[318] Without it, false information is unlikely to cause the subversive effect required to rise to the level of coercion.

Actor Anonymity

Differentiating lawful persuasion from unlawful coercion often coincides with the overtness or covertness of an information operation.[319] Unknown actor identities should carry the presumption of unduly influencing the audience and manipulating their capacity to reason.[320] Such covertness also indicates deceptive and manipulative intent,[321] especially when the information is false.

---

314 Jane Wakefield, 'Deepfake president used in Russia-Ukraine war' BBC (18 March 2022) Accessed 22 November 2024.

315 Helal (n 14) 73.

316 Kilovaty (n 15) 134, 136-137; Daniel Susser, Beate Roessler and Helen Nissenbaum, 'Online Manipulation: Hidden Influences in the Digital World' (2019) 4 Georgetown Law Technology Review 1, 9-11; Wojciech Mazurczyk *et al.*, 'Disinformation 2.0 in the Age of AI: A Cybersecurity Perspective' (2024) 67 Communications of the ACM 3.

317 Chesney and Citron (n 234) 1771-1784.

318 Helal (n 14) 73.

319 *Ibid.* 115; Schmitt (n 15) 51; Australia's National Security Legislation Amendment (Espionage and Foreign Interference) Act 2018, No 67, 2018, framing foreign interference as 1) carried out by, or on behalf of a foreign actors; 2) coercive, covert, deceptive, clandestine; 3) contrary to the sovereignty, values and national interests of Australia.

320 Moynihan (n 12) 42; Baade (n 49) 1363 citing Jürgen Habermas, *Theorie des Kommunikativen Handelns* (1981) Volume 1 at 47, 52-53; Sender (n 243) 13; Koh (n 313) 450; Rodriguez (n 313) 170.

321 Corn (n 16) 226.

In the context of the 2016 US elections, Schmitt states that '[a]rguably, the covert nature of the troll operations 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.'[322] These scenarios eliminate the recipients' ability to evaluate the trustworthiness of the information,[323] while transparent attribution arguably preserves the freedom to make autonomous decisions and have the option of informed dismissal.[324]

The overt-covert distinction guides the differentiation between deception and influence, and although deception and coercion are not identical,[325] deception is indicative of coercion. Circling back to Joseph Nye's differentiation between lawfully persuasive and unlawfully coercive information operations, he argues that covertness of the operation – in particular when the person behind the creation, production or dissemination of the information remains hidden – is the decisive element.[326] Yet, while exposure of the identity of involved actors may indeed afford States and people 'the opportunity to question the genuineness of the disseminated information',[327] caution is due not to place an excessive reliance on source identity in the assessment of information genuineness. It is only one of many factors that people – consciously and unconsciously – consider in evaluating information.[328]

Artificial Personae

More deceptive than identity concealment is identity falsification and/or fabrication. Impersonation, "sock puppetry" and false personae are recognised hallmarks of coercive disinformation. The use of sock puppets or bot networks presume coercion 'because, in these circumstances, it is clear that the foreign power wants to manipulate the domestic debate, but also that it wants the population to believe that political discussions were not subject to outside interference.'[329] Subversive disinformation containing such 'false flag operations' thus often amounts to hidden coercion, manipulating individuals without them ever being aware of this influence.[330]

In sum, the analysis of disinformation's coercive potential reveals that its qualification as prohibited intervention depends on an accumulation and balancing exercise of several 'coercion indicators': deliberately false or synthetic

322 Schmitt (n 17) 51.
323 Baade (n 49) 1364; Terry (n 56) 96.
324 Wheatley (n 29) 184.
325 Ohlin (n 57) 82.
326 Nye (n 266).
327 Helal (n 14) 115.
328 Section 1.3.2 'Persuasion and Manipulation Techniques'.
329 Wheatley (n 29) 194-195.
330 Kilovaty (n 13) 144; Susser, Roessler and Nissenbaum (n 312) 4; Martha Finnmore and Duncan B Hollis, 'Beyond Naming and Shaming: Accusations and International Law in Cybersecurity' (2020) 31 The European Journal of International Law 3, 976.

content, systematic dissemination strategies, and deceptive practices through anonymity or identity falsification.

