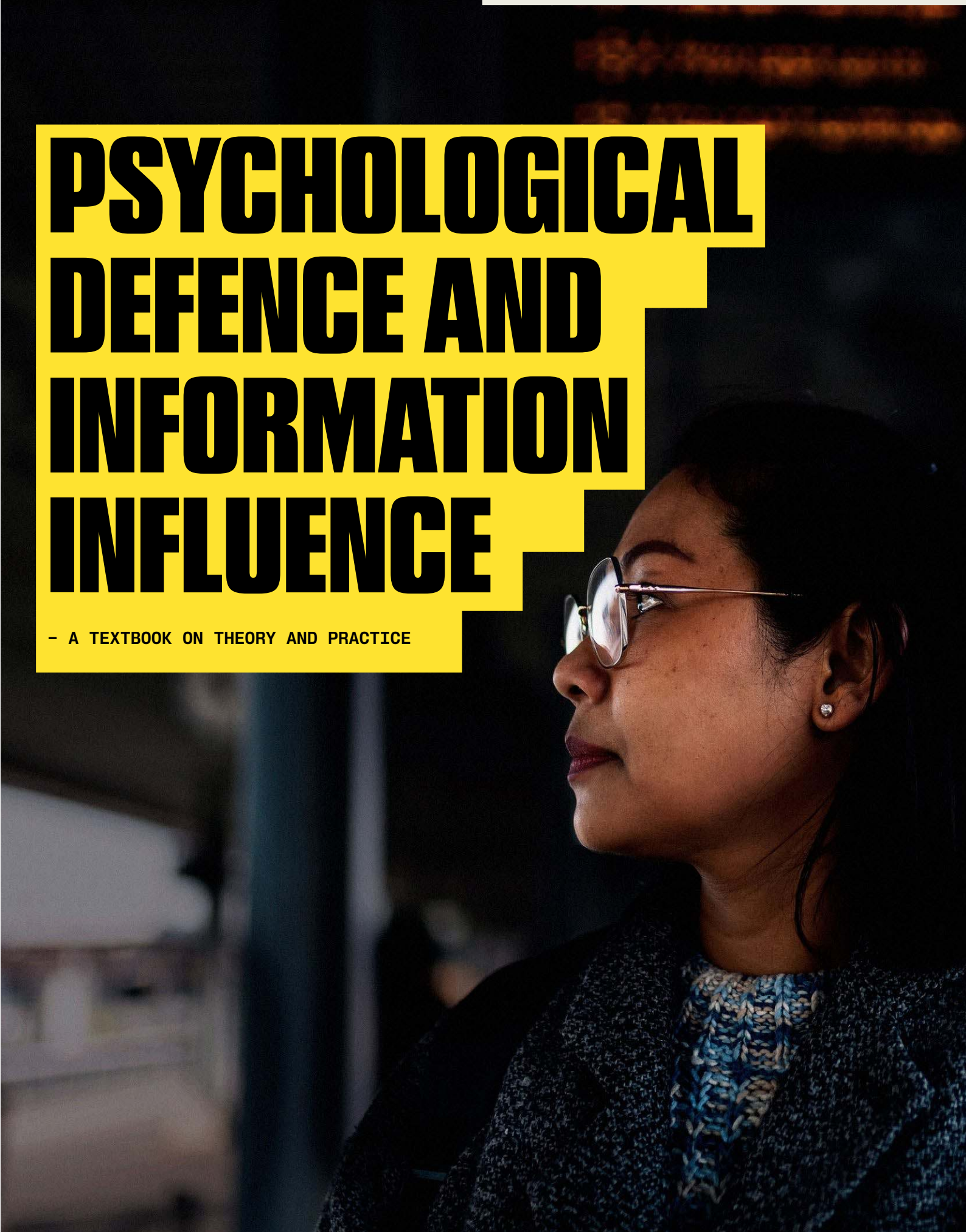
JESPER FALKHEIMER
& JAMES PAMMENT, EDS.

Psychological Defence Agency

LUNDS UNIVERSITET

PSYCHOLOGICAL DEFENCE AND INFORMATION INFLUENCE

– A TEXTBOOK ON THEORY AND PRACTICE



Psychological Defence
MPF 2025
ISBN: 978-91-989646-8-4
Production: Gritti
Print: Scandinavian Print Group

TABLE OF CONTENTS

FOREWORD

In an age defined by rapid information flows and shifting security landscapes, the resilience of societies rests not only on military strength or technological capacity, but equally on the ability of individuals and institutions to withstand psychological influence and manipulation. Psychological defence is therefore not merely a technical field, it is a civic responsibility and a cornerstone of democratic resilience.

This book provides knowledge, practical guidance, and reflection on how psychological defence can be understood and applied in different contexts. It explores the threats we face – such as disinformation and propaganda – and the tools available to counter them, ranging from critical thinking and communication strategies to institutional preparedness.

The overarching aim of this book is not only to raise awareness of psychological threats, but also to empower readers with the knowledge and confidence to respond effectively and responsibly. By fostering trust, openness, and critical engagement, psychological defence contributes to safeguarding the values of free and democratic societies.

It is my hope that this book will serve as both a source of insight and as a practical companion for all who recognize the importance of building psychological resilience – whether as professionals, decision-makers, or engaged citizens.

It is the authors themselves who are responsible for the content of the book.

MAGNUS HJORT
DIRECTOR GENERAL

INTRODUCTION

This textbook is the first of its kind: a comprehensive overview of key issues related to psychological defence and information influence written by leading scholars on each topic. As fields of knowledge, psychological defence and information influence are multi-disciplinary. This becomes even more clear editing this volume, where disciplines such as history, strategic communication, political science, psychology, media studies, art history and security studies are represented. As fields of applied knowledge, psychological defence and information influence are of core importance for defending our democracy – the open society with free speech, fair elections and informed citizens. Threats to this open society are not new, but the technological and geo-political developments of the past decade have increased the breadth and intensity of these threats.

This textbook is an anthology with contributions reflecting the broad debate in Sweden (all scholars are Swedish, based in Sweden, or have worked there). The chapters themselves are short summaries (around 5,000 words) of the more advanced works of many of these scholars, adapted from books, reports, and academic articles in a concise and approachable format. The knowledge field is growing fast, not least in Sweden and the Nordic countries. A research review[1] on how dis- and misinformation have been studied in Denmark, Finland, Iceland, Norway, and Sweden between 2014 and 2024 confirms the cross-disciplinary features of the field. The review also identifies five main topics: security and Russia's actions, the role of news media and fact-checking, health aspects, media literacy, and social media. It is well-known that the Nordic region has a high level of resilience towards antagonistic information influence campaigns and dis- and misinformation due to high societal trust, robust media systems and low polarisation. But this does not mean that we can sit back. The increased level of conflict in our external environment is also spilling over into our countries. The development of AI is creating new threats and the vulnerability of certain social groups to disinformation has increased. The mentioned research review also identified some knowledge gaps for the future. One of these gaps concerns the need for more studies of the psychological and cognitive effects of dis- and misinformation on individuals and groups; another gap concerns how media literacy and critical thinking may be developed in relation to children and youth.

This textbook is a starting point for students and researchers to explore the topic and themes; each chapter includes discussion questions and further reading. The book is directed towards a wide group of readers – undergraduate and graduate students as well as participants in training courses and continuing education

[1] Grahn, H., Kalsnes, B., Isaksson, E., Mayerhöffer, E., Ólafsson, J. G., Falkheimer, J., Henriksen, F. M., Kristensen, J. B., & Saari, D. (2025). Mapping research on disinformation and misinformation across the Nordic countries: An integrative review. *Nordicom Review,* 46(S1), 175–220.

focusing on psychological defence and information influence and their related issues. The book is also directed towards professionals in communications, civil defence, and military defence at all levels. A few chapters are mainly of interest for Swedish readers while most of the chapters are relevant for international readers.

The book is divided into four main sections. First, we dive into the *fundamentals of psychological defence.* This section consists of eight chapters that together introduce the broad issues that shape the field of psychological defence: the democratic values that this work seeks to protect such as an open and free debate, the governing legal frameworks, and the security architecture of total defence, civil defence, and intelligence, and a chapter on the past, present and future of psychological defence.

Second, we dig into aspects on *understanding malign information influence.* The eight chapters in this section offer insight into the concept of malign information influence. The contributions give an overview of key terminologies, an examination of how information influence exploits media systems and encourages political polarisation, and chapters on cognitive biases, rhetorical devices conspiracy theories, gendered disinformation, and the impact of artificial intelligence.

Third, the book offers *case studies in information influence.* The first three chapters focus on China and Russia, detailing some of the main trends in their information influence activities. The remaining chapters explore key issues including the outsourcing of government-led influence campaigns, election interference, the LVU (social services) campaign targeting Sweden, the role of arts and culture, and the role of video games.

Finally, the last section focuses on the *measures* that can be taken to mitigate, counteract, or push back on malign information influence campaigns. The first chapter draws upon psychological mechanisms as techniques, and is followed by examples of education initiatives, crisis communication as a countermeasure, and lessons learned from Ukraine's experiences of dealing with Russian information influence. The last chapters explore the concepts of deterrence and attribution, and the section concludes with discussion on the ethics of countermeasures.

The field of psychological defence and information influence is developing fast both in research and practice. Our aim as editors is therefore to treat this textbook as a 'living document' that can be updated as regularly as needed, with the overall goal is to become a standard work in the field both nationally and internationally. We want to thank all participating authors for their important efforts, and the Swedish Psychological Defence Agency for support and guidance.

Jesper Falkheimer & James Pamment, editors

I. FUNDAMENTALS OF PSYCHOLOGICAL DEFENCE

This section consists of eight chapters that together introduce the broad issues that shape the field of psychological defence. The section begins by highlighting the democratic values that this work seeks to protect and introduces some of the reasoning around Swedish efforts to protect open and free debate in the context of evolving threats to national security. It then discusses the governing legal frameworks, before introducing the security architecture of total defence, civil defence, and intelligence, into which psychological defence fits. Finally, it discusses the past, present and future of psychological defence in terms of its developing role and purpose.

1. DEFENDING DEMOCRACY

OLOF PETERSSON

SUMMARY

Psychological warfare has many different weapons at its disposal. But psychological defence possesses many tools as well. Democracy, however, must choose its means of defence with discernment. Defending democracy with methods that harm democracy is out of the question. Defence methods that are neutral in their effect are acceptable. The best protection is a psychological defence that strengthens democracy. Such a psychological defence can increase the understanding of the fundamental values of democracy and strengthen democratic citizenship.

While the history of psychological warfare is as long as the history of wars, and thus as old as humanity, psychological defence is a more recent phenomenon. With universal suffrage the people became a political force. Political leaders could now be removed because of a general election. In a democracy, the power of the state rests on popular acceptance. An antagonistic actor can weaken democratic support by psychological warfare. Psychological defence is an integral part of the self-defence of democracy. Democracy is a method of peaceful conflict resolution. Conflicts are managed through dialogue, debate, negotiation, compromise and voting. An open society is based on freedom of expression. Dissent, criticism and opposition are vital to democracy. Government by discussion requires a free and open public sphere. Compared to other forms of government, democracy has both strengths and weaknesses. A dictatorship is based on fear and a costly system of surveillance and reprisals. A democracy is based on trust and citizen support. Democracy is a flexible and resilient system. But an open society is inherently vulnerable to attacks that can disrupt public discourse and undermine trust. All systems, including political systems, have mechanisms to protect their survival. Dictatorships and autocracies have means at their disposal that democracy does not: censorship, prohibition, repression of dissent and the use of violence without regard for human rights. The challenge for democracy is to defend itself exclusively by democratic means.

Freedom of expression in a democracy is broad but not unlimited. The European Convention on Human Rights recognises that the exercise of the freedom of expression carries with it duties and responsibilities and may be subject to restrictions as prescribed by law and considered necessary in a democratic society, as, for instance, in relation to national security. In this way, democracy can defend the public sphere by prohibiting certain forms of expression. But even legally permitted information can harm democracy. Disinformation, fake news and scare tactics are examples of methods that may disrupt public discourse and undermine trust. Political scientist Robert Dahl has formulated some basic requirements for a well-functioning democracy (Dahl, 1989). One such requirement is control over the agenda. In a democracy, the people and their elected representatives should be able to define what is important, what issues should be high on the political agenda. An antagonistic actor may try to distort attention and

steer the public debate towards issues that favour their own interests. Another requirement is enlightened understanding. Every citizen should have an equal opportunity to consider the implications of the various proposals under debate. Psychological warfare may focus on disrupting this knowledge formation through irrelevant, misleading and false messages.

Anti-democratic influence operations feature certain recurring themes. The worldview is often conspiratorial. Democratic practice, such as discussion and consensus-building, are scorned in favour of force and action. Arguments are often based on a sharp dichotomy. Anti-democratic messages are framed as an antagonistic relation between “us” and “them”. In far-right propaganda, the “other” may be Jews, immigrants or foreigners in general. Islamist propaganda attacks Muslim groups that are not considered to be living in an orthodox way, often with anti-Semitic overtones. Left-wing extremism can also be based on polarisation, where opposition to sexism, homophobia, racism and fascism is used to justify undemocratic messages and practices. Democratic self-defence faces a problem. Democratic states are threatened by several adversaries, in particular from Russian expansionism, Chinese great power ambitions and Islamic fundamentalism. Democracies may be lured into a rhetoric based on a black and white worldview. However, democracy is not a perfect system. Its strength lies in its ability to openly discuss its weaknesses and failures.

Psychological defence is a concern for the whole of democratic society, particularly schools, civil society, business, media, culture, research and, in particular, individual citizens. Through social media, every individual is nowadays armed with powerful tools in the public sphere. Power should be always matched by responsibility. Individuals should ensure that the messages they formulate or pass on are not part of an antagonistic influence operation. The government has a particular responsibility to enable an effective psychological defence. In a democratic state based on the rule of law, specific restrictions apply to government agencies. These restrictions impose limitations but bring with them opportunities as well. The rule of law is often interpreted such that everything that is not prohibited is authorised. However, this rule does not apply to the state itself. On the contrary, a government agency must always have a legal foundation for its activities and may only act within its mandate. Public administration must also respect the equality of all before the law and observe the principles of objectivity and impartiality. This is why a government agency is not allowed to control the public sphere. A Ministry of Truth has no place in a democracy. To some extent, however, a government agency enjoys within its legal framework unique opportunities in psychological defence. By gathering intelligence, it can detect foreign influence operations and warn those concerned. An agency can also improve awareness of various types of threats by means of analysis and research. As there are many different actors, both public and private, involved in psychological defence, an agency also has an important coordinating role. Finally, an agency may help to strengthen the resilience of the population.

REPRESENTATIVE DEMOCRACY

Representative democracy is essentially a process of communication. Voters send signals to their elected representatives. Elected politicians send signals to the public administration through legislation. Various agencies implement the political decisions. Citizens interact with the public sector in the form of regulations and the provision of services of various kinds. Like any communication system,

representative democracy is susceptible to disruption. Representative democracy is based on universal and equal suffrage. Elections must be free (no one else may decide what the voter will vote for), secret (the voter is protected from disclosing his or her vote) and direct (the elected representatives in parliaments and councils are appointed by the voters). Special rules and control bodies ensure that the election procedure is carried out correctly.

Elections can be manipulated during the election campaign by spreading false news and misleading information. Electoral fraud during the election itself can include threatening voters, creating unrest on election day, sabotaging polling stations, falsifying ballot papers, and bribing and threatening election officials. After the election, the results can be manipulated by disrupting the vote count by cyber-attacks and manipulating the calculations when totalling local election figures into national results. Attacks on representative democracy can also take the form of threats and violence against elected representatives. Cyber-attacks and influence operations pose threats to the people's elected representatives. Leading politicians who oversee sensitive or politically controversial issues are more vulnerable than others. Women and immigrant politicians are particularly exposed. Attacks on the integrity of elected representatives occur at the national level, but this also applies in relation to regional and municipal councils.

Several government agencies are specifically charged with detecting and preventing attacks on representative democracy. *The Swedish Election Authority* is responsible for ensuring that general elections are conducted in a secure manner. The *Electoral Appeals Board* is a body within parliament in charge of examining whether the election has been conducted correctly. In the event of serious errors, a re-election might be called for. *The Swedish Security Service* counteracts unauthorised interference with the basic functions of democracy, particularly crimes aimed at influencing the decision-making of politicians and representatives of public authorities or the professional practice of journalists. The activities aim to prevent unauthorised influence on political decision-making, the implementation of political decisions and free debate. *The Swedish Psychological Defence Agency* protects against foreign disinformation aimed at weakening Sweden's resilience and the willingness of the population to defend itself, or at unduly influencing people's perceptions, behaviours and decision-making. The agency develops methods and techniques for identifying hostile foreign actors and their methods of influence. The purpose is to safeguard freedom of expression and an open and democratic society. The task of *The Swedish National Council for Crime Prevention* is to contribute to the development of knowledge in the field of criminal policy and to promote crime prevention by monitoring threats against elected politicians, for example.

CITIZEN RESILIENCE

While government agencies play an important role in the defence of the democratic system, informed and critical citizens are the ultimate guarantee for the survival of democracy. Citizen resilience can be analysed by a model of civic voluntarism developed by political scientist Sidney Verba and colleagues (Verba, Schlozman & Brady, 1995). Strong resilience indicates ability, willingness and belonging to a community. Resilience, therefore, is conditioned by three factors: *resources, commitment and networks* (Petersson, 2023a).

Citizen resources are factors that can be expected to contribute to increasing the abilities of the citizenry and thus strengthening the resilience of the population.

Such resources are education, occupation and other experiences that empower participation in society. Material and economic assets provide resources that can be translated into influence. Education provides knowledge and cognitive resources that increase the ability to assert oneself in debates and decision-making contexts. Other types of life experience can also enhance assertiveness. Conversely, having few or weak resources can lead to powerlessness and disempowerment. Civic skills are the abilities required to participate as an active and responsible citizen in a democracy. More specifically, they can include language skills, critical thinking skills, speaking, listening, communication and co-operation skills. A special instance of civic skills is media and information literacy, the ability to take a critical approach in the media landscape, particularly in relation to social media.

Citizen engagement concerns the mental commitment of an individual. It refers to the psychological conditions underpinning the resilience of the population. Resilience depends on the mental preparedness of an individual citizen to take responsibility in a crisis situation. Ultimately, it is about how individuals view their responsibility as citizens in society. Research has shown that two main dimensions can be distinguished: duty-based citizenship and engaged citizenship. Trust also plays a central role in citizen engagement. A well-functioning society requires that people can trust each other. One aspect relates to interpersonal trust, that is whether other people can be trusted. Another aspect is concerned with trust in political decision-makers, public administration and other social institutions. Trust in quality media outlets is also an important condition.

Citizen networks are another element of the resilience of the population. A member of society who is an active part of the community is better equipped to face crises than an isolated individual. Popular movements and political parties with a broad membership base have historically played a crucial role in the democratic infrastructure of a country like Sweden. Nevertheless, official membership figures underestimate the existence of informal networks. There is no evidence that people in Sweden have become more socially isolated. On the contrary, surveys indicate that interpersonal contacts in the local community are lively. Social media facilitate communication within such informal networks. The official brochure *If Crisis or War Comes* points out that the responsibility for the safety and security of our country is shared by everyone who lives here. “One of our most important assets when something threatens us is our willingness to help each other.” It is this willingness to help each other that is at the heart of the informal resilience network.

SPIRIT OF RESISTANCE

Today democracy is under attack. What Putin fears most is democracy, two scholars commented after Russia’s full-scale invasion of Ukraine (Person & McFaul, 2022). Defending democracy ultimately requires citizens to be prepared to take up arms to defend the territorial integrity of their democratic state. The willingness to defend one’s own country is a concept as crucial as it is difficult to define. There may be a broad agreement on its general meaning. Willingness to defend is closely related to the spirit of resistance. But specifying this general concept to measure it more precisely has proved much more difficult.

The World Value Survey includes a simple question about whether people are prepared to defend their own country. Sweden, together with the other Nordic countries, stands out for a relatively high willingness to do so (Inglehart, Puranen & Welzel, 2015). Another study asked Swedish respondents about their defence willingness: ”In the event of a crisis, I am willing to contribute to the military

defence", where respondents could answer on a five-point scale (Persson & Widmalm, 2023). The Swedish Defence Research Agency has initiated a survey on the public's attitude to total defence. The questions included assessments of Sweden's preparedness in various areas, concern about the possibility of various events occurring, views on defence and personal crisis preparedness. A general question concerned how reasonable people thought it was to have personal responsibility in crises and war. Nine out of ten responded that it was reasonable. Two questions concerned their own willingness to participate in Sweden's defence (Wedebrand, 2018).

Sweden is an example of how considerable resources have been devoted to measuring defence readiness. Already in the early 1950s, an interview question was asked that has been repeated for over half a century. The question was formulated as follows: "Suppose Sweden is attacked. Do you think we should offer armed resistance even if the outcome seems uncertain to us?" But these measurements do not appear to have been used much, either for total defence planning or for social science research. The main reason is a fundamental problem. This way of measuring the willingness to defend does not capture the concept of spirit of resistance. The spirit of resistance refers to the mental readiness of an individual citizen to take responsibility in a crisis. Willingness to defend as defined in these surveys is about whether "we", implicitly we in Sweden, should resist in the event of an armed attack. What is measured is probably primarily the attitude to official Swedish defence policy. These empirical studies on the willingness to defend one's country rely on cross-sectional surveys where respondents answer questions about what to do in the event of a military attack on Sweden from a foreign power. Whether these responses are realistic or not remains a fundamental validity problem with this research approach. Scenario-based tools have been suggested to improve our knowledge about the determinants of the willingness to defend (Cancino, Svenonius & Michélsen Forsgren, 2023). Several studies have thus been carried out that shed light in various ways on the public's willingness to defend itself and its preparedness for crises. However, few have directly analysed the spirit of resistance. Thus, there is a gap in the current state of knowledge.

"Well-functioning crisis preparedness requires awareness of personal responsibility and knowledge at all levels of how to act quickly and effectively in the event of a crisis. Individuals form the foundation on which society's crisis management rests." These are the words of a government spokesperson (Petersson, 2023b). It is a clearly formulated norm that the individual citizen has a personal responsibility in a crisis. A study of citizens' spirit of resistance can provide valuable information on how well this norm is anchored as an idea and how it is observed in practice. The spirit of resistance is part of the norms that guide us as citizens in a democratic society. Research on democratic citizenship shows that it is possible to measure civic norms in a systematic and useful way (Petersson, Hermansson, Micheletti, Teorell & Westholm, 1998). The key question is what an individual considers important to be a good citizen. Examples of such norms are paying taxes, obeying laws, voting in public elections and forming independent opinions. The spirit of resistance can be mapped with this methodology. The citizen norms relevant here are about defence preparedness, psychological defence and individual preparedness for crises and accidents. Interview questions have also been asked about the extent to which people consider that other citizens comply with civic norms. In addition, there are also questions that measure how a person assesses their own compliance with these citizen norms. It turns out, for example,

that many people have the perception that other people try to evade taxes. Such skewed perceptions risk undermining solidarity in society. Why should I follow the norms when others do not seem to?

Ideally, the spirit of resistance is at its best when three conditions are met. Firstly, there must be a widespread norm in society that it is a citizen's duty to ultimately defend the country with arms. Second, people must believe that other citizens live by this norm. Thirdly, it is necessary for citizens themselves to live up to the demands they make on others. Thus, the spirit of resistance and the willingness to defend one's country form an integral part of democratic citizenship.

DISCUSSION

- How can democracy defend itself? Which methods are most effective? Can self-defence harm democracy itself?
- Who is responsible for the defence of democracy? Discuss the roles of individual citizens, the government, and other actors in society.
- What is meant by citizen resilience and spirit of resistance? How are these two concepts interrelated? What do they mean in an actual crisis situation?

OLOF PETERSSON *is a political scientist, formerly professor at the University of Uppsala. He has authored several books on democracy and government and has also participated in government commissions on institutional reform in Sweden.*

REFERENCES

Cancino, S., Svenonius, O. & Michélsen Forsgren, M. (2023). *Vem vill försvara landet? En scenariobaserad ansats till studiet av försvarsvilja.* Stockholm: Totalförsvarets forskningsinstitut, FOI-R--5428—SE.

Dahl, R. A. (1989). *Democracy and Its Critics.* Yale: Yale University Press.

Inglehart, R. F., Puranen, B. & Welzel, C. (2015). Declining willingness to fight for one's country. The individual-level basis of the long peace. *Journal of Peace Research,* (52)4, 418–434.

Person, R & McFaul, M. (2022). What Putin fears most. *Journal of Democracy,* (33)2, 18–27.

Persson, T. & Widmalm, S. (2023). Upon entering NATO. Explaining defence willingness among Swedes. *European Security,* (33)4, 690–710.

Petersson, O., Hermansson, J., Micheletti, M., Teorell, J. & Westholm, A. (1998). *Demokrati och medborgarskap. Demokratirådets rapport 1998.* Stockholm: Studieförbundet Näringsliv och Samhälle.

Petersson, O. (2023a). *Motståndskraft.* MPF skriftserie, 1. Karlstad: Myndigheten för psykologiskt försvar.

Petersson, O. (2023b). *Motståndsanda.* MPF skriftserie, 2. Karlstad: Myndigheten för psykologiskt försvar.

Verba, S., Schlozman K. L. & Brady, H. E. (1995). *Voice and Equality. Civic Voluntarism in American Politics.* Cambridge: Harvard University Press.

Wedebrand, C. (2019). *Allmänheten och totalförsvaret. Resultat från en enkät utskickad hösten 2018.* Stockholm: Totalförsvarets forskningsinstitut, FOI-R--4771—SE.

2. COMMUNICATION & OPEN SOCIETY

HOWARD NOTHHAFT

SUMMARY

- The idea of 'open society' is a core concept integral or at least complementary to liberal democracy. Taking its cue from the works of Karl Popper, the article outlines the advantages and disadvantages of societal openness.
- The reasoning attempts to explain the resonance of criticism that Western society has become 'too open' but also makes the case that open societies provide unprecedented quality of life and prosperity for its citizens.
- The article ends with implications for psychological defence.

In global surveys of democracy like the ones conducted by the V-Dem project, Freedom House or The Economist, Sweden consistently ranks among the most democratic countries in the world. With authoritarianism on the rise all over the world, Sweden is upheld as a model democracy, i.e., a functioning democratic system wedded to a truly open, tolerant and inclusive society. However, while it is perhaps safe to say that the majority still considers 'openness' a desirable quality, there is increasing concern over mass immigration, loss of national identity and societal dysfunction. As a general trend, 'populist' parties all over Europe have been successfully suggesting that Western democracies have become *too* open: materially, in the sense of porous borders and high levels of immigration; culturally, in the sense of excessive deference to alien ways of life brought to the country; politically, in the sense of misguided tolerance for dysfunctional, anti-democratic or simply criminal elements in society. And, in the current climate of conflict and confrontation, another argument has come to the fore: that overly open societies are vulnerable. Some groups hold that overly open societies are in fact already unravelling. With 'cancel culture' stifling honest debate at universities and elsewhere, governments utilizing hate-speech legislature to repress free speech, and political correctness placing the trump cards in the hands of anti-democratic elements, liberal democracies are said to be collapsing under their own ideals. Whether this is true or not, these internal background conditions constitute a challenge for psychological defence against deliberate communicative attack by external foreign actors.

OPEN SOCIETY AND LIBERAL DEMOCRACY

The system of government in the Western World is generally called 'liberal democracy'. The term liberal indicates here that the freedom of the individual ranks relatively high in the value order of society; the government is understood as serving the people, not the people the government; the powers of the state vis-à-vis citizens are limited. Constitutions of liberal democracies guarantee the individual relatively far-reaching rights – above all, the right to vote the government

into and out of power. Liberal democracy also obliges the state to defend these rights against encroachment by majorities or the government itself. To guarantee individual freedom not only in theory but in practice, liberal democracies have evolved institutions such as free and secret elections, separation of power with an independent judiciary, press and journalism (and today watchdogs and activists) separate from the state, freedom and independence of academic research, and tolerance towards artistic expression. Historically, liberal democracy also goes hand in hand with market economy, as it was the capitalist entrepreneurs which forced liberalization. The degree to which markets are 'free' varies, of course. The U.S. is traditionally seen as tending towards *laissez-faire* capitalism while many European states are agreed that markets need to be regulated, capitalism 'tamed': the economic regime in Germany is called 'social market economy', for example. In addition to institutional safeguards, there are also typical cultural patterns of liberal democracies, like e.g., an active civil society.

OPEN AND CLOSED SOCIETY

A characteristic not visible on diagrams of political institutions is that liberal democracies tend to be 'open' as opposed to 'closed' societies. The concept of 'open society' (*société ouverte*) was developed in the 1930s by French philosopher Henri Bergson, but the theorist mainly associated with it is the philosopher of science Karl Popper. Popper wrote *Die offene Gesellschaft und ihre Feinde* (The Open Society and its Enemies) in 1945 after emigrating in the wake of Austria's annexation by Nazi Germany. The 'Open Society Foundation' instigated by financier George Soros takes its name from Popper's book. 'The Open Society and its Enemies' (Popper, 2005) is a work of political philosophy. In Popper's conception, the open-closed-dichotomy is *epistemic* in character, i.e., knowledge is the key dimension. The enemies of the open society treated in Popper's book are not dictators like Hitler, Stalin or Mao, but philosophers, like Plato, Hegel and Marx, who, in Popper's not uncontroversial opinion, laid the groundwork for totalitarianism. Popper criticized the prevailing *historicist* conception of history, i.e. the idea that history 'marches on' irrespective of human actions towards the rule of the proletariat, doomsday, or some other conclusion of human history. Thus, in Popper's view, a closed society is not necessarily one which prevents foreigners from entering the country – like Japan did during its 'sakoku' period until 1853 – but one that discourages individual responsibility and prevents *new ideas* from taking hold. Simply speaking: In a totally closed society, everything that needs to be known is known. As the course of history is determined, the great leader, the party, the junta or the church have all the answers.

Today's conception of open society leans on Popper's ideas, but greater emphasis is placed on the interplay of morals, culture, politics and institutions. The connection between culture and migration is acknowledged: New ideas take hold in a country, amongst other reasons, because new people take hold. Trade is acknowledged as a central factor, furthermore, and geography regarded as vital. Japanese 'sakoku' isolationism was enabled by the fact that the Shogunate was a rather centralized realm on a relatively isolated, by and large self-sufficient group of large islands. In contrast, classical Greek culture flourished because of the geography of the Aegean with long continental coastlines and numerous small islands, which were not self-sufficient, had to trade. It is exposure to other ways of life that stimulated philosophical thought and defeated the dogmatism of traditional communities. Xenophanes, who shrewdly remarked that if horses had

gods they would look like horses, and cattle gods like cattle (Xenophanes, 1992, p. Fragment 15), was an itinerant philosopher, a widely travelled man.

The contemporary conception of open vs. closed is more straightforward than Popper's philosophical conception then. Whether they regard history as determined or not, closed societies normally are characterized by wariness towards foreigners and 'foreign' ideas. Closed societies tend to be more insistent, for example, that visitors follow their customs. When Western female politicians visit Islamic countries, there is considerable pressure that the representatives cover their hair. In contrast, there is very little pressure on Saudi sheiks to wear 'proper' suit and tie at international events in the West.

The matter is not one of black and white, of course, and one should compare *overly* open with *overly* closed society. Political scientists have offered the open-close spectrum as an additional dimension in conjunction with the classic distinction between left and right. Although chancellor of Germany Angela Merkel and US president Donald Trump were both considered conservatives, Merkel's conservativism in Germany was of the open, globalist kind. In contrast, Donald Trump's conservativism emphasized return to 'American' values. Political scientists also note a number of 'hybrid regimes' that do not display the classic pattern of market economy, democracy and cultural openness going hand in hand (Linz, 2000). China offers perhaps the prime example of a political system that allows considerable economic leeway but severely curtails political freedom. Hale refers to Putin's Russia to point out that hybrid regimes are not simply 'half-way categories', however: "hybrid regimes have their own distinct dynamics that do not simply amount to half of what we would see in a democracy plus half of what we would see in an autocracy." (Hale, 2010, p. 34)

OPEN SOCIETY AND COMMUNICATION

Despite more multi-faceted, tangible understandings, the epistemic, or perhaps 'communicative' dimension, remains at the core. In the end, the question is what the state and society are supposed to be about. In totalitarian states, the most extreme case of a 'modern', i.e., non-traditional closed society, society is built around a dominant narrative of a greater cause to which everything is subjected. The greater cause then allows categorization of individuals into those who 'make a contribution' and those who do not. In Hitler's national-socialist fascism, individuals were regarded as valuable insofar they contributed the *volk's* struggle for *lebensraum*. The regime of the Communist Party in the Soviet Union subjected the individual life to a similar totalitarian rationale. As Alexander Aleksandr Solzhenitsyn's *Gulag Archipelago* showed, rule of law had no meaning when you were branded an enemy of the party (Solzhenitsyn, 1973). In recent times, the advent of the 'caliphate' of ISIS or the rule of the Taliban in Afghanistan marked the return of political bodies governed by Qur'an scripture. By way of contrast, there is no overriding rationale or revelation at the centre of open society except the idea that individuals themselves are responsible for their lives.

Even in the most totalitarian regimes, people find ways to 'speak their minds', either privately, among trusted friends, or in coded ways, between the lines. However, one of the major defining criteria for truly open society is the depth and breadth of public debate and public life. Depth means that public debate impacts political decisions, furthering 'substantive' as opposed to merely 'procedural' democracy – in genuine democracies, political leaders cannot ignore public opinion for fear

of punishment at the polls. As for breadth, in truly open societies large minority groups will not only be tolerated but will have their voices heard: not only the government's Islam expert will talk on television, but the representative of the Muslim community as well.

Historically, the struggle for individual freedom began not with the mind but with the body and the purse. The *Magna Charta Libertatum* (Great Charter of Freedoms) signed in 1215 on the fields of Runnymede is often hailed as a medieval guarantee of human rights, but it is predominantly concerned with safeguarding against arbitrary arrest (*habeas corpus*) and exploitative taxation of the aristocracy. After the reformation, matters of the soul came to the fore as well – with religious freedom, not only tolerance, high on the agenda. But there are other issues associated with freedom as well. For large segments of conservative society in the US, 'freedom' means that the 'right of the people to keep and bear arms shall not be infringed' (US Constitution, 2nd Amendment). In today's media society, however, 'communicative freedom' – freedom of thought and opinion, and especially freedom of speech – are perhaps the most-discussed individual liberties.

In John Stuart Mill's classic conception, espoused in the essay 'On Liberty' (Mill, 1859), two of the three basic liberties of individuals are associated with public life and communication: a) liberty of thought and opinion; and b) liberty to join other like-minded individuals for a common purpose that does not hurt anyone (the third is, c) liberty of tastes and pursuits, i.e., the freedom to plan one's own life). For classic liberals like Mill, liberty of thought and opinion was far-reaching. Mill held that no expression of opinion should ever be silenced, as that would rob the public of the opportunity to learn: "If all mankind minus one, were of one opinion, and only one person were of the contrary opinion, mankind would be no more justified in silencing that one person, than he, if he had the power, would be justified in silencing mankind." (Mill, On Liberty).

In the Victorian Age, people's opinions were confined to coffee-houses, salons and their immediate circles, of course; access to the reading public at large was controlled by newspaper editors. With today's social media, everyone has potential access to everyone. But the epistemic core of liberalism, shared by proponents of deliberative democracy and libertarians today, lies in the conviction that open and public debate will reveal the truth and expose the lie. The common conception, which is also reflected in law in most countries, is that freedom of speech protects the right to freely express one's *opinion*, be it false or correct, honest or not. The protection of *false factual statements*, such as "scientific research shows that the earth is flat" – always has been controversial. In extremely interconnected 'network societies' (Castells, 1996), where ostensible scientific proof for almost every conceivable opinion is on offer, the difference quickly breaks down, however. It is here that openness and closedness reappear. As open society does not allow for ultimate arbiters of truth, there is no authority that could declare the earth round once and for all (and everyone who claims otherwise a heretic like the Catholic Church did with Galileo). The solution open societies seem to have found to deal with overload is to replace epistemic requirements with *moral* ones. No attempts to disprove flat earthers are made, but they are morally disqualified as 'conspiracy-theorists' and 'cancelled', i.e., unfollowed and not given a platform.

The problem with moral criteria is, of course, that they are underpinned by emotions and subject to mood swings in society as well as group dynamics. For strategic or opportunistic reasons, or simply for self-protection against cognitive

overload, cancellation quickly moves from outrageous and rationally indefensible standpoints to perfectly defensible ones that in-groups simply do not find agreeable. In other words, overly open society, paradoxically, may lead to closed ideological dogmatism in subgroups. When aggressive social movements instigate irrational moral trials of individuals in the court of public opinion, these witch hunts are not only at odds with liberal ideals but with the rule of law altogether.