2.4.2.3 Intent and Purpose

That coercion requires intentionality is well recognised in State practice, jurisprudence and scholarship.[331] This principally derives from the *Nicaragua* case, where the ICJ noted that

> 'in international law, if one State, *with a view* to the coercion of another State supports and assists armed bands in that State whose purpose is to overthrow the government of that State, that amount to an intervention by the one State in the internal affairs of the other, whether or not the political objective of the State giving support and assistance is equally far-reaching [emphasis added]'.[332]

A State should thus have the intent to coerce, even if, when third parties are involved, there is no shared objective between the State and the respective proxies.[333] Drawing on the *Nicaragua* case, some authors argue that coercion and coercive behaviour is inherently intentional, and that any clarifying effort of non-intervention has to take into consider an 'intention to compel an outcome or conduct'.[334] Emphasis on a specific intent finds support in primary authorities, including the Friendly Relations Declaration and the OAS Charter, which prohibit coercion that seeks to achieve specific *aims*.[335] Intent also constitutes a central element of coercion in the digital realm: the Tallinn Manual explicitly confirms that 'intent […] is a further constitutive element of a violation of the prohibition of intervention.'[336] Especially online, where information operations are 'factually indistinguishable, and their efforts permeate borders unintentionally',[337] an intent requirement prevents the principle from 'being spread too thinly.'[338] One author also rightfully argues that '[b]ecause States

331 New Zealand Foreign Affairs and Trade (n 259) para 9b ('coercive intention of the state actors is a critical element of the rule'); Costa Rica (n 158) 7 ('[…] it suffices that a State intend to coercive another State, employs coercive methods, or eventually causes coercive effects in another State'); Norway (n 259) 68 (implied through the inclusion of '*with the intent* of altering election results in another State').

332 *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (n 7) para 241; Wingfield and Wingo (n 14) 584.

333 Tallinn Manual 2.0 (n 43) Rule 66, Commentary 28; Tsagourias (n 150) 55; On the role of proxies, see Tim Maurer, 'States, Proxies, Offensive Operations' in Paul Cornish (ed), *The Oxford Handbook of Cyber Security* (2021 Oxford University Press) 543-556.

334 Moynihan (n 12) 32.

335 Helal (n 14) 63 (such as the 'subordination of the exercise of [a State's] sovereign rights and to secure from its advantages of any kind, "depriving" peoples of their national identity,' and causing 'the violent overthrow of the regime of another State').

336 Tallinn Manual 2.0 (n 43) Rule 66 Section 27.

337 Tsagourias (n 150) 55.

338 Tay (n 151) 47; Milanovic (n 13) 645; Urs (n 103) 31.

are always adjusting their policies in response to the conduct of others, the element of intent ensures that no State can claim coercion simply because it feels compelled to react to another State going about its own business.'[339] No State, after all, 'would resort to disinformation if the truth and actual course of events were operating in its favor.'[340]

Academic discourse and State practice have attributed different terms to this requirement, explicitly emphasising 'intent'[341] or arguing in favour of determining the legality of cross-border communications based on their purpose,[342] design,[343] aims,[344] or objectives.[345] The dominant interpretation nevertheless consistently rejects requiring intent to produce a specific outcome or effect.[346] While disinformation in the form of a threat may aim for specific actions or decisions, this insufficiently captures the essence of coercion-as-control, operating through broader mechanisms of influence and manipulation.[347] It also risks confusing intent with motive.[348] Rather, as doctrinal standard, intent encompasses the coercing State's aim to deprive the victim

339 *Ibid.*

340 Martin (n 30) 47.

341 Wheatley (n 29) 194; Milanovic (n 13) 644-645; including the Netherlands, Finland, New Zealand (coercion requires intention, and with coercion-as-control this would be 'an intention to deprive the target state of control over matters falling within the scope of its inherently sovereign functions') Canada and Germany, in Milanovic (n 13) 619; Czech Republic ('Coercion is an activity intended to deprive, either directly or indirectly, the State of its ability to exercise control or govern matters within its internal and external affairs') para 9(b); Germany ('the acting State must intend to intervene'); Estonia 'cyber operations that *aim* to force another nation to act…'; Italy ('influence activities *aimed*, for instance, at undermining a State's ability to safeguard public health during a pandemic, or at manipulating voting behaviour'); Poland ('an intervention must include the element of coercion that aims at influencing the state's decisions belonging to its *domaine réservé*') para 3.