ADVANTAGES AND DISADVANTAGES

Closed societies can be surprisingly robust and resilient under the right circumstances. For example, over the course of almost 200 years, traditional-religious, quasi-tribal collectives living on remote territories of today's Afghan state (although not necessarily *in* the Afghan state) ousted the respective superpowers of the time: the British, the Soviet Union, and the US-led coalition. By and large, however, closed societies, whether congruent with a state or not, suffer from severe disadvantages. Viewed outside-in, the perhaps greatest downside is that closed societies simply do not provide the economic prosperity and quality of life pluralistic and open societies offer. In their attempts to keep society closed, regimes tend to suppress individual initiative and create a culture of fear which hampers economic growth and stifles cultural life. Viewed inside-out, closed societies are under constant threat for the very same reason: as soon as citizens glimpse other ways of life, life at home appears unattractive – it is said, tongue-in-cheek, that the socialist German Democratic Republic foundered because the West had bananas, East Germany did not. In order to manage dissent, the closed society has to become more and more repressive, hence more unattractive. North Korea is a contemporary example.

Another disadvantage of closed societies is that they cannot adapt to changing circumstances as quickly. As every change has to conform to preordained wisdom, the mere fact of 'closedness' hampers innovation. It was not the Mujahideen fighters who developed the Stinger surface-to-air missile instrumental in breaking Soviet air dominance; it is doubtful they would have prevailed over the Red Army without modern weapons delivered by the U.S. Similarly, accounts of the great variety of long-term viable political systems should be viewed with some scepticism: During the Cold War, many authoritarian regimes – e.g., in Africa or Latin America – only existed because they were propped up as superpower proxies. Today, superpower confrontation has taken a backseat to the politics of resource extraction.

Although instruments like the Ipsos Global Happiness Report, the World Happiness Report, the World Population Review and other quality of life-rankings are essentially Western constructions, comparative research suggests very strongly that life in open societies tends to be 'better' for the average citizen. While criticisms of capitalism and its unfair and exploitative aspects remain valid, modern market economies have created unprecedented levels of prosperity and standards of living. Excessive openness creates problems and disadvantages of its own, however.

A common criticism, both from the inside and out, is that modern Western states are 'weak' (e.g. van Creveld, 2016). Because of excessive tolerance for all kinds of misguided causes, it is suggested, Western citizens have become soft and disinterested in the common good – Western cultures turned into 'victimhood cultures' (Campbell & Manning, 2018). Thus, once the rest have caught up with the West's technological edge, overly open societies will become easy prey to more

homogenous, better disciplined regimes. Since 'modern' democracies are relatively new and the technological gap has been closing only recently, it is too early to tell. On the one hand, Western-led coalitions had to withdraw from Vietnam, Afghanistan and Iraq with their aims only partly achieved, mainly because citizens at home were not willing to carry the burden of interventionism. On the other hand, autocrats banking on the unwillingness of open societies to defend their values have again and again found, to their ruin, that the 'decadent' West is capable of very robust response.

Perhaps the key problem of open society is a reduced ability to culturally, emotionally and spiritually integrate individuals. East German propaganda denigrating West Germany emphasized the misery of drug users, for example, i.e., of society's cast-outs. Evolutionarily speaking, homo sapiens evolved as a small group-species living in bands of about 50-150 hunter-gatherers – a political system described by social anthropologist Ernest Gellner as 'tyranny of cousins' (Fukuyama, 2011, p. 53). It is not surprising, therefore, that there is a lingering longing for cultural homogeneity and a well-ordered, bounded society. Anthropologists stress religion as the glue that held prehistoric communities together, furthermore, so the yearning for a spiritual community united by a great cause is not surprising either (Norenzayan, 2013).

A corollary of reduced cultural and spiritual integration is reduced political engagement in the sense that large parts of society do not participate in politics anymore. Political theory has always been sceptical about the average citizens' ability to politically engage to the degree liberal democracy requires. Walter Lippman captured the haphazard fashion of public opinion-formation metaphorically: "The public will arrive in the middle of the third act and will leave before the last curtain, having stayed just long enough perhaps to decide who is the hero and who the villain of the piece." (Lippmann, 2005 [1925], p. 38).

With overly open society, the concern goes beyond incompetence and apathy. Political scientists like e.g. Colin Crouch (2004, 2020) warn that Western societies are turning into *post-democracies*, amongst other reasons because of excessive openness – or pseudo-openness, really. Overly open societies offer so many ways of engaging in narrow, personal issues – identity politics is the keyword here – that the political process becomes fractured, a cacophony, not a debate. Because the identity-driven citizen does not solidarize with groups large enough to be politically relevant – i.e. the interests of 'labourers' or 'farmers' as opposed to vegan, gluten-intolerant foodies with attention disorder – strategically focused elites behind the scenes successfully divide and conquer, engineering political decisions in their favour. Crouch says: "A post-democratic society is one that continues to have and to use all the institutions of democracy, but in which they increasingly become a formal shell. The energy and innovative drive pass away from the democratic arena and into small circles of a politico-economic elite." (Carrigan, 2013; Crouch, 2004).

The post-democratic criticism, to be clear, is not that citizens have no political power in overly open society – only that they have far less say about *relevant matters* than they are led to believe. Crouch's and similar analyses are not only concerned about the democratic deficit – which closed societies have to a far larger degree – but about deceptive, hypocritical structures. Despite the rhetoric of openness for public consumption, post-democratic analyses imply, society is not open because everyone is genuinely committed to the constant,

unprejudiced search for better, fairer solutions. The real reason is that moral and cultural relativism, i.e., constantly shifting values, benefit *plutocracy* – i.e., the rule of a wealthy few who largely stay in the background. The globalism displayed by modern market economies is not due to universal appreciation of humankind, thus, but reflects capitalist requirements: the free flow of cheap labour must not be hampered. In essence, ruling elites keep up the illusion of an open society by allowing all kinds of opinions where it does not matter. Where it matters, in contrast, the corridor of opinions and 'realistic' choices is strategically engineered and remains relatively limited.

Whether correct or not, the post-democratic analysis highlights several conceptual issues. Firstly, fairly representing the people and undogmatically adapting to a changing environment are not the same thing. Viewed from the outside-in, democratic societies are not only faced with their own peculiar problem of political justice but face the problem non-democratic societies face as well: viable political *order* with living standards acceptable to the population. For theorists of deliberative democracy like e.g. Jürgen Habermas, the one comes with the other. It is assumed that decisions made by way of open, unprejudiced and rational debate will not only be fair but 'good', i.e., deliver the desired result (Habermas, 1984; Habermas, 1996). Given the complexity of a globalised economy, other theorists are sceptical. Samuel Huntington repeatedly pointed out during his career that political order without justice can be viable while political justice is impossible without order (see e.g.Crozier et al., 1975).

A second noteworthy issue lies in the fact that Crouch's academic criticism and other similar accounts by political theorists are by and large identical with narratives commonly dismissed as populist. When political scientists like Crouch postulate a plutocracy pulling the strings behind the scenes, their assertions are seriously debated; yet the very same assertion made in some internet forum is dismissed as crackpot conspiracy theory. The fact remains that there is substantial overlap between serious academic analysis, conspiracy theory and the narratives employed in psychological attacks seeking to undermine open society. On the positive side, this can be regarded as typical for open society, which is almost defined by the fact that it not only allows but encourages its own criticism. On the negative side more relevant for psychological defence, it means that open society itself forges the weapon with which to attack it.

Thirdly and finally, the inherent disadvantage in openness is, of course, that it is non-defensive. Critics of open society win both ways. If criticism is permitted and the intolerant tolerated, it might do some damage. If criticism is not repressed and actions against the intolerant taken, the adversary simply points out how hollow the commitment to open society really is. The coming years will show how open society deals with its fundamental disadvantage.

DISCUSSION

- Some groups see dangers in Western democracies becoming too open? What are the dangers associated with overly open society?
- What does it mean to say that Popper's concept of an open society is predominantly epistemic in character?
- Mill thinks no expression of opinion should ever be silenced, because one can also learn from being exposed to wrong opinions. After the experience of totalitarian propaganda machines in the 20th century, theorists have become more cautious and guarded here. What do you think?
- Crouch does not claim that we live in post-democratic societies but that we are moving towards pseudo-democracy. Do you agree?
- What are the advantages of open society?

HOWARD NOTHHAFT, *PhD, is Associate Professor at the Department for Communication (IKO at Lund University), and Visiting Professor at University of Johannesburg. His main interests lie at the intersection of strategic communication, psychology and cognitive science.*

REFERENCES

Campbell, B., & Manning, J. (2018). *The Rise of Victimhood Culture - Microaggressions, Safe Spaces, and the New Culture Wars.* PalgraveMacMillan.

Carrigan, M. (2013). *British Politics and Policy: Five minutes with Colin Crouch*. http://blogs.lse.ac.uk/politicsandpolicy/five-minutes-with-colin-crouch/

Castells, M. (1996). *The Rise of the Network Society: The Information Age: Economy, Society and Culture, Volume I.* Blackwell.

Crouch, C. (2004). *Post-Democracy.* Polity.

Crouch, C. (2020). *Post-Democracy: After the Crises.* Polity.

Crozier, M., Huntington, S., & Watanuki, J. (1975). *The Crisis of Democracy: Report on the Governability of Democracies to the Trilateral Commission.* New York University Press.

Fukuyama, F. (2011). *The Origins of Political Order: From Prehuman Times to the French Revolution.* Profile Books

Habermas, J. (1984). *A Theory of Communicative Action.* Heinemann.

Habermas, J. (1996). *Between Facts and Norms. Contributions to a Discourse Theory of Law and Democracy. Translated by William Rehg.* The MIT Press.

Hale, H. E. (2010). Eurasian Polities as Hybrid Regimes: The Case of Putin's Russia. *Journal of Eurasian Studies,* 1(1), 33-41. https://doi.org/10.1016/j.euras.2009.11.001

Linz, J. J. (2000). *Totalitarian and Authoritarian Regimes*. Lynne Rienner.

Lippmann, W. (2005 [1925]). Excerpt from The Phantom Public. . In J. Gripsrud, H. Moe, A. Molander, & G. Murdock (Eds.), *The Idea of the Public Sphere. A Reader.* (pp. 24-41). Lexington Books.

Mill, J. S. (1859). *On Liberty.* https://archive.org/details/onlibertyxero00milluoft

Norenzayan, A. (2013). *Big Gods. How Religion Transformed Cooperation and Conflict.* ([Kindle Edition] ed.). Princeton University Press.

Popper, K. (2005). *The Open Society and Its Enemies: Hegel and Marx.* Routledge. (1945)

Solzhenitsyn, A. (1973). *The Gulag Archipelago.*

van Creveld, M. (2016). *Pussycats: Why the Rest keeps beating the West.* DLVC Enterprises.

Xenophanes. (1992). *Fragments. A Text and Translation with a Commentary. Edited by J. H. Lesher.* University of Toronto Press.

3. THE LEGAL FRAMEWORK IN SWEDEN

ANNA ANDERSSON, SALLY LONGWORTH & PONTUS WINTHER

SUMMARY

- Individuals have the right to freedom of opinion and freedom of expression without undue interference in accordance with, inter alia, the Swedish Instrument of Government and the European Convention on Human Rights.
- Expressions in the press and in radio, television, film and certain similar media enjoy extended legal protection by the Swedish Freedom of the Press Act and the Fundamental Law on Freedom of Expression.
- Censorship, that is restriction of expressions, in advance is strictly regulated in Swedish law and in the European Convention on Human Rights. It is, as a main rule, prohibited.
- There must always be an legal basis for Swedish public institutions to act in countering malign influence operations.
- Swedish public institutions may provide correct information in reaction to already publicised information that concerns their area of activities and which is deemed factually incorrect.

This chapter outlines the fundamental parts of the legal framework applicable to Swedish public institutions in addressing malign influence activities through information measures.[2] Malign influence campaigns have been shown to have impact on the right to freedom of expression.[3] The chapter therefore focuses on the parameters of this right under international and Swedish law during peacetime.[4]

The right to freedom of expression holds an important place within the constitutional framework and history of Sweden (Nordin, Giles and Graves, 2023; Axberger, 2019). It is protected in three of the four constitutional instruments:

- The Instrument of Government (*Regeringsformen (1974:152),* RF);
- The Freedom of the Press Act (*Tryckfrihetsförordningen (1949:105)*, TF); and
- The Fundamental Law on Freedom of Expression (*Yttrandefrihetsgrundlagen (1991:1469)*, YGL).

Moreover, the European Convention on Human Rights 1950 (ECHR) includes this

[2] The chapter is based on the findings in Andersson, 2023.
[3] Freedom of expression has been identified as a "touchstone" right, meaning the enjoyment of which is often indicative of the enjoyment of all other human rights. Breaches of freedom of expression therefore often lead to a chain of breaches of other human rights. As such, it may not be the only right impacted by malign influence, but provides a useful starting point for analysis. See UN General Assembly (1946) *Resolution 59(1);* and Hussain, 1994, para. 14.
[4] Official translations of Swedish law have been used where possible. On the legal framework applicable during armed conflict, see Longworth, 2022; and Winther, 2019.

right under Article 10.[5] The ECHR is directly incorporated as Swedish law *(lag 1994:1219)* and has a special status (RF Chapter 2 Art. 19), according to which public institutions and courts must not apply law or regulations that are in conflict with the ECHR or the constitution (RF Chapter 11 Art. 14 and RF Chapter 12 Art. 10). The right to freedom of expression is included in numerous other international treaties that Sweden is a party to (CPD Art. 21, CERD Art. 5, CRC Arts. 13 and 17).[6] Since Sweden is, in principle, a dualist state, international and domestic law are generally treated as separate systems. Public institutions and courts should nevertheless interpret Swedish law in light of applicable international law (Bring, Klamberg, Mahmoudi, Wrange, 2020, 63-64).

The right to freedom of expression is broad and includes a number of distinct individual rights. It covers the right to freedom of opinion without undue interference (ICCPR Art. 19(1); ECHR Art. 10(1)), the right to express oneself (ICCPR Art. 19(2); ECHR Art. 10(1)), the means of expression (ICCPR Art. 19(2); Schabas, 2015, 455-456), the right to seek and receive information, ideas and opinions (ICCPR Art. 19(2); ECHR Art. 10(1)), and the right to access information from public institutions (ICCPR Art. 19(2); ECtHR, *Magyar Helsinki Bizottság v. Hungary, 2016*). In the Swedish constitution, freedom of expression and freedom of information are established as two separate rights (RF Chapter 2 Art. 1, pt. 1 and 2).

The right to freedom of expression is a right "regardless of frontiers" (ICCPR Art. 19(2); ECHR Art. 10(1)), meaning that individuals have the right to share, seek and receive information, ideas and opinions beyond the territory of the State. It is the only right under the ECHR that provides that individuals have duties and responsibilities in exercising it (ICCPR Art. 19(3); ECHR Art. 10(2)). The exercise of this right has meaning for both the individual and society in enabling public debate through shared news, information and opinions.[7] This chapter focuses on those aspects of the right shown to be vulnerable to malign influence activities.[8]

FREEDOM OF OPINION WITHOUT UNDUE INFLUENCE

Individuals have the right to freedom of opinion without undue interference (ICCPR Art. 19(1); ECHR Art. 10(1)). ECHR Art. 10(1) provides this is a right "without interference from public authority", whereas no such limitation is included in ICCPR Art. 19(1). The right covers all forms of opinions, the right to change opinions and the right not to have an opinion (HRC, General Comment No. 34, para. 9; Schabas, 2015, 580). The right to freedom of opinion is linked to the right to freedom of thought and conscience protected under ICCPR Art. 18(1) and ECHR Art. 9(1). Whether something amounts to a violation of one or the other will depend on the facts (Hussain, 1994, para. 25; Schabas, 2015, 457; HRC, *Yong-Joo Kang v. Republic of Korea*, 2003).

[5] As a member of the European Union (EU), Sweden is also bound by the EU Charter on Fundamental Rights and Freedoms 2010 when Sweden implements and applies EU law. This includes Art. 11 on freedom of expression and information.
[6] Note as the CRC has been incorporated into Swedish law through countering malign influence operations directed against children should therefore take into account a child rights perspective, see lag (2018:1197) om Förenta nationernas konvention om barnets rättigheter.
[7] The work of journalists has been identified as encapsulating both dimensions of the right. See IACtHR, *Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism,* Advisory Opinion, 1985, paras. 30-34 and 70-72. See also ECtHR, *Axel Springer AG v. Germany,* 2012, para. 79; ECtHR, *The Sunday Times v. the United Kingdom (no. 2)*, 1991, para. 50; and ECtHR, *Bladet Tromsø and Stensaas v. Norway,* 1999, paras. 59 and 62.
[8] It is acknowledged that other aspects of the right to freedom of expression may also be vulnerable to unlawful interference from malign influence activities. One example includes abuse of the right to freedom of information.

No individual should suffer an impairment of their rights based on their actual, perceived or supposed opinions (HRC, General Comment No. 34, para. 9). Criminalising holding an opinion would therefore be a breach of this right. As a right to a "freedom", individuals can choose whether or not to express their opinions and should not be coerced to do so (HRC, General Comment No. 34, para. 10). Under Swedish law, RF Chapter 2 Art. 2 provides that "no one may in his or her relations with the public institutions be coerced to divulge an opinion in a political, religious, cultural or other such connection".

Not all efforts to influence the opinion of individuals are "undue", however, and the exchange of information, ideas and opinions is a core function of the societal dimension of freedom of expression. Indeed, States may be obliged to provide information in order to protect other human rights, such as public health information (see further below). A careful balance is therefore essential so as not encroach on the right to freedom of opinion, and the right to seek and receive information, so that the State does not monopolise the information environment (Hussain, 1994, para. 26; Nowak, 2005, 441). States are not the only actors that may exert influence over the opinions of individuals. Ensuring the individual's knowledge, awareness, voluntary engagement and choice is therefore a key component (Khan, 2021, paras. 66 and 80).

THE RIGHT TO EXPRESS ONESELF AND THE RIGHT TO INFORMATION

Freedom of expression includes the right to express and impart information, ideas and opinions of all kinds (ICCPR Art. 19(2); ECHR Art. 10(1)). "Expression" has been defined broadly in international law (HRC, General Comment No. 34, paras. 11-12). RF Chapter 2 Art. 1 pt. 1 similarly defines freedom of expression broadly as "the freedom to communicate information and express thoughts, opinions and sentiments...". RF Chapter 2 Art. 1 pt. 2 further guarantees the freedom to procure and receive information and otherwise acquaint oneself with the utterances of others.

The right covers "every form of idea and opinion capable of transmission to others" (HRC, General Comment No. 34, para. 11), including expression that may be regarded as "offensive, shocking or disturbing" (ECtHR, *Handyside v. The United Kingdom*, 1976, para. 49). Spreading false information and factually incorrect information can also be protected under the right (ECtHR, *Salov v. Ukraine*, 2005; Andersson, 2023, 28). In *Lingens v. Austria*, 1986, para. 46, the European Court of Human Rights (ECtHR) noted that the "existence of facts can be demonstrated, whereas the truth of value-judgements is not susceptible to proof", and requiring individuals to prove the proof of value judgments is an infringement on freedom of opinion (see further *Makraduli v. The Former Yugoslav Republic of Macedonia,* 2018, para. 62). Untrue factual statements disguised as value judgments, however, are not protected under the right. The right protects also the means used to express oneself. As such oral, written, printed, artistic, digital and any other form of media chosen are also protected. ICCPR Art. 19(2) lists these means of expression with the general clause to cover "any other...media", and has been applied to digital means of expression (see HRC, General Comment No. 34, para. 12). ECHR Art. 10(1) does not include a similar list, but the jurisprudence of the ECtHR recognises protection for the different means listed in ICCPR Art. 19(1) (Schabas, 2015, 455-456; ECtHR, *Delfi AS v. Estonia*, 2015; ECtHR, *Ahmet Yildirim v. Turkey*, 2012).

Under Swedish law, RF provides that the freedom to communicate information and express thoughts, opinions and sentiments, whether orally, pictorially, in writing, or

in any other way is protected (RF Chapter 2 Art. 1, pt. 1). The TF is only applicable to expressions in the press and the YGL only to expressions in radio, television, film and certain similar transmissions and recordings (RF Chapter 2 Art. 1, para. 2). While only applicable to specified forms of media, the TF and YGL provide an extended guarantee to freedom of expression in these media forms (Andersson, 2023, 43).

Freedom of expression also includes the right to seek, receive and access information, ideas and opinions. States uphold this right by ensuring plural and diverse information space, protecting the independence of the media, facilitating digital literacy, providing access to public libraries, digital means and government archives, among others. Human rights bodies have emphasised the importance of expressions on matters of public interest, including critique of the government, journalism, political and academic expressions (ECtHR, *Feldek v. Slovakia*, 2001, para. 83; ECtHR, *Sürek v. Turkey* (No. 1), 1999, para. 61. See also ECtHR, *Bladet Tromsø and Stensaas v. Norway*, 1999, para. 62; HRC, General Comment No. 34, paras. 34 and 37–49). In addition, individuals have a specific right to request access to information held by the State (see HRC, General Comment No. 34, 2011, paras. 18 and 19),[9] and States have a positive obligation to provide information on their activities to the public (UN Secretary General, 2022, para. 27; UN Human Rights Council, Resolution 49/21, 2022, preamble para. 16).

THE OBLIGATIONS OF THE STATE

The right to freedom of expression is a right of individuals. States are obliged to respect, protect and facilitate this right within their jurisdiction (ICCPR Art. 1; ECHR Art. 1; RF Chapter 2 Art. 1 pt. 1). Any act by Swedish authorities aimed at countering malign influence activities must therefore uphold these requirements (Andersson, 2023, 52).

The obligation to respect is a negative obligation requiring that States do not act in a way that would violate the right to freedom of expression. Accordingly, acts aimed at countering malign influence activities must not themselves constitute a breach of the right to freedom of expression, or any other human rights (Andersson, 2023, 25 and 38). States are also obliged to protect individuals within their jurisdiction from interferences with the exercise of their human rights. This includes protection from interferences from other individuals within and outside the State's jurisdiction, and from interferences from other State actors (HRC, General Comment No. 31, 2004, para. 8).[10] The obligation to protect requires establishing clear regulatory frameworks, carrying out adequate investigations into alleged interferences with individual's rights, and taking measures to address unlawful interferences (HRC, General Comment No. 34, para. 23).

The obligation to facilitate requires States to implement measures so that individuals are able to exercise their rights. This requires legislative, administrative, budgetary, judicial and other actions available to the State (Schabas, 2015, 453-454; HRC, General Comment No. 34, paras. 15, 16, 19, and 20; OSCE Joint

[9] This right relates to information held by public authorities. Individuals may have a right to access to certain information held by private actors through the exercise of the right to privacy and data protection. See further ECtHR, 2020.

[10] An important example in the context of malign information interferences is the obligation of States to protect individuals from "hate speech", that is advocacy and incitement to violence, discrimination or hostility. See BrB (1962:700) Chapter 16 Art. 8; ICCPR Art. 20(2); and ECtHR, *Nepomnyashchiy and other v. Russia,* 2023. See further CERD Art. 4 and Convention on the Prevention and Punishment of the Crime of Genocide 1948, Art. 3(c). See also HRC, General Comment No. 34, paras. 50–52; Kaye, 2019; CERD Committee, 2013.

Declaration 2017, Principle 1a(i)–(v) and Principle 1b(i)-(ix)). Protecting and facilitating the right to freedom of expression have been identified as essential for countering malign information campaigns (UN General Assembly, Resolution 76/227, 2022, UN Secretary General, 2022, paras. 27–40 and 57; OSCE Joint Declaration, 2017, Principle 3a), such as ensuring pluralistic information sources and an environment where others can call out disinformation. This corresponds with the objective of the Swedish psychological defence (SOU 2020:29, 77).

RESTRICTIONS ON THE RIGHT TO FREEDOM OF EXPRESSION

Freedom of expression is not an absolute right. It is recognised in international law and provided in Swedish law that restrictions may be introduced by the State under specific circumstances (ICCPR 19(3); ECHR 10(2); RF Chapter 2 Art. 20 pt.1). Indeed, restrictions may themselves be a means of protecting this and other rights, such as with protections against hate speech and regulation of the revocation of broadcasting licenses for the spread of such content (for example, ECtHR, *NIT S.R.L v. Moldova*, 2022).

Restrictions must be prescribed by law,[11] pursue a legitimate aim, be necessary in a democratic society, and proportionate to achieve the aim pursued (ICCPR Art. 19(3); ECHR Art. 10(2); RF Chapter 2 Arts. 20-23 and 25; HRC, General Comment No. 34, paras. 21–36). Note that there are differences between the limitation grounds permitted in the different treaties. The ECHR contains more restricting grounds than the ICCPR. The need for restrictions must be established in relation to the precise threat to be addressed in specific and individualised fashion (HRC, General Comment No. 34, para. 35; ECtHR, *Stoll v. Switzerland*, G 2007, para. 101). Human rights bodies have held that there is little scope for restrictions on political speech or debates on questions of public interest (ECtHR, *Feldek v. Slovakia*, 2001, para. 83; ECtHR, *Sürek v. Turkey* (No. 1), 1999, para. 61; HRC, General Comment No 34. paras. 34, 37–38), and the measures must be the least intrusive (HRC General Comment No. 34, para. 34). Legal restrictions on the right to freedom of expression under Swedish law can only be introduced through legislation adopted by the Swedish parliament *(Riksdag)* (RF Chapter 8 Art. 2). Examples of Swedish restrictions includes the criminalisation of hate speech (BrB *(1962:700)* Chapter 16 Art. 8), and slander (BrB Chapter 5 Art. 1).

Restrictions should not themselves breach the right to freedom of expression, or breach other rights (HRC, General Comment No. 34, para. 26). This is explicitly provided in RF, according to which restrictions may not extend so far that they represent a threat to the free shaping of opinion as one of the foundations of democracy, and no restriction may be imposed solely on grounds of political, religious, cultural or other such beliefs (RF Chapter 2 Art. 21). The TF and YGL contain specific requirements for restricting expression covered in the press, radio, television, film and similar transmission and recordings (see further below).

PROVIDING INFORMATION TO COUNTER DISINFORMATION

States are obliged to proactively provide information on their activities to the

[11] The ECtHR has consistently held that this not only requires that the measure should have a legal basis in domestic law, but it should also be accessible to the person concerned and foreseeable as to its effects. This requires that the measures be formulated with sufficient precision to enable people to regulate their conduct. See ECtHR, *Mariya Alekhina and others v. Russia,* 2018, paras. 253–255.

public as part of facilitating and protecting human rights (IACtHR, *Case of Perozo et al. v. Venezuela*, 2009, para. 151). This also applies to countering of malign influence operations. States are also obliged to engage in public debate, particularly where the State's activities in protecting human rights are called in to question (ECtHR, *Steel and Morris v. the United Kingdom*, 2005 para. 89; ECtHR, *Magyar Helsinki Bizottság v. Hungary*, 2016, paras. 166 and 168). Public institutions' information measures may protect or facilitate the right to information and other individual rights (Andersson, 2023, 14 and 38).

To counter malign influence operations, Swedish public institutions must identify and act within an applicable legal basis (Andersson, 2023, 39–40 and section 5 above). This follows from the fundamental principle of legality (RF Chapter 1 Art. 1 para. 3), and that all activities undertaken by Swedish public institutions must have a basis in the legal order (Administrative Procedure Act *(Förvaltningslag (2017:900)*, FL) Art. 5; Lebeck, 2018). For information measures, this may be found in legislation, regulations, decrees, instructions by government or other acts based in the legal order (Prop. 2016/17:180, 289). Note that for certain other types of measures taken by public institutions, such as decisions, the legal basis must be found in legislation (SOU 2010:29, 148; Sterzel, 2020, 87–88; Lebeck, 2018, 109-110).

As a general staring point, FL Arts. 6-7 obliges all public institutions to provide information relating to their services and availability. This may also serve as a legal basis to proactively provide information and to reactively respond to malign influence operations that relate to the public institution's area of activities (Andersson, 2023, 44). Note, however, that FL is subsidiary to other more specific law and hence special acts takes precedence over FL (FL Art. 4).

Institutions may only act upon malign influence operations that relate to their area of activities (Andersson, 2023, 45–46). The information provided must also constitute "pure information", advice or guidance. This means that information must not be conflated with injunctions or decisions, which require a different legal basis (SOU 2010:29, 207; Lebeck, 2018, 189). Where malign influence operations targets the area of activities of several public institutions, affected institutions need to coordinate their counter measures (FL Art. 8). In doing so, institutions may seek support from the Psychological Defence Agency (*Myndigheten för psykologiskt försvar,* MPF), see below.

Some public institutions have other information assignments in special law that involves active dissemination of information to specific groups of society about their services. For example, the municipal social services are obliged to provide information to families and individuals about its services and to inform specifically about possibilities on ways out of abuse under the Social Welfare Act (socialtjänstlag (2001:453), SoL Chapter 3 Arts. 1 and 7). This can be considered as a legal basis to counter malign influence operations in light of the objective of the psychological defence, provided that malign influence activities are directed against the relevant groups and services. This was the case with the malign information operation directed against the Swedish social services and the Care of Young Persons Act *(lag (1990:52) med särskilda bestämmelser om vård av unga,* LVU). (Andersson, 2023, 43).

A few institutions have specific, permanent assignments associated with malign influence. This includes the Swedish Institute (*Svenska Institutet,* SI), and MPF. SI is tasked to share information and knowledge about Sweden, promote Swedish

interests abroad, and to analyse disinformation that relates to Sweden *(förordning (2015:152)* Art. 1 and Art. 2 pt. 2). MPF has a specific obligation to identify, analyse and be able to provide support in the countering of malign influence operations directed against Sweden or Swedish interests *(förordning (2021:936)* Art. 2 pt.1). Accordingly, MPF has an essential role to support other public institutions that may face malign influence operations, for example by disseminating knowledge on psychological defence and promoting coordination *(förordning (2021:936)* Art. 2 pt. 1-2 and 5; SOU 2020:29, 196).

Other public institutions have on occasion been tasked to counter specific malign influence campaigns through information measures. For example, during the Covid-19 pandemic, the government tasked several institutions specifically to provide information on vaccination (Government Decision 17 December 2020, dnr. S2020/09533). The government also tasked the National Board of Health and Welfare *(Socialstyrelsen)* to counter the malign information operation against the social services and LVU, through dialogue and information (Government Decision 7 July 2022, dnr. S2022/03244. See further Andersson, 2023, 42).

In formulating and disseminating information and communication, public institutions must ensure that their measures are not themselves a breach of the freedom of expression, but respect and safeguard this right. Information should be formulated so that it is not perceived as inhibiting critique against it or a free and open debate but rather protects the freedom of expression and a free exchange of information (Andersson, 2023, 51-52; SOU 2020:29, 77). Institutions must not make, sponsor, encourage or further disseminate statements which they know or reasonably should know to be false or inaccurate (IACtHR, *Case of Perozo et al. v. Venezuela*, 2009, para. 151; OSCE Joint Declaration, 2017, Principle 2(c) and 2(d)).[12]

Information provided by public institutions must further be objective, impartial, and non-discriminatory (RF Chapter 1 Art. 9; FL Art. 5 para. 2; ECHR Art. 14; Andersson, 2023, 52-55). The institutions must ensure that the information they provide is factually correct and reliable, and has a solid foundation in facts (Andersson, 2023, 54). Only circumstances that are relevant should be considered and included in information measures (Bull, 2020, 107). The information should be formulated and presented in a correct and respectful manner. Derogatory language should not be used (JO 2016-06-01, dnr 678-2015). To be impartial, the information should be comprehensive and objective, and institutions and their employees should act in an unbiased manner without conflict of interests (JO 2013-11-21, dnr. 5875-2012; JO 2010-02-16, dnr. 4935-2009; Bull, 2020, 105).

Public institutions should also generally avoid to seek to form opinions in society in ways that could be perceived as the institution taking a stand in political issues that should be decided by parliament or government (Prop. 2009/10:175, 40). There are, however, exceptions as some authorities are tasked to “affect knowledge, attitudes and behaviour” within their field, for example relating to public health issues (Prop. 2009/10:175, 40; Andersson, 2023, 56.).

Finally, any measure must be proportional and necessary (FL Art. 5, para. 3). Hence, institutions should consider whether a response to malign influence operations is adequate in the situation at hand. This is particularly relevant if the response would be directed against information disseminated in TF and YGL

[12] In armed conflict it is not unlawful under international humanitarian law for State parties to the conflict to spread misinformation as a ruse of war. See Longworth, 2022, 425-436.

protected media, as the media is key to a free exchange of opinions (Andersson, 2023, 51). However, if malign influence campaigns are disseminated through traditional media on a large scale the effect may be extended and enhanced (SOU 2016:80, 397). This may then motivate a proportionate response from public institutions to ensure individuals are aware and can access correct information (Andersson, 2023, 51).

THE PROHIBITION OF CENSORSHIP

Measures aimed at addressing malign influence campaigns must also comply with the prohibition of censorship. Censorship can be direct or indirect and can take place prior to the imparting or after the expression has been shared (HRC, General Comment No. 34, paras. 13 and 20). Examples of indirect censorship include broadly defined legislation introduced and implemented to silence critics of the government, and creating environments or situations that lead to individuals practicing self-censorship (ECtHR, *Dareskizb Ltd v. Armenia*, 2021; ECtHR, *Khadija Ismayilova v. Azerbaijan (No. 2)*, 2020). Prior censorship, or prior restraint, is the most extreme form of censorship, but is not prohibited by under ECHR Art. 10.[13] However, it is scrutinised extremely careful by the ECtHR. The prior banning of publications, websites or similar where the content is unknown at the time of the decision to ban would be an unlawful interference with the right to freedom of expression (ECHR, *Ahmet Yildirim v. Turkey*, 2012, paras. 47–56; ECtHR, *Ürper and Others v. Turkey*, 2009, para. 44).

Under Swedish law there are specific prohibitions of censorship for media protected by the TF and YGL. Accordingly, public authorities may not examine or censor in advance information in books, newspapers, television shows or radio programs (TF Chapter 1 Art. 8; YGL Chapter 1 Art. 11). Authorities are also prohibited under TF Chapter 1 Art. 8; YGL Chapter 1 Art. 11 from preventing the printing, publication or dissemination or otherwise making publicly available information in media protected by the TF and YGL. Information protected by the TF or YGL may thus, as a main rule, only be acted upon after it has been made publicly available, and then only to the extent and in the manner explicitly stipulated in the TF and YGL (TF Chapter 7; YGL Chapter 5).

The prohibition of censorship does not prevent public institutions from providing correct information in reaction to already publicised information that concerns their area of activities and which is deemed factually incorrect (see further ECtHR, *Lingens v. Austria*, 1986). This is a central measure to counter malign influence activities (Andersson, 2023, 44). In the assessment of whether such measures are appropriate, specific consideration should be given to news media's important role in scrutinising and countering malign influence operations. The key role that news media plays in the free exchange of information and opinions and the objective of psychological defence being to secure freedom of opinion and free news media are further relevant factors (Ds. 2017:66, 105 and 108; Andersson, 2023, 34).

However, proactive measures aimed at countering malign influence campaigns must not constitute threats of reprisals or threats to take repressive measures against TF or YGL-protected media because of a suspicion of false or misleading

[13] In contrast, the ACHR, Art. 13(2) provides that the exercise of freedom of expression shall not be subject to prior censorship, and Art. 13(3) provides explicit protection against indirect censorship. Art. 13(4) provides, however, that is public entertainments may be subject by law to prior censorship for the sole purpose of regulating access to them for the moral protection of childhood and adolescence.

information (Andersson, 2023, 33–34). Public institutions must generally refrain from attempts to influence publicist decisions and they should not make statements about the appropriateness of such decisions (JK 23 May 2012, dnr. 3391-14-30 and 3696-14-30; JK 24 March 2006, dnr. 1319-06-21).[14]

CONCLUSIONS

Efforts to counter malign influence operations can in themselves be a means to ensure and protect the exercise of freedom of expression. That said, care and attention needs to be taken so that counter measures do not themselves encroach on the right to freedom of expression. Facilitating the right is also itself an important measure in countering malign influence activities.

Swedish authorities are required to proactively provide and communicate correct and clear information on their activities. For the many public institutions that have a role in the Swedish psychological defence, these information obligations form the legal basis for affected institutions to respond to malign influence operations within their area of activity.

Public institutions must ensure that their information measures have a legal basis, respect freedom of expression, are factually correct and carried out in an objective, proportionate and impartial manner. Upholding these measures involves a careful balancing of different interests. Achieving this balance, however, is not only a counter to malign influence operations, but is itself a means of strengthening protection for freedom of expression and ultimately the exercise of democracy within society.

DISCUSSION

- What types of media are entitled to special protection under TF and YGL?
- What are the differences in how censorship is regulated under RF, TF, YGL, and article 10 ECHR?
- In what circumstances under Swedish law can Swedish authorities issue information measures in response to malign influence operations?

ANNA ANDERSSON, *LLM, is a senior researcher at the Swedish Defence Research Agency and a PhD candidate in international law at the University of Oslo. Anna has a master with a major in law from Örebro University and a LL.M in international humanitarian law and human rights from the Geneva Academy of International Humanitarian Law and Human Rights. Previously, she worked as a Lecturer in international law, public law and social welfare law at Örebro University and the Linneaus University.*

DR SALLY LONGWORTH, *LLM, is a researcher at the Swedish Defence Research Agency and a Postdoctoral researcher at Stockholm University. Previously, she worked as a Senior Lecturer in international law at Stockholm University Faculty of Law and as a Lecturer and Researcher in international law at the Swedish Defence University. Sally defended her PhD thesis on the relationship between international human rights law and international humanitarian law in regards to the right to freedom of expression in armed conflict in October 2022. She has an*

[14] The Chancellor of Justice (*Justititekanslern,* JK) is a non-political civil servant appointed by the government. One of JK's main tasks is to ensure that the limits of the freedom of the press and other media are not transgressed, TF Chapter 9 Art. 1. JK acts as sole prosecutor in cases concerning offences against the freedom of the press and the freedom of expression under TF and YGL, though with some possibility to delegate prosecutorial responsibilities of certain crimes under YGL to public prosecutors, TF Chapter 9 Arts. 2-3 and Chapter 7 Arts. 1-2 YGL.

LL.M in international human rights law from Lund University, an LL.B in European, comparative and international law from the University of Sheffield, and was called to the Bar of England and Wales in 2010.

DR PONTUS WINTHER, *LLM, is Deputy Research Director at the Swedish Defence Research Agency. His research interest lies in international law related to military operations and information operations. He defended his PhD thesis on the protection of civilians against communication influence activities during armed conflict in 2019. He has previously been Director of Studies in law at the Swedish Defence University, a Legal Adviser (reserve) in the Swedish Armed Forces, and an officer in the Swedish Navy. He holds the degrees of Master of Laws and Doctor of International Law from Uppsala University.*

REFERENCES

BOOKS AND REPORTS

Andersson, A. (2023). *Rättsligt ramverk för bemötande av informationspåverkan: En studie av det rättsliga ramverket för bemötande av informationspåverkan genom informationsåtgärder.* Swedish Defence Research Agency, FOI-R--5443--SE.

Axberger, HG. (2019). *Yttrandefrihetsgrundlagarna, Yttrandefrihetens gränser efter 2019 års grundlagsreform.* 4th ed. Norstedts Juridik.

Bring, O, Klamberg, M, Mahmoudi, S, and Wrange, P. (2020). *Sverige och folkrätten.* 6th ed. Norstedts Juridik.

Bull, T. (2020) 'Objektivitetsprincipen', in Marcusson, L (ed.). *Offentligrättsliga principer.* 4th ed. Iustus.

European Court of Human Rights. (2020). *Guide to the Case-Law of the European Court of Human Rights: Data protection.* 1st ed.

Hussain, A. (1994). *Promotion and Protection of the Right to Freedom of Expression: Report of the Special Rapporteur, Mr Abid Hussain, pursuant to Commission on Human Rights resolution 1993/45.* UN Doc. E/CN.4/1995/32, UN Commission of Human Rights.

Kaye, D. (2019). *Promotion and protection of the right to freedom of opinion and expression: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.* UN General Assembly, UN Doc. A/74/486.

Khan, I. (2021). *Disinformation and freedom of opinion and expression: Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan.* UN Human Rights Council, UN Doc. A/HRC/47/25.

Lebeck, C. (2018). *Legalitetsprincipen i förvaltningsrätten.* Norstedts Juridik.

Longworth, S. (2022). *Freedom of expression in armed conflict: The silences between the spaces.* Stockholm University.

Nordin, J, Giles, I (TRANS.), and Graves, P, (TRANS.). (2023). *The Swedish Freedom of the Press Ordinance of 1766: Background and Significance.* Kungliga biblioteket.

Nowak, M. (2005). *U.N. Covenant on Civil and Political Rights: CCPR commentary.* 2nd ed. NP Engel.

Schabas, W. A. (2015). *The European Convention on Human Rights: A Commentary.* Oxford University Press.

Sterzel, F. (2020) 'Legalitetsprincipen', in Marcusson, L. (ed.) *Offentligrättsliga principer.* 4th ed. Iustus, 79-100.

Winther, P. (2019). *International Humanitarian Law and Influence Operations: The Protection of Civilians from Unlawful Communication Influence Activities During Armed Conflict.* Uppsala University.

UN Secretary General. (2022). *Countering disinformation for the promotion and protection of human rights and fundamental freedoms: Report of the Secretary-General.* UN General Assembly, UN Doc. A/77/287.

LEGAL SOURCES

Swedish law

Constitution

Kungörelse (1974:152) om beslutad ny regeringsform (Instrument of Government, RF)

Tryckfrihetsförordning (1949:105) (Freedom of the Press Act, TF)

Yttrandefrihetsgrundlag (1991:1469) (Fundamental Law of Freedom of Expression, YGL)

Statutes

Brottsbalk (1962:700) (Swedish Criminal Code, BrB)

Förvaltningslag (2017:900) (The Administrative Procedure Act, FL)

Lag (1990:52) med särskilda bestämmelser om vård av unga (Care of Young Persons Act, LVU)

Lag (1994:1219) om den europeiska konventionen angående skydd för de mänskliga rättigheterna och de grundläggande friheterna) (European Convention on Human Rights and Fundamental Freedoms Act)

Lag (2018:1197) om Förenta nationernas konvention om barnets rättigheter (United Nations Convention on the Rights of the Child Act)

Socialtjänstlag (2001:453) (Social Services Act, SoL)

Regulations

Förordning (2015:152) med instruktion för Svenska institutet (Ordinance with instruction for the Swedish Institute)

Förordning (2021:936) med instruktion för Myndigheten för psykologiskt försvar (Ordinance with instruction for the Psychological Defence Agency)

Swedish Government decisions

Swedish Government, *Uppdrag att genomföra nationella informationsinsatser om vaccination mot covid-19,* Government Decision 17 December 2020, dnr. S2020/09533

Swedish Government, *Uppdrag att motverka ryktesspridning och desinformation om socialtjänsten,* Government Decision 7 July 2022, dnr. S2022/03244

Justitiekanslerns beslut (Decisions by the Chancellor of Justice, JK)

JK 24 March 2006, dnr. 1319-06-21

JK 23 May 2012, dnr. 3391-14-30 and 3696-14-30

Justitieombudsmannens beslut (Parliamentary Ombudsmen decisions, JO)

JO 2010-02-16, dnr. 4935-2009

JO 2013-11-21, dnr. 5875-2012

JO 2016-06-01, dnr 678-2015

SWEDISH PREPARATORY WORKS

Proposition (Government bill, Prop.).

Prop. 2009/10:175. *Offentlig förvaltning för demokrati, delaktighet och tillväxt*

Prop. 2016/17:180. *En modern och rättssäker förvaltning – ny förvaltningslag*

Statens offentliga utredningar (Swedish Government Official Reports, SOU)

SOU 2010:29, *En ny förvaltningslag*

SOU 2016:80, *En gränsöverskridande mediepolitik. För upplysning, engagemang och ansvar*

SOU 2020:29. *En ny myndighet för att stärka det psykologiska försvaret*

Departementsserien (Ministry Publications Series, Ds.)

Ds. 2017:66. *Motståndskraft Inriktningen av totalförsvaret och utformningen av det civila försvaret 2021–2025*

INTERNATIONAL TREATIES

American Convention on Human Rights 1969 (ACHR), O.A.S. Treaty Series No. 36, 1144 UNTS 123

Charter on Fundamental Rights of the European Union (EU Charter), Official Journal of the European Union 2012/C 326/02, 26 October 2012, 391–407

Convention for the Protection of Human Rights and Fundamental Freedoms 1950 (ECHR), ETS 5, as amended by Protocols Nos. 11, 14 and 15, and as supplemented by Protocols Nos. 1, 4, 6, 7, 12, 13 and 16

Convention on the Elimination of All Forms of Racial Discrimination 1969 (CERD), 660 UNTS 195

Convention on the Prevention and Punishment of the Crime of Genocide 1948, 78 UNTS 277

Convention on the Rights of Persons with Disabilities 2006 (CPD), 2515 UNTS 3

Convention on the Rights of the Child 1989 (CRC), 1577 UNTS 3

International Covenant on Civil and Political Rights 1966 (ICCPR), 999 UNTS 171

INTERNATIONAL CASE LAW

European Court of Human Rights (ECtHR)

Ahmet Yildirim v. Turkey, Application no. 3111/10, Judgment, 18 December 2012

Axel Springer AG v. Germany, Grand Chamber Judgment, Application no. 39954/08, 7 February 2012

Bladet Tromsø and Stensaas v. Norway, Grand Chamber Judgment, Application no. 21980/93, 20 May 1999

Dareskizb Ltd v. Armenia, Application no. 61737/08, Judgment, 21 September 2021

Delfi AS v. Estonia, Application no. 64569/09, Grand Chamber Judgment, 16 June 2015

Donaldson v. The United Kingdom, Application no. 56975/09, Decision on Admissibility, 25 January 2011

Feldek v. Slovakia, Application no. 29032/95, Judgment, 12 July 2001

Handyside v. The United Kingdom, Application no. 5493/72, Judgment, 7 December 1976

Khadija Ismayilova v. Azerbaijan (No. 2), Application no. 30778/15, Judgment, 27 February 2020

Lingens v. Austria, Application no. 9815/82, 8 July 1986

Magyar Helsinki Bizottság v. Hungary, Grand Chamber Judgment, Application no. 18030/11, 8 November 2016

Makraduli v. The Former Yugoslav Republic of Macedonia, Application nos. 64659/11 and 24133/13, Judgment, 19 July 2018

Mariya Alekhina and others v. Russia, Application no. 38004/12, Judgment, 17 July 2018

Nepomnyashchiy and other v. Russia, Applications no. 39954/09 and 3465/17, Judgment, 30 August 2023

NIT S.R.L v. Moldova, Application no. 28470/12, Grand Chamber Judgment, 5 April 2022

Salov v. Ukraine, Application no. 65518/01, Judgment, 6 September 2005

Steel and Morris v. the United Kingdom, Application no. 68416/01, Judgment, 15 February 2005

Stoll v. Switzerland, Grand Chamber Judgment, Application no. 69698/01, 10 September 2007

Sürek v. Turkey (No. 1), Application no. 26682/95, Grand Chamber Judgment, 8 July 1999

The Sunday Times v. the United Kingdom (no. 2), Judgment, Application no. 13166/87, 26 November 1991

Ürper and Others v. Turkey, Application nos. 14526/07, 14747/07, 15022/07, 15737/07, 36137/07, 47245/07, 50371/07, 50372/07 and 54637/07, Judgment, 20 October 2009

Inter-American Court of Human Rights (IACtHR)

Case of Perozo et al. v. Venezuela, Preliminary Objections, Merits, Reparations, and Costs, Judgment of January 28, 2009, Series C No. 195

Compulsory Membership in an Association Prescribed by Law for the Practice of Journalism (Arts. 13 and 29 American Convention on Human Rights), Advisory Opinion OC-5/85 of November 13, 1985, Series A No. 5

UN Human Rights Committee individual communications

Yong-Joo Kang v. Republic of Korea. (2003). Communication No. 878/1999 UN

Doc. CCPR/C/78/D/878/1999

OTHER INTERNATIONAL DOCUMENTS

United Nations General Assembly (UNGA)

UNGA. (2021). *Resolution 76/227: Countering disinformation for the promotion and protection of human rights and fundamental freedoms.* UN Doc. A/RES/76/227

United Nations human rights bodies

UN Committee on the Elimination of Racial Discrimination (CERD Committee). (1993). *General Recommendation 15 on article 4 of the Convention.* Adopted in the forty-second session

UN Human Rights Committee (HRC). (2004). *General Comment No. 31, 'The Nature of the Legal Obligation Imposed on State Parties to the Covenant.* UN Doc. CCPR/C/21/Rev.1/Add. 13

UN Human Rights Committee (HRC). (2011). *General Comment No. 34, 'Article 19: Freedoms of opinion and expression'.* UN Doc. CCPR/C/GC/34

UN Human Rights Council. (2013). *Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.* Appendix to the Report of the United Nations Commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred. UN Doc. A/HRC/22/17/Add.4

UN Human Rights Council. (2022). *Resolution 49/21: Role of States in countering the negative impact of disinformation on the enjoyment and realization of human rights.* UN Doc. A/HRC/RES/49/21

Organization for Security and Co-operation in Europe (OSCE)

OSCE, 'Joint Declaration on "Fake News", Disinformation and Propaganda', adopted by the UN Special Rapporteur on Freedom of Opinion and Expression, the OSCE Representative on Freedom of the Media, the Organization of American States Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression and Access to Information in Accra, on 3 March 2017

4. THE SWEDISH TOTAL DEFENCE SYSTEM

RIKARD BENGTSSON

SUMMARY

- Sweden resumed planning for 'total defence' in 2015, after having dismantled it after the end of the Cold War.
- The overall objective of total defence is to deter actors from armed or other aggression against Sweden and maintain the ability to defend the country against such acts to safeguard freedom of action and protect fundamental values, such as human rights, freedom of expression and democracy.
- Total defence is an integrated approach to national security and defence comprising civil and military dimensions in a mutually constitutive approach.
- As a whole-of-society approach to national security and defence, total defence comprises an array of societal sectors and functions, including the armed forces, national government agencies, regional and local entities, the business sector, civil society organizations and the citizens.
- At the individual level, a key element of the total defence system is the duty of total defence service covering everyone living in Sweden aged between 16 and 70. This duty comprises three different tracks: military conscription, civilian service and general compulsory national service.
- As a member of the EU and NATO, Sweden develops its total defence policy not in isolation from others but rather in a context of multilateral European and transatlantic security institutions.

The deteriorating security situation in Europe over the last decade has led Sweden to return to a 'total defence' model for national security and defence. The idea of total defence – a whole-of-society approach to deterrence and defence organization – was the cornerstone of Swedish defence policy during the Cold War but was successively dismantled in the 1990s and early 2000s.

In formal terms, 'total defence' encompasses all activities that are needed to prepare Sweden for an armed attack or war. The total defence system consists of two areas of activity – the military defence and the civil defence. A key feature of the total defence idea is that the two areas are mutually reinforcing. The military defence is readily observable and organizationally explicit in the form of the Armed Forces (including the Home Guard) and a limited number of adjacent government agencies. In contrast, civil defence is not a single organizational entity but a broad set of activities and operations by public and private actors contributing to realizing the goals for civil defence (see further below). Importantly, it also involves civil society and the individual citizens. While the role of the military defence is to

defend the territorial integrity of Sweden, civil defence concerns the resilience of society to withstand and recover from a broad array of threats and challenges.

In 2015, the government called on the Armed Forces and the Swedish Civil Contingencies Agency (MSB) to resume planning for total defence (Regeringen 2015). Successively, defence spending has increased, the military defence organization has expanded and in the wake of Russia's invasion, Sweden has joined NATO, which gives new preconditions for total defence. Also, as of 2022, a new organization of the civil defence has been put into operation, aiming at a clearer distribution of responsibilities and a better fit with the military side of the total defence. However, it is still too early to evaluate effects of these reforms and key issues remain concerning the governance structure as well as the role of the business sector in the total defence system.

THE GOALS OF THE TOTAL DEFENCE

The overall objective of total defence is to deter actors from armed or other aggression against Sweden and maintain the ability to defend the country against such acts in order to safeguard freedom of action and protect fundamental values, such as human rights, freedom of expression and democracy (for elaboration, see Försvarsberedningen (2023) and the National Security Strategy (Regeringen 2024)). The Defence Commission further specifies the goals of the military and civil defence as follows:

The goals of the military defence are to:

- defend Sweden against armed attack
- uphold Sweden's territorial integrity
- safeguard rights and national interests outside Swedish territory in accordance with international law
- within the framework of NATO's collective defence and other tasks, fulfil Sweden's commitments as a member of NATO
- promote Sweden's security and prevent and manage conflicts and wars by conducting peacetime operations on Swedish territory, in the neighbouring region, and by participating in international peace support operations; and
- protect society and its functionality by using existing capabilities and resources to assist the rest of society.

The goals of the civil defence are to:

- ensure the most important societal functions
- contribute, within the framework of NATO's collective defence and other tasks, to military defence capabilities
- protect the civilian population; and
- maintain the will to defend and society's resilience to external pressures.

THE MAIN ELEMENTS OF THE SWEDISH TOTAL DEFENCE SYSTEM

As a whole-of-society approach to national security and defence, total defence comprises an array of societal sectors and functions. In the following, focus is on the armed forces, the structure of the civil defence system, the importance of the business sector and the role of the citizens.

The military defence sector is organized around the Swedish Armed Forces, consisting of the Army, the Navy, the Air Force and the Home Guard, as well as supporting services. The Armed Forces are geographically divided into five military regions (since 2019, when Region Gotland became a region of its own) to coordinate support and war efforts in the region as well as interaction with civil authorities. Other government agencies in the military defence sector include the Swedish Defence Material Administration (FMV), the National Defence Radio Establishment (FRA) and the Swedish Defence Research Agency (FOI).

The last decade has seen a gradual increase in defence spending as a response to the deteriorating security situation. In the period 2020–2024, military defence expenditures nearly doubled and for the defence bill for 2025–2030, further increases are planned, to clearly surpass the NATO threshold of 2 % of GDP. Naturally, these developments have also led to an expanding defence organization, for instance visible in the growth of the number of Army regiments and in a fourth air force base, and the recruitment of new officers, soldiers and sailors (through the reactivation of the conscription system, see below).

Turning to civil defence, as was noted above, this does not denote a single organization. Instead, civil defence includes activities and actors in several sectors and layers of government as well as in the private sector. This approach necessarily comes with a degree of uncertainty regarding distribution of responsibilities, competing expectations, plurality of conceptualizations etc (Bengtsson and Brommesson 2023). Complexities also arise because of the Swedish administrative model, with organizationally independent government agencies as well as autonomous municipalities and regions. In order to address some of these challenges and find a more apt organizational model from a total defence perspective, a new system for civil defence was set up by October 2022. The new system contains two novel features. One is the introduction of a higher regional system of governance in addition to the existing regional level (which comprises 21 regions). The higher regional level consists of six civil defence regions, led by one of the county administrative boards in each such region. The civil defence regions are of limited importance in peace-time crisis situations but have a coordinating role to play among the counties and not least in relation to the military sector in time of heightened state of alert. The other novelty concerns the government agencies, in the form of an administrative framework comprising twelve preparedness sectors.[15] Each sector has a designated agency as the lead agency for the sector, coordinating the work and being responsible for the development of civil preparedness in the sector. Moreover, the Swedish Civil Contingencies Agency has a key role as the national coordination agency and the focal point of civil defence. Notably, the Psychological Defence Agency lies outside the system of civil preparedness agencies. For an overview of the new system, see MSB 2025.

Civil defence relates to crisis preparedness. It can be noted that as the civil defence was dismantled following the end of the Cold War, emphasis was instead placed on (municipal) crisis preparedness and management. It has been a recurrent question how to align crisis preparedness and civil defence – as planning for the civil defence was taken up a decade ago, the initial idea was that the civil defence was to rest on the existing crisis preparedness system. Over time there

[15] The ten sectors are: economic security; energy supply; electronic communications and postal services; financial services; basic data; health, medical care and welfare; food supply and drinking water; public order and security; civil protection; transport; foreign trade; and industry, building and commerce.

is an increasing realization that civil defence is potentially broader involving also structural dimensions and simultaneously more proactive (capability-enhancing) and less events-driven in orientation than conventional crisis preparedness.

The Swedish Civil Contingencies Agency uses the term 'civil preparedness' to denote the combined area of crisis preparedness and civil defence. While the term is not formally recognized (as a legal concept) in official documents, it underlines the co-constitutive nature of the two areas.

The business sector has a key role in the modern total defence. Societal changes have given new preconditions – privatization and outsourcing has meant that a vast array of production of goods and services of relevance to total defence are in the hands of private companies, sometimes with foreign ownership. From a public sector perspective, there is thus a substantial degree of dependence on private suppliers, which in turn also yields a need for involving the private sector in the planning for total defence. In principle, this is nothing new – also during the Cold War, there were private companies of key importance for the military defence as well as for the civilian side. A system with central companies in contractual relations with the government, obliged to maintain or redirect production to meet national needs (so-called *k-företag*) was at work. A corresponding such system would be harder to institutionalize today given more complex ownership and production structures.

At the individual level, a key element of the total defence system is the duty of total defence service covering everyone living in Sweden aged between 16 and 70. This duty comprises three different tracks: military conscription, civilian service and general compulsory national service. Military conscription was dismantled in 2010 after two decades of diminishing size but was reactivated in 2017 and expanding ever since. Civilian service concerns maintaining functionality in vital sectors of society and if activated by the government, people with training in such fields (irrespective of current place of work) are covered by compulsory civilian service and will get a war posting in that sector. In January 2024, the government activated civilian service for municipal rescue services and the electricity supply sectors. General compulsory national service, finally, potentially encompasses everyone and could include contributing with transports, work in the medical and health care sectors etc to make sure that society functions as normal as possible also in times of war or threat of war. General compulsory national service can only be activated if the government declares a heightened state of alert.

A key element of individual responsibility for contributing to total defence is to be prepared for times when society does not function in a normal way; i.e. when services cannot be upheld, and supply of goods, medicines and even water may be disrupted. In order to prepare citizens for this eventuality and also to raise awareness concerning disinformation, the government distributes the brochure "If crisis or war comes" to all households in Sweden. After the resumed planning for total defence, such a brochure was distributed in 2018, with a new version in late 2024. This rests on an old tradition – the first such brochure (entitled "If war comes") was distributed in 1943 and repeated throughout the Cold War (from the 1970s, the information was also printed the telephone books).

AN INTERNATIONAL OUTLOOK

The idea of total defence is not unique to Sweden. During the Cold War it was primarily associated with non-aligned small countries like Sweden and Finland.

Today, a whole-of-society approach to security and defence is to be found in an increasing number of countries, not least in all of the Nordic and Baltic countries, albeit under somewhat different labels – while most speak of total defence, also notions of 'comprehensive security' and 'comprehensive defence' can be found (see further Wrange, Bengtsson and Brommesson 2024).

The total defence idea also appears in the EU and NATO. The EU has for quite some time had various arrangements for cooperation related to crisis management and civil defence, such as the EU Civil Protection Mechanism, and further cooperation on civil preparedness is a priority of the new European Commission. In NATO, the establishment of the NATO seven baseline requirements in 2016 and work in the Resilience Committee (related to Article 3 in the Washington Treaty) are of relevance here.

DISCUSSION

- What are major challenges for the successful implementation of a total defence policy?
- Civil defence, an integral part of total defence, centers around the notion of resilience. How can this concept be understood in a security context?
- Individual citizens have an important role to play in total defence. How can citizens be motivated to contribute to society's defence efforts?
- Sweden recently joined NATO. What implications may NATO membership have for Sweden's total defence policy?

RIKARD BENGTSSON *is an Associate Professor at the Department of Political Science, Lund University, Sweden. His research interests include global and regional order, security and defence policy, and EU external action. He has published articles in journals such as European Journal of International Security, Foreign Policy Analysis, Journal of European Integration, European Security, Journal of Contingencies and Crisis Management, Cooperation and Conflict and Politics and Governance. During the period 2012-2016 he worked with strategic analysis as a Senior Advisor in the Prime Minister's Office and the Ministry for Foreign Affairs in Stockholm.*

REFERENCES

Bengtsson, R. & Brommesson, D. (2023). Styrningsuppfattningar, förtroende och hotbilder hos det civila försvarets genomförare (Perceptions of Governance, Trust and Threats among Swedish Civil Defence Practitioners). *Statsvetenskaplig tidskrift*, 125(3): 767-794.

Försvarsberedningen (2023). Kraftsamling. *Inriktningen av totalförsvaret och utformningen av det civila försvaret* (the Swedish Defence Commission's report on the future direction of the total defence and the composition of the civil defence), Ds 2023:34. Stockholm: Government Offices of Sweden, https://www.regeringen.se/contentassets/0decd61162c24c73a9ca443328ccd9dd/sammandrag-pa-engelska-av-kraftsamling-ds-202334.pdf

MSB (2025). *Det svenska civila beredskapssystem*et (the Swedish civil preparedness system). Karlstad: The Swedish Civil Contingencies Agency, https://www.msb.se/sv/amnesomraden/krisberedskap--civilt-forsvar/det-svenska-civila-

beredskapssystemet/

Regeringen (2015). ”Uppdrag till Försvarsmakten och Myndigheten för samhällsskydd och beredskap avseende totalförsvarsplanering”, 2015-12-10, https://www.regeringen.se/globalassets/regeringen/dokument/forsvarsdepartementet/regeringsbeslut/regeingsbeslut-5-2015_12_10-uppdrag-till-forsvarsmakten-och-msb-avseende-totalforsvarsplanering.pdf

Regeringen (2024). *National Security Strategy.* Stockholm: Government Offices of Sweden, https://www.government.se/contentassets/dee95d002683482eba019df49db2801f/national-security-strategy_.pdf

Wrange, J., Bengtsson, R., & Brommesson, D. (2024). Resilience through total defence: Towards a shared security culture in the Nordic-Baltic region? *European Journal of International Security,* 2024, 1-22.

5. THE SWEDISH INTELLIGENCE AND SECURITY SERVICES

PER THUNHOLM

SUMMARY

Russia's 2022 invasion marked a turning point for Sweden and Europe, often described as the gravest security crisis since the Second World War. Sweden's foreign and security policy has thus gained renewed importance, underpinned by the intelligence and security services. These services have evolved through Cold War tensions, terrorism, cyber threats, and malign influence operations, becoming increasingly complex. While secrecy remains central, greater transparency has emerged, with insights into threats like Russian sabotage, hybrid warfare, and Chinese industrial espionage shaping current debates.

Russia's full-scale invasion in 2022, while deeply concerning, was not entirely unforeseen. For years, there were unmistakable signs of an increasingly assertive Russia and a growing great power competition that challenged the stability of the rules-based world order. Despite these warnings, February 2022 became a pivotal moment for Sweden and Europe's security landscape. The Swedish government described the resulting security situation as the most serious since World War II, with an expectation that it could persist for the foreseeable future or even deteriorate further (Regeringens skrivelse 2023/24:163, p.6). Within this context, Sweden's foreign and security policy has taken on renewed significance as a primary mechanism to counter such antagonistic threats, relying heavily on an effective and well-coordinated intelligence and security apparatus.

The evolution of Sweden's intelligence and security services has been shaped over decades by changes in the security environment—from post-war tensions and Cold War dynamics to increased globalization, terrorism, malign influence operations, espionage, subversion, and cyber threats. Today, intelligence activities are more complex and multifaceted than ever, reflecting the intricate nature of modern security challenges. Furthermore, the intelligence field carries a unique mystique, often surrounded by myth and speculation, with its reputation shifting according to public perceptions and media portrayals. It is inherent to the nature of intelligence services to operate covertly, as states often wish to keep certain types of information secret. This information, in turn, becomes the target of other states, which employ covert methods to gather it (Lowenthal, 2009, pp. 1ff). Despite its largely opaque nature, recent decades have seen a gradual but notable increase in transparency regarding intelligence work, both within Sweden and internationally (Intelligence and Security Committee of Parliament, 2015, p. 8; Christopher Andrew, 2019, p. xviii). While many remains shrouded in secrecy, glimpses into the intelligence community are now more accessible through various channels such as the media, academia, think tanks, and governmental reports. These insights range widely, from analyses of Russia's unlawful invasion of Ukraine and the acts of sabotage and assassinations attributed to Russian intelligence in Western

Europe, to coverage of malign influence campaigns, hybrid warfare strategies, and industrial espionage, including Chinese acquisitions of critical infrastructure.

This chapter offers an overview of the evolution, mission, and organizational structure of Sweden's intelligence and security services, concluding with key opportunities and challenges shaping this field.

THE EVOLVING NATURE OF INTELLIGENCE WORK

The period from the end of the Cold War to the present day has been marked by a transition away from the remnants of a bipolar world order and the static threat assessments of the Cold War era to a broader, increasingly complex threat landscape. This evolution—viewed from both Eastern and Western perspectives—has resulted in what some have described as a new form of Cold War (Karaganov, 2018; Mclaughlin, 2020; Zhao, 2019; Conradi, 2018). Unlike its predecessor, today's Cold War is shaped by a dynamic interplay of great power competition, characterized by threats that seamlessly traverse politics, economics, and military power. The effects of globalization and the information technology revolution further challenge traditional intelligence practices, as smaller states and non-state actors become capable competitors. The erosion of the intelligence monopoly once held by state agencies illustrates this shift; in some respects, non-state actors surpass state intelligence agencies in information gathering, big data analysis, and speed of processing (Zegart, 2022, pp. 228–252).

These post-Cold War developments have driven significant transformations in strategic intelligence in Sweden (Eriksson, 2013, pp. 79–81) and globally (Gill, Marrin, and Phythian, 2009). In Sweden, this transformation has expanded intelligence needs both geographically and thematically. To address a broader and more complex threat landscape, the intelligence community has had to rapidly adapt in terms of technology, functionality, and overall strategy (SOU 1999:37, 1999; Försvarsdepartementet, 1999; Ds 2005:30, 2005; Försvarsdepartementet, 2007a; Prop. 2024/25:34). The Swedish defense intelligence services have shifted from serving a narrow alarm function with a primarily military focus to encompassing a diversified function that monitors and interprets a complex array of security challenges in an unpredictable global environment (Försvarsdepartementet, 1999, pp. 5–7).

The September 11, 2001, attacks and the subsequent "War on Terror" further reshaped intelligence services, both in Sweden and internationally, fostering new collaborative frameworks and a greater emphasis on information sharing (Lowenthal, 2009, pp. 361–374). With a more complex threat landscape and increasing overlap between internal and external threats, the distinction between civilian and military challenges became less relevant. As the unipolar world order of the early 2000s gave way to today's tripartite power dynamic—centered on competition among the U.S., China, and, to a lesser extent, Russia—Sweden, like other countries, has adapted its intelligence capabilities accordingly. Since 2022, Sweden has, for the first time, appointed a National Security Adviser and associated staff to coordinate and guide the government's national security efforts, enabling a holistic approach to identifying, analyzing, and managing both internal and external threats. In 2024, the Swedish government further strengthened this adaptation by appointing a special investigator to conduct a comprehensive review of the national intelligence system. The investigator is tasked with describing the implications for operations in terms of collection, processing, analysis, and dissemination, based on the current security policy

situation, NATO membership, and technological developments, as well as ensuring that a relevant selection of actionable intelligence is provided to the government, state authorities, and other relevant actors[16].

THE SWEDISH INTELLIGENCE SYSTEM

The Swedish intelligence model is heavily influenced by the British model, focusing on "secret intelligence," where the intelligence service provides complementary value rather than providing a final assessment. This approach contrasts with the American model, where the intelligence community often delivers consolidated and definitive judgments (Ds 2005:30, 2005, p. 91; Zegart, 2022, p. 31; Davies, 2002; Hjelm).

Swedish defense intelligence is conducted to "support Swedish foreign, security, and defense policy, as well as map external threats to the nation" (SFS 2000:130, 2000). However, it is not served as a general decision-making support function; rather, it serves a purpose of certain character and specific importance within the context of intelligence activities both nationally and internationally. As Sherman Kent[17] noted in the U.S., intelligence work goes beyond simply producing general knowledge; it focuses on creating insights that are "vital for national survival" (Kent, 1966, p. vii). In Sweden, this concept is echoed in the notion that intelligence must be "essential to the security of the state." Additionally, the responsibility for the final analysis rests with the recipient, while intelligence acts as a complementary resource that enriches and informs their broader assessments (Försvarsdepartementet, 2007a, p. 54).

A notable characteristic of Sweden's intelligence landscape has been a deeply ingrained culture of excessive secrecy, which historically has been more pronounced compared to other comparable nations. This secrecy is rooted in Sweden's policy of neutrality during World War II and its position of non-alignment during the Cold War. Although the country transitioned to a policy of solidarity in its security outlook, elements of this secrecy culture have persisted and even thrive today. Compounding this has been Sweden's distinct administrative structure, characterized by a relatively small Government Offices and large, autonomous agencies. This structure, further shaped by a juridification of operations, has been reinforced by a prevailing fear of mistakes, as well as changes to secrecy legislation in the 1980s that moved away from Axel Oxenstierna's legacy of inter-agency cooperation. The subsequent era of New Public Management exacerbated these issues, and the influence of what has been termed the "realm of the meritocracy" now manifests through excessive coordination and resource-draining efforts that often detract from core functions[18].

In Sweden, there is separate legislation governing foreign intelligence, referred to in a direct translation as defense intelligence. The term "defense" in this context signifies total defense, rather than solely referring to the Armed Forces, as might be commonly assumed. The authorities authorized by law and further specified by the government to conduct foreign intelligence include the Military Intelligence and Security Service (Must), the National Defense Radio Establishment (FRA), the Defense Materiel Administration (FMV), and the Swedish Defense Research

[16] Dir. 2023:150, Översyn av underrättelseverksamheten
[17] The late Sherman Kent is often described as the father of modern intelligence studies, particularly with a focus on intelligence analysis. Sherman Kent was a professor at Yale and also had a background with the American OSS/CIA.
[18] DN Debatt, Välkommen till mittokratin – NPM:s oheliga efterträdare, 2024-07-06, Johan Alvehus, Gustaf Kastberg Weichselberger

Agency (FOI). However, foreign intelligence merely serve as one contributor to the comprehensive security policy assessments conducted by the Government Offices. Other significant contributors to security policy analysis include diplomatic reporting, along with information and intelligence from the Security Service, the Police, the Swedish Civil Contingencies Agency (MSB), the Psychological Defense Agency (MPF), and a broad range of other relevant actors.

INTELLIGENCE AUTHORITIES

THE SWEDISH MILITARY INTELLIGENCE AND SECURITY SERVICE (MUST)

Must, or the Swedish Military Intelligence and Security Service, operates within the Swedish Armed Forces but holds a distinct role with its director appointed by the government. Must contributes analyses of foreign and security policy developments, as well as military developments on tactical, operational, and strategic levels. It involves collecting, processing, analyzing, and disseminating intelligence, with a focus solely on foreign matters. This includes integrating its own intelligence gathering with input from other domestic and international services. The intelligence products it produces aim to support decision-making at the highest political levels and enhances situational awareness within the Armed Forces. Must also manages Sweden's military attachés and is tasked with preventing, detecting, and countering security threats against the Armed Forces, both domestically and abroad, through security intelligence, protective security, and signal security. Additionally, Must plays a crucial role in international intelligence cooperation, collaborating closely with partners and allies. Domestically, Must works with the Security Service (Säkerhetspolisen), particularly in counterterrorism, counterespionage, and managing hybrid threats. Given today's complex threat landscape, Must serves a growing number of stakeholders within Sweden's total defense and national security sectors, particularly regarding hybrid threats and malign influence campaigns.

THE NATIONAL DEFENCE RADIO ESTABLISHMENT (FRA)

The FRA, or the National Defence Radio Establishment, is Sweden's authority for signals intelligence and IT security, also housing the National Cybersecurity Center (NCSC). Although it has existed as a civilian authority since 1942, all signals intelligence operations require a permit from the Foreign Intelligence Court (Fud). The Law on Signal Intelligence in Foreign Intelligence (Lag 2008:717 om signalspaning I (försvarsunderrättelseverksamhet) outlines the so-called purpose catalog regulating what FRA may conduct surveillance against, including external military threats, participation in peace-promoting missions, terrorism, serious cross-border crime, weapons of mass destruction proliferation, threats to critical infrastructure, and foreign intelligence activities. FRA has extensive partnerships that are vital for effective signals intelligence. Within its IT security mission, FRA supports civilian agencies and organizations in both the public and private sectors.