342 Russell Buchan and Nicholas Tsagourias, 'The Crisis in Crimea and the Principle of Non-Intervention' (2017) 19 International Community Law Review 165-193.

343 African Union positions that coercion 'as a policy designed to impose restraints on the will of a foreign State', without the need of such conduct to 'rise to the level of completely depriving a foreign States of its freedom or choice […]' in African Union Peace and Security Council, 'Communique of the 1196th meeting considering the Draft Common African Position on the Application of International Law to the Use of Information and Communication Technologies in the Cyberspace' (29 January 2024); Denmark ('[t]o be coercive the effort to intervene must be *designed* to have a decisive impact'); United Stated ('[b]ecause the principle of non-intervention prohibits "actions *designed* to coerce a State … in contravention of its rights'); the Netherlands ('with a view to employing coercion').

344 Roscini (n 100) 399 ft 231; Canada (n 121) para 22 ('the activities *aim* to interfere'); Sender (n 243) para 13.

345 Wheatley (n 29) 194-195.

346 Milanovic (n 13) 644.

347 Jamnejad and Wood (n 143) 381; Schmitt (n 49) 51 ('[a]t its core, a coercive action is intended to cause the State to do something, such as take a decision that it would otherwise not take, or not to engage in an activity in which it would otherwise engage').

348 Milanovic (n 13) 644; UNGA A/RES/2625 (n 170) (no State has 'the right to intervene, directly or indirectly, for any reason whatever, in the internal or external affairs of any other State').

state of its ability to control its reserved domain – to 'compromise or subordinate the will of the other state.'[349] It therefore does not matter whether, *inter alia*, Russia intended to interfere with the US elections by investing $2 million in disinformation to promote a specific candidate, or by microtargeting a group of voters; the fact that it intentionally flooded the information environment with significant amounts of false and misleading information to undermine the US's sovereign prerogatives to conduct its elections free from outside interference and to manipulate the public opinion *per se*, suffices.[350]

Regardless of the precise standard, evidentiary challenge in proving intent remain either way. When a coercing State engages directly in conduct on foreign territory without consent of that State, Roscini opines that intent to coerce may be presumed.[351] Certain statements or involvement of State operatives may similarly be considered as factual and demonstrable evidence proving subversive intent.[352] In the absence hereof, resorting to circumstantial evidence and 'surrounding circumstances' is necessary and legitimate.[353] One author, while acknowledging that one cannot with certainty know the motivations or intent of others, argues that 'voluntary actions are presumably motivated'.[354] He adds that 'with respect to election interference, given the expenditure of time, money and the risk of condemnation if discovered, it is implausible to conclude that a foreign State would […] engage in a sustained influence operation, for any reason other than to decisively influence the outcome of the vote.'[355]

Introducing more nuance and tangible guidance, Wheatley and others also suggest the inclusion of 'a reasonable observer' standard.[356] Would a reason-

---

349 Schmitt (n 49) 52 ('[i]t is only required that they be intended to have a coercive effect with respect to a *domaine reserve*') Lahmann (n 49) 424 ('[f]or example, the *Tallinn Manual* stipulates that, for a threat made by one state against another to violate the non-intervention principle, it does not need to successfully compel the target state to take some action against its will, but it must at least be sufficiently 'coercive in nature', which would seem to exclude obviously empty or implausible threats); Buchan and Tsagoursias (n 342) 165-193.

350 Hongju Koh, The Trump Administration and International Law (2017) 45 Washburn Law Journal 413, 45.

351 Roscini (n 100) 158; Stephan Lewandowsky *et al.*, 'Liars know they are lying: differentiating disinformation from disagreement' (2024) 11 Humanities and Social Science Communications 986.