THE SWEDISH DEFENSE RESEARCH AGENCY (FOI)

The Swedish Defense Research Agency, FOI, conducts research on total defense topics, such as defense analysis, military equipment, CBRN (chemical, biological, radiological, and nuclear) warfare, electronic warfare, and cyber defense. FOI describes itself as a knowledge carrier and expert support in defense and security. The agency separates its research activities from its defense intelligence mission, with its primary consumers being the government, the Government Offices, the

Armed Forces, and the Defense Materiel Administration (FMV). FOI also serves as a critical incubator for defense innovation.

THE SWEDISH DEFENCE MATERIEL ADMINISTRATION (FMV)

The Swedish Defence Materiel Administration, FMV, is another key defense intelligence authority with the primary mission of procuring, developing, and delivering equipment and services for Sweden's defense. To support these processes, FMV conducts intelligence collection, processing, analysis, and dissemination, focusing on analyzing weapons systems, technical solutions, and related areas to inform defense procurement and development.

THE SWEDISH SECURITY SERVICE (SÄKERHETSPOLISEN)

The Swedish Security Service operates as a security service with a police mandate, tasked to prevent and detect threats to Sweden's security, combating terrorism and protect the country's central Government. Its work spans over six main areas: Counter-subversion, dignitary protection, counter-espionage, counter-terrorism, counter-proliferation and protective security. While these areas are interdependent Counter-subversion, counterterrorism, and counterespionage are particularly intelligence-heavy. The overlap between Counter-subversion protection and counterterrorism or counter-espionage is especially significant. Counter-subversion aims to protect national security and Sweden's democracy through counteract and prevent unlawful and malign influence on political decision-making, policy implementation, or public debate." Such malign influence may come from terrorists, extremists, or foreign powers engaging in sabotage, subversion, or other destabilizing acts. Cooperation between the Swedish Security Service, Must, and FRA is essential, as external threats frequently manifest internally, particularly in the context of hybrid threats.

OTHER RELEVANT AUTHORITIES

In addition to the primary intelligence community, numerous agencies contribute to Sweden's security through intelligence-based work. These include the Police, the Psychological Defense Agency (MPF), Customs, the Coast Guard, the Swedish Economic Crime Authority (EBM), the Tax Agency, the Swedish Financial Supervisory Authority, and the Inspectorate of Strategic Products (ISP). This extensive list demonstrates the critical need for effective collaboration, intelligence sharing, and information exchange to identify and mitigate hybrid threats and broader security challenges facing Sweden today. Such cooperation is crucial to manage higher levels of conflict and, ultimately, national defense capabilities in wartime.

CHALLENGES FOR SWEDEN'S INTELLIGENCE AGENCIES

The Swedish intelligence community, supporting foreign, security, and defense policy, must be examined from a governance perspective to assess its adaptability to shifts in the security environment. Two recent examples illustrate the complexities of intelligence governance in Sweden. The 2007/2008 debate around the National Defense Radio Establishment (FRA) and the response to Russia's acts of aggression against Ukraine in 2014 and 2022 represent two contrasting cases. The FRA debate was marked by high public and political tensions, with significant disagreements among political factions. Conversely, the response to Russia's escalation—from a relatively low-intensity conflict following the illegal annexation of Crimea in 2014 to the full-scale invasion in 2022—showcased near-universal political agreement on measures for Sweden's security and support for Ukraine,

with a few exeptions. Exceptions included the Left Party's opposition to the initial arms delivery to Ukraine[19] and diverging views from the Left and Green parties on NATO membership as the optimal path for enhancing Swedish security (Sveriges Riksdag, 2022b, p. 2; Sveriges Riksdag, 2022a, pp. 10–12, 17–19). Today, however, there is broad consensus among all parties on the importance of supporting Ukraine.

The FRA debate also underscored institutional rivalries within the intelligence and security sector. Government departments and their agencies became near-adversaries over issues of mandates, legislation, and ultimately budgetary allocations. It is rare in Sweden for one department (e.g., the Ministry of Justice) to supplement another's (e.g., the Ministry of Defense) submission of a legislative proposal to the Council on Legislation with an independent procedure and divergent viewpoints, as seen in the handling of certain consultations (Försvarsdepartementet, 2007b). In contrast, the political response to Russia's full-scale invasion of Ukraine was characterized by a high degree of unanimity on immediate measures for national security and support for Ukraine. The public debate over whether Russia's invasion represented an intelligence failure and what warnings political leaders received from intelligence services, both domestically and internationally, continues to be a point of contention. The United States and the United Kingdom stood out by declassifying intelligence to publicly warn of the impending Russian assault. However, outside the sphere of secrecy, the available narratives often conflict regarding who knew what and when concerning Russia's actions in 2014 and 2022 (Agenda, 2014; Godmorgon världen, 2022; Söndagsintervjun, 2022; Aftonbladet, 2022; Washington Post, 2022).

The Defense Commission has acknowledged these challenges, emphasizing repeated misjudgments regarding Russia's intentions and risk appetite, as well as a lack of understanding of the intentions and motivations driving an authoritarian and imperialist regime like Russia's. However, it remains unclear where this lack of understanding primarily resides. Criticism appears to be partially directed at the intelligence services, as evidenced by calls for accurate and timely reporting to enable the government's analytical and decision-making capabilities. The Commission also highlights the importance of reporting from diplomatic missions, the Security Service, and other relevant sources. Additionally, there is an implicit critique regarding the interaction between the services, the government, and the Government Offices, emphasizing the need for mutual understanding, well-functioning coordination, and the requisite expertise for processing and utilizing intelligence reporting (Försvarsdepartementet, 2007a, p. 54; Försvarsdepartementet, 2024, pp. 201–202[20]).

THE WAY FORWARD: OPPORTUNITIES AND CHALLENGES

Sweden now faces numerous challenges and potentially fundamental changes that demand careful attention. The nation is situated in a volatile and unpredictable security environment where the rules-based world order is under serious threat, and in some cases, replaced by great power competition in a dog-eat-dog world. Time has become an increasingly scarce resource, as the volume of available information continues to grow exponentially. To navigate these conditions, Sweden's intelligence agencies and consumers must master modern technology and keep pace with rapid innovation. Meeting these demands will require enhanced

[19] Riksdagens snabbprotokoll, 2021/22:75, måndagen den 28 februari. Vänsterpartiet ändrade därefter ståndpunkt och har i de efterkommande besluten bifallit även det militära stödet och vapenleveranserna till Ukraina.
[20] Ds 2024:6

cooperation across agencies, the public sector, the private sector, and academia. As former MI6 Chief Sir Alex Younger remarked, "The twin drivers of technological change and international complexity mean that we must keep adapting if we are to be as effective at spying in the future as we are today. There will be a dividing line between those Intelligence Services that grasp this, as the UK agencies have, and those services that don't." [21]

Cooperation between intelligence services and their stakeholders must continue evolving. As the state's executive authority, the government must clearly define agency mandates through formal steering mechanisms. Within these mandates, maintaining a well-functioning dialogue is essential—particularly when managing internal and external threats holistically. Sharing the same information across multiple simultaneous purposes also requires robust communication channels. When properly utilized, informal contacts and dialogue between Government Offices and agencies provide essential support to both governmental direction and agency operations. As highlighted in the Swedish Parliament's Committee on the Constitution's 2012/13 report, a smooth-functioning administrative apparatus is hard to envision without informal contacts. In today's complex threat landscape, effective information sharing is critical to ensuring that the right information reaches the right recipients. While not everyone needs access to all intelligence, its primary purpose is to be available to those who require it. As Hirdman's 1993 inquiry observed, "Irrelevant intelligence is worthless; relevant intelligence that goes unused is equally worthless."

In conclusion, intelligence work in its narrower sense aims to protect Sweden's freedoms, rights, and independence. At the same time, intelligence is part of the state's monopoly on force, inevitably impacting the personal integrity of adversaries when national security demands it. As established by Article 8 of the European Convention on Human Rights:

1. Everyone has the right to respect for their private and family life, their home, and their correspondence.
2. There shall be no interference by a public authority with the exercise of this right except as is in accordance with the law and as necessary in a democratic society in the interests of national security, public safety, or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others (European Court of Human Rights, 2021, para. 8).

Therefore, while intrusions into personal integrity for national security purposes are internationally accepted, they must always be proportionate and necessary, serving a purpose of "essential importance for the security of the state."

DISCUSSION

- How did Russia's 2022 full-scale invasion of Ukraine alter Sweden's security outlook, and in what ways did it confirm or challenge previous threat assessments?
- To what extent should intelligence agencies balance secrecy with transparency, especially in liberal democracies where public trust and legitimacy are crucial?

[21] Speech, MI6 'C' speech on fourth generation espionage, Foreign & Commonwealth Office, Secret Intelligence Service and Alex Younger, Published 3 December 2018 https://www.gov.uk/government/speeches/mi6-c-speech-on-fourth-generation-espionage

- How does Sweden's intelligence model—based on providing complementary "secret intelligence"—compare with the American model of consolidated assessments, and what are the advantages and disadvantages of each?
- What challenges does Sweden face in coordinating intelligence across multiple agencies, and how might inter-agency rivalries or differing mandates undermine national security?

PER THUNHOLM *is a strategic adviser at the Swedish Defence University (Försvarshögskolan). His work focuses on hybrid warfare, asymmetric conflict, and grey zone security. Thunholm co-edited the anthology Hybrid Warfare: Security and Asymmetric Conflict in International Relations (2021), Security Challenges in the Grey Zone: Hybrid Threats and Hybrid Warfare.*

REFERENCES

Aftonbladet (2022). *No Title, Höjd Beredskap.* Available at: https://www.aftonbladet.se/podcasts/ab/program/1185 (Accessed: 19 August 2022).

Agenda (2014). 'Intervju med Sveriges utrikesminister Margot Wallström'. Sverige: SVT2, 2 november. Available at: https://www.aftonbladet.se/podcasts/ab/program/1185.

Agrell, W. (2012). *Essence of assessment: methods and problems of intelligence analysis.* Edited by S. Moores. Stockholm: National Defence College, Center for Asymmetric Threat Studies CATS.

Bang, M. (2017). *Military intelligence analysis: institutional influence.* dissertation. National Defence University.

Christopher Andrew, R.J.A. (2019). *Secret Intelligence*. 2nd edn. Florence: Routledge. doi:10.4324/9780429029028.

Conradi, P. (2018). *Who lost Russia? : how the world entered a new Cold War.* Paperback. London: Oneworld.

Davies, P. (2002). 'Ideas of intelligence: Divergent national concepts and institutions.(Intelligence)', *Harvard international review.*, 24(3), p. 62.

Davies, P.H.J. (2013). *Intelligence Elsewhere Spies and Espionage Outside the Anglosphere.* Edited by K. Gustafson. Washington, DC: Georgetown University Press.

Ds 2005:30 (2005). *En anpassad försvarsunderrättelseverksamhet.* Stockholm. Available at: http://www.regeringen.se/rattsdokument/proposition/2007/03/prop.-20060763/.

Engelbrekt, K. (2016). *High-Table Diplomacy: The Reshaping of International Security Institutions.* Washington: Georgetown University Press.

Eriksson, G. (2013) .*The intelligence discourse : the Swedish military intelligence (MUST) as a producer of knowledge*. dissertation. Örebro universitet.

Eriksson, J. (2004). *Kampen om hotbilden: rutin och drama i svensk säkerhetspolitik.* Stockholm: Santérus.

European Court of Human Rights (2021). *European Convention on Human Rights.* Available at: https://www.echr.coe.int/Documents/Convention_ENG.pdf (Accessed: 19 September 2022).

Försvarsdepartementet (1999). *Lag om Försvarsunderrättelseverksamhet (Regeringens proposition 1999/2000:25).* Stockholm: Regeringskansliet. Available at: https://www.regeringen.se/rattsliga-dokument/proposition/1999/11/prop.-1999200025/.

Försvarsdepartementet (2007a). *En anpassad försvarsunderrättelseverksamhet (Regeringens proposition 2006/07:63).* Stockholm: Regeringskansliet. Available at: https://www.regeringen.se/rattsliga-dokument/proposition/2007/03/prop.-20060763/ (Accessed: 9 May 2022).

Försvarsdepartementet (2007b). *Fö2005 1912/RS - underlag.pdf.*

Försvarsdepartementet (2015). *Försvarspolitisk inriktning – Sveriges försvar 2016–2020 (Regeringens proposition 2014/15:109).* Available at: https://www.regeringen.se/contentassets/266e64ec3a254a6087ebe9e413806819/proposition-201415109-forsvarspolitisk-inriktning--sveriges-forsvar-2016-2020.

Gill, P., Marrin, S. and Phythian, M. (2009). *Intelligence theory key questions and debates.* Edited by P. Gill, S. Marrin, and M. Phythian. London: Routledge (Studies in intelligence series).

Gioe, D. V. (2018). 'Cyber operations and useful fools: the approach of Russian hybrid intelligence', *Intelligence and national security,* 33(7), pp. 954–973. doi:10.1080/02684527.2018.1479345.

Godmorgon världen (2022) .'Intervju med Utrikesminister Ann Linde (S)'. Sweden: Sveriges radio P1, 1 maj.

Hart, T.G. (1976). 'The Cognitive Dynamics of Swedish Security Elites: Beliefs about Swedish National Security and How They Change', *Cooperation and conflict*, 11(2), pp. 201–219. doi:10.1177/001083677601100204.

Herman, M. (2009) *Intelligence Power in Peace and War.* Cambridge University Press.

Holmström, M. (2022). 'Sverige missade ryska invasionen - DN.SE', *Dagens Nyheter,* 15 June. Available at: https://www.dn.se/sverige/kallor-sa-missbedomde-sverige-det-ryska-krigshotet/ (Accessed: 21 March 2023).

Hughes-Wilson, J. (2004). *Military intelligence blunders and cover-ups.* Carroll & Graf Publishers. Available at: https://www.defence.lk/upload/ebooks/John Hughes-Wilson-Military Intelligence Blunders-Carroll & Graf Publishers (2000).pdf.

Intelligence and Security Committee of Parliament (2015). *Privacy and Security: A modern and transparent legal framework.* Available at: https://isc.independent.gov.uk/wp-content/uploads/2021/01/20150312_ISC_PSRptweb.pdf.

Jervis, R. (2010). *Why Intelligence Fails, Why intelligence fails: lessons from the Iranian Revolution and the Iraq War.* Ithaca: Cornell University Press (Cornell studies in security affairs). Available at: https://web-p-ebscohost-com.proxy.annalindhbiblioteket.se/ehost/detail/detail?vid=0&sid=6924c406-ab28-46fc-979ff36ea46a9e4c%40redis&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3D%3D#AN=673724&db=e000xww.

Johnson, D.H. (1991). 'CRIMINAL SECRECY: THE CASE OF THE ZANDE "SECRET SOCIETIES"', *Past & present,* 130(1), pp. 170–200. doi:10.1093/past/130.1.170.

Jones, B.D. (2005). *The politics of attention : how government prioritizes problems.* Edited by F.R. Baumgartner. Chicago: University of Chicago Press.

Karaganov, S. (2018). 'The new Cold War and the emerging Greater Eurasia', *Journal of Eurasian Studies*, 9(2), pp. 85–93. doi:10.1016/j.euras.2018.07.002.

Kent, S. (1966). *Strategic intelligence for American world policy.* Princeton, New Jersey: Princeton University Press (Princeton Legacy Library). doi:10.1515/9781400879151.

Lowenthal, M.M. (2009). *Intelligence : From Secrets to Policy.* 4th ed. Washington, D.C: CQ Press.

M. Weible, C. (2018). *Theories of the Policy Process.* 4th edn. New York: Routledge.

Mclaughlin, G. (2020). *Russia and the media : the makings of a new Cold War.* Ed. G. Mclaughlin. London: Pluto Press.

Moore, R. (2021). *C's speech to the International Institute for Strategic Studies - GOV.UK, gov.uk.* Available at: https://www.gov.uk/government/speeches/cs-speech-to-the-international-institute-for-strategic-studies (Accessed: 22 September 2022).

Petersson, O. (2018). *Den offentliga makten.* Femte upplagan. Lund: Studentlitteratur.

Prop. 1995/96:12 (1996). *Totalförsvar i förnyelse.* Available at: https://rkrattsbaser.gov.se/prop?ar=1995/96&dok=P&dokid=12 (Accessed: 5 September 2022).

Regeringens skrivelse 2023/24:163 Nationell säkerhetsstrategi

Ruffa, C. (2018). *Military Cultures in Peace and Stability Operations.* Philadelphia: University of Pennsylvania Press.

SFS 2000:130 (2000). *Lag (2000:130) om försvarsunderrättelseverksamhet.* Available at: https://rkrattsbaser.gov.se/sfst?bet=2000:130.

Söndagsintervjun (2022). 'Radiointervju med Lena Hallin, chef för den Militära underrättelse- och säkerhetstjänsten, Must'. Sweden: Sveriges radio P1, 1 maj.

SOU 1923:16 (1923). *Försvarsrevisionens betänkande III, betänkande och förslag, Revision av Sveriges försvarsväsende.* Available at: https://weburn.kb.se/metadata/973/SOU_1590973.htm.

SOU 1976:19 (1976). *Den militära underrättelsetjänsten - Betänkande av 1974 års underrättelseutredning.* Stockholm: Försvarsdepartementet. Available at: https://weburn.kb.se/metadata/247/SOU_7258247.htm.

SOU 1999:37 (1999). *Underrättelsetjänsten - en översyn.* Stockholm.

Sveriges regering (2022a). *Regeringsförklaring 18 oktober 2022.* Available at: https://www.regeringen.se/4a99d4/contentassets/d6ad6308cc984aa1903d9a542ce1421c/regeringsforklaringen-2022.pdf (Accessed: 15 November 2022).

Sveriges regering (2022b). *Utrikesdeklarationen 2022 - Regeringens deklaration vid 2022 års utrikespolitiska debatt onsdagen den 16 februari 2022.* Available at: https://www.regeringen.se/globalassets/regeringen/dokument/utrikesdepartementet/utrikesdeklarationen-2022.pdf (Accessed: 18 March 2023).

Sveriges Riksdag (2022a). *Riksdagens protokoll 2021/22:114 Måndagen den 16 maj.* Available at: https://data.riksdagen.se/fil/266A4AAE-D60D-4906-AB48-EE9116EB3203 (Accessed: 19 September 2022).

Sveriges Riksdag (2022b). 'Riksdagens protokoll 2021/22:75 Måndagen den 28 februari'.

Taylor, A. (2023) .'What we've learned from the leaked Pentagon documents - The Washington Post', *The Washington Post*, 10 April. Available at: https://www.washingtonpost.com/world/2023/04/10/faq-leaked-pentagon-documents/ (Accessed: 18 April 2023).

Washington Post (2022). *As Russia prepared to invade Ukraine, U.S. struggled to convince Zelenskyy, allies of threat - Washington Post.* Available at: https://www.washingtonpost.com/national-security/interactive/2022/ukraine-road-to-war/ (Accessed: 19 August 2022).

Wirtz, J. (2004). 'Paradoxes of Strategic Intelligence: Essays in Honor of Michael I. Handel', in Richard K. Betts, T.M. (ed.). London: Routledge.

Wohlstetter, R. (1962). *Pearl Harbor : warning and decision.* Stanford, Calif: Stanford Univ. Press.

Zegart, A.M.Y.B. (2022). Spies, Lies, and Algorithms. Princeton University Press. doi:10.2307/j.ctv1t8q8tp.

Zhao, M. (2019). 'Is a New Cold War Inevitable? Chinese Perspectives on US–China Strategic Competition', *The Chinese journal of international politics,* 12(3), pp. 371–394. doi:10.1093/cjip/poz010.

6. OPEN-SOURCE INTELLIGENCE (OSINT)

HEDVIG ÖRDÉN & KIRA VRIST RØNN

SUMMARY

- Due to an increasing availability of open-source information, state intelligence services no longer hold the monopoly on intelligence collection, analysis and dissemination.
- Civil society open-source intelligence actors, such as Bellingcat, play an increasing role in informing policymakers and the public on security matters such as foreign information influence.
- This demonopolization of intelligence poses important questions concerning the responsibility of state intelligence services as guardians of society and concerning the expertise of emerging civil society intelligence actors.
- When perceived as independent experts, civil society intelligence actors can have an advantage over state actors in the countering of foreign influence campaigns; especially when such campaigns exploit domestic political vulnerabilities.
- The transparent civil society OSINT methodology also fills an important function in countering information influence as it creates a deeper understanding among the public of how threat actors operate.
- Civil society intelligence actors might nevertheless lack the necessary intelligence tradecraft, and the contextual knowledge required for producing high-quality intelligence analyses.
- Blurred lines of responsibility between civil society and state actors in intelligence may work to undermine the credibility of civil society actors grounded in integrity and independence and erode the responsibility of governments for intelligence and security matters.

In recent decades, rapid developments in information and communication technologies, and a rise of social media, online blogs, and forums have created a 'world of weak gatekeepers' (Farrell and Schwartzberg, 2021, p. 212). Whereas, in the past, journalists and editors working at news outlets controlled much of the information environment, today anyone can communicate with the public. In liberal democratic states, this new information environment has impacted the threat and security landscape in complex ways. On the one hand, possibilities for creating and sharing personalized content creates opportunities for malicious actors – both state and non-state actors – to engage in destructive information influence campaigns targeting democratic populations and institutions. On the other hand, the vast volume of publicly available information has opened the doors for a range

of new civil society actors – 'fact checkers' and Open-Source Intelligence (OSINT) activists – who harness the possibilities of collecting and verifying information, thereby acting as public knowledge brokers and exposing disinformation and influence operations.

This chapter engages with the democratic countermeasures offered by a weak gatekeeping structure. Assessing the role of civil society OSINT actors in contemporary 'whole-of-society' approaches to information influence, it asks: What are the promises and pitfalls of civil society actors in producing and sharing OSINT? In so doing, the chapter draws on recent debates within intelligence studies (IS) on the emergence of an 'open-source intelligence ecosystem' (Zegart, 2021) and a subsequent 'demonopolization' of intelligence from state to non-state actors (Bigo 2019; Rønn & Ördén, forthcoming).

The increasingly important role played by civil society OSINT actors in the countering of disinformation and information influence is part of a broader transformation within the intelligence domain where access to open-source information generates novel possibilities for both civil society (Petersen & Tjalve, 2017) and private actors in intelligence (Moesgaard, 2013; Van Puyvelde, 2019). This entrance of a broader public into intelligence collection, assessments and sharing (Petersen & Rønn, 2019) in turn challenges traditional forms of security and intelligence expertise, undermines established divisions of labour, and introduces questions about responsibility (Rønn & Ördén, forthcoming). For instance, questions emerge over 'who' is recognized as an expert, and 'where expertise resides' (Petersen & Tjalve, 2017: 22), but also about the necessary skills and forms of knowledge pertaining to OSINT expertise (Zegart, 2021; Van Puyvelde & Rienzi, 2025). Both questions are relevant to consider in the context of civil society OSINT actors involved in the countering of foreign information influence.

This chapter begins by introducing the concept of OSINT. Second, it outlines the role of civil society OSINT in countering disinformation and information influence and gives some insight into the current landscape. Third, the chapter examines the grounds for credibility embraced and mobilized by civil society OSINT actors. Fourth, the debates on civil society OSINT skills and knowledge are outlined. In the fifth and concluding part, the chapter considers the possibilities and potential limitations of relying on civil society OSINT in countering disinformation and information influence.

WHAT IS OSINT?

OSINT is intelligence from *non-secret* sources. The concept originates from a 1990 article by Robert Steele, but information from open sources has played a crucial role in intelligence practices for a long time. It is commonplace to associate the origins of open-source collection for intelligence purposes with the 1939 founding of the BBC Monitoring Services (Calkins, 2011), but Block (2024) shows that the history of OSINT goes back as far as the mid-19th century. In recent decades, however, the growing amount of publicly available and easily accessible digital data has increased the relevance and importance of non-secret sources in intelligence (Zegart, 2022).

While several competing definitions exist, the US Director of National Intelligence defines OSINT as: 'intelligence produced from publicly available information that is collected, exploited, and disseminated in a timely manner to an appropriate

audience for the purpose of addressing a specific intelligence requirement' (Williams & Blum, 2018: 1). This includes a diverse range of publicly available information such as social media material, newspaper articles, media content and online forum discussions. As is clear from the above definition, there is an important difference between OSINT and publicly available information (PAI). In short, OSINT is information collected and processed with a *purpose*, to meet a *specific need*: information collected systematically, and carefully analysed with a specific audience in mind (Miller, 2018).

OSINT plays an important role for state intelligence actors countering information influence. While state intelligence services are commonly associated with secret forms of collection and methods which either require specialized technologies, such as signal's intelligence (SIGINT), or humans for intelligence collection (HUMINT), they rely on publicly available information as part of their assessments. Hulnick (2002, p. 566) even calls OSINT 'the lifeblood of intelligence'. Foreign information influence often involves highly *public* campaigns, allowing for a monitoring of the threat landscape to create situational awareness, using a range of open sources (Pamment & Isaksson, 2024: 42-43). Nevertheless, when state actors monitor and analyse threats, OSINT is often employed in combination with other forms of intelligence (Hribar et al., 2014; Janjeva et al., 2022). Drawing on secret sources to establish threat actors' identities and intentions while also using open-source material which provides important contextual information allows state actors to gauge the spread and potential impact of influence campaigns.

In sum, OSINT is not a 'new' form of intelligence (Lomas, 2023; Van Puyvelde & Rienzi, 2025). What is novel in the contemporary context is the scope of publicly available information that can be used for OSINT and how access to this information empowers new actors (Zegart, 2022; 2023). This in turn raises questions about the potential consequences of civil society actors in intelligence.

EMERGING CIVIL SOCIETY OSINT ACTORS

Today, civil society OSINT plays a central part in 'whole-of-society' approaches to counter disinformation and information influence. Threat activities in the information domain might range from coordinated disinformation campaigns, which may be what policymakers describe as 'lawful but awful', to more serious threats where information influence is conducted in concert with other forms of hybrid interference (Pamment & Isaksson, 2024, p. 24-26). Due to this wide range of potential threats, there is often a division of labour between state and civil society intelligence actors. Whereas state actors primarily focus on illegal activities and threats to national security, a substantial amount of the work surrounding other forms of information influence, disinformation and misinformation is 'outsourced' to civil society actors.

The European Digital Media Observatory' (EDMO) highlights how civil society OSINT investigations on disinformation and other forms of influence are 'of extreme value and importance' in a European Union context (EDMO, 2024). EDMO points to the established Institute for Strategic Dialogue (ISD), an organization dedicated to 'safeguarding human rights and reversing the rising tide of polarisation, extremism and disinformation worldwide' (ISD, 2024a). Using OSINT techniques to pinpoint and analyse disinformation and election interference, ISD sometimes partners with governments and provides 'evidence-based data to media, policy and civil society representatives' (ISD, 2024b). The

ecology of civil society OSINT actors active in a European context is however highly diverse, both in terms of interests and in terms of expertise.

The broader civil society open-source information ecology is made up of well-established independent groups. The most recognized example is perhaps Bellingcat whose stated aim is to investigate 'subjects of public interest' (Bellingcat, 2024). Bellingcat provides evidence-based investigations, and sometimes back up government claims – as in the buildup to the Russian full-scale invasion of Ukraine in 2022 (Bellingcat Investigations Team, 2022) – but also hold authorities accountable (Leise, 2024). Some other, less publicly recognised, examples of civil society OSINT actors are *The Centre for Information Resilience* (CIR) and *Darksight Analytics* (DA) driven by an aim to expose human right violations and threats to democracy via OSINT analysis (CIR) and to assist in deriving publicly available data on emerging online security threats (DA).

In addition to such established organizations, the OSINT community has recently seen a rise in individual activists or hobbyists, sometimes embracing the role of open source 'influencers' (Okholm, 2022). These actors are more diverse, and they might be characterised as highly motivated amateurs driven by a will to contribute to intelligence collection in ongoing conflicts via their specific expertise. More critical voices point to personal attention as a key driver for investigations (Okholm, 2022).

As previously stated, the emergence of an eclectic community of civil society OSINT groups and activists generates a set of questions regarding expertise (Rønn & Ördén, forthcoming). Such questions can be broken down into two distinct, but interrelated, debates on, first, *who is recognized* as a credible expert on OSINT and, second, what *kinds of knowledge* and *skills* should be tied to OSINT expertise?

EXPERTISE AS CREDIBLE KNOWLEDGE

If we regard the role of civil society OSINT actors as emerging knowledge brokers in the public realm, a core question is whether there is public *trust* in, and *recognition* of, their role as credible experts.

A key reason why democratic governments work with, and support, civil society actors in countering information influence is the position of such actors as credible voices in the public domain (Ördén, 2019: 431). While government actors may be 'viewed with suspicion, both by domestic and (especially) international audiences', and lack the right standing to counter disinformation, non-state actors can have authority in the eyes of the public due to their perceived political independence (Bjola, 2018, p. 310).

This independent position of civil society actors might be particularly valuable when countering information influence in relation to domestically sensitive cases where public attribution and exposure constitutes a political risk for democratic governments (Hedling & Ördén, 2025). The work of the ISD can be seen as an illustrative case-in-point in this regard. By focusing on 'the complex relationship between foreign state and transnational non-state actors attempting to undermine democracy' (ISD, 2024), the organization is driven to engage in highly polarizing issues which might be difficult for government actors to navigate.

OSINT actors' credibility as independent experts is, however, closely tied to their incentives and motivations. A key aspect for anyone active in countering information influence is to show *'integrity'*; a consistency between 'stated

objectives' and 'actions' (Bjola, 2018, p.310. Orig. italics). For a well-respected organization like Bellingcat, credibility is linked to their stated aim of acting in the interest of the public, as an 'intelligence agency for the people' (Higgins, 2021). The impactful role the organization holds today has been shaped by their past actions of successfully holding authorities to account, and acting as providers of truth, rather than being tied to government interests (Janjeva et al., 2022. p. 11). Through such public manifestations of integrity, Bellingcat introduced a shift in the intelligence landscape where government stakeholders were no longer 'the ultimate arbiters of the public's access to the intelligence cycle' (Janjeva et al., 2014. p. 11).

However, the increasingly audience-driven OSINT analysis we see today can also introduce problems regarding credibility. In an information ecology which favours speed and constant responses to the 'permanent now' (Ford & Hoskins, 2022, p. 48), incentives might shift away from integrity and independence. For instance, OSINT groups and activists acting as 'influencers' can be rewarded for 'sensationalist analyses' (Okholm, 2022) and analysts driven by speed and clicks might focus on 'sexy topics' while failing to produce necessary, but 'mundane', insights (Lomas, 2023).

Problems can also arise in the context of state-civil society cooperation due to the diverging motivations and sources of integrity of the actors involved. For civil society OSINT actors, cooperation with governments can lead to a questioning of their independence and, as a result, of their integrity and credibility (Ördén, 2019). When tasked to co-produce intelligence for state purposes, or when providing information on security matters to state authorities, civil society actors might also become responsible for matters traditionally belonging to governments (Petersen & Tjalve, 2017). This form of 'responsibilization' may in turn contribute to an erosion of state responsibility (Petersen & Tjalve, 2013) and generate dilemmas in relation to democratic control and accountability (Petersen & Tjalve, 2017). In a tense security environment or ongoing conflict, civil society actors taking on intelligence tasks traditionally belonging to governments may even lead to a securitization of such actors, resulting in a change of status from civilians to active agents in an antagonistic setting (Diderichsen, 2019; Saugmann, 2019).

EXPERTISE AS A SPECIALIZED SKILLSET

If we instead view expertise as a set of specialized skills, the knowledge and skill sets of civil society actors vary widely within the community, which is nevertheless unified by a focus on methodological transparency.

As a group spearheading OSINT investigation, Bellingcat were early promoters of a transparent methodology where evidence is presented publicly. OSINT has the advantage over classified information in that it can be openly shared (Hribar et al., 2014). Using this feature to their own advantage, Bellingcat offers comprehensive appendixes and teaching sessions where both methods and detailed investigative designs are shared with the public. This approach to intelligence analysis, and the results produced through it, gains legitimacy by being open to scrutiny. In comparison to insights produced by state intelligence actors who generally need to be restrictive with methods and sources, the transparent civil society OSINT methodology can offer a distinct advantage in countering information influence as it creates a deeper understanding of how threat actors operate.

However, the transparent methodology can also prove highly problematic – especially when actors lack the appropriate skills. Intelligence scholars like Zegart (2022) generally warn against the lack of intelligence tradecraft among certain civil society OSINT actors. She notes how the community is currently fragmented, lacking 'formal qualifications, rules, or standards' (Zegart, 2021). Without the right analytical skills and subject-matter knowledge, publicly available information will not provide useful or actionable intelligence (Hribar et al., 2014). Consequently, even though civil society actors have entered the world of OSINT, it might be a common myth that intelligence has then become a 'common property' or 'competency' of everyone (Petersen, 2024). Like any form of intelligence analysis, useful OSINT analysis requires technical expertise, analytical rigour and context specific knowledge.

What is more, the consequences of 'bad' OSINT can be serious. Amateur OSINT analysts embracing methodological transparency may for instance misattribute an event or actor (Janjeva et al., 2022). In the past, the public nature of such misattributions has given rise to mob behaviors and vigilantism (Zegart, 2021). Even when analysts provide correct attributions, the practice of drawing on data uploaded to personal social media accounts may put innocent individuals at risk (Saugmann, 2019). While a risk of mistakes (and a risk of purposeful deception) is present also among state intelligence actors relying on publicly available information (Hulnick, 2002), a key difference between such actors and civil society OSINT groups is the rigorous procedure in place as well as the persistent secrecy culture permeating state intelligence services.

CONCLUDING REMARKS

Reliance on civil society OSINT actors is considered part of a *whole-of-society approach* for countering disinformation and foreign influence. While this concept has resurfaced as a common catch phrase in government strategies in the current security landscape, the involvement of a range of different actors may also introduce challenges. Civil society OSINT comes with benefits. Intelligence can be publicly shared, methodologies elaborated on, and civil society actors can be more credible public messengers. At the same time, when state intelligence actors collaborate with civil society actors, questions may arise about the quality of analytical skills and motivations. For instance, state actors relying on civil society OSINT might need to check the validity of contributions and consider both the interests behind, and motivations of, such actors. The same is true for civil society OSINT groups cooperating with state actors who rely on political independence for credibility. In general, the muddiness of personal, organizational and state interests can be difficult to separate when operating in a landscape which involves different types of intelligence actors.

What is more, through *whole-of-society approaches* to threats like information influence, new types of actors become 'responsibilized' in new ways for safeguarding the public. This tendency might in turn lead to a 'securitization' of new civil society actors who, willingly or unwillingly, become a part of the state security apparatus and hereby change status from civilians to actors in ongoing conflicts (Ford & Hoskins, 2022; Saugmann, 2019). Consequently, in addition to questions about the right competences to conduct solid OSINT analysis, questions concerning a potential erosion of responsibility among state actors for safeguarding societies become equally important.

DISCUSSION

- Are some intelligence and security tasks inherently governmental?
- What are the potential risks and benefits of an increasingly demonopolized intelligence community?
- How can roles, tradecraft and responsibilities for counteracting current security threats such as malign influence operations best be distributed between state and non-state actors?

HEDVIG ÖRDÉN *is a researcher at the Psychological Defence Research Institute, Lund University, and a postdoc at the Department of Political Science and Public Management, the University of Southern Denmark. She is also an affiliated researcher at the Europe Programme at the Swedish Institute for International Affairs. Her work is situated within critical security studies and critical intelligence studies. She publishes on topics related to security and foreign information influence, intelligence and security expertise, and intelligence and liberal democracy.*

KIRA VRIST RØNN *is an associate professor and head of section at the Department of Political Science and Public Management, University of Southern Denmark. Her research interests cover intelligence, policing, ethics, and national security. She is PI of the research project IntelHub which seeks to voice Scandinavian scholars in intelligence studies, and she recently edited the collected volume "Intelligence Practices in High-Trust Societies" (2025) Routledge's book series, New Studies In Intelligence.*

LINKS TO CIVIL SOCIETY OSINT ACTORS

Bellingcat: https://www.bellingcat.com/

The Centre for Information Resilience (CIR): https://www.info-res.org/

Darksight Analytics (DA): https://www.darksightanalytics.com/

REFERENCES

Bellingcat (2024). Who we are, *Bellingcat,* https://www.bellingcat.com/about/who-we-are/

Bellingcat Investigation Team (2022). 'Documenting and Debunking Dubious Footage from Ukraine's Frontlines', *Bellingcat*, February 23, https://www.bellingcat.com/news/2022/02/23/documenting-and-debunking-dubious-footage-from-ukraines-frontlines/

Bigo, D. (2019). Shared secrecy in a digital age and a transnational world, *Intelligence and National Security,* 34(3), 379-394.

Block, L. (2024). The long history of OSINT. *Journal of Intelligence History,* 23(2), 95-109.

Calkins, L. M. (2011). Patrolling the Ether: US–UK Open Source Intelligence Cooperation and the BBC's Emergence as an Intelligence Agency, 1939–1948. *Intelligence and National Security,* 26(1), 1-22.

Diderichsen, A. (2019). Spreading Intelligence, *Intelligence and National Security,* 34(3), 409-420.

Dylan, H. & Maguire, T. J. (2022). Intelligence and Public Diplomacy in the Ukraine War, *Survival,* 64(4), 33-74.

Farrell, H. & Schwartzberg, M. (2021). The democratic consequences of the new public sphere. In Bernholz, L., Landemore, H., & Reich, R. (Eds.). *Digital technology and democratic theory* (pp.191-218). Chicago: University of Chicago Press.

Ford, M. & Hoskins, A. (2022). *Radical Wars,* Hurst Publishers.

Gibson, S. D. (2014). Exploring the Role and Value of Open Source Intelligence. In edited by Hobbs, C., Moran, M., & Salisbury, D. (Eds.) *Open Source Intelligence in the Twenty-First Century: New Approaches and Opportunities* (pp. 9-23). Palgrave Macmillan UK. https://doi.org/10.1057/9781137353320_2

Gill, P. & Phythian, M. (2018). *Intelligence in an Insecure World.* Cambridge: Polity.

Ganguly, M. (2022). *The Future of Investigative Journalism in the Age of Automation, Open-Source Intelligence (OSINT) and Artificial Intelligence (AI).* University of Westminster.

Hanham, M. (2022). *Setting Your Moral Compass: A Workbook for Applied Ethics in OSINT,* available at: https://stanleycenter.org/publications/osint-applied-ethics-workbook/

Hedling, E. & Ördén, H. (2025). 'Disinformation, Deterrence and the Politics of Attribution', International Affairs, 101(3), 967–986.

Hribar, G., Podbregar, I. & Ivanuša, T. (2014). OSINT: A "Grey Zone"?, *International Journal of Intelligence and CounterIntelligence*, 27(3), 529-549.

Higgins, E. (2021). *We are Bellingcat: An intelligence agency for the people.* London: Bloomsbury Publishing.

Hulnick, A. S. (2002). The Downside of Open Source Intelligence. *International Journal of Intelligence and CounterIntelligence,* 15(4), 565–579.

ISD, (2024a). 'About us', Institute for Strategic Dialogue, https://www.isdglobal.org/about/

ISD, (2024b). 'Disinformation', Institute for Strategic Dialogue, https://www.isdglobal.org/disinformation/

Janjeva, A., Harris, A. & Byrne, J. (2022). The Future of Open Source Intelligence for UK National Security. Royal United Services Institute for Defence and Security Studies. ISSN 2397-0286.

Leise, A. (2024). Fighting the spread of online disinformation: an interview with Bellingcat founder Eliot Higgins, *Vox,* 17 April, https://www.voxweb.nl/en/fighting-the-spread-of-online-disinformation-an-interview-with-bellingcat-founder-eliot-higgins

Lomas, D. (2023). The Death of Secret Intelligence? Think Again. RUSI, July 5. https://rusi.org/explore-our-research/publications/commentary/death-secret-intelligence-think-again.

Block, L. (2023). The Long History of OSINT. *Journal of Intelligence History* 23 (2): 95–109.

Miller, B. (2018). Open Source Intelligence (OSINT): An Oxymoron?, *International Journal of Intelligence and CounterIntelligence,* 31(4), 702-719.

Moesgaard, C. (2013). Private military and security companies- from mercenaries to intelligence providers, DIIS Working Paper 2013:09. https://www.diis.dk/files/media/publications/import/wp2013-09_moesgaard_web.pdf

Okholm, C.S. (2022). OSINT er blevet en kultdiciplin, men dens faldgruber bliver glemt i begejstringen, *Olfi,* https://olfi.dk/2022/11/04/osint-er-blevet-en-kultdisciplin-men-dens-faldgruber-bliver-glemt-i-begejstringen/

Omand, S. D., Bartlett, J., & Miller, C. (2012). Introducing Social Media Intelligence (SOCMINT). *Intelligence and National Security.* 27(6), 801–823.

Pamment, J., & Isaksson, E. (2024). *Psychological Defence: Concepts and principles for the 2020s,* MPF Report Series 6/2024, Psychological Defence Agency.

Petersen, K. L. (2023). Ukraine og enden på den private sektors uskyld. Politica, 55(1), 74–85. https://doi.org/10.7146/politica.v55i1.135829

Petersen, K. L., & Tjalve, V.S. (2013). (Neo) republican security governance? US homeland security and the politics of "shared responsibility". *International Political Sociology,* 7(1), 1-18.

Petersen, K. L. & Tjalve, V.S. (2015). En offentlig hemmelighed: Når sikkerhedspolitik går fra statsmandskunst til allemandseje, *Politik,* 18(3), 13-23.

Petersen, K. L., & Tjalve, V.S. (2017). Intelligence expertise in the age of information sharing: public–private 'collection' and its challenges to democratic control and accountability. *Intelligence and National Security,* 33(1), 21–35. https://doi.org/10.1080/02684527.2017.1316956

Petersen, K. L. & Rønn, K. V. (2019). Introducing the special issue: bringing in the public. Intelligence on the frontier between state and civil society, *Intelligence and National Security,* 34(3), 311-316.

Rønn, K. V. & Søe, S. O. (2019). Is social media intelligence private? Privacy in public and the nature of social media intelligence, *Intelligence and National Security,* 34(3), 362-378.

Rønn, K. V. & Ördén, H. (forthcoming). No More Secrets? De-monopolizing Intelligence. *Canadian Foreign Policy Journal.*

Saugmann, R. (2019). The civilian's visual security paradox: how open source intelligence practices create insecurity for civilians in warzones, *Intelligence and National Security*, 34(3), 344-361.

Steele, R. D. (1990). Intelligence in the 1990's: Recasting national security in a changing world. *American Intelligence Journal,* 11(3), 29-36.

Van Puyvelde, D. & Tabárez Rienzi, F. (2025). The rise of open-source intelligence. European Journal of International Security. Published online 2025:1-15. doi:10.1017/eis.2024.61

Van Puyvelde, D. (2013). Intelligence accountability and the role of public interest groups in the United States. *Intelligence and National Security*, 28(2), 139-158.

Williams, H., & Blum, I. (2018). *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise.* RAND Corporation https://doi.org/10.7249/RR1964

Zegart, A. (2023). Open Secrets Ukraine and the Next Intelligence Revolution, *Foreign Affairs* January/February 2023.

Zegart, A. (2022). *Spies, Lies and Algorithms. The History and Future of American Intelligence.* Princeton University Press.

Zegart, A. (2021). Spies Like Us: The Promise and Peril of Crowdsourced Intelligence, *Foreign Affairs,* July/August 2021

Ördén, H. (2019) .Deferring substance: EU policy and the information threat. *Intelligence and National Security,* 34(3), 421-437.

7. THE HISTORY OF PSYCHOLOGICAL DEFENCE

NIKLAS H. ROSSBACH

SUMMARY

- Modern propaganda, the effort to manipulate mind-sets and opinions, came about during the First World War as an effort by the Western allies to breach the stalemate on the Western front, by breaking the will of the Germans to continue fighting.
- Although the term propaganda dates back to the 17th century and is again in vogue, it encompasses a lot of terms, some of which have tried to replace 'propaganda', such as information wars, psyops, active measures and so forth.
- Propaganda has always made use of the latest technologies, the printing press for books and pamphlets, and later newspapers, and then radio, film and television, and now social media.
- During the Cold War Sweden set up an agency to protect Sweden from foreign propaganda in case war broke out. Against the backdrop of a worsening international situation, Sweden has again set up an agency for psychological defence – against propaganda. But this agency must now be active in peace time.

To paraphrase an old joke: you must be Swedish to come up with a term like 'psychological defence'. The term is only a description of protection against psychological warfare, i.e. hostile propaganda. After the Second World war the term propaganda was associated with Nazi manipulation of public opinion. Yet, after war the West had to be prepared to counter propaganda from another authoritarian regime, the Soviet Union. The West also had to be able to respond in kind. Hence, the West preferred what it considered a more modern terminology. 'Psychological warfare' was meant to describe how to undermine enemy soldiers' morale and how to win the 'hearts and minds' of a local population. Sweden, which identified with the West, but officially stated that it was non-aligned in the Cold War, preferred the term 'psychological defence' to underline Sweden's defensive stance.

Psychological warfare is only one type of propaganda, but all propaganda dates back to ancient times. It was, and is, often vital to successful deception in military campaigns. Even in a democracy open propaganda, for example a campaign poster, where the publisher is acknowledged, is an essential ingredient in modern day elections. Such propaganda – where the publisher publicly stands behind the message – is usually referred to as 'white propaganda', as opposed to covert propaganda. In propaganda where the origins of a message are hidden or even misrepresented it is labelled 'black propaganda'. Sometimes, an adversary may find that it is enough to create uncertainty about the origins of a message. That approach is called 'grey propaganda'. However, propaganda has not remained

relevant simply by being relabelled time and again when used in a new context. For example, it was called 'public relations' in the American business world of the 1920s. Since then, public relations and propaganda have developed on different trajectories. The focus of the former is about selling and promoting. Whereas the focus of the latter is about influencing politics and decision-making. But propaganda has remained a social phenomenon because it has always been adapted to the latest technology, such as newspapers, radio, film and, in the 2010s, social media online.

As a result, it is no surprise that the modern-day propaganda came about during the first modern industrial war, the First World War. It demanded all of an industrialised society's resources. Hence, some later called such a major war, a 'total war'. The leader of the British propaganda effort at the end of the war, the newspaper magnate Lord Northcliff, claimed that the allied propaganda effort had been successful enough to shorten the duration of the war by a whole year. However, after the war the West's propaganda efforts came under scrutiny as having exaggerated German atrocities and mislead the public in the West about the war effort.

Sweden neither participated in the First World War, nor learnt the lessons regarding 20th century propaganda. Eventually, Sweden had to do so fast, in order to avoid being dragged into the Second World War, which began in 1939. During the war the government had to manage foreign propaganda as well as domestic extremists and activists who, for very different reasons, wanted to Sweden to side with one of the warrying parties – though not the same one.

The Swedish government's approach of learning-by-doing when tackling foreign propaganda and managing domestic information, during the Second World War, was ham-handed. It involved: the threat of introducing censorship; official but not public reminders about what the press should publish; a tight control of radio (which was controlled but not actually owned by state); and the request that local dignitaries report on the public mood. The restrictions were coordinated by the State Information Board. The government also tried to inoculate the public against foreign propaganda through awareness campaigns about the dangers of enemy propaganda and the value of being Swedish. Eventually, such public information efforts were also meant to boost morale. Such efforts were the task of the People's preparedness effort (*Folkberedskapen*). It eventually involved a lot of people. It should be noted that the term people (in Swedish 'folk') is innocuous. It has no political connotation, at least none similar to how the term people is used by collectivists regimes, either on the extreme left, such as naming communist dictatorships a 'people's republic', or on the extreme right, as Nazi references to race, using people or in German ('*Volk*').

As the tides of war began to favour the allies, in 1943, the hap-hazard government efforts at countering foreign propaganda and controlling information discredited itself by reckless secret efforts, directed at the public, to 'balance' foreign propaganda from the Western allies and Nazi Germany. It was also tainted by having threatened the press with censorship. This was a sensitive issue in a young democracy, such as Sweden. Overall, the efforts were regarded more as being directed against the public, rather than involving it. In fact, the effort was disbanded even before the war was over in 1945. Also, a few years after the war, new legislation came in place to strengthen the free press.

PSYCHOLOGICAL DEFENCE A PILLAR OF SWEDEN'S TOTAL DEFENCE

After the Second World War, Sweden, like all western European countries pursued economic growth and the new idea of establishing a welfare state. Sweden, somewhat tone deaf, stuck with its interwar years rhetoric of establishing a 'people's home'. However, in contrast to before the war all democratic parties supported re-armament. Officials knew that Sweden could not count on being lucky enough to avoid a third European conflagration. Up until the early 1960s, a third world war between the West, lead by the US, and the Soviet Union, and its communist satellite states, seemed a very real risk. Hence, the Swedish government decided to establish a 'total defence' in response to the risk of a new 'total war'.

Total defence was a comprehensive whole-of-society effort that put a lot of emphasis on preparedness in peace time. The main part of the total defence effort was of course the military forces. Sweden invested heavily in the air force and even planned for a nuclear weapons programme. The three others were civilian: civil preparedness, such as shelters; economic defence, such as supply-security; and psychological defence, to sustain 'the will to defend' the country. While the psychological defence was the smallest of the four pillars that made up the total defence effort it was at the same time inconceivable that Sweden would continue to fight if its people and its forces had been demoralised.

To counter enemy propaganda in case of war the minister of the interior Eije Mossberg, in 1950 requested a blue-ribbon study on how this should be done. Three year later, in 1953, the report on 'Psychological defence' was published. The following year the National Preparedness Commission for Psychological Defence (also: 'the Preparedness Board for Psychological Defence', Beredskapsnämnden för psykologiskt försvar) was established. The report itself became known as 'the Mossberg report' after he had taken over as chairman over the study. It remained the basis for training for decades. Mossberg had also been instrumental in establishing Sweden's secret stay-behind resistance in case of Soviet occupation. However, much of the report was the result of a young expert in pedagogics, Torsten Husén, who already, during the war, had published a book about the threat of psychological warfare.

Sweden would not invest in an offensive capability but opted instead to have only a defensive capability in term of psychological warfare. This fitted the official policy of Sweden being non-aligned in peace time, intending to be neutral in a war. Officials also argued that an offensive psychological warfare capability would be too costly. However, some research indicates that Sweden from early on in the Cold War was a de facto partner to the new western alliance, NATO. Hence, one reason for Sweden not having an offensive propaganda capability may simply have been that Sweden expected NATO to conduct all the necessary psychological warfare in case of a European war.

The new psychological defence sought to avoid repeating the mistakes of the people's preparedness effort during the Second World War. It focused almost entirely on preparing a wartime organisation. During the 1950s, the new psychological defence effort tried to be as transparent as possible, and invited key political figures to its exercises on how media and communications would be protected in wartime. Its leaders gave speeches and interviews about how it was

meant to function in wartime. For example, it stated that it wanted to ensure that newspapers would continue to be printed and that a press office would make sure the government would be able to communicate at home and abroad. The openness was probably essential for its success since much of the organisation was in fact a modified blueprint of the wartime State Information Board and People's preparedness efforts. However, the hierarchy was clearer, and it relied on trust and less on top-down control. In the worst-case scenario, the miniscule peace-time psychological defence commission would mobilise and rise 'as a bird phoenix'.

The information campaign was a success, and many journalists vied for being assigned to the psychological defence, in case they were called up. Of course, training also gave them better access to the defence and policy circles. The total defence was not only focused on the government's ability to communicate with the public. In a war, the expertise of public relations professionals would also be needed to help produce counterpropaganda.

Central to its success was the first head of the psychological defence, the political science professor Gunnar Hecksher. He also headed Sweden's foreign public relations effort, the Swedish Institute. In addition, he was also a leading conservative politician and as such belonged to a party that would be out of power for most of the Cold War. However, his political affiliation served the purpose of the left-centrist coalition government that sought to underline that everyone was behind the total defence effort, across the political spectrum. His role in managing the image of Sweden abroad might also have tempered any ideas about making the psychological defence less open and more like other western psychological warfare efforts, which were often aligned with intelligence operations.

THE FALL OF PSYCHOLOGICAL DEFENCE

After the Cuban missile crisis, in 1961, between the two superpowers, the US and the Soviet Union, it was clear to everyone what was at stake in a case of a third world war – civilization itself. Hence, slowly but surely the superpowers moved to mitigate the risk of an accidental war and began negotiations about arms limitation treaties, and the size of their nuclear weapons stockpiles. This continued up until the end of the Cold War in the early 1990s, but it was not a linear process. Initially, from the 1960s onwards Sweden's psychological defence adapted to the Cold War as a steady state of international affairs, and not as in the 1950s, a prelude to war. In its statutes, it had been task with researching propaganda. From the 1960s onwards the psychological defence would finance several studies on various aspects of propaganda. The intention was a more advanced version of the propaganda vaccination than during the Second World War. The idea was that a better understanding of propaganda would improve society's ability – including the media – to recognise foreign and hidden propaganda.

The media's interest in propaganda increased in 1970s because of the increased critique of society, which at the time was often associated with the political left. Ironically, instead of immediately improving the media's scrutiny of propaganda in international affairs, the psychological defence itself was seen, by some journalists, as a 'propaganda machine'. Of course, the psychological defence did not have the means to manipulate the public, nor did it intend to be active in peace time. However, the political fashion was to be critical of defence efforts and the psychological defence was an easy target. Exercises that had once helped the psychological defence present its activities instead arose suspicions among journalists. Some of them argued that the psychological defence wartime

organisation amounted to a competitor to the free press. Nevertheless, unlike other Swedish political scandals in 1970s related to intelligence and domestic surveillance, the critique of the psychological defence never really became a major cause celebre. Because even if little had changed little since the early Cold War years, its inability to adjust to the political fashion was also proof that the psychological defence had no intention to manipulate the contemporary press.

There was one notable success, and that was the measurement of the willingness of the Swedish public to defend the country in case of war. From the 1960s the surveys of the 'will to defend' (*försvarsvilja*), which in Swedish is one single intuitive word, became more regular. At the time the research was seen as more scientific than would be the case today. Especially, in view of the carful phrasing of the questionnaire. It circumvented any language that could generate existential angst in the respective respondent. As a result, the surveys always reported that the Swedish public to a considerable degree, of about 70 percent, was always willing to defence their country. There may have been some hope that the result would have a deterrent effect on an aggressor. However, the results could also in effect have been a kind of domestic propaganda, as it bolstered the willingness to defend the country rather than reflect the actual willingness of people to endure a war.

Against the backdrop of an increased East-West tensions in the 1980s and a political willingness that considered the critique the psychological defence was put on a new organisational footing in 1985, as the National Board of Psychological Defence (*Styrelsen för psykologiskt försvar,* SPF). In effect, this was a new round of even more openness. However, after only a few years, the Cold War came to an end. Sweden, from the mid-1990s, when it was clear that the Soviet Union really had collapsed and that the new Russia did not pose an immediate threat, began a draw down on defence. By the 2000s the total defence had effectively been abolished. With globalisation and the introduction of the internet there was hope that trade as well as easier and cheaper international connections would cement global peace between countries. Consequently, the agency looked more at natural disasters and similar crises. Nevertheless, the opinion polls on the willingness of the Swedish public to defend Sweden continued. Since, Sweden assumed this to be the new steady state of affairs, the psychological defence increasingly seemed like an anachronism. Shortly after its 50 years jubilee it was closed. As so often the case, at least in Sweden, when the government discreetly seek to abolish an agency, it does so by merging it with two other agencies that were meant to remain. The few remaining psychological defence tasks, such as measuring 'the will to defend', were transferred to a new Civil Contingencies Agency (*Myndigheten för Samhällskydd och Beredskap,* MSB) in 2009.

THE RETURN OF PSYCHOLOGICAL DEFENCE

Just like the Cold War globalisation did not turn out to be a steady state. It soon became evident that the new Civil Contingencies Agency, needed some capability to deal with foreign propaganda. Hence, the agency established a Counter Information Influence Section at the Swedish Civil Contingencies Agency. Unlike the National Board of Psychological Defence it had an active peacetime role, such as helping to protect the integrity of Swedish elections.

Globalisation, as Sweden had understood it, began to unravel in the 2010s, especially after Russia's initiated a war against Ukraine in 2014. That contributed to Sweden relaunching its total defence. It was initially a slow process. But

Sweden was more forward leaning compared to many other states in the West, which were in denial about the risk of return of great power rivalry. Yet, there were signs of Russia, amongst others, adapting its Soviet era propaganda to the new tools enabled by the internet and cyberattacks. For example, as a result Russian interference abroad increased, for example in election campaign in western countries.

The need for Sweden to once again have a psychological defence was part and parcel of restarting the total defence in 2015. However, once again this required a commission and a report. The study was launched in 2019, and completed a year later. One major advantage was that the personnel that had maintained and developed modern psychological defence capabilities at the Civil Contingencies Agency, could be transferred to a new agency. The new Psychological Defence Agency was set up in 2022, just as Russia launched its full-scale invasion of Ukraine. This underlined that the activities of the 21st century psychological defence would be very different from those of its predecessors. It had to focus on being active in what was formally peace time and only thereafter focus on preparing its war time organisation and engaging in funding relevant research. It had the benefit of a secretary of the official report, who had also studied the Cold War psychological defence effort, taking over as the head of the new agency, in 2022, shortly after it had been established.

The modern psychological defence is a response to propaganda, which has again gained currency as the term that highlights the dangers of foreign influence operations and similar activities directed at undermining Sweden's free and open society. Unlike, the Cold War, when Sweden was non-aligned Sweden has been an EU-member since 1995 and a member of NATO since 2024. In contrast to the Cold War era Sweden has been involved with several multilateral efforts, at countering propaganda, such as the HybridCoE centre in Helsinki, jointly funded by the EU and NATO as well as establishing relations with the NATO's Strategic Communications Centre of Excellence for, in Riga, Latvia. While, Sweden still has a total defence concept, Sweden today is in many ways much more like other European nations, and unlike the Cold War, it is clear to the public that a major conflict in Europe, will engulf all European countries.

DISCUSSION

- Is, and if so how, is propaganda today different from how it was conducted during the 20th century?
- In an open and free society, what is the best way for a government and the authorities to engage with the public regarding the need to defend the country?
- Should the media be involved in the psychological defence – or similar efforts – to protect a country from foreign propaganda – and if so under what conditions and in what way?
- The Cold War was essentially about two narratives, the West democratic and free society vs. the East's autocratic one-party system. Are today's narratives in international politics similar or different?

NIKLAS H. ROSSBACH, *is a senior analyst at the Swedish Defence University and a senior researcher at the Swedish Institute of International Affairs. His focus is on the impact of geopolitics on key national security issues, such as energy security and narratives, as well as the consequences of new technologies for strategic intelligence. He has published on Trans-Atlantic relations, Russia's energy politics, critical minerals and other related issues. He holds a PhD in History and Civilization from the EUI and is a certified future strategist. He was an expert in the commission that set up the Swedish psychological defence agency.*

REFERENCES

Garth, S. J. and O'Donnell, V. (2018). Propaganda & Persuasion. London: Sage.

Hamilton, J. M. (2024). Manipulating the Masses: Woodrow Wilson and the Birth of American Propaganda. Baton Rogue: LSU press.

Rossbach, N. H. (2017). Fighting Propaganda – The Swedish Experience of psychological warfare and Sweden's psychological defence. Stockholm: Axess Publishing AB.

Tye, L. (2002). The Father of Spin: Edward L. Bernays and the Birth of Public Relations. London: Picador.

8. THE PRESENT AND FUTURE OF PSYCHOLOGICAL DEFENCE

JAMES PAMMENT

SUMMARY

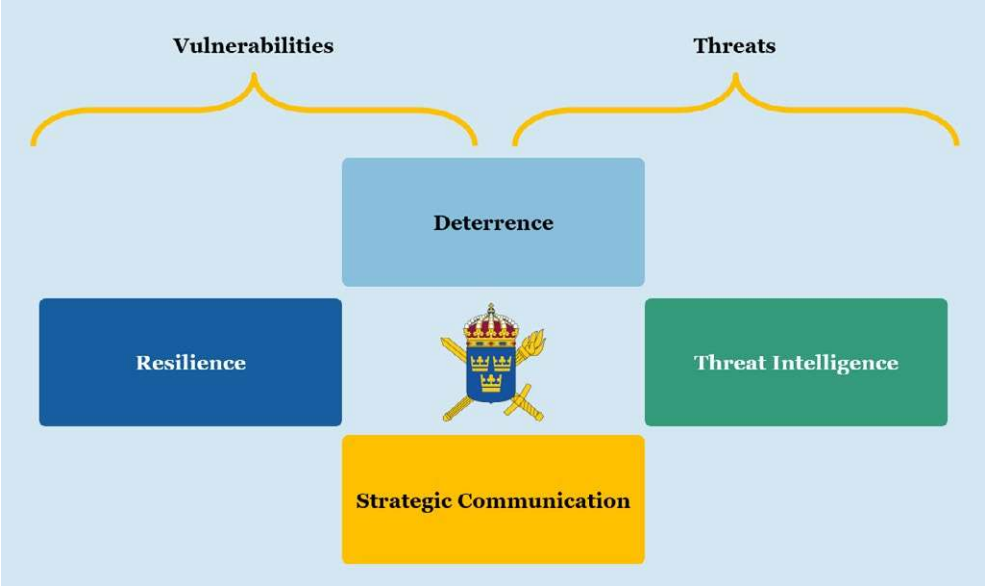
Since the Psychological Defence Board had its mandate significantly reduced in 2002 before finally being disbanded in 2008, much has changed in geopolitics, technology, and information consumption. This is reflected in a policy area which now uses terms such as disinformation, influence operations, information manipulation, FIMI, hybrid threats, and foreign interference. What, then, is the value of a concept such as psychological defence in the 2020s? What does it offer to an already overloaded terminological apparatus beyond yet another poorly defined concept to use? And how does Sweden's new psychological defence agenda fit within the contemporary international counter-disinformation and counter-hybrid fields?

During the Cold War, psychological defence had a reasonably stable meaning intimately bound to Sweden's modest place in the post-War, and later bipolar, geopolitical order. When that geopolitical order broke down in the 1990s, the value of total defence, and with it psychological defence, diminished in favour of alternative capabilities such as crisis management. What does the present geopolitical context, and the overlapping concepts that have thrived in recent years, mean for the new psychological defence? In what ways do the new threat actors and manipulation techniques, new societal vulnerabilities and grievances, new security policy instruments and alliances, and new communication ecosystems prompt a reformulation of psychological defence's core principles? This chapter outlines an agenda for the new psychological defence.

The contemporary understanding of psychological defence should acknowledge two poles: *vulnerabilities* and *threats*. Vulnerabilities are the weaknesses that exist in our society, our institutions, and in ourselves. Threats are the negative events, acts, or actors that exploit those weaknesses in ways that can cause us significant problems. Current interpretations of Swedish legislation have operationalised these terms to mean *domestic vulnerabilities* and *external threats*. In other words, problems that are domestic in origin are vulnerabilities for national security since they belong to "us". Problems from abroad are categorised as threats. Psychological defence is mandated to try to reduce vulnerabilities and has some powers to tackle threats.

The present adoption of psychological defence can be conceptualised as comprising of four principles. The first two are *Resilience* and *Threat intelligence*, which correspond to the work on reducing vulnerabilities and analysing threats. In the middle, where the understanding of vulnerabilities and threats meet, are two further principles, *Deterrence* and *Strategic Communication*. Deterrence refers broadly to security policy work that seeks to develop policy positions for

shaping the behaviour of adversaries in line with our understanding of the damage that their threat activities can have on society. Strategic communication is the understanding of the communication ecosystem as it relates to both threats and vulnerabilities and involves the development of strategies and tactics to implement security policy measures.



RESILIENCE

Hybrid threats, that is to say, so-called grey-zone subversion activities under the threshold of war, target vulnerabilities in society, such as insufficient defences, under-resourced or under-developed capabilities, societal fissures and grievances, as well as gaps between institutional responsibilities (NATO 2024; Giannopoulos et al., 2021). The emergent term for describing work on vulnerabilities within psychological defence is resilience. Resilience focuses on one's own proactive and defensive capabilities for minimising risk to a society. At its core, resilience has traditionally been about the capacity to endure and manage change, regardless of the circumstances. This is often referred to as the ability to 'bounce back' from sudden shocks, to adapt to changing realities, and to quickly restore some form of normality in the face of significant disruption. It is closely associated with protection of the critical infrastructure that keeps a society functioning.

More recently, resilience has come to be repositioned as part of the holistic understanding of a society's resolve in the face of hybrid threats as well as traditional military threats and unexpected crises. It refers to the *routines, processes, and practices that empower the whole-of-society to participate in collective security* (Pamment & Palmertz 2023). It is centred on shared responsibility for security between a country's population, its public institutions, civil society, and private sector, and hence is a core facet of the long-term Swedish doctrine of *total defence*.

Resilience from a psychological defence perspective is agnostic about the source or nature of the threat; its focus is on the self. A society with better preparedness, fewer vulnerabilities, and effective crisis management is a less attractive target to hostile actors. Resilience is therefore not just a defensive concept since its

resources simultaneously act as deterrents and raise the overall costs of efforts to disrupt a society. Recurrent themes therefore include a strengthening of:

- *Will to defend* and *spirit of resistance*. The concepts of "will to defend" (*försvarsvilja*) and spirit of resistance (*försvarsanda*) encompass more than just the individual readiness to protect oneself. Rather, the concepts are often viewed as being deeply entwined with the perception of the society one lives in, usually boiling down to fundamental questions such as: Is my way of life worth defending? It is often linked to citizens' trust in the state, authorities, and democratic institutions, which can differ based on various factors such as personal experiences, economic conditions, cultural background, and political beliefs. The hint of a spiritual dimension points to these factors as building upon nationalism and faith as well as logic.
- *Civil defence.* Civil defence "encompasses the whole of society and comprises the collective resilience in the event of war or danger of war" (Government Office of Sweden 2024). It encompasses a wide range of activities, including emergency preparedness, disaster response, crisis management, and recovery efforts, aimed at safeguarding lives, property, and essential services.

Efforts to protect society's critical functions are determined through a prioritisation based on risk and vulnerability assessments. It is common for a country to monitor high priority societal vulnerabilities and to develop thresholds to inform about evolving threats. For example, adversaries might test sensitive computer networks at regular intervals, looking for avenues to infiltrate classified systems. Most countries quietly monitor these efforts and establish thresholds that would be triggered in case of a sudden intensification of hostile activity. The ability to create country-wide, and even internationally recognised, capabilities in these areas contributes to coherence, interoperability between different actors, and a common view of problems and solutions. Key capabilities include:

- *Risk assessment, vulnerability assessment* and *crisis contingency planning* for critical infrastructure and other crucial public services.
- *Monitoring* and *early warning capabilities* based around civil contingencies and crisis response.
- *Recognised security certifications* for organisations dealing with sensitive systems and other essential processes.
- *Training and exercises* in crisis management coordination and response.
- *Whole-of-society participation* in shared capability development.

A fundamental focus of psychological defence since its origins has been on protection of free public debate. Since the Mossberg Report, it has been assumed that "results can be achieved by teaching people to recognise propaganda, to be critical of rumours, and to distinguish between false and genuine messages." (SOU 1953:27, p. 63). Resilience is, in other words, centred on developing people and institutions so that they can embody and enact a society's resolve. This remains a key principle of contemporary psychological defence. Contemporary approaches to improving resilience from a psychological defence perspective include:

- *Public awareness-raising campaigns,* such as public information campaigns about foreign propaganda and propaganda methods.
- *Efforts to improve media literacy*, for example by providing education or training in how to critically interpret media and especially content shared on social media.
- *Efforts to improve source criticism,* by encouraging people to critically evaluate

information sources.

- *Support of credible journalism,* to foster a critical and independent media system based on established journalistic ethical norms.
- *Support of fact-checking initiatives,* by providing independent, nonpartisan reviews of mediated content for factual errors.
- *Support of debunking initiatives,* involving the targeted review of mediated content on specific topics or from certain sources to expose particularly politically motivated falsehoods.
- *Intelligence disclosures for the purpose of inoculation,* for example the "prebunking" conducted by the US and UK prior to the 2022 Russian invasion of Ukraine, which prepared the public for anticipated disinformation about the premise of the war.

Hence, the Swedish strategy for maintaining a robust psychological defence revolves around promoting a free media, bolstering citizens' resilience, and strengthening trust in state authorities (The Swedish Agency for Public Management 2017:5; MSB 2018). Much of this work is about strengthening democratic participation through media and information literacy. Raising awareness of the risk of cyberattacks and information influence campaigns is part of that work. For example, the campaign Tänk Säkert[22], endorsed by the MSB, the Police, and Stöldskyddsföreningen (SSF), strives to educate individuals on information and cyber security. The ongoing "Doṅt Be Fooled!" ("Bli Inte Lurad!")[23] initiative seeks to promote awareness and empower individuals to tackle deceptive and inaccurate information.

THREAT INTELLIGENCE

In recent years, the cybersecurity sector has established a form of intelligence work based around analysis of digital signals, behavioural markers, and contextual factors for the purpose of tracking threats. Often referred to as *threat intelligence*, this approach has provided much of the inspiration and language that has informed the burgeoning field of influence operations analysis. This includes some key concepts, institutional structures, data collection and analysis methods, as well as community standards. While cybersecurity is not the only field to have inspired contemporary approaches to influence operations analysis, there are many implicit adoptions, including:

- *A focus on identifying and tracking threat actors.* In cybersecurity, these are referred to as Advanced Persistent Threats, or APT. In analysis of influence operations, efforts are made to attribute campaigns to threat actors based on their known capabilities and interests.
- *Use of activity classifiers.* Cybersecurity analysis draws upon a series of standardised classifiers designed to facilitate data sharing within the defender community. Known as Tactics, Techniques and Procedures, or TTP, the classifiers enable coding of manipulation techniques in a manner that can for example reveal the connections between activities that comprise a cyberattack. In influence operations analysis, classifiers such as DISARM are currently in the testing phase and are high on the international agenda (See e.g., Newman 2022; The European Union Agency for Cybersecurity 2022).

[22] https://sakerhetskollen.se/
[23] https://www.bliintelurad.se/

- *Situational awareness informs strategic interventions.* Both fields rely on a strategic approach to monitoring threats, which may involve allowing threat actors to establish an infrastructure to better understand their goals and methods, compromise the threat actors, and/or remove all hostile assets at the same time. Usually, defensive and offensive counteroperations are clearly distinguished, and may even be conducted by entirely different teams.

Threat intelligence from the perspective of psychological defence may therefore be described as monitoring foreign propaganda and developing effective methods for analysing, investigating, and sharing insights about trends. On the one hand, it is heavily focused on understanding threat vectors, such as the technical opportunities, behaviours, and contexts that are used to undermine the information environment. On the other, it is focused on understanding specific threat actors, their intentions, capabilities, opportunities, and resources, and ensuring that these profiles are kept up to date. Overall, this implies the sharing of situational awareness between relevant societal stakeholders, including the Government Offices, public agencies, local government, the private sector, civil society, media and journalism, and the public. Exactly what is shared is different depending on the audience.

Threat intelligence can inform a variety of countermeasures (See e.g., Pamment 2022a) including:

- *Counterintelligence:* A specialism in identifying domestic proxies who conduct information influence on behalf of hostile foreign states. For example, the FBI[24] includes disinformation alongside other aspects of foreign interference as part of its counterintelligence work.
- *Intelligence disclosures:* Making conclusions or assessments from secret intelligence public in order to inform about and/or attribute threat activities.
- *Network disruption:* Use of cyber capabilities to disrupt an adversary's network. For example, during the 2018 midterm elections, the US allegedly disrupted the internet access of the notorious St. Petersburg troll farm behind the 2016 election interference, the Internet Research Agency (Nakashima 2019).
- *Offensive operations:* Run covert, coordinated influence operations abroad against a hostile state or its agents. MPF has the mandate to conduct offensive influence operations in the event of war.

DETERRENCE

Deterrence refers to coordinated activities that aim to shape adversaries' perceptions of cost and benefits to dissuade threatening behaviour (Keršanskas 2020). According to Schelling (1966, p. 2), the development of the nuclear deterrent during the Cold War contributed to a geopolitical environment in which "the art of coercion, of intimidation, and deterrence" became a core facet of military thinking. In the early-21st century, deterrence theory – traditionally seen as state-centric with Mutually Assured Destruction at its core – was applied to nonstate actors such as terrorist groups (Davis & Jenkins 2002), paving the way for its more recent application in areas such as cybersecurity and hybrid (see e.g. Pamment & Agardh-Twetman 2019).

In its modern application, deterrence is usually divided into two areas: denial and

[24] https://www.fbi.gov/investigate/counterintelligence/foreign-influence

costs. *Deterrence by denial* involves the reduction or removal of an adversary's capabilities and/or their intended effects. The three main denial areas are:

- *Denial of benefit,* reducing or removing the rewards anticipated from adversarial behaviour.
- *Denial of capabilities,* restricting or nullifying the threatening capabilities that the adversary can bring to bear.
- *Denial by punishment,* levelling punitive measures upon the adversary.