352 Tsagourias (n 150) 55; Wheatley (n 29) 183.

353 *Ibid.*

354 Wheatley (n 29) 183.

355 *Ibid.;* However, scholars have pointed out a plurality of reasons for States to engage in disinformation, in Herbert Lin and Jaclyn Kerr's characterisation of 'chaos-producing operations' in Herbert Lin and Jaclyn Kerr, 'On Cyber-Enabled Information Warfare and Information Operations' in Paul Cornish (ed), *The Oxford Handbook of Cyber Security* (Oxford University Press 2021) 257-258.

356 *Ibid.* 191-194; Riccardo Pizzili-Mazzeschi, 'Due Diligence Rule and the Nature of the International Responsibility of States' in René Provost (ed), *State Responsibility in International Law* (Routledge 1992) 9, 12; Urs (n 103) 30.

able observer conclude that the disinformation campaign 'was intended to create confusion about the facts of the situation and/or undermine the faith of the local population in the democratic system'? Would they 'judge that the communication was intended to influence the target's decision-making to such an extent that they would be left without a meaningful choice about what to think, and therefore what to do?'[357] While helpful, this measurement would benefit from Tsagourias' suggested inclusion that:

> 'one can look into whether the confidentiality, integrity or availability of the information has been breached […] For example, in the case of deep fakes or leaked e-mails, it is the authenticity, integrity and confidentiality of the disseminated information that is breached but even in the case of true information, it is its integrity and authenticity that is encroached if it is mixed with false information or if presented in a false or fabricated context or if it related to partial truths. Other factors to consider to establish intent are the political and ideological competition that exists between states, the strategic or other interests served by the operations, the timing of the operations, the intensity and widespread nature of the operations.'[358]

This argument aligns with conclusions that certain tools and methods used in the creation and dissemination of subversive disinformation are indicative of intent. Thus, while proving intent presents inherent challenges in the digital realm, a comprehensive analysis of circumstantial evidence – including operational patterns, technological methods, and contextual factors – suggest a practical framework for establishing coercive intent in contemporary disinformation campaigns is feasible and within reach.

2.4.2.4 Knowledge

Unlike foreseeable harm and malign intent, victim State knowledge remains contested as a requirement for coercion.[359] Coercion traditionally implies 'conscious unwilling acts on the part of the victim',[360] but this consciousness is often absent in the digital realm due to the identified anonymity and actor falsification.[361] The majority of the Tallinn Manual experts, as well as academic discourse and the limited available State practice,[362] therefore maintain

357 Wheatley (n 29) 192-194.
358 *Ibid.*
359 Tsagourias (n 150) 55.
360 Wheatley (n 29) 163.
361 Zach Bastick, 'Would You Notice if Fake News Changed Your Behavior? An Experiment on the Unconscious Effects of Disinformation' (2021) 116 Computers in Human Behaviour 106633, 1-10.
362 Terry (n 56) 90 citing Paul C Ney Jr, 'Remarks at US Cyber Command Legal Conference' (US Department of Defense, 2 March 2020) and Jeremy Wright, 'Cyber and International Law in the 21st Century' (UK Government, 23 May 2018).

that knowledge of interventionist operations is not required to breach the non-intervention rule.[363]

This position logically follows earlier observations that intervention does not have to succeed in order to be unlawful, thus eliminating the need for knowledge of the operation on the side of the targeted state.[364] This proves particularly relevant to disinformation, as Kilovaty notes that 'victim States would rarely be aware in real time that they or their citizens' decisions are being affected by manipulation, disruption, or disinformation.'[365] The growing sophistication of disinformation technologies and gradual deployment patterns complicate both detection and attribution. Even if a State knows an intervention is occurring, timely identification of the source has proven convoluted. Therefore, as Milanovic observes, '[i]t would make no sense to regard the most successful (and therefore most harmful) operations – those that go undetected – as failing to meet the requirements of intervention.'[366]

While retrospective awareness may support a claim that intervention has taken place,[367] requiring knowledge contradicts the prophylactic nature of the non-intervention rule,[368] and exacerbates existing power asymmetries between States.[369] States with advanced intelligence services and technological capabilities to acquire such knowledge could more readily invoke a violation, while those lacking such resources stand at a disadvantage. Thus, recognising these evidential challenges, victim State knowledge should be treated as probative evidence of intervention rather than a constitutive element, ensuring the principle's effectiveness in addressing modern forms of covert interference while maintaining its protective function for all States, regardless of their technological capabilities.