These *denial* options contribute to an overall approach to *deterrence by imposition of costs*. Deterrence by imposition of costs is a mindset and form of strategising based upon the assumption that the collective impact of denial measures on the adversary's cost/benefit analysis will lead them to conclude that their aggressive actions are not worth it. In essence, it asks the question, *can we make this type of attack more costly to carry out?* Those costs might for example be in terms of a need for increased resources to carry out the harm (e.g., more people, advanced tools, and work hours are required), unanticipated costs to reputations (e.g. an increased risk of attribution and exposure (Pamment & Smith 2022), or highly damaging countermeasures (e.g. likelihood of offensive responses). While it is not always the case that the adversary acts rationally, at its core, deterrence by imposition of costs tries to make the harmful activity more costly (both metaphorically and actually) than the rewards that the adversary anticipates from its disruptive behaviour.

It is therefore essential to have some understanding of the adversary's decision-making processes, the resources they consider proportionate, and their own red lines or limitations. In other words, security policy experts need some understanding of the underlying historical and cultural context and frame by which the adversary views the world and interprets the strategies and actions of themselves and others. Some adversaries are more sensitive to certain types of costs than others; for example, it is often assumed that oligarchs are sensitive to economic sanctions and travel bans because of the expectation that wealth enables a luxurious international lifestyle. Others are acutely sensitive to attribution since they like to maintain a strongly positive reputation in public perceptions. Terrorists might be entirely unmoved by punitive or financial measures but respond to theological reasoning. Understanding what makes the adversary tick requires intelligence about their psychological make-up, motivations, information sources, and decision-making processes (Pamment & Palmertz 2023; Pamment & Agardh-Twetman 2019).

Deterrence in the context of influence operations draws upon a toolbox of countermeasures designed to shape the behaviour of a threat actor. It is intimately connected to resilience, in the sense that known vulnerabilities provide a level of insight into what behaviours can be accepted and which must be averted; to threat intelligence, in the sense that all knowledge of adversary intentions, resources, opportunities, and behaviours feed into an assessment of risk and priority; and to strategic communication, in the sense that many deterrence actions are directly or indirectly communicated to an adversary, whether through e.g. signalling, attribution, intelligence disclosures, or awareness raising. Deterrence is, in other words, the security policy function that draws together and provides strategic direction to the psychological defence apparatus. Countermeasures can include:

- *Signalling:* communicating to a hostile actor awareness of their behaviour or sending a message that their behaviour will not be tolerated, through indirect means. For example, an intelligence agency stating that a foreign actor is seeking to undermine the national interest, without naming that actor, signals awareness without escalation.
- *Deterring:* coordinated efforts to influence a hostile state's calculus. For example, during the Sweden's NATO accession period, the UK signed a mutual protection pact and sent the HMS Queen Elizabeth to the Baltic to demonstrate resolve.
- *Attribution:* technical and political capabilities to assign blame to states and their proxies. For example, following the Salisbury Poisoning, Prime Minister Theresa May stated in Parliament that it was "highly likely" that Russia was behind the poisoning.
- *Legislation:* specific laws that empower government agencies to act proactively. For instance, Australia has the National Security Legislation Amendment (Espionage and Foreign Interference) Act 2018 and Singapore has the Foreign Interference (Countermeasures) Act.
- *Sanctions:* levy costs upon hostile state and its agents. Example: in March 2022, the EU imposed sanctions on Russian state media, including RT and Sputnik, in response to disinformation spread about Ukraine prior to and during the invasion (Council of European Union 2022).

STRATEGIC COMMUNICATION

To operationalise resilience, deterrence, and threat intelligence into cohesive countermeasures, strategic communication plays a key role. This role is so fundamental that communication should be considered a principle of psychological defence in its own right, in addition to being a tool of implementation. Building societal resilience relies upon a continuous and inclusive dialogue between Government, civil society, the private sector, and individuals. Threat intelligence involves continuous scanning of the information environment, categorisation of threats, and information exchange between countries, government agencies, civil society, and the private sector across intelligence stovepipes. Deterrence implies a continuous dialogue between antagonists and one's own government, as well as within and between allied governments for coordination purposes. This is more than simply communication, and rather encompasses a commitment to being communicative.

From this perspective, many countermeasures are essentially strategic communication interventions designed to shape behavioural change of some kind. For example, countermeasures associated with mis-, dis- and mal-information often take the form of information campaigns. This may include factchecking low-level disinformation and developing public education initiatives, with the goal of changing how groups and individuals consume information, or building trust and developing or repairing reputations. Countermeasures associated with influence operations emphasise social listening, audience insights, and developing powerful counternarratives for the purpose of shedding light on clandestine influence campaigns. Countermeasures associated with foreign interference emphasise information sharing to support a purposeful dialogue with threat actors, with the aim of changing the calculus of adversaries. Planned and coordinated communication is the key to an effective response.

For example, guidance from MSB (2018) and the UK Government Communication Service (Pamment 2021) outline a series of proactive and reactive communication tools such as:

- *Inoculation:* communication interventions designed to proactively "pre-bunk" false messaging before it has become widely spread.
- *Awareness raising:* efforts to proactively shape public debate about issues likely to be subjected to mis- and disinformation.
- *Information campaigns:* a planned sequence of communications and interactions that uses compelling narratives over time to deliver a defined and measurable outcome, such as behaviour change.
- *Network building:* shaping networks of likeminded allies and organisations to provide a safe space for solving problems together.
- *Counter-branding:* a range of communicative activities that collectively seek to ensure a reputational cost to actors who persistently spread false, misleading and harmful information.
- *Resilience building:* the aim of resilience building and media literacy initiatives is to empower people to better understand how false information can be spread on and offline, so that they can more effectively engage with what they see, read, and hear.
- *Debunking:* when false or manipulated information is circulating and you wish to counteract the impact of the false information by asserting the truth.
- *Counter-narratives:* countering narratives involves exposing falsehoods and contradictions in how important issues are explained to different audiences and where possible replacing them with a more truthful narrative.
- *Crisis communication:* managing reputations and ensuring that accurate information reaches target audiences as it becomes available.

Countermeasures can also include a broad array of activities that involve policy innovations, physical interventions, symbolic actions, deterrence acts such as signalling, good governance programmes, and better coordination. The strategic communication component emphasises that all such activities should be planned from a communicative perspective. This is challenging in most countries due to the difficulties of cross-governmental coordination, as well as the multiple layers of cooperation required with civil society and the private sector in civil defence. In Sweden, cross-governmental coordination has an additional level of complication due to the unique public agency system, which does not allow for ministerial rule and hence at times makes coordination more challenging. MPF currently runs the governmental cooperation structure which is based upon voluntary participation In addition, the Government Offices – as well as individual ministers – often communicate on psychological defence issues in coordination with MPF. For example, in January 2023, Sweden's Prime Minister, and the head of the Psychological Defence Agency (MPF) Operations Department disclosed an ongoing foreign influence campaign aimed at manipulating public opinion and decision-making processes regarding Sweden's potential NATO membership (Prime Minister's Office 2023).

Other areas of activity relevant to strategic communication include developing and maintaining an up-to-date understanding of how traditional media, social media platforms, evolving technologies such as Artificial Intelligence (AI) and Machine Learning (ML) (Fredheim & Pamment 2024), and non-traditional platforms such as video games (See e.g., Pamment, Falkheimer & Isaksson 2023) function and behave. This may involve regular contact with industry. The aim is to understand the information environment infrastructures used both by individuals and by threat actors, and potentially how they can also be used to disseminate trustworthy

information. It also involves understanding media systems, their governing policies, and the consumption habits of their users. This is fundamental to understanding audiences.

DISCUSSION

- Who should be responsible for understanding and mitigating domestic vulnerabilities? What are the main hindrances to doing this effectively?
- In an era often characterised by borderlessness, is the distinction between domestic and external still viable? What are the main benefits and drawbacks?
- What are the main points of contention between the governing legal mandates and the operational requirements of psychological defence?
- Given Sweden's membership of the European Union and NATO, should any parts of the psychological defence work be outsourced to these organisations? What are the main opportunities and weaknesses?

JAMES PAMMENT *is Director of the Lund University Psychological Defence Research Institute. His main research interest is in the role of strategic influence in international relations, both its legitimate sides (e.g., public diplomacy and aid) and illegitimate (e.g., propaganda and hostile foreign interference). Previous affiliations include the Carnegie Endowment for International Peace, Swedish Defence University, the EU-NATO Hybrid Threats Center of Excellence, and the University of Texas at Austin.*

REFERENCES

Council of the European Union. (2022.) *Council conclusions on a Framework for a coordinated EU response to hybrid campaigns.* Retrieved from https://www.consilium.europa.eu/en/press/press-releases/2022/06/21/council-conclusions-on-a-framework-for-a-coordinated-eu-response-to-hybrid-campaigns/

Davis, P., & Jenkins, B. (2002). *Deterrence and Influence in Counterterrorism: A Component in the War on al Qaeda,* RAND Corporation, Retrieved from http://www.rand.org/pubs/monograph_reports/MR1619.html.

Fredheim, R. & Pamment, J. (2024) Assessing the risks and opportunities posed by AI-enhanced influence operations on social media. *Journal of Place Branding & Public Diplomacy*

Giannopoulos, G., Smith, H., & Theocharidou, M. (2021). *The landscape of Hybrid Threats: A conceptual model.* The European Centre of Excellence for Countering Hybrid Threats. Retrieved from https://www.hybridcoe.fi/publications/the-landscape-of-hybrid-threats-a-conceptual-model/

Government Offices of Sweden. (2024). *This is civil defence.* Retrieved from https://www.government.se/government-policy/civil-defence/this-is-civil-defence/

Keršanskas, V. (2020). DETERRENCE: Proposing a more strategic approach to countering hybrid threats. *Hybrid CoE Paper 2*. Retrieved from https://www.hybridcoe.fi/wp-content/uploads/2020/07/Deterrence_public.pdf

MSB. (2018). *Countering information influence activities – A handbook for communicators.* Retrieved from https://www.msb.se/siteassets/dokument/publikationer/english-publications/countering-information-influence-activities---a-handbook-for-communicators.pdf

Nakashima, E. (2019, February 27). US Cyber Command Operation Disrupted Internet Access of Russian Troll Factory on Day of 2018 Midterms. *The Washington Post.* Retrieved from https://www.washingtonpost.com/world/national-security/us-cyber-command-operation-disrupted-internet-access-of-russian-troll-factory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9-af5b-b51b7ff322e9_story.html

NATO. (2024). *Countering hybrid threats.* Retrieved from https://www.nato.int/cps/en/natohq/topics_156338.htm

Newman, H. (2022). Foreign information manipulation and interference defence standards: Test for rapid adoption of the common language and framework 'DISARM'. *Hybrid CoE Research Report 7.* Retrieved from https://www.hybridcoe.fi/wp-content/uploads/2022/11/20221129_Hybrid_CoE_Research_Report_7_Disarm_WEB.pdf

Pamment, J., & Agardh-Twetman, H. (2019). Can there be a deterrence strategy for influence operations? *Journal of Information Warfare*, 18(3), Article 123-135.

Pamment, J. (2021) *RESIST 2.* London: Government Communication Service

Pamment, J. (2022a). *A Capability Definition and Assessment Framework for Countering Disinformation, Information Influence, and Foreign Interference.* Riga: NATO Strategic Communications Centre of Excellence

Pamment, J. (2022b) How the Kremlin circumvented the EU sanctions on Russian state media in the first weeks of the illegal invasion of Ukraine. Journal of Place Branding & Public Diplomacy

Pamment, J. & Palmertz, B. (2023). Deterrence by Denial and Resilience Building, in Arcos, R., Chiru, I., & Ivan, C. (Ed) *Routledge Handbook of Disinformation and National Security.*

Pamment, J. & Smith, V. (2022) *Attributing Influence Operations: toward a community framework.* Riga: NATO Strategic Communication Centre of Excellence & EU-NATO Hybrid Centre of Excellence

Prime Ministers Office. (2023). *Pressträff om Sveriges säkerhet.* Retrieved from https://www.regeringen.se/pressmeddelanden/2023/01/presstraff-om-sveriges-sakerhet/

Schelling, T. (1966.) *Arms and Influence,* New Haven: Yale University Press.

SOU (Government Official Inquiries). (1953:27). Psykologisk försvar.

The European Union Agency for Cybersecurity (ENISA). (2022). *Foreign Information Manipulation and Interference (FIMI) and Cybersecurity- Threat Landscape.* Retrieved from file:///Users/el2105is/Downloads/Foreign%20Information%20Manipulation%20and%20Interference%20(FIMI)%20and%20Cybersecurity%20-%20Threat%20Landscape%20(2).pdf

The Swedish Agency for Public Management 2017:5. *Myndigheternas arbete med psykologiskt försvar.* Retrieved from https://www.statskontoret.se/publicerat/publikationer/publikationer-2017/myndigheternas-arbete-med-psykologiskt-forsvar/

II. UNDERSTANDING MALIGN INFORMATION INFLUENCE

The eight chapters that make up this section offer insight into the concept of malign information influence. The section begins with an overview of key terminologies, before examining how information influence exploits media systems and encourages political polarisation. It discusses cognitive biases, and the ways in which rhetorical devices are used to influence publics, including through conspiracy theories and gendered disinformation. The section concludes by assessing the current and potential future impact of artificial intelligence on the field.

9. MALIGN INFORMATION INFLUENCE: DEFINITIONS AND CONCEPTUALISATION

JAMES PAMMENT & JESPER FALKHEIMER

SUMMARY

This chapter introduces one of the key terms related to psychological defence in the Swedish conceptualisation of the propaganda field: malign information influence (otillbörlig informationspåverkan). It does so in the context of other, more widely used terms, in order to explain the strengths and weaknesses of the term.

This chapter has two main aims. First, to describe, define and clarify what is meant with concepts such as malign information influence, psychological defence, disinformation, misinformation, malinformation, information operations and foreign interference. Second, to show and argue for why the concepts psychological defence and information influence have stronger relevance from a policy and institutional perspective. But before we delve into the conceptual discussion, it is appropriate to describe the development of information influence from a broader societal perspective.

Information influence is basically about persuasive communication, where one actor has the well-thought-out aim of influencing another actor. All forms of persuasive communication occur in different social, cultural and political contexts that determine the reach and effects on attitudes, feelings and behaviour of individuals or groups of people. In addition to the importance of social contexts, the effects of persuasive communication are governed by several individual factors, psychological mechanisms, and characteristics of the person who pays attention to and interprets the content. In this book, we intend to contribute insights from all the perspectives that can be assumed to be important for influence - from cognitive bias, rhetorical strategies, to various contextual aspects and factors.

In the contemporary social debate, it is easy to get the impression that information influence, disinformation, and conspiracy theories have never had as much influence as they do today. But of course, this is not true. Persuasive communication – which is a form of human communication and not per se the same as manipulation, propaganda, or false information dissemination – is something that humanity has lived with since time immemorial. Through new media technologies, the power and scope of information influence has undeniably increased radically - during the 20th century, first through radio and TV, then through the Internet, and today through AI. However, although this of course varies in different societies, resistance in the form of informed, educated, and critically thinking citizens has also increased. Rational knowledge and critical thinking are the fundamental factors that create resistance to false and manipulative information.

From a historical standpoint it is worthwhile to introduce the concept of propaganda. Propaganda has been defined in several different ways but is usually described as "(...) the deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behaviour to achieve a response that furthers the desired intent of the propagandist" (Jowett & O'Donnell 2015, p. 7). The definition illustrates the difficulties in making simple conclusions about the role of propaganda in a democratic society. In an open society, it is permitted to produce and disseminate one-sided propaganda in several forms, such as political propaganda, advertising, PR campaigns and so on, even if there are obviously judicial limits that propagandists need to acknowledge (such as marketing regulations and laws against slander or threats to national security) which differs between states.

Historically, *propaganda* (from Latin, meaning to propagate or to sow) was used as a concept in 1622 in the Vatican for the *Sacra Congregatio de Propaganda Fide*, with a positive mission to propagate the faith of the Roman Catholic Church. The positive notion of propaganda prevailed into the early 20th century as used by the Soviet Union, China as well as the pioneers (e.g. Edward Bernays) of early public relations in the USA. In this thinking, propaganda was viewed as a legitimate tool for steering and controlling the "masses" and the public opinion, avoiding riots, protests and social concerns.

Reflecting the importance of transparency to the concept, propaganda is often divided into different categories. One of the most simple and clear divisions is between black, grey and white propaganda. Black propaganda is like disinformation – the information is false, and importantly, the source is also hidden. Grey propaganda is a mixing of false and true information, with the latter being used to increase effect. White propaganda is true information that is based on accurate facts and with accurate attribution to the source, but where the content is biased one-sidedly and often exaggerated through various rhetorical strategies.

CONTEMPORARY TERMINOLOGY

One of the points of concern of current debates about digital propaganda is a lack of conceptual clarity. While popular culture characterised the new media landscape as corrupted by "fake news", the emerging international policy area was initially characterised by the term "disinformation". However, over time it has become clear that use of the term disinformation as a catch-all is both problematic and misleading. The establishment of psychological defence as an alternative conceptualisation for the policy area in the Swedish context is therefore a significant development of the policy area, with potential consequences for international partners. It is these consequences that the following section explores in greater detail.

We argue here that use of the term disinformation has been shorthand for three overlapping groups of problems. The first group of problems is the spread of false information, whether deliberately or by accident, by individuals communicating through traditional and social media. Debates in this area are fundamentally about the quality of deliberation in the public sphere, as well as protection of fundamental freedoms such as expression. The second group of problems is the more complex phenomena of influence campaigns driven by motivated organisations capable of coordination, using multiple communication tools, and conducting clandestine activities. Such coordinated activities are often referred to as operations or campaigns to emphasise the complex nature of the planned influence effort. The third group of problems is foreign interference, which takes place in the context of other hybrid, cyber, and espionage activities that hostile states conduct. It

positions influence campaigns within the broader bilateral diplomatic relationship with a hostile foreign power. In our view, these groups of problems can overlap, but are distinct to the degree that different actors should be involved in monitoring, educating, and responding to the threats.

The first group of problems (mis-, dis-, and mal-information, or MDM) can be characterised by an emphasis on specific pieces of information content spread by individuals who are exercising their freedom of speech but are factually incorrect. NATO defines disinformation as the 'deliberate creation and dissemination of false and/or manipulated information with the intent to deceive and/or mislead' (NATO 2020). Although often mistakenly used to cover the entire policy area related to information-based interference, disinformation is best understood as part of a group of closely related terms focusing on two factors: the factualness (or truthfulness) of a message, and the likely intent behind the creation of the message. Misinformation refers to verifiably false information that is shared without an intent to mislead, whereas malinformation refers to true or partially true information that is twisted or taken out of context to support false interpretations (Pamment 2021). Together, the three terms cover many of the problematic issues associated with a digital public sphere, in which false information circulates without the checks and balances that traditional media provide.

Disinformation & associated concepts

Term	Misinformation	Disinformation	Malinformation
Definition	False information spread unintentionally	False information created intentionally	Factual information distorted intentionally
Operational components	Truth/factualness of content Intent of content creator		
Counter-measure capabilities	Content correction capabilities Public resilience-building capabilities		

- Content correction
- Fact-checking
- Debunking
- NGO networks "Elves"
- Public resilience
- Media literacy
- Public awareness campaign
- Prebunking

Source: Based on Pamment, J. (2022a), p. 15 and Pamment & Isaksson, 2024.

For many countries including Sweden, dis-, mis- and mal-information lack a legal or institutional basis and should be considered descriptors of a type of content for an analytical purpose. In the US Department of Homeland Security, for example, mis/dis/mal is referred to collectively as MDM (See e.g., Department of Homeland Security 2022). Disinformation is the more widely recognised term, albeit often as a synonym for the legal term 'information influence' (Sweden) or its international equivalents (see below), or as a more general reference to false or misleading content with a meaning gliding between mis/dis/mal. Disinformation has, for example, been used in this manner in official government documents such as the regulatory letters outlining the mandates of MPF and the Swedish Institute, as well as other government statements (See e.g., Dir 2018:80; SFS 2015:152). In more recent international debates, MDM increasingly represents an approach to factchecking and debunking from a health of democratic debate perspective,

which in practice means an emphasis on increasing public participation and reducing political polarisation. Countermeasures are often seen as the remit of civil society, such as journalists, nongovernmental organisations, tech platforms, think tanks, and academia in order to avoid the perception that government acts as the arbiter of truth. Tasks such as factchecking, debunking, content moderation, media literacy education, and source criticism are seen as key methods for improving the overall health of the public sphere from MDM (Pamment & Lindvall Kimber 2021).

The second group of terms focuses on more complex campaigns in which a hostile actor coordinates a variety of illegitimate communication techniques to influence target groups to their benefit. Encouraging the spread of mis-, dis- and mal-information may be among the methods used. The Swedish term malign information influence (*otillbörlig informationspåverkan*) is used to encapsulate efforts to influence democratic processes using illegitimate, but not necessarily illegal, methods to the benefit of a hostile external power. It emphasises the communication techniques that make up a coordinated effort to influence a society, their manipulative components, and the objectives of those conducting them (Pamment et al., 2018). Similarly, terms such as information manipulation (used by France and EU institutions) and influence operations (preferred by tech companies) define coordinated efforts to influence that often make use of clandestine techniques, and that ultimately seek to benefit the source and/or cause harm to others (Jeangéne Vilmer et al., 2019).

Information influence & associated concepts

Term	Information Influence	Information Manipulation	Influence Operations
Definition	Illegitimate communication intended to influence society to the benefit of hostile foreign powers	Coordinated efforts involving the diffusion of false or distorted information with the intent to cause political harm	Coordinated efforts to manipulate or corrupt public debate for a strategic goal
Operational components	Intent to cause harm to the benefit of hostile actor Use of multiple illegitimate communication techniques Negative interference in public debate Covert coordination		
Counter-measure capabilities	Analysis and identification capabilities Strategic communication capabilities		

- Analysis & identification
- Monitoring
- Investigation
- OSINT
- Strategic communication
- Counter-narrative
- Counter-brand
- Published analysis

Source: Based on Pamment, J. (2022a), p. 20 and Pamment & Isaksson, 2024.

This group of terms has stronger policy and institutional support insofar as they have a clearer legal basis. France has adopted information manipulation into law (Guillaume 2019), and the EU integrated the most important principles into its Foreign Information Manipulation & Interference (FIMI) policy (Council of the European Union 2022). Tech companies refer to influence operations and coordinated inauthentic behaviour in their policies for content removal and attribution (See e.g., Facebook 2021). The Swedish term “malign [or undie] information influence” covers influence efforts with a connection to foreign powers (state or nonstate) and provides the legal basis for government institutions to take countermeasures toward such campaigns where there is a clear external

dimension (Andersson, 2023). The fundamental principle of this group of activities is the idea of a concerted, often clandestine campaign with objectives that benefit the source and/or seek to cause harm to others. It is no longer a question of factchecking or debunking individual messages, but rather of understanding how messages fit within broader narratives and developing the means to counter those narratives. It heavily emphasises the capability to analyse, uncover, and counteract the coordinated behaviour of threat actors in the information environment.

The third group of concepts focuses explicitly on foreign interference. This area builds upon the themes covered in the concepts of MDM and influence campaigns by adding two additional factors. First is the assumption that the activities within this category, no matter who conducts them, are carried out on behalf of a hostile foreign power, ultimately with some form of state backing. Second, the communication activities broadly considered to be under foreign interference go beyond information per se and overlap with the categories of hybrid, cyber, or other state threats including espionage (Ördén & Pamment 2021). This imparts an additional layer of complexity upon information influence that positions the communication activities within a set of (often) covert tools for generating geopolitical influence.

Foreign interference

Term	Foreign Interference
Definition	Disinformation, information influence, and other hybrid influence methods conducted by or on behalf of a hostile state actor
Operational components	Intent to cause harm to the benefit of hostile actor Use of multiple illegitimate communication techniques Negative interference in public debate Covert coordination; Deployment in coordination with other hybrid influence methods
Counter-measure capabilities	Intelligence: collecting, processing, and use capabilities Security Policy: actor-specific capabilities

- Intelligence
- All-source
- Intelligence sharing
- Counterintelligence
- Security policy
- Deterrence
- Attribution
- Legislation

Source: Based on Pamment, J. (2022a), p. 25 and Pamment & Isaksson, 2024.

Foreign interference has the strongest legal and institutional support insofar as it is tied to military and civilian intelligence, counterintelligence, protection of critical infrastructure, and bilateral relations with hostile states. NATO and the EU have developed significant tools to deal with sub-threshold activities, most prominently cyber and hybrid, with the EU cyber sanctions regime and NATO announcements that cyberattacks and hybrid interference can trigger Article 5 (NATO 2024). The EU sanctions against Russian state media during the Ukraine invasion indicate the intensification of legal countermeasures to foreign interference through malign information influence (Council of the European Union 2022; Pamment 2022b).
The EEAS' FIMI policy attempts to combine the three groups referred to in this section in a manner which helps to broaden the policy area from "disinformation" to influence campaigns and hybrid foreign interference (European Union External Action Service 2021b). However, it does so in a broad manner that does not necessarily capture the distinctions between problem groups.

While the distinctions are not always neat (for example, it is often unclear who is behind MDM activities or how they fit into broader campaigns), we argue that distinguishing these three problem sets is necessary to explaining how and why psychological defence can become a key concept for this field in the coming years. With its staunch focus on external threats, malign information influence avoids the risk of meddling in domestic debates, instead viewing any interventions as the work of educators, journalists, and interest groups. The focus is instead solely on unravelling complex campaigns with clandestine and coordinated dimensions, and where necessary developing deterrence strategies to dissuade those actors from continuing with their illegitimate activities. At a time when the whole "disinformation" debate risks becoming a pawn in partisan freedom of speech debates, separating out the national security aspects of the problem is essential.

DISCUSSION

- If we imagine three groups of overlapping problems within the contemporary propaganda landscape, who should take responsibility for improving the quality of public debate in each problem area?
- What is the best way for organisations and countries to coordinate their activities in this area if they use different terms and have a different legal basis for countermeasures?
- Which term(s) are most viable in the long-term to develop this field?

JAMES PAMMENT *is Director of the Lund University Psychological Defence Research Institute. His main research interest is in the role of strategic influence in international relations, both its legitimate sides (e.g., public diplomacy and aid) and illegitimate (e.g., propaganda and hostile foreign interference). Previous affiliations include the Carnegie Endowment for International Peace, Swedish Defence University, the EU-NATO Hybrid Threats Center of Excellence, and the University of Texas at Austin.*

JESPER FALKHEIMER, *PhD, is Professor of Strategic Communication, Department of Communication, Lund University. He is national coordinator for the research network Communication and Media in Crisis and War (Campus Totalförsvar) and a researcher at the Lund University Psychological Defence Research Institute. He is also Professor II at Kristiana University of Applied Sciences, Norway, Honorary Professor at Hong Kong Polytechnic University, Visiting Professor at University of Johannesburg and Editor-in-Chief for Journal of Communication Management. His research interests are strategic communication in general, and crisis communication, disinformation and communication management.*

REFERENCES

Andersson, A. (2023). *Rättsligt ramverk för bemötande av informationspåverkan – En studie av det rättsliga ramverket för bemötande av informationspåverkan genom informationsåtgärder.* FOI-R--5443--SE. Retrieved from https://www.foi.se/rapportsammanfattning?reportNo=FOI-R--5443--SE

Council of the European Union. (2022.) *Council conclusions on a Framework for a coordinated EU response to hybrid campaigns.* Retrieved from https://www.consilium.europa.eu/en/press/press-releases/2022/06/21/council-conclusions-on-a-framework-for-a-coordinated-eu-response-to-hybrid-campaigns/

Department of Homeland Security. (2022). *DHS Needs a Unified Strategy to Counter Disinformation Campaigns.* Retrieved from https://www.oig.dhs.gov/sites/default/files/assets/2022-08/OIG-22-58-Aug22.pdf

Dir. 2018:80. *En ny myndighet för psykologiskt försvar.* Retrieved from https://www.regeringen.se/contentassets/b4b90c231b4144e683d5b4a594fe27b1/en-ny-myndighet-for-psykologiskt-forsvar-dir.-201880

European Union External Action Service. (2021a). *Tackling disinformation: Information on the work of the EEAS Strategic Communication division and its task forces (SG.STRAT.2).* Retrieved from https://www.eeas.europa.eu/countering-disinformation/tackling-disinformation-information-work-eeas-strategic-communication-division-and-its-task-forces_und_en?s=2803

European Union External Action Service. (2021b). *Tackling Disinformation, Foreign Information Manipulation & Interference.* Retrieved from https://www.eeas.europa.eu/eeas/tackling-disinformation-foreign-information-manipulation-interference_en

Facebook. (2021). *Threat Report The State of Influence Operations 2017-2020.* Retrieved from https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

Guillaume, M. (2019). *Combating the manipulation of information – a French case.* —(Hybrid CoE Strategic Analysis 16). Helsinki: Hybrid COE. Retrieved from https://www.hybridcoe.fi/wp-content/uploads/2020/07/HybridCoE_SA_16_manipulation-of-information_.pdf

Jeangène Vilmer, J., Escorcia, A., Guillaume, M., & Herrera, J. (2018). *Information Manipulation: A Challenge for Our Democracies.* Retrieved from https://www.diplomatie.gouv.fr/IMG/pdf/information_manipulation_rvb_cle838736.pdf

Jowett, G.S. & Donnell, C. (2015. *Propaganda and Persuasion.* 6th ed. Sage.

NATO. (2020). NATO's approach to countering disinformation: a focus on COVID-19. Retrieved from https://www.nato.int/cps/en/natohq/177273.htm#intro

Pamment, J. (2021). *RESIST 2.* London: Government Communication Service.

Pamment, J. (2022a). *A Capability Definition and Assessment Framework for Countering Disinformation, Information Influence, and Foreign Interference.* Riga: NATO Strategic Communications Centre of Excellence.

Pamment, J. (2022b). How the Kremlin circumvented the EU sanctions on Russian state media in the first weeks of the illegal invasion of Ukraine. *Journal of Place Branding & Public Diplomacy*. 19(2), 200-205.

Pamment, J. & Isaksson, E. (2024). *Psychological Defence: Concepts and principles for the 2020s*. Solna: Psychological Defence Agency. MFP report series 6/2024.

Pamment, J. & Lindvall Kimber, A. (2021). *Fact-checking and debunking: a best practice guide to dealing with disinformation.* Riga: NATO Strategic Communication Centre of Excellence.

Pamment, J., Nothhaft, H., Agardh-Twetman, H., & Fjällhed, A. (2018). *Countering Information Influence Activities: The State of the Art.* Swedish Civil Contingencies Agency (MSB). Stockholm: MSB.

SFS 2015:152. Instruktion för Svenska institutet.

Ördén, H. & Pamment, J. (2021). What is so Foreign about Foreign Influence Operations? Washington DC: *Carnegie Endowment for International Peace,* https://carnegieendowment.org/2021/01/26/what-is-so-foreign-about-foreign-influence-operations-pub-83706

10. MEDIA SYSTEMS AND RESILIENCE TOWARD INFLUENCE OPERATIONS AND DISINFORMATION

JESPER STRÖMBÄCK

SUMMARY

- While virtually all countries have transformed into high-choice media environments where it has become easier than ever to conduct malicious influence and disseminate propaganda and disinformation, not all countries and media systems are equally susceptible to the ills of propaganda, disinformation and misinformation.
- The analysis in this chapter suggests that what is most important for the degree to which different countries are resilient toward propaganda, disinformation, and misinformation is the supply of, use of, reliance on, and trust in traditional news media guided by journalistic norms and values compared to the use of, reliance on, and trust in political alternative and social media.

Digitalization and the transition from low-choice to high-choice media environments have fundamentally transformed virtually all processes of political communication and reshuffled the relationships between different sets of political actors, media actors, and people in general (Strömbäck, Boomgaarden, et al., 2022; Van Aelst et al., 2017; Young, 2023). Among other things, digital and social media have lowered the entry barriers to the public sphere, allowing anyone to publish and disseminate messages online. At the same time, traditional news media have lost much of their old function as gatekeepers, whose task it is to verify and select what information should be transformed into news and thereby reach the public sphere.

Traditional news media have also lost much their old function as a key intermediary between different sets of political actors and citizens at large. In contemporary media environments, political actors can bypass traditional news media and communicate directly with ordinary people, while ordinary people likewise can bypass traditional news media and get information elsewhere. This includes different social media and online forums, but also political alternative media that are driven by political goals and that seek to function as "correctives" of traditional news media and their news coverage (Holt et al., 2019).

Importantly, these processes have created much better opportunity structures for propaganda and malicious influence operations, whether by foreign or domestic actors. Here, propaganda refers to "deliberate, systematic attempt to shape perceptions, manipulate cognitions, and direct behavior to achieve a response the furthers the desired intent of the propagandist" (Jowett & O'Donnell, 1999, p. 6). In some cases (*white propaganda*), the source of the information is identified and the information accurate, but in other cases (*black propaganda*), information

is credited to a false source and includes lies, fabrications, and deceptions. Then there are also cases (*gray propaganda*) where the source may or may not be identified and the information may be false and misleading (Jowett & O'Donnell, 1999, p. 12-15). The includes, for example, when politicians spread false and misleading information.

Hence, propaganda and malicious influence operations includes spreading rumors, fake news, conspiracy theories as well as false and misleading information, and this may be done to convince as well as to sow uncertainty and distrust, and to fuel antagonism and hostility between groups (Benkler et al., 2018; Bradshaw et al., 2021). Increasingly, this is done using bots, algorithms, artificial intelligence, fake accounts, and human curation to produce, manage, and distribute false and misleading information (see chapter by Heath in this book).

As a consequence, the supply of false and misleading information has increased substantially, whether (*disinformation*) or not (*misinformation*) it *originally* is intended to mislead (Lecheler & Egelhofer, 2022). This has blurred the line between true and false information and increased the risk that people develop misperceptions and false beliefs. This has resulted in increasing politicization and polarization of facts (Rekker, 2021; Strömbäck, Wikforss, et al., 2022; Young, 2023) as well as a post-truth situation "in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief" (McIntyre, 2018, p. 5). Given that democracy require that people agree on a large set of shared facts, this development threatens to undermine liberal democracy.

EXPLAINING RESILIENCE TOWARD INFLUENCE OPERATIONS, DISINFORMATION AND MISINFORMATION

While this overall development is quite universal across countries, and in particular across liberal democracies, equally important is that not all countries are equally susceptible to the ills of malicious influence operations, disinformation and misinformation. Instead, research suggests that some countries are more resilient than others. For example, countries that are characterized by a high degree of political and affective polarization, and by strong populist parties, are generally more susceptible to malicious influence operations, disinformation and misinformation (Humprecht et al., 2020), as such factors both create a stronger demand for attitude-consistent information regardless of its veracity and more divisions that strategically can be exploited by domestic or foreign actors.

This chapter will however focus on factors related to media systems at the more general level (for more general analyses of media systems, see Hallin & Mancini, 2004; Humprecht et al., 2022). More specifically, it will focus on the role and importance of traditional news media, political alternative media, and social media.

THE ROLE AND IMPORTANCE OF TRADITIONAL NEWS MEDIA

In brief, traditional news media refer to news media like newspapers, radio and television, in their offline or online formats, that are guided by journalistic norms and values. These include that the purpose of journalism is to provide people with the kind of information they need to be free and self-governing (Kovach & Rosenstiel, 2021). Toward that end, other important journalistic norms and values are that news journalism should be free and independent of political or other interests, that all information should be verified before published, that news journalism should be impartial and describe reality as truthful as possible, and that

news journalists should act as watchdogs, scrutinizing those in power on behalf of the citizenry (Kovach & Rosenstiel, 2021). By doing this, the news media will provide information that people can trust and use to inform and orient themselves.

Of course, the extent to which news media and journalism lives up to its norms and values varies between news media both within and across countries (de Vreese et al., 2017). In reality, news journalism is shaped not only by journalistic norms and values, but also by their production routines, their formats, the structural relationship with the political system, the competition for audience attention, and economic pressures and considerations (Hallin & Mancini, 2004; Hamilton, 2004; Shoemaker & Reese, 2014). Most news media are commercial, meaning that they need to balance between journalistic and commercial goals and values. The exception is public service media that, depending on their political independence and resources, have greater opportunities to let journalistic norms and values guide their reporting. Consequently, research suggests that news reporting is more informative in public service compared to commercial news media, but also that people learn more from following public service news (Aalberg & Curran, 2012; Bjerling, 2022; Strömbäck, 2017). Even so, generally speaking, politically independent traditional news media are less likely to produce or disseminate disinformation and misinformation than other information sources (Humprecht et al., 2020).

Of key importance is hence the supply of news media that are politically independent and that provide high-quality news reporting, largely guided by journalistic norms and values. This includes public service media, that are of special importance since they are free to use and thus have greater opportunities to reach out broadly with high-quality, informative news coverage.

Of importance is however also the extent to which people use and trust the different news media (Strömbäck et al., 2020). The supply of high-quality news does not matter much if traditional news media are not widely used, or if people do not trust the information that they are provided by these news media. In this context, three challenges – aside from those related to political alternative media and social media that will be addressed later – are increasing differences in news media use depending not least on age and political interest (Espeland, 2024), increasing politicization of news media trust (Andersson, 2024), and in some countries (albeit not Sweden) decreasing news media trust (Newman et al., 2023).

Summing up, this review suggests that resilience toward information operations and disinformation is strengthened by (a) wide supply of traditional and politically independent news media and (b) public service media that (c) provide high-quality and truthful news reporting, (d) widespread use of traditional news media, and (e) widespread trust in traditional news media.

THE ROLE AND IMPORTANCE OF POLITICAL ALTERNATIVE NEWS MEDIA

One key change following from digitalization is the emergence and increasing importance of political alternative (or partisan) media. In brief, this concept refers to media, usually digital-only, that are guided by political or ideological goals and values and that are opposed to mainstream news media (Cushion, 2024; Holt et al., 2019). Political alternative media may or may not have formal or informal alliances with certain political parties, but more important is that their coverage is influenced by their political or ideological goals and values. Also important is that the degree of alternativeness may differ (Freudenthaler & Wessler, 2022; Staender

et al., 2024). While some largely follow traditional journalistic norms and values in how they do their work but focus on other issues than mainstream news media, others are more politicized in both what they cover and how they cover it. Some are also linked to foreign interests, such as the Russian RT and Sputnik (Yablokov & Chatterje-Doody, 2022).

Differences between political alternative media aside, a common denominator is that they pursue political or ideological goals and values. This holds in particular for right-wing alternative media, that are furthermore often very hostile toward mainstream media and seek to compete by affirming the beliefs and political identities of their users rather than by providing verified and truthful news (Benkler et al., 2018; Cushion, 2024; Figenschou & Ihlebaek, 2021). Because of this, political alternative media are more likely than traditional news media to be receptive to malicious information operations, but also to produce and disseminate false and misleading information that support their political goals and values (Strömbäck, 2023). Research also shows that use of political alternative media is associated with misperceptions and greater beliefs in conspiracy theories (Hmielowski et al., 2020; Strömbäck et al., 2023; Theorin et al., 2023).

From this perspective, what is important in terms of resilience toward information operations and disinformation is (a) the supply, (b) use of, (c) trust in, and (d) reliance on political alternative media. This holds in particular for right-wing alternative media, that typically display a greater alternativeness compared to and hostility toward traditional news media than left-wing alternative media (Hmielowski et al., 2020; Strömbäck et al., 2023; Theorin et al., 2023). Important though is to make a distinction between the *use of* and *reliance* on political alternative media. In most cases, such media do not offer comprehensive news coverage, and from that follows that they are often used in addition to traditional news media. This implies that the degree to which people rely on and trust political alternative media differ is important in the context of resilience toward information operations and disinformation. More specifically, it is less problematic if people now and then check them up and use them in combination with traditional news media than if they rely on and trust political alternative media more than traditional news media.

A special subset of "alternative media" that also needs to be recognized, in the context of malicious information operations, propaganda, and disinformation, are fake news sites (Egelhofer & Lecheler, 2019). In some cases, the line between political alternative media and fake news be blurry, since both seek to resemble real news in format and style, and since both may spread false and misleading information (Egelhofer & Lecheler, 2019). An important difference though is that content on fake news sites may be fully fabricated, either manually or by using artificial intelligence, and this is not typical for political alternative media.

THE ROLE AND IMPORTANCE OF SOCIAL MEDIA

On social media, virtually anyone can post, share, react to or comment on posts from people they know as well as people they do now know and anonymous accounts. Consequently, a large share of information on social media constitutes false and misleading information. Not surprisingly, research also shows that social media facilitate the dissemination of false and misleading information, not only because anyone can post virtually anything (although some content moderation occurs) but also because frequent dissemination signals popularity which the algorithms pick up and which contributes to further dissemination (Åkerlund, 2022; Vosoughi et al., 2018). One study focusing on Twitter also found that "falsehood

diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information" (Vosoughi et al., 2018, p. 1147).

Social media are also ideally suited for computational propaganda, that is, the use of bots, algorithms, artificial intelligence, fake accounts, and human curation to disseminate and promote disinformation and misinformation, and to shape the perceived popularity of certain posts or accounts (Bradshaw et al., 2021; Hagen et al., 2022; see also chapter by Heath in this book). Research also shows that a range of tools are used by political parties, governments and others to strategically influence the dissemination of false and misleading information, either domestically or to influence beliefs and political discussions in other countries (Bradshaw et al., 2021).

Although research suggests that differences in – among other things – technical affordances and content moderation means that not all social media are equally conducive to the dissemination of false and misleading information and malicious information operations, on a general level, there is hence little doubt that social media are problematic. Like what holds for political alternative media, what matters is however not only the prevalence of different social media platforms. Rather, what is important in terms of resilience toward information operations and disinformation is the (a) use of, (b) trust in, and (c) reliance on social media platforms for information, and (d) the willingness among users to share information that is suspected of being false and misleading (Humprecht et al., 2020, 2023).

CONCLUSION

What this discussion shows is that several factors related to media systems in different countries are important to understand the degree of resilience toward malicious information operations, disinformation, and misinformation. Summing up, what ultimately may be most important for the degree to which different countries are resilient toward malicious information operations, disinformation, and misinformation is the supply of, use of, reliance on, and trust in traditional and politically independent news media guided by journalistic norms and values compared to the use of, reliance on, and trust in political alternative and social media. While traditional news media certainly have a lot of flaws and at times clearly fail at providing people with the information that they need to be free and self-governing, they are still more trustworthy and informative than other types of media, less vulnerable to malicious information operations, and less likely to disseminate false and misleading information.

Given this, it is problematic that people increasingly turn away from using traditional news media, either by just tuning out or replacing them with increasing use of and reliance on political alternative media and social media. It is also problematic that traditional news media increasingly have become the target of politically motivated attacks, which contributes to both decreasing media trust and an increasing politicization thereof.

DISCUSSION

- How can people be encouraged to use and rely more on traditional news media instead of using and relying on political alternative and social media?
- How can politically motivated attacks on traditional news media be discouraged and counteracted?
- What policies and institutional behaviors by public authorities might strengthen or support traditional news media without infringing on their political independence?

JESPER STRÖMBÄCK, *PhD, is professor in journalism and political communication at the Department of Journalism, Media and Communication, University of Gothenburg. Among other things, his research focuses on the linkages between media use, media trust, beliefs in misinformation and conspiracy thinking. One of his most recent books is Knowledge resistance in high-choice information environments (Routledge, 2022), edited together with Åsa Wikforss, Kathrin Glüer, Torun Lindholm and Henrik Oscarsson.*

REFERENCES

Aalberg, T., & Curran, J. (Eds.). (2012). *How media inform democracy. A comparative approach.* London: Routledge.

Åkerlund, M. (2022). *Far right, right here: Interconnections of discourse, platforms, and users in the digital mainstream.* Umeå: University of Umeå.

Andersson, U. (2024). Ideologi vs användning – om dynamiken bakom förtroendet för public service. In U. Andersson, B. Rönnerstrand & A. Carlander (Eds.). *Inferno* (pp. 151–173). Göteborg: SOM-institutet.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda. Manipulation, disinformation, and radicalization in American politics.* New York: Oxford University Press.

Bjerling, J. (2022). *Public service. En svensk kunskapsöversikt.* Göteborg_ SOM-institutet.

Bradshaw, S., Bailey, H., & Howard, P. N. (2021). *Industrialized disinformation. 2020 global inventory of organized social media manipulation.* Oxford: Oxford Internet Institute.

Cushion, S. (2024). *Beyond mainstream media. Alternative media and the future of journalism.* London: Routledge.

de Vreese, C., Esser, F., & Hopmann, D. N. (2017). *Comparing political journalism.* London: Routledge.

Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: A framework and research agenda. *Annals of the International Communication Association,* 43(2), 97–116.

Espeland, E. (2024). The dynamics of political interest and news media avoidance: A generational and longitudinal perspective. *Journalism Studies,* 1–22.

Figenschou, T. U., & Ihlebaek, K. A. (2021). Media criticism from the far-right: Attacking from many angles. *Journalism Practice,* 13(8), 901–905.

Freudenthaler, R., & Wessler, H. (2022). How alternative are alternative media? Analyzing speaker and topic diversity in mainstream and alternative online outlets. *Digital Journalism,* 1–21.

Hagen, L., Neely, S., Keller, T. E., Scharf, R., & Vasquez, F. E. (2022). Rise of the machines? Examining the influence of social bots on a political discussion network. *Social Science Computer Review,* 40(2), 264–287.

Hallin, D. C., & Mancini, P. (2004). *Comparing media systems. Three models of media and politics.* New York_ Cambridge University Press.

Hamilton, J. T. (2004). *All the news that's fit to sell. How the market transforms information into news.* New Haven: Princeton University Press.

Hmielowski, J. D., Hutchens, M. J., & Beam, M. A. (2020). Asymmetry of partisan media effects? Examining the reinforcing process of conservative and liberal media with political beliefs. *Political Communication,* 37(6), 852–868.

Holt, K., Ustad Figenschou, T., & Frischlich, L. (2019). Key dimensions of alternative news media. *Digital Journalism,* 7(7), 860–869.

Humprecht, E., Castro Herrero, L., Blassnig, S., Brüggemann, M., & Engesser, S. (2022). Media systems in the digital age: An empirical comparison of 30 countries. *Journal of Communication,* 72(2), 145–164.

Humprecht, E., Esser, F., Aelst, P. V., Staender, A., & Morosoli, S. (2023). The sharing of disinformation in cross-national comparison: Analyzing patterns of resilience. *Information, Communication & Society,* 26(7), 1342–1362.

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. *The International Journal of Press/Politics,* 25(3), 493–516.

Jowett, G. S., & O'Donnell, V. (1999). *Propaganda and persuasion* (3rd ed.). Thousand Oaks: Sage.

Kovach, B., & Rosenstiel, T. (2021). *The elements of journalism. What newspeople should know and the public should expect* (4th ed.). New York: Crown.

Lecheler, S., & Egelhofer, J. L. (2022). Disinformation, misinformation, and fake news: Understanding the supply side. In J. Strömbäck, Å. Wikforss, K. Glüer, T. Lindholm & H. Oscarsson (Eds.), *Knowledge resistance in high-choice information environments* (pp. 69–87). London: Routledge.

McIntyre, L. (2018). *Post-truth.* Cambridge: MIT Press.

Newman, N., Fletcher, R., Eddy, K., Robertson, C. T., & Nielsen, R. K. (2023). *Reuters Institute Digital News Report 2023.*

Rekker, R. (2021). The nature and origins of political polarization over science. *Public Understanding of Science*, 30(4), 352–368.

Shoemaker, P. J., & Reese, S. D. (2014). *Mediating the message in the 21st Century.* New York: Routledge.

Staender, A., Humprecht, E., & Esser, F. (2024). Alternative media vary between mild distortion and extreme misinformation: Steps toward a typology. *Digital Journalism,* 1–21.

Strömbäck, J. (2017). Does public service TV and the intensity of the political information environment matter? *Journalism Studies,* 18(11), 1415–1432.

Strömbäck, J. (2023). Political alternative media as a democratic challenge. *Digital Journalism,* 11(5), 880–887.

Strömbäck, J., Boomgaarden, H., Broda, E., Damstra, A., Lindgren, E., Tsfati, Y., & Vliegenthart, R. (2022). From low-choice to high-choice media environments: Implications for knowledge resistance. In J. Strömbäck, Å. Wikforss, K. Glüer, T. Lindholm & H. Oscarsson (Eds.), *Knowledge resistance in high-choice information environments* (pp. 49–68). London: Routledge.

Strömbäck, J., Broda, E., Bouchafra, S., Johansson, S., Rettenegger, G., & Lindgren, E. (2023). Conspiracy thinking and the role of media use: Exploring the

antecedents of conspiratorial predispositions. *European Journal of Communication,* 38(3), 255–271.

Strömbäck, J., Tsfati, Y., Boomgaarden, H., Damstra, A., Lindgren, E., Vliegenthart, R., & Lindholm, T. (2020). News media trust and its impact on media use: Toward a framework for future research. *Annals of the International Communication Association,* 44(2), 139–156.

Strömbäck, J., Wikforss, Å., Glüer, K., Lindholm, T., & Oscarsson, H. (Eds.) (2022). *Knowledge resistance in high-choice information environments.* London: Routledge.

Theorin, N., Johansson, S., Strömbäck, J., Johansson, B., & Oscarsson, H. (2023). Allmänhetens syn på oegentligheter vid svenska val. *Statsvetenskaplig Tidskrift,* 125(4), 1009–1034.

Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stępińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheafer, T., & Stanyer, J. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association,* 41(1), 3–27.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science,* 359(6380), 1146–1151.

Yablokov, I., & Chatterje-Doody, P. N. (2022). *Russia Today and conspiracy theories. People, power and politics on RT*. London: Routledge.

Young, D. G. (2023). *Wrong. How media, politics, and identity drive our appetite for misinformation.* John Hopkins University Press.

11. OPINION FORMATION AND POLITICAL POLARIZATION

HANNA BÄCK & NILS GUSTAFSSON

SUMMARY

- Public opinion is essential for democratic societies, but social media changes how opinion is formed.
- Affective polarization – the phenomenon that individuals develop strong emotional attachments to their own political group and strong dislike and bias toward political outgroups – is growing in many democratic societies.
- Affective polarization may make individuals more susceptible to believe in and share political misinformation, suggesting that this type of polarization may be detrimental for information processing and information sharing in democratic societies.
- Another potential consequence of affective polarization is that it may influence what political opinions individuals take on various issues, and that it may increase ideological polarization in the electorate, for example on issues like climate change.

In many contemporary societies, there is a concern that the polarization of public opinion is undermining democratic governance and threatening societal stability. In this chapter, the aim is to present a research-based perspective on this issue. We start out by describing what public opinion is and why it is important for democratic societies. Next, we discuss key theories for how public opinion is formed and transformed. Subsequently, political polarization and how it can affect opinion formation is explained. The chapter concludes with a reflection on the extent to which polarization might constitute a significant threat to society.

WHAT IS OPINION FORMATION?

Public opinion can be said to be the opinions and views of the majority of the population in a society on political matters and other things of societal relevance, formed through private and public debates and discussions. Public opinion is sometimes conceived as an aggregation of the views of all individuals in a society, but on the one hand, people form opinions in a social context, and on the other, public opinion only comes into existence after it has been published. Indeed, people might not hold measured views on everything of importance unless someone asks them what they think (Zaller, 1992, p. 1).

In democratic societies, public opinion is of immense consequence to elected leaders. A political candidate or a political party that does not consider public opinion will fail getting elected. Often, a distinction is made between well-considered and informed opinions shaped by citizens through thoughtful deliberation in a democratic public sphere, and more emotionally driven reactions from a group perhaps influenced by populist rhetoric. This line is however not easy to draw.

An attitude can be said to have three components: cognition, affect and behaviour, where cognition refers to what you believe about the world, affect refers to what you feel about it, and behaviour refers to how you will act. This means that an attitudinal change demands either a change in values or a redefinition of perceived reality. The process of how public opinion comes into being is called opinion formation. This research tries to connect the psychological processes in attitude formation in individuals with sociological processes of the formation of opinions by the collective, and how these processes are mediated through communication. It also seeks to reconcile the observed shifts in collective opinion with the relative lack of empirical evidence for lasting changes in individual attitudes (cf. Druckman, 2022).

In a historical perspective, research has evolved from the main view that the crowd is impulsive and easy to manipulate by propaganda, to a consensus that attitudes and values in individuals are stable and difficult to change in the long perspective. In the current era characterized by a small number of algorithmically driven and powerful social media platforms, there is a debate whether digital and computational propaganda might have stronger effects, but there is still no clear empirical evidence of this (Budak et al., 2024).

One way of explaining this seeming paradox has been to point to the influence of social norms on what opinions are publicly expressed. Collective opinion changes are often driven by social conformity and situational factors, which do not necessarily translate into enduring individual attitude changes. Two theories of opinion formation with a long-lasting legacy focus on the direct and individual effects versus the indirect and collective effects of media on opinion formation. Whereas the so-called hypodermic needle model postulates that there are direct effects of media and communication on individuals, the two-step flow model, which sees media effects as indirect, points to the importance of opinion leaders who in turn influence their social networks (Lazarsfeld, 1948).

An important aspect of opinion formation is communication. The field of media and communication studies tries to understand the way that various types of media shape our opinions through *priming* us for what issues we should care about and *framing* them in a way that shapes what we think about them (Scheufele & Tewksbury, 2007). In the past decade, one main interest has been to study the influences of social media on opinions, including worries about how the algorithmic presentation of content might create “filter bubbles” that limit our scope of information – a worry that seems to be exaggerated according to empirical research (cf. Jones-Jang & Chung, 2024). Not least, social media has been hypothesized to be one of the main culprits of the seeming increase in political polarization that has been seen in many societies around the world.

WHAT IS POLITICAL POLARIZATION?

The study of political polarization has a long tradition within political science. Early research on this topic focuses mainly on *ideological polarization*, which refers to the idea that individuals or political groups have vastly different opinions on some political issues, resulting in that there is little overlap between for example parties in ideological positions. This can include differences in policy positions between individuals or groups on a left-right dimension or some other salient policy dimension. More recently, the study of so-called *affective polarization* has become much more prevalent, originating in the study of conflicts between Republican and Democrat supporters in the US (see e.g. Mason 2018; Iyengar et al. 2012).

In short, affective polarization refers to when individuals develop strong emotional attachments to their own political group and strong dislike and bias toward political outgroups that one does not identify with. The field of studies analysing affective polarization outside of the US, for example focusing on the multiparty context of European countries, is growing rapidly (see e.g., Wagner 2021).

There are some studies in the Swedish context that show that voters are affectively polarized. For example, Renström and colleagues (2020) show, based on representative survey data collected among Swedish citizens, that supporters of most of the parties have negative feelings and are biased against supporters of the Sweden Democrats. Sweden Democrat supporters harbour similar negative feelings and biases against supporters of the other parties, especially toward supporters of the Left party and the Greens. When asked about their feelings toward their own party's supporters, voters are much more positive.

An important debate in the literature on affective polarization is the question of what the relationship is between ideological and affective polarization. One view of this relationship, which has been called the *party-over-policy hypothesis,* argues that because political group attachments may function as a social identity, strengthening group identification increases intergroup differentiation, which manifests as polarization between supporters of different parties. Another view, the so-called *policy-over-party hypothesis* instead argues that policy preferences drive affective polarization, suggesting that if individuals hold more extreme political attitudes, they become more affectively polarized. Based on several experimental studies, Dias and Lelkes (2021) find support in favor of the party-over-policy hypothesis, concluding that partisan identity is the principal mechanism of affective polarization.

Another important conclusion in this field of research is that political elite communication influences affective polarization in the electorate. This type of research draws on a long-standing literature which has stressed the role of so-called *party cues*. The argument in this literature is that citizens can make reasoned choices by making use of information shortcuts and that the actors behind a political issue are an important shortcut in politics (see e.g., Nicholson 2012). Receiving a cue that some political information comes from an ingroup, or outgroup party representative may according to this literature determine how citizens interpret political messages and are persuaded by them, which might be interpreted as a version of the two-step hypothesis with opinion leaders described above.

In the field of affective polarization, it has been hypothesized that such party cues are important to consider when explaining how political elites' social media communication polarizes the electorate. Based on a survey experiment performed in Sweden, Bäck and colleagues (2023) show that individuals who received a factual political message with a source cue from an in- or outgroup representative exhibited higher affective polarization compared to a control group who did not receive any party cue when reading the same political message. This effect was especially pronounced when the individuals already held strong partisan affinities. This suggests that individuals who are already polarized are more likely to be biased when interpreting elite communication online, which in turn may lead them to become even more polarized when reacting to such communication, contributing to a polarization of public opinion.

Finally, the news media play a crucial role in shaping perceptions of polarization by presenting polarizing depictions of others' behaviours (Hoewe & Peacock, 2020;

Farjam et al., 2024). According to the theory of presumed media influence, people tend to believe that others are more affected by the media and perceive others as more extreme than themselves (Levendusky & Malhotra, 2016). This holds especially true for social media, since social media algorithms tend to reinforce content that drives emotional responses, showing us in effect a distorted image of the world (Bail, 2020).

HOW DOES POLARIZATION INFLUENCE OPINION FORMATION?

It is only recently that scholars have started to analyze the political effects of affective polarization. In an important study based on Twitter data, Osmundsen and colleagues (2021) show that individuals who report hating their political opponents, and who are highly polarized, to a larger extent share fake news. Belief in false information and in conspiracy theories has also been linked to affective polarization. For example, a study based on survey data collected in the US shows that affective polarization influences cognitive processing, such that people become less critical of information from their own party and thus are more likely to believe false information (Jenke 2024). Hence, some research suggests that the more affectively polarized a person is, the more likely he or she is to believe in and share political misinformation, suggesting that this type of polarization may be detrimental for information processing and information sharing in democratic societies.

Another potential consequence of affective polarization is that it may influence what political opinions individuals take on various issues, and that it may increase ideological polarization in the electorate. More research is needed on this topic but there are some important studies focusing on this potential consequence of affective polarization on opinion formation.

For example, in a large-scale survey study analysing political attitudes over time in the US during the outbreak of COVID-19, Druckman and colleagues (2020) found a strong association between individuals' partisan animosity and their attitudes about the pandemic and their willingness to engage in preventative behaviours, showing clear gaps between Democrats and Republicans in attitudes toward the virus and policies to prevent the spread of it. Druckman et al. (2020) stress the role of *partisan-motivated reasoning*, which refers to the idea that partisans process information and form political attitudes with the "goal of confirming their partisan identities". The authors also stress the role of party cues, where in the US context, the Democratic party representatives consistently expressed a greater concern about the virus and supported more restrictive policies than Republican politicians. Hence, individuals may become polarized on policy issues because they are given cues from parties or other political groups or actors that they identify with to take on certain positions.

Another important example of how affective polarization may influence opinion formation relates to the issue of climate change and how to tackle it, which has shifted from being primarily a scientific concern to becoming a highly polarized issue. In Sweden, the relationship between partisanship and attitudes toward climate change and climate mitigation measures has become much stronger in recent years, with supporters of left-leaning parties reporting higher beliefs in human-induced climate change, greater concern, and support for climate policies compared to right-leaning supporters (Axelsson & Jönsson 2023).

This type of opinion polarization is likely to be due to a *sorting* of policy issues, where individuals who identify with right wing populist parties are given cues about where to stand on the climate issue. Starting as parties that mainly focused on immigration policy, right-wing populist parties' stances on climate change have incorporated anti-environmentalism into their core ideologies, including scepticism toward the scientific consensus on climate change and opposition to mitigation policy solutions. In line with such an argument about sorting of policy preferences, Chen and colleagues (2023) analyse Finnish Twitter endorsement networks and show that the climate issue has become more aligned with the immigration issue. This suggests that the right-wing populist True Finns who stress an anti-environmentalist agenda give their voters cues about climate policy resulting in that people who identify with this party hold both anti-immigrant views and take negative positions on climate mitigation policy. More research is needed to analyse such a sorting of policy preferences which is likely to be a result of affective polarization.

It is important to emphasize that ideological polarization in the elite or in the electorate can be seen as an aspect of a healthy and competitive democracy. A society that is characterized by intensive affective polarization, can hypothetically experience a number of threats to democratic stability. Identity-based animosity can lead to political violence, as well as to a decline in trust of democratic institutions. The opportunities for the citizenry to form enlightened and informed opinions might be severely hampered by lacking trust in public and commercial sources of information and arenas for debate. This might in turn be exploited by foreign hostile agents.

DISCUSSION

- How do social media and the algorithms that determine what content citizens encounter online influence affective polarization?
- How do political elites and the traditional news media contribute to increasing affective polarization in the electorate?
- What are the consequences of affective polarization for the spread of misinformation?
- What are the consequences of affective polarization for opinion formation and the policy attitudes that citizens hold on various issues?

HANNA BÄCK, *PhD, is a professor of political science at Lund University. Her main research interests are political parties and political behavior. She has published extensively in the field of political psychology and much of her recent contributions focus on understanding the causes and consequences of affective polarization. Her work has appeared in journals such as British Journal of Political Science, Comparative Political Studies, New Media & Society, Political Psychology and Political Science Research and Methods.*

NILS GUSTAFSSON *holds a PhD in political science and is Senior Lecturer in Strategic Communication at Lund University, specializing in political communication and social media. His research focuses on the impact of social media on political participation, political parties, activism, and civil society. His work has appeared in journals such as New Media & Society, Social Media + Society, Political Behavior, and Organization.*

REFERENCES

Axelsson, S., & Jönsson, E. (2023). *Miljö-och klimatopinion i Sverige 2022.* SOM-rapport nr 40. Göteborg: SOM-institutet.

Bail, C. (2022). *Breaking the social media prism: How to make our platforms less polarizing.* Princeton University Press.

Bäck, H., Carroll, R., Renström, E., & Ryan, A. (2023). Elite communication and affective polarization among voters. *Electoral Studies,* 84, 102639.

Budak, C., Nyhan, B., Rothschild, D.M. et al. (2024). Misunderstanding the harms of online misinformation. *Nature 630,* 45–53. https://doi.org/10.1038/s41586-024-07417-w

Chen, T. H. Y., Salloum, A., Gronow, A., Ylä-Anttila, T., & Kivelä, M. (2021). Polarization of climate politics results from partisan sorting: Evidence from Finnish Twittersphere. *Global Environmental Change,* 71, 102348.

Dias, N., & Lelkes, Y. (2022). The nature of affective polarization: Disentangling policy disagreement from partisan identity. *American Journal of Political Science,* 66(3), 775-790.

Druckman, J. N. (2022). A framework for the study of persuasion. *Annual Review of Political Science,* 25(1), 65-88.

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). Affective polarization, local contexts and public opinion in America. *Nature human behaviour,* 5(1), 28-38.

Farjam, M., Bruhn, T., Gustafsson, N., & Segesten, A. D. (2024). The uses of the term polarisation in Swedish newspapers, 2010–2021. *Nordicom Review*, 45(1), 1-34.

Hoewe, J., & Peacock, C. (2020). The power of media in shaping political attitudes. *Current Opinion in Behavioral Sciences,* 34, 19-24.

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public opinion quarterly,* 76(3), 405-431.

Jenke, L. (2024). Affective polarization and misinformation belief. *Political Behavior*, 46(2), 825-884.

Jones-Jang, S. M., & Chung, M. (2024). Can we blame social media for polarization? Counter-evidence against filter bubble claims during the COVID-19 pandemic. *New Media & Society,* 26(6), 3370-3389.

Lazarsfeld P., Berelson B., Gaudet H. (1948). *The People's Choice*, New York: Columbia University Press.

Levendusky, M., & Malhotra, N. (2016). Does media coverage of partisan polarization affect political attitudes?. *Political Communication,* 33(2), 283-301.

Mason, L. (2018). *Uncivil agreement: How politics became our identity.* University of Chicago Press.

Nicholson, S. P. (2012). Polarizing cues. *American journal of political science,* 56(1), 52-66.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review,* 115(3), 999-1015.

Renström, E., Bäck, H., & Schmeisser, Y. (2020). Vi gillar olika. Om affektiv polarisering bland svenska väljare. In Andersson, U., Carlander A., & Öhberg, P. (eds.) *Regntunga skyar. SOM-undersökningen 2019.* Göteborg: SOM-institutet.

Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication,* 57(1), 9-20.

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies, 69,* 102199.

Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge University.

12. BIAS AND COGNITIVE INFLUENCE MECHANISMS

JOHAN ÖSTERBERG

SUMMARY

- Cognitive biases are mental shortcuts that can lead to flawed judgments and decision-making errors, as the brain simplifies information processing. These biases arise from the brain's use of heuristics, which are mental shortcuts, and are rooted in theories such as the dual-process theory, where quick, intuitive thinking can cause errors.
- Some common biases include confirmation bias, the tendency to favour information that confirms existing beliefs, and anchoring bias, the tendency to rely too heavily on initial information when making decisions.
- Other biases include the availability heuristic, where easily recalled events are seen as more frequent, hindsight bias, where past events seem more predictable than they were, and overconfidence bias, where individuals overestimate their own abilities.
- The framing effect shows how the presentation of information can influence choices, the sunk cost fallacy describes continuing with losing endeavours due to past investment, and groupthink, the desire for group conformity, can lead to the spread of disinformation.
- Disinformation campaigns often intentionally exploit cognitive biases to manipulate perceptions and drive behaviour. For example, they may use emotionally charged content to trigger strong reactions or leverage confirmation bias to ensure that the message aligns with targeted individuals' beliefs

Understanding cognitive biases is essential for psychological defence because these biases can significantly influence how we perceive, interpret, and react to situations, often leading to distorted thinking. Cognitive biases are mental shortcuts our brains use to make quick judgments, but they can also lead to errors in reasoning and decision-making. By being aware of these biases, individuals can improve their ability to think critically, make more rational decisions, and avoid self-deception. This awareness serves as an influential tool for psychological defence, as it helps people recognize when their judgments are being clouded by bias and allows them to maintain a more balanced and correct view of themselves and the world around them. In turn, this can foster emotional resilience and reduce the likelihood of falling prey to disinformation.

Cognitive biases can be described as systematic patterns of deviation from standard or rationality in judgment, where individuals create their own subjective reality based on their perception (Haselton et al, 2015, Blanco, 2016). These biases arise due to the brain's attempt to simplify information processing, leading to

flawed judgments, memory errors, and decision-making shortcuts, or heuristics. The study of cognitive biases is critical in fields such as, e.g., psychology, behavioural economics, and decision-making. These cognitive biases reflect how our thinking can be influenced by irrational patterns, even when we strive for logical decision-making. Cognitive biases stem from various mental shortcuts our brains use to process vast amounts of information quickly. Recognizing these biases can help individuals and organizations mitigate their potential negative impacts.

Cognitive biases have roots in several psychological theories, as for example, the dual process theory (Kahneman, 2011), who suggests that human cognition operates on two systems: system 1 (fast, intuitive, emotional) and system 2 (slow, deliberate, analytical). Cognitive biases often arise from the quick, intuitive thinking of system 1. Heuristics and biases program (Tversky & Kahneman, 1974): This significant work established the idea that heuristics, or mental shortcuts, often lead to biased judgments. Furthermore, there is the bounded rationality (Simon, 1957), which argue that humans are only rational within the limits of the information available to them, the time they have, and their cognitive processing capacity. This limitation can often lead to biased decisions.

There are different types of cognitive biases, and they can be categorized into types based on their mechanisms.

- *Memory biases* affect how we recall past experiences. For example, self-serving bias leads people to attribute positive events to their own actions and negative events to external factors (Shepperd et al, 2008).
- *Decision-Making biases,* which influence how people make choices. For instance, the framing effect shows that the way information is presented can significantly affect decisions, the framing-effect will be further explained in this chapter.
- *Social biases* arise in social contexts, for instance, in-group bias causes individuals to favor members of their own group over others.

The rapid technological development, the increasing mental workload an individual experience, as well as biological and social factors makes us more vulnerable for disinformation through cognitive bias. While these heuristics are efficient, they can often lead to errors. Some of the most common cognitive biases will be described in the following section.

COGNITIVE BIASES

Confirmation bias, is the tendency to search for, interpret, and remember information in a way that confirms one's pre-existing beliefs or hypotheses, while giving disproportionately less consideration to alternative possibilities (Nickerson, R., S. 1998). It can be exemplified as a person with a specific political viewpoint might only read news articles that support their opinions and ignore contradictory evidence. For example, individuals engaged in political debates often favor news sources and data that support their political ideology, dismissing alternative viewpoints even if they are more accurate or balanced. As Nickerson (1998) explains, confirmation bias can contribute to polarization in society, as it entrenches people deeper into their own beliefs without critically examining opposing arguments. This can be particularly prevalent what one emphasizes and how one assesses information on social media. Another common bias is the anchoring bias, in which individuals rely too heavily on the first piece of information they receive (the "anchor") when making decisions. This initial anchor serves as a reference point, and subsequent judgments are made by adjusting away from it,

even when the anchor is arbitrary or irrelevant. A typical example of anchoring bias is perceived in negotiations, when a seller sets a high price for a product, that initial price becomes the anchor, influencing how buyers perceive future price offers. Tversky and Kahneman (1974) demonstrated that even arbitrary anchors, such as randomly generated numbers, can influence participants' estimates in experiments, illustrating how pervasive this bias is.

The availability heuristic refers to the tendency to overestimate the likelihood of events based on how easily examples come to mind. This cognitive bias is particularly evident when people evaluate risks. Events that are more recent, vivid, or emotionally charged tend to be more accessible in memory, making them seem more frequent or probable than they really are. For instance, after high-profile airplane crashes are reported in the media, many people overestimate the risk of air travel, even though statistically, it remains one of the safest modes of transportation. Tversky and Kahneman (1973) highlighted the availability heuristic as a crucial factor in understanding how people assess probabilities in uncertain situations. or control, such as driving, investing, or decision-making in professional contexts. Availability heuristic can lead to distorted perceptions of risk, causing people to overreact to rare but dramatic events, such as terrorist attacks, while underestimating more common but less extraordinary risks, like car accidents or chronic diseases.

Hindsight bias, often summarized by the phrase "I knew it all along," occurs when people perceive past events as being more predictable than they were at the time. This bias can lead individuals to overestimate their ability to have foreseen an event's outcome, especially after the event has occurred. E.g., after a stock market crash, investors may claim they "saw it coming" or that it was "obvious" in hindsight, even though such events are difficult to predict accurately. Roese and Vohs (2012) describe hindsight bias as a retrospective distortion of reality, which can make it difficult for people to learn from past experiences, as they believe the outcome was inevitable all along. Furthermore, hindsight bias can reduce the perceived need for rigorous planning and analysis, as people may believe they can rely on intuition or common sense to predict future events. This is also prevalent online, where antagonists aim at pointing out certain groups or occurrences.

Overconfidence bias occurs when individuals overestimate their knowledge, abilities, or the accuracy of their judgments (Moore & Healey, 2008). This bias is particularly common in areas where people feel they have expertise or control, such as driving, investing, or decision-making in professional contexts. Overconfidence can lead to poor decision-making in high-stakes environments, such as financial markets or leadership roles, and leaders who underestimate challenges may make inadequate preparations (West & Stanovich, 1997). Closely related to overconfidence bias is the Dunning-Kruger effect, which describes how individuals with low competence in a given area tend to overestimate their abilities. Meanwhile, those with high competence may underestimate their expertise. Kruger and Dunning (1999) explain that this bias arises because the skills required to perform well in a domain are often the same skills needed to assess one's own performance. The Dunning-Kruger effect can lead to overconfidence in situations where competence is critical, such as professional decision-making or problem-solving. It can also lead to people rejecting expert advice, believing they are more knowledgeable than they truly are, which relates to many debates on social media. Loss aversion is the tendency for individuals to prefer avoiding losses over

receiving equal gains. Kahneman and Tversky (1984) described loss aversion as central to prospect theory, which explains that people feel the pain of loss more acutely than the pleasure of gain. For example, losing $1000 typically feels worse than gaining $1000 feels good. Loss aversion can explain why individuals and organizations are resistant to change, as they fear the potential losses associated with new decisions or strategies, even when the gains might outweigh the risks. The result might be that innovation and progress stagnates.

The framing effect describes how people's decisions are influenced by the way information is presented, rather than by the information itself. Tversky and Kahneman (1981) demonstrated that people respond differently to choices depending on whether they are framed as gains or losses. For instance, when a medical treatment is described as having a "90% survival rate", people are more likely to choose it than when the same treatment is described as having a "10% mortality rate", even though both statements convey the same information. The Asian disease experiment (Tversky & Kahneman, 1986) illustrate this.

Tversky and Kahneman presented participants with the following scenario, which involves a hypothetical outbreak of a disease expected to kill 600 people:

Positive Frame:

- **Program A:** 200 people will be saved.
- **Program B:** There is a one-third probability that 600 people will be saved and a two-thirds probability that no one will be saved.

Negative Frame:

- **Program C:** 400 people will die.
- **Program D:** There is a one-third probability that no one will die and a two-thirds probability that 600 people will die.