2.4.3 Interim Conclusion

As technology expands, so do the pathways for intervention. The first part of the analysis has demonstrated that while the traditional concept of *domaine réservé* presents challenges in the digital age, a reorientation towards protecting the object of the doctrine – the ability to make free choices on sovereign

363 Tallinn Manual 2.0 (n 43) Rule 66 Commentary 25.
364 Moynihan (n 12) 33.
365 Kilovaty (n 150) 109.
366 Milanovic (n 13) 645.
367 Tsagourias (n 150) 56.
368 Corn (n 16) 225.
369 Roscini (n 100) 397; Willmer (n 124) 509-510; Wu (n 170); Helal (n 14) 57 citing John Linarelli, 'An Examination of the Proposed Crime of Intervention in the Draft Code of Crimes Against the Peace and Security of Mankind' (1995) 18 Suffolk Transnational Law Review 1, 24, 31, 36; Alexandra Hofer, 'The Developed/Developing Divide on Unilateral Coercive Measures: Legitimate Enforcement or Illegitimate Intervention' (2017) 16 Chinese Journal of International Law 175.

matters – enables structural alignment with contemporary threats, including disinformation. This shift responds to the evolving nature of influence operations and the growing role of non-State actors, while maintaining the essential protective function of the non-intervention framework. By emphasising the framework's connection to self-determination, the analysis strengthened its applicability, recognising that interference with a population's ability to form independent opinions and make autonomous decisions constitutes intervention in State affairs.

Moreover, maintaining the relevance of the non-intervention doctrine requires 'expanding our understanding of coercion, […] and acknowledging the role of manipulation, disinformation and disruption.'[370] The second part therefore substantiated that subversive disinformation could constitute coercion under the non-intervention principle in two distinct ways: as a credible threat (coercion-as-extortion) or, more commonly, as manipulative control. While neither inherently coercive nor automatically unlawful, disinformation can rise to the level of prohibited intervention. This determination requires a holistic assessment of three dimensions: the operation's foreseeable consequences and intensity; the tools, features, and methods employed; and the presence of coercive intent.

Regarding the foreseeable consequences and intensity, disinformation must have the capacity to produce significant effects that substantially impair State independence, measured through its scale, reach, and intensity – applying McDougal and Feliciano's 'consequentiality' test, weighting the importance and number of values affected, the extent of impact, and the number of participants affected. In assessing the tools, features and methods employed, certain operational characteristics should be taken as an indication of coercion: the use of demonstrably false information – particularly when artificially generated, coordinated and strategic dissemination patterns, actor anonymity, and the deployment of false identities. Finally, while coercive intent must be established, it need not aim for specific outcomes. Instead, intent encompasses the broader purpose of compromising the target State's autonomous decision-making. Notably, victim State awareness, valuable as evidence, is not a prerequisite for establishing a breach of the non-intervention principle.

2.5 Conclusion

This chapter has examined subversive disinformation, analysing its relationship to State sovereignty and the non-intervention framework. Subversive disinformation, the chapter demonstrated, is directly governed by two frameworks

370 Michael N Schmitt, 'Introduction to the Research Handbook on International Law and Cyberspace' in Nicholas Tsagourias and Russell Buchan (eds), *Research Handbook on International Law and Cyberspace* (Edward Elgar Publishing 2021) 3.

under international law: State sovereignty and the principle of non-intervention. It revealed that contemporary discourse on disinformation and sovereignty disproportionately – though understandably – associates 'subversion' with electoral interference. This focus, however, insufficiently acknowledges the broader spectrum of subversive disinformation that encompasses economic, scientific, cultural and even judicial domains. Encompassing this diverse typology, the chapter positioned 'subversive disinformation' as a sophisticated evolution from traditional propaganda, characterised by dual-track targeting, advanced manipulation strategies, and digital system exploitation that enables subtle, but disruptive, forms of State destabilisation.