Even though the two sets of choices are logically identical (saving 200 people is equivalent to 400 dying, given 600 people), people's responses are strongly influenced by how the choices are framed:

- In the **positive frame**, most people choose the **certain option** (Program A: 200 people saved).
- In the **negative frame**, most people choose the **risky option** (Program D: a one-third chance no one dies).

This experiment shows how individuals tend to be risk-averse when a positive frame (lives saved) is presented and risk-seeking when a negative frame (deaths) is presented. Tversky and Kahneman used this experiment to support their development of prospect theory, which describes how people make decisions based on potential gains or losses rather than final outcomes, and how they weigh certain outcomes more heavily than probabilistic ones. Furthermore, people are generally more sensitive to losses than to gains. The framing of the problem emphasizes this — the prospect of losing lives tends to make people more willing to take risks to avoid a loss. The framing effect is frequently exploited in marketing, political campaigns, and public health messaging to influence people's choices. Understanding how framing impacts decision-making is crucial for ensuring ethical communication in many different areas. The sunk cost fallacy occurs when individuals/organizations continue to invest in a losing endeavour only because they have already invested time, money, or effort into it. This is irrational since

future decisions should be based on potential outcomes, not past investments. For instance, someone might stay in a failing business or a bad relationship because they have already invested years into it, even though leaving would be more beneficial in the long run. Arkes and Blumer (1985) found that people are more likely to continue with a losing investment when they feel that they have already put significant resources into it. The sunk cost fallacy can lead to poor decision-making in both individual and professional contexts. It may cause organizations to continue with unprofitable projects rather than cut their losses, leading to wasted resources and missed opportunities.

Finally, the well-established concept of Groupthink (Janis, 1972), where the desire for conformity within social groups often leads people to adopt the beliefs and attitudes of those around them. This makes disinformation more likely to spread within echo chambers, where contradictory viewpoints are minimized or excluded. The pressure to conform within the group can result in the suppression of objections, self-censorship, and the illusion of unanimity, even when some group members privately disagree with the group's course of action.

CONCLUSION

Cognitive biases are an inherent part of human thinking, shaping how we perceive the world and make decisions. While these biases help us navigate a complex environment, they also introduce systematic errors that can lead to irrational judgments and poor decisions which affects our perceptions of information. By recognizing and understanding cognitive biases, individuals and organizations can take steps to mitigate their effects, improve decision-making, and promote more rational behaviour. Understanding these biases is not only an academic issue; it has practical implications in every facet of life, from individual relationships to business and politics, as well as within the information society.

Disinformation campaigns often intentionally exploit cognitive biases to manipulate perceptions and drive behaviour. For example, they may use emotionally charged content to trigger strong reactions or leverage confirmation bias to ensure that the message aligns with targeted individuals' beliefs. The combination of cognitive biases and the viral nature of disinformation, especially on social media, creates an environment where falsehoods can easily spread, even in the presence of contradictory evidence. In addition, cognitive biases significantly contribute to the effectiveness of disinformation. They create mental shortcuts that lead people to accept, believe, and share misleading information without critically evaluating its accuracy. Combating disinformation requires not only fact-checking and education but also awareness of how cognitive biases influence thinking and decision-making.

DISCUSSION

- How do cognitive biases shape our perception of reality, and to what extent can we trust our own judgment when making decisions?
- In what ways do cognitive biases contribute to the persistence of stereotypes and misinformation, both at an individual and societal level?
- Are some cognitive biases more harmful than others in specific contexts (e.g., in legal, medical, or business decision-making)? How can we mitigate their impact?
- Can education or awareness of cognitive biases reduce their influence, or do we need more structural changes in our environments to prevent biased thinking?

JOHAN ÖSTERBERG *works as a research and training coordinator at the Swedish psychological defence agency. He holds a PhD in Psychology from Karlstad's university and the Swedish defence university. He current work focuses on the intersection of psychology, security, and national defence. With experience in cognitive psychology and behavioral science, Dr. Österberg works with the development of strategies aimed at understanding and mitigating psychological threats and disinformation. Furthermore, he conducts training and exercises on the topic of information influence. Österberg also participates in several international networks and cooperations.*

REFERENCES

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes,* 35(1), 124-140.

Blanco, F. (2016). Positive and negative implications of the causal illusion, Consciousness and cognition. Doi: 10.1016/j.concog.2016.08.012.

Haselton, M. G., Nettle, D., & Andrews, P. W. (2015). The evolution of cognitive bias. The handbook of evolutionary psychology, 724-746.

Janis, I. (1972). *Victims of groupthink.* Boston, Houghton Mifflin.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist,* 39(4), 341-350.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology,* 77(6), 1121-1134.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Farrar, Straus and Giroux.

Mitte, K. (2008). Memory bias for threatening information in anxiety and anxiety disorders: a meta-analytic review. Psychological bulletin, 134(6), 886.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review,* 115(2), 502-517.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology,* 2(2), 175-220.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science,* 7(5), 411-426.

Simon, H. A. (1957). *Models of Man: Social and Rational.* Wiley.

Shepperd, J., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bias. Social and Personality Psychology Compass, 2(2), 895-908.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology,* 5(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science,* 185(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science,* 211(4481), 453-458.

Tversky, A., & Kahneman, D. (1986). The framing of decisions and the evaluation of

prospects. In Studies in Logic and the Foundations of Mathematics (Vol. 114, pp. 503-520). Elsevier.

West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. Psychonomic Bulletin & Review, 4(3), 387-392.

13. RHETORICAL STRATEGIES FOR DISINFORMATION

ORLA VIGSØ

SUMMARY

- Rhetorical strategies for disinformation can be two things: trying to persuade people about what you know not to be true, or making people distrust everybody.
- Disinformation often use emotional appeals (ethos, pathos), but even rational arguments (logos).
- Often, disinformation presents emotional arguments as if they were facts.
- Disinformation can deal with both the past, the present, and the future.
- Disinformation strategies are often long term, creating a belief in order to use it in later argumentation.

Rhetorical strategies for disinformation are planned ways of communicating in order to persuade a certain audience about something which the sender themselves know to be in contradiction of facts. The definition of rhetoric presented by Aristotle, one of the founding fathers of rhetoric, is the ability in a particular case to see the available means of persuasion (Aristotle 2020: I. 2.1355b20), so when applied to disinformation it includes all measures which can be used to either make people believe counterfactual claims or to strengthen their existing belief in such claims. Thus, rhetorical strategies are arguments for claims which can be proven false but are believed to be true by an audience.

The reason this works is because argumentation in real life does not follow the laws of the logical syllogism “all X are Y, Z is X, thus Z is Y”. The model for argumentation used by most rhetoricians is focused on what is needed to convince people, which is not undisputable truths but rather claims which are probable to the audience. And what is deemed probable by a given audience has to do with already existing views in this group; people are most willing to accept an argument if it fits in with what they already believe (Toulmin 2020, McCroskey 2006).

Therefore, rhetorical strategies of persuasion must take as their starting point something which is accepted by the public to make an argument which supplies the audience with new beliefs. The further away from the existing truths the new claims are, the more steps can be needed to make the public accept them. In other words, disinformation campaigns must be understood as possibly performed in multiple steps over a period, leading people from A (the accepted), via B and C to D (the new claim). (A very short overview of the main points of rhetorical theory can be found in Vigsø 2018.)

THE PAST, THE PRESENT, AND THE FUTURE

Rhetoric distinguishes between three "genres" used when trying to persuade people to accept something: the judicial genre, which deals with what has happened and who is responsible; the deliberative genre, which deals with what we ought to do to make things better in the future, and the demonstrative genre, which deals with the present situation, trying to make the audience accept that something is the case. The three genres often work together, for instance when a politician from the opposition describes the present situation as horrible, blames the government for mistakes which have led to this, and proposes a solution for the future ("vote for me"). But even if these genres work together, they can also be used singularly for disinformation, thus paving the way for later arguments supported by the earlier claims.

One strategy regarding the past presented during the last decade or so, is the depiction of the Social Democratic Party as having a racist past, with forced sterilizations, the foundation of the institute for racial research, cooperation with Hitler, the refusal to accept Jewish refugees, etc. (Larsmo 2021). This has been spread by numerous sources within or associated with the Sweden Democrats, and all attempts at correcting this view have been vehemently attacked as propaganda. Internationally, the same rewriting of history has been observed in many countries, and in all cases, it becomes clear that rewriting history is a means to present argumentation about the present. In the case of Sweden, the blaming of the Social Democrats came as a counterattack against the recurring pointing out of the Sweden Democrats' background in the neo-Nazi movement. The goal was thus to diminish the blame put on the Sweden Democrats, to shift the blame to the Social Democrats, and to portray these as liars and hypocrites while diverting the focus from one party to the other. A similar strategy was shown by the Conservative Party's (Moderaternas) Secretary Sofia Arkelsten back in 2011, when she claimed that her party had been part of the introduction of general suffrage, when in fact the party had been opposed to this.

Disinformation about the present is often done through false claims about what is happening at some place in Sweden. To be fair, there is a grey zone between partisan interpretation of facts and blatant disinformation, with the difference being that the legitimate partisan interpretation can present the warrants behind the arguments in such a way as to discuss them, while disinformation builds on "accepted views", what rhetoric calls *doxa*: that which people in a certain group believe without being conscious about believing it. One such claim which gathered international attention was when President Donald Trump in 2017 referred to Sweden:

> *During another ferocious attack on the media on Saturday evening, US President Donald Trump cited a non-existent incident in Sweden, baffling many – not least Swedes. "You look at what's happening in Germany, you look at what's happening last night in Sweden. Sweden, who would believe this. Sweden. They took in large numbers. They're having problems like they never thought possible," the new US president told a crowd of supporters at a rally in Florida. (BBC 2017)*

The disinformation about an event (a terror attack, most likely) is here linked with a fact about the past (a huge number of immigrants), presenting an argument: what happened was due to the immigration. Trump clearly expected his public to know nothing about what happened in Sweden, but he used an accepted interpretation of facts (that the number of immigrants in Sweden was huge), together with an accepted truth in this group that immigration causes trouble, to make the claim about something happening plausible. This example also shows the intricate relations between claims and beliefs: using a lie to argue that what people believe is true, strengthens the belief that it is true as well as the belief that something happened.

Disinformation about the future mostly has to do with false predictions about the development to be expected (unless action is taken). Even here, there is a grey zone between pessimistic forecasts and deliberate disinformation. This can e.g. be seen in the Eurabia conspiracy theory, which claims that the rising number of Muslim immigrants in Europe is part of a globalist conspiracy, led by French and Arab powers, with the aim of islamizing Europe, thus weakening its existing culture and undermining its previous alliances with the United States and Israel. The claim is that there is a "great replacement" of peoples which eventually will lead to "ethnic Europeans" becoming a minority in Europe, thus paving the way for a covert Islamization.

This conspiracy theory is also an example of how disinformation argumentation can be construed: the fact that people from Muslim countries form a growing part of immigrants (the present) is used to make predictions about the future, which are interpreted as not only negative but as parts of a strategy from those having opposed stricter immigration laws in the past. The warrants are without any substantial proof, but they are inserted as "logical" steps uniting the past and the future.

FACTS AND FEELINGS

From a rhetorical point of view, persuasion is depending on the interaction between three different appeals. One is *logos*, which has to do with reason and logical relations such as cause-effect, before-after, etc. This is what is generally meant when people use the term "argue" as synonymous with presenting facts to support one's claim. But apart from this, there are also two emotion-based appeals, *ethos* and *pathos*. Ethos is the trust in the sender (as competent, benevolent, and a moral person) as the basis for accepting their claims. Pathos, on the other hand, is the feelings supported in the audience as support for the sender's claims ("think of the children", or "have you experienced this?") Numerous studies of propaganda have shown how the emotional appeals are the most common ones here, but the same is true of disinformation. Citing sources which the public hold in high regard (whether true or false) helps construct the claims as plausible, as do the awakening of emotions in the public. Outrage, hatred, envy, and other negative emotions are presupposed as existing in the audience and are intensified by the communication, which also emphasizes their acceptability as grounds for political stances (and actions).

The use of ethos is for instance evident in the rise of many populist politicians, who often transform fame within one area (business, entertainment) into general ethos and thus political capital. "You know me" becomes a support for claims which to the public is stronger than the opposition's logos-based argumentation and refutation of the claims. The same goes for the argument "if it feels right, it is right", putting subjective sentiment before objective proof. And once these "truths"

have been established, the use of facts to try and dissuade them is often in vain.

NON-PERSUASIVE STRATEGIES

The strategies mentioned above are concerned with contrafactual persuasion, that is trying to convince people about something the communicators themselves know to be false. But another strategy is not directed at convincing people that something is the case, but rather to create disbelief and confusion as to what is true. This strategy is also rhetorical, as it aims at influencing people's beliefs, but in a negative way. When the goal is not to make people accept something as true, but rather to doubt everything everybody is saying, the strategy is not to present plausible arguments but to make sure that there is a plethora of possible arguments and "facts" out there. The arguments should preferably be so different as to create a general state of disbelief and make people reject all arguments. One could say that the aim is the suspension of belief and the nivellation of all sources, regardless of their initial trustworthiness.

To obtain this effect, several approaches are helpful. First, the more the difference in arguments and claims, the higher the confusion. Secondly, the higher the number of sources, the stronger the feeling will be that none can be trusted. These two in combination lead to a proliferation of different user profiles on social media spreading all kinds of claims and arguments, often easily rejected or contrary to common sense. The very activity of some readers to try and counter the arguments in the comments is part of the strategy, as it supplies even more contradictory claims. So, in fact the attempts at proving the communicators wrong will in the end help them in their attempts at confusing the public.

WHO CAN BE PERSUADED? AND WHY?

As has been pointed out above, the "success" of disinformation is dependent on the audience, which must be open to the arguments presented, to find them plausible. But equally important is a distrust in other sources, which the disinformation contradicts. A common belief has thus been that high-trust societies like Sweden, where people in general believe that the authorities and one's fellow citizens are not trying to deceive them are immune to this. Societal trust is thought to work as an antidote to disinformation, but as this trust is lower in certain groups in society, these groups are also more vulnerable to disinformation. Rhetorically speaking, disinformation is never aimed at the population at large but rather targets those groups which are more prone to accepting claims deviating from what is considered the general doxa (the accepted truths). This also means that disinformation often is diversified to suit the profiles of different groups, rather than sticking to one single strategy. Often, one can see confusion or ridicule when scholars are confronted with what to them are clearly false claims, e.g. in relation to conspiracy theories, but these claims were never directed at society, but instead to those groups where each claim fits into established distrust and sentiments. Instead, the overall idea that there is a conspiracy by the elite against the people can function as a frame for several different false claims, each attracting its own public.

One example of how this works was the disinformation campaign about the Swedish social services removing children of Muslim families, not because of concern for the health of the children but as part of a forceful christianization of Muslims – described and analysed in a later chapter.

DISCUSSION

- What does the sender take for granted that the public believes? Are there any direct references to things "you already know", or cues for the public to agree (like a rhetorical question "Am I right?").
- What in the text is directed at reasoning (logos), and what is directed att feelings (ethos, pathos)?
- What is used as "evidence" in the attempt to influence people? Where does it come from?

ORLA VIGSØ, *PhD, Professor of Media and Communication Studies at the University of Gothenburg, Sweden. Previously professor of Rhetoric at Södertörn University. Writes about rhetoric, satire, scandals, and crises.*

REFERENCES

Aristotle (2020). *The Art of Rhetoric.* Translated by J.H. Freese. Cambridge: Harvard University Press.

BBC (2017). *Sweden to Trump: What happened last night?* Published 19 February 2017. https://www.bbc.com/news/world-us-canada-39020962

Larsmo, O. (2021). *Tio lektioner i svensk historie.* Stockholm: Kaunitz-Olsson.

McCroskey, J. C. (2006). *An Introduction to Rhetorical Communication.* 9th edition. Boston: Pearson.

Toulmin, S. E. (2020). *The Uses of Argument.* 2nd Edition. Cambridge: Cambridge University Press.

Vigsø, O. (2018). Rhetoric. In: Heath, Robert L. & Johansen, Winni (red.): *International Encyclopedia of Strategic Communication.* Hoboken: Wiley Blackwell.

14. CONSPIRACY THEORIES AS VECTORS OF FOREIGN INFORMATION MANIPULATION AND INTERFERENCE (FIMI)

ANDREAS ÖNNERFORS

SUMMARY

- Conspiracy theories as meaning-making narratives are powerful and persuasive due to a clear structure that can be replicated in many contexts
- People are susceptible to conspiracy theories since they cater to various needs to understand the nexus between inputs/cause and outputs/effects (causal relationships)
- In a malign information influence campaign during the 1980s, AIDS was portrayed as a consciously designed bioweapon used by the US against minorities and African people – this narrative was recycled during the COVID-19-pandemic
- Conspiracy theories about the maltreatment, abduction, abuse and killing of children evoke particularly strong emotions over time and in various contexts, for instance as precursors of the Swedish 'LVU-campaign'
- Taken together, conspiracy theories are ideally constructed to convey strategic messages of FIMI in the ongoing 'battle of perceptions'

Conspiracy theories are expressions of a particular way of thinking, creating meaning about the world and human existence. As a rule, conspiracy theories assume that nothing happens by chance, they discover patterns that show that everything is connected and discover dark actors who, with evil intentions behind our backs, have conspired in secret and plan to carry out actions that affect us negatively. For this conspiracy there is 'evidence' that is either withheld from us or covered up and manipulated by the real, hidden powers. The world and historical development are trapped in an eternal struggle between evil and good forces. Negative events are someone else's fault, someone who consequently also can be blamed. Not only that, but the culprits are also genuinely evil demons and if we do nothing, society as we know it will perish. The only salvation is the truth tellers and those who are on their side. They have seen through the evil plans of the conspirators and gather the victims for a final and decisive battle. Conspiracy narratives are communicated with images, text and sound that seek to express the dark drama unfolding before our eyes, how we are controlled, led and manipulated, entangled in the tentacles of the octopus, threatened by men in cloaks, controlled by the strings of the omnipotent puppeteer.

Over the last decade, conspiratorial world explanations have entered the mainstream of political communication both in domestic and international contexts and with various purposes: to justify warfare as in Ukraine, to claim that the COVID-19-pandemic was nothing but a global power grab or to incite political and terrorist violence (Christchurch 2019, storming of the Capitolium in 2021, racist riots in Britian in 2024). Undoubtedly, they have played a considerable role in fueling the dynamics between online and offline radicalization, but how and to what extent have conspiracy theories been vectors of Foreign Information Interference and Manipulation (FIMI)? Since FIMI seeks to "negatively impact values, procedures and political processes" (Hénin, 2023) in various domestic settings, conspiracy theories have been used to transport narratives undermining political authority, the rule of law, expert knowledge and independent media coverage and to stoke societal unrest. But why are humans susceptible to such narratives?

Social psychologist Jan-Willem van Proijen (2018) has proposed five critical ingredients for statements to qualify as conspiracy theories:

1. *Patterns.* Conspiracy theories explain events by creating nonrandom connections between actions, objects and people. Nothing happens through coincidence.
2. *Agency.* Events are caused on purpose by an intelligent and powerful actor. A carefully and intentionally designed plan is carried out in detail.
3. *Coalitions.* The conspiracy is always carried out by a coalition of multiple actors ('they') and directed against a target community ('us').
4. *Hostility.* The goals of the conspiracy are always evil, selfish and not in the public interest. The actions of the plotters are always guided by malicious intent.
5. *Continued secrecy.* The coalitions behind conspiracies are operating in secret and their actions cannot be proven until real and hard evidence is provided.

To van Proijen, "these five qualities distinguish belief in conspiracy theories from many other beliefs people may hold" (2018, 8). Apart from van Proijen's five qualities, a host of other definitions have been proposed to clarify the difference between conspiracy theories and other plausible statements about the world (Önnerfors 2021a and 2024). Additionally, it is important to highlight the focus of conspiracy theories on

6. *'Proof'.* There is so-called evidence to prove that the conspiracy is true.
7. *Dichotomies.* Conspiracy theories are guided by a black-and-white binary world view.
8. *Scapegoats.* The aim of conspiracy theories is to project (moral) guilt upon someone else.
9. *Demonization.* The actors behind conspiracies are frequently portrayed as (metaphysically) evil and are dehumanized.
10. *Apocalypticism.* Unless the conspiracy is defeated, society and humanity as we know them will collapse.
11. *Truth-tellers.* Despite the powerful manipulation we are exposed to, a selected few have seen through the conspiracy and can guide the blind crowd.
12. *Mediatizations.* Text, image and sound are deployed to communicate the sense of doom and urgency inherent in the conspiracy. Social media play a constitutive part in the dissemination of conspiratorial content.

These twelve ingredients of conspiracy theories form the basis for a narrative pattern, dramaturgical design or in the true sense of the word, a 'plot' with a clear division of characters such as perpetrators and victims, villains and heroes, traitors and renegades who to various degrees are exposed to or complicit in the evil master plan.

Psychologists Lantian, Wood and Gjoneska (2020) have suggested that humans are susceptible to conspiracy narratives since they cater to a range of cognitive, psychological and existential human needs. We want to understand causes of negative events, channel collectively shared anger and fear about them (in community with others) and grasp more metaphysical explanations of why 'bad things happen to good people', as it already was suggested in classical philosophy. This combination of epistemological, emotional and existential needs charges conspiracy narratives with a powerful salience since they provide with explanations of causality on all three levels. Yet it is important to stress that conspiracy theories cannot be proven since they establish fictional relations between causes and effects even if they generally are constructed around 'grains of truth'. Theories about conspiracies (in a legal sense: an agreement with the intention to commit a crime) on the other hand must provide objectively verifiable facts backed by solid evidence to establish criminal liability.

All this explains why conspiracy theories are a particularly productive tool in the toolbox of FIMI, drawing from the construction of powerful and divisive enemy images (Steiner and Önnerfors, 2024). Let us now look at one prominent example where a conspiracy narrative was consciously planted in the global media landscape: the attempt to pin the blame for the outbreak of the 1980s AIDS-pandemic on US biological warfare. It is not only an informative case as such but foreshadowed also other more recent conspiratorial narratives such as about the claimed 'lab leak' at the Wuhan Institute of Virology (releasing the Sars-CoV-2-virus) or the purported existence of secret American biolabs in Ukraine (Önnerfors, 2021a, p. 10; Robinson et al, 2022).

'OPERATION DENVER': THE NARRATIVE ABOUT THE INTENTIONAL DESIGN AND DISSEMINATION OF HIV

In July 1983, a letter to the editor of the Indian English-language newspaper *The Patriot* proclaimed that "AIDS may invade India" and that the "mystery disease [was] caused by US experiments" (*The Patriot,* 1983). It was purportedly submitted by someone pretending to be a well-known American scientist and anthropologist. Already in the first paragraph, accusations are made against the United States. It is stated that AIDS is believed to be a result of the Pentagon's experiments to develop new, dangerous biological weapons. Furthermore, it is claimed that the infection no longer can be controlled and that attempts are underway to spread it to countries susceptible to US pressure, pointing out Pakistan to become the next testing ground for these "experiments". The virus is described as a time bomb that invades the body and weakens the immune system to cause people to die of the flu. The second part of the article is about the secret development of biological weapons in the US military laboratory Fort Detrick in the United States. According to the letter, it was here AIDS, and other viral diseases had been created. They were tested on people from developing countries, drug addicts and homosexuals to cause intentional harm. In the last paragraph, "the unknown virus" is portrayed as a global threat to the world population.

Research agrees that the text in *The Patriot* was planted by the Soviet intelligence service KGB and that 'Operation Denver' was a coordinated campaign carried out in collaboration between East German, Czechoslovak and Bulgarian security services with the aim to amplify the narrative (Selvage and Nehring, 2014). Why the codename 'Denver' was chosen remains unclear, but Douglas Salvage, who extensively has researched the case, suggests it might have been a simple misunderstanding (mixing up 'Detrick' with 'Denver') or a reference to a popular 1980s TV-show ('Denver Clan'/Dynasty) in Western Germany (Kramer, 2020).

Russian magazine *Literaturnaja Gazeta* took up the story of AIDS in 1985 in an article based on the *The Patriot* and expanded it with a report on and pictures of Fort Detrick. The article was translated and distributed in several countries worldwide, including Sweden. The campaign was recycled a year later in Soviet media and an attempt was made to give it a scientific veneer by referring to East German doctor Jakob Segal's report *Aids – USA-home made evil [sic!]; NOT from AFRICA.* Segal speculated that the virus had been created by the US military at Fort Detrick.

'Operation Denver' supposes that a malevolent actor (the US) intentionally designed a virus in order to harm people across the globe through a secret cabal between a military laboratory, a federal agency and US allies. The program is designed, and the evil plan carried out in secrecy. 'Proof' is provided by insinuations only. The reader is unable to independently verify facts or evidence. Secret and harmful US agency fits into the construction of a bipolar world order, where blame for the AIDS-pandemic is projected upon demonized wirepullers, cementing existing enemy images. Thus, it could be integrated into larger narratives of cold-war-polarization and strategic antagonism, for instance by stoking tensions between Pakistan and India. The narrative disseminated in 'Operation Denver' is one of an imminent and existential threat to India (and mankind as a whole). The anonymous "scientist and anthropologist" referred to and even more the testimony of Segal serve as legitimate truthtellers exposing the malicious plot. When it comes to amplification through news media of the narrative, the most likely entirely faked 'letter to the editor' in an English-language outlet and later articles serve the purpose to create an impression of authenticity and of scientific validity through the Segal-report. Once inserted into the global flow of information, the story could be rewritten and repackaged.

The central claim of Segal's report, that the HIV-virus was not transmitted by zoonosis (animal to human) but that its features had been artificially manipulated, resurfaced as a defining feature of the 'lab-leak'-theory during the COVID-19-pandemic. The scientific and intelligence communities are still undecided concerning the origins of Sars-CoV-2 (Robertson, 2023). It is however interesting to observe that the unscientific dissemination and amplification of the 'lab leak'-theory served various unscientific purposes: mostly to place blame on either the Chinese communist state, biological research(ers), international research funding and collaboration or the pharmaceutical industry ('big pharma') as a whole, frequently portrayed as entangled deeply in a global cabal with the aim to introduce a world dictatorship (Önnerfors, 2021b; Önnerfors and Hamrud 2024). A more recent example of how central claims in 'Operation Denver' were recycled was when Russian foreign minister Lavrov accused the US to run secret labs for biological warfare in Ukraine (Ling, 2022). The story merits further attention since the claim of the existence of such labs was pushed by global right-wing social media networks and personalities and used as 'proof' of the dark motives of warfare in Ukraine.

THE PLACE OF CONSPIRACY THEORIES IN BATTLES FOR STRATEGIC NARRATIVE DOMINANCE

Naturally, there are many more examples of how conspiracy theories have been integrated into the playbook of FIMI:s tactics, techniques and procedures (TTP:s) for manipulation. The LVU-campaign, described in a separate chapter in this handbook, builds upon an age-old narrative on the abduction and abuse of vulnerable and innocent children, from accusations of blood libel to #pizzagate 2016 and thereafter, 'Save the Children' during the pandemic and support for the feature film 'Sound of Freedom' (2023). Antisemitic myths from the Middle Ages to modern times accused Jews of kidnapping and killing Christian children to use their blood for Passover Matzah flatbread. The #pizzagate conspiracy theory alleged that Democrats secretly convened at an Italian restaurant in Washington D.C. to abuse children in secret underground spaces. During the pandemic, conspiratorial anti-vaccine sentiments were disseminated under the cover of protecting children and the movie 'Sound of Freedom' portrays a lone, male and brave US-hero's fight to liberate children trafficked by criminal Latin American gangs.

Yet already almost a decade before the exceptional information influence campaign against Swedish social services unfolded and in an almost identical narrative, Norwegian counterparts were singled out in Russian state media in stories amplified across Eastern Europe aimed to demonstrate the moral decay and depravation of the West (Astapova et al., 2021). Children as victims of sinister crime evoke particularly strong emotional responses, fueling the violent street riots of the British far-right in summer of 2024 stoked by inauthentic accounts with Russian ties and transnational support in social media (Kivi, 2024). The ongoing 'battle of perceptions' (Zinzone and Cagnazzo, 2020) which characterizes the informational domain in increasingly hybrid war theatres across the globe will most likely exacerbate the use of conspiracy theories as vectors of FIMI and destabilize the rule of law as much as international law alike. To strengthen psychological resilience on both individual and societal levels requires the ability to decode and debunk their destructive narratives. Interventions will have to be designed, taking the epistemological, emotional and existential dimensions of conspiratorial meaning-making into account.

DISCUSSION

- Why do you believe that people are susceptible to conspiracy theories and where have you encountered them?
- Explain the narrative structure of a conspiracy theory! What are the ingredients needed to produce a sinister and persuasive story about malign plots?
- Compare the case of the AIDS-conspiracy theory ('Operation Denver') and what you know about various theories circulating about the origin of COVID-19 – do you see similarities or differences?
- In what sense are we particularly emotionally affected by narratives related to the abuse of children? What other examples can you come up with?
- After having read the chapter – what do you believe is necessary to build resilience against conspiracy theories in society? Is it needed or important? What kind of interventions would you prefer?

ANDREAS ÖNNERFORS, *PhD, is professor of intellectual history and project manager at Fojo Media Institute at Linnaeus university, where he trains journalists in combatting disinformation. His main interests over the last decades are ideological drivers of radicalization and conspiracy theories. Önnerfors has worked on the nexus between terrorist manifestos and conspiratorial imagination, but also on the history of ideas of conspiracy and its mediatization. In 2024, Önnerfors published a monograph in Swedish, Konspirationsteorier. Meningsskapande berättelser i historia och nutid, Conspiracy theories, Meaning-making narratives in history and present times and worked together with Kristian Steiner on the first textbook on Enemy images.*

REFERENCES

AIDS may invade India, *The Patriot,* 16.7.1983

Astapova, A. et al. (2021). *Conspiracy Theories and the Nordic Countries.* London: Routledge.

Hénin, N. (2023). *FIMI: towards a European redefinition of foreign interference.* Brussels: EU Disinfo Lab.

Kivi, E. (2024). How dubious website Channel3NOW fueled misinformation about Southport suspect in the U.K.. *Logically Facts,* 6 August 2024, URL: https://www.logicallyfacts.com/en/analysis/how-dubious-website-channel3now-fueled-misinformation-about-southport-suspect-in-the-u.k [accessed 8 August 2024].

Kramer, M. (2020) Lessons From Operation "Denver," the KGB's Massive AIDS Disinformation Campaign, *The MIT Press Reader,* https://thereader.mitpress.mit.edu/operation-denver-kgb-aids-disinformation-campaign/ [accessed 13 February 2025].

Lantian, A.; Wood, M. & Gjoneska, B. (2020). Personality traits, cognitive styles and worldviews associated with beliefs in conspiracy theories. In P. Knight & M. Butter (Eds.), *Routledge Handbook of Conspiracy Theories* (pp. 155–167). London: Routledge.

Ling, J. (2022). How U.S. Bioweapons in Ukraine Became Russia's New Big Lie. *Foreign Policy,* 10 March 2022, URL: https://foreignpolicy.com/2022/03/10/bioweapons-ukraine-russia-disinformation/ [accessed 8 August 2024].

Önnerfors, A. (2021a). *Conspiracy theories and COVID-19: The mechanisms behind a rapidly growing societal challenge.* Stockholm: Myndigheten för samhällsskydd och beredskap.

Önnerfors, A. (2021b). COVID-19: The Lab Leak Theory Makes a Comeback. *Fair Observer,* 2 September 2021, URL: https://www.fairobserver.com/coronavirus/andreas-onnerfors-covid-19-lab-leak-origins-report-far-right-conspiracies-news-14421/ [accessed 8 August 2024].

Önnerfors, A. (2024). *Konspirationsteorier. Meningsskapande berättelser i historia och nutid.* Lund: Nordic Academic Press.

Önnerfors, A. & Hamrud, A. (2024). Digitala viskningslekar: konspirationsteorier och det journalistiska berättandet som ett hot mot demokratin. In M. Hagevi (Ed.) *En ifrågasatt demokrati* (pp. 121–153). Göteborg: Makadam.

Proijen, J.-W. (2018). *The psychology of conspiracy theories.* London: Routledge.

Robertson, L. (2013). Still No Determination on COVID-19 Origin. *FactCheck.org,* 2 March 2023, URL: https://www.factcheck.org/2023/03/scicheck-still-no-determination-on-covid-19-origin/ [accessed 8 August 2024].

Robinson, O.; Sardarizadeh, S. & Horton, J. (2022) "Ukraine war: Fact-checking Russia's biological weapons claims", BBC, https://www.bbc.com/news/60711705 [accessed 13 February 2025].

Selvage, D. & Nehring, C. (2014). Die AIDS-Verschwörung. Das Ministerium für Staatssicherheit und die AIDS-Desinformationskampagne des KGB, *BF informiert,* 33/2014, URL: https://www.stasi-unterlagen-archiv.de/informationen-zur-stasi/publikationen/publikation/die-aids-verschwoerung/ [accessed 8 August 2024].

Steiner, K. & Önnerfors, A. (2024). *Enemy images: Emergence, Consequences and Counteraction.* London: Routledge.

Zinzone, F. and Cagnazzo, M. (2020). *The Art of War in the Post-Modern Era: The Battle of Perceptions.* Milano: Zinzone & Cagnazzo.

15. GENDER AND DISINFORMATION

ELSA HEDLING & MARTINA SMEDBERG

SUMMARY

- Gendered disinformation refers to the deliberate spread of harmful and false content targeting women, gender non-conforming individuals and marginalized groups, with the aim of advancing political agendas.
- It exploits polarization around gender equality and identity politics to weaken democratic systems and erode societal cohesion.
- Gendered disinformation leverages strategies such as amplifying misogyny and homophobia, spreading sexualized content, and orchestrating targeted harassment to increase societal divisions and silence public voices, especially those of individuals in leadership and decision-making roles.
- The widespread normalization of gendered hate speech, coupled with the ambiguous boundaries between free speech and hate speech, makes detecting, regulating, and attributing gendered disinformation particularly difficult.
- By undermining representation and pluralism, gendered disinformation threatens the foundations of democratic systems. Tackling this issue requires comprehensive strategies to mitigate its effects and uphold democratic integrity.

Gendered disinformation or "gender-based disinformation" are terms interchangeably used by governments, organisations and researchers to refer to the intersection of disinformation and gender. While definitions vary in the scope of their included elements, they focus on the social and political effects when women, gender non-conforming individuals and marginalised groups, including the LGBTQ+ communities, are disproportionately subjected to targeted harassment through the dissemination of hateful or deceitful content about them, often on social media platforms. Importantly, this form of disinformation differs from other forms of gender-based violence online in its use as a deliberate strategy for foreign information manipulation by threat actors (Hedling, 2024). Ultimately, gendered disinformation weaponizes gender-based violence online, which is both a domestic and an international problem stemming from the blurry lines between hate speech and free speech in the digital sphere (Howard, 2019). This exploitation of a democratic vulnerability is partly why gendered disinformation is difficult to trace and attribute.

Over the past decade, the growing recognition of information influence as a threat to democratic processes and electoral integrity has led to a more nuanced and sophisticated understanding of the array of disinformation tactics and their consequences. Given the perception that disinformation could weaken

democratic processes (most notably by influencing election outcomes), research efforts and government responses have increasingly focused on understanding how disinformation exploits existing political, social, and economic divisions (Hameleers, 2023). These efforts aim to uncover how disinformation deepens societal cleavages, intensifies polarization, and undermines social cohesion, posing a serious risk to essential democratic processes and effective governance. In exploiting societal divisions and vulnerabilities, disinformation campaigns often focus on identity-based differences, such as class and race, to fuel polarization and sow distrust for political advantage. This manipulation operates alongside broader relations of power in societies, such as white supremacy, heteronormativity, and patriarchy (Reddi et al., 2023). While attacks on gender thus belong to a broader set of social identities prone to polarization in Western societies, "gendered disinformation" alone has been lifted to a security concern of its own standing (U.S. Department of State, 2023). This chapter introduces the connection between gender and disinformation and explains why it is relevant in the broader framework of psychological defence.

The chapter opens by exploring the premise of gender (in)equality as a vulnerability that can be exploited, positioning it as integral to psychological defence debates. It then outlines common patterns and tactics in the operation of gendered disinformation. The chapter concludes with a concise summary.

GENDER (IN)EQUALITY AS DEMOCRATIC VULNERABILITY

Democracies are fundamentally characterized by pluralism and the processes of deliberation and negotiation among diverse groups that represent a variety of socio-political perspectives. Debates around gender norms and equality engage political options and meaningful policy alternatives available for public debate and collective decision-making. The true strength of liberal democracy lies in its ability to sustain these processes, enabling societies to resolve