The legal analysis examined how the obligations to respect State sovereignty and the corollary prohibition of intervention – both recognised frameworks that delineate the category of 'prohibited subversive propaganda' – govern contemporary disinformation. It explored how these paradigms demarcate the boundary between lawful persuasion and unlawful manipulation of States and their population, demonstrating that while these established legal frameworks remain theoretically applicable to disinformation, their practicality is constrained by technological obfuscation, growing power asymmetries between State and non-State actors, and expanding tools for transnational influence. Therefore, this remains a 'grey area' of international law that struggles to adequately address the technological, cognitive, and societal dimensions of twenty-first-century interference and intervention.

This gap's legal implications are revealed by examining subversive disinformation in light of States' international obligations not to interfere with the territorial integrity or inherently sovereign functions of other States. In theory, disinformation influences all areas of State sovereignty falling within the scope of 'inherently sovereign functions' and although the doctrine has long maintained a focus on physical intrusions, the chapter illustrated the growing recognition of non-material harm and online operations as threats to States' sovereignty. Within this widening scope, disinformation, nevertheless, presents a distinct challenge through its indirect and cognitive effects, exercising subversive influence through dual-track targeting of both decision-makers and public opinion. It exploits digital systems and spans domains far beyond its popular characterisation as election interference – including economic destabilisation, environmental governance disruption, judicial interference and public health subversion.

The sovereignty analysis established that while disinformation rarely violates territorial integrity directly, it can interfere with a State's 'inherently sovereign functions' when it systematically undermines State's capacity for autonomous decision-making. This reconceptualisation – shifting the focus from the protected domain to the protected function – drew from cyberlaw scholarship and emerging State practice. While providing clarity, this approach still faces the absence of clear doctrinal thresholds for establishing sovereignty violations. To address this gap, the chapter favoured inclusion of a threshold

to differentiate between from permissible and impermissible interference. To evaluate disinformation as impermissible interference, three factors were proposed: the type of disinformation and targeted State function, the audience and the degree for behavioural change, and the operation's timing and scale.

Regarding non-intervention, the analysis demonstrated that disinformation can easily interfere with a State's internal or external affairs, yet only under limited circumstances does this interference amount to coercion. Traditional interpretations limiting 'coercion' to dictatorial force artificially excludes new forms of coercive behaviour, including contemporary disinformation that achieves involuntary behavioural change through cognitive manipulation rather than explicit threats.

Subversive disinformation interferes with States' autonomous decision-making across multiple domains that were traditionally considered beyond the reach of foreign interference. Addressing this evolution and improving alignment between new forms of interference and the rule's protective functions, requires doctrinal reorientation: from traditional domain-based approaches towards protecting states' ability to make autonomous decisions. This reorientation influences the meaning of coercion. Synthesising emerging scholarship and State practice, the chapter suggested that the coercive nature of disinformation can be identified through a contextual assessment of:

1. The scale, intensity and magnitude of the disinformation operation or campaign, which indicates the likelihood of successful intervention.
2. The tools and methods of creation and dissemination, which suggest both potential coercive effect and intent.
3. The intent as a requisite element of intervention, applicable to both coercion-as-extortion and coercion-as-control.

Although determining foreseeable harm and malicious intent requires case-specific evaluation, certain recurring characteristics should be accepted as presumed coercion in subversive disinformation: (1) creation of false information, often artificially generated; (2) that is disseminated in a coordinated a calculated manner; (3) by anonymous actors; (4) that use false, artificially created, identities.

The application of these two international legal frameworks to subversive disinformation ultimately reveals a troubling tension: their object and purpose to prevent, restrict, and ultimately sanction foreign influence is undermined by their inherent open-endedness, which, while deemed indispensable to safeguard flexibility and adaptability, now impedes their practical effectiveness in addressing such interference. Nevertheless, despite this tension between flexibility and effectiveness, these frameworks retain significant potential as governing instruments, provided they evolve to address contemporary challenges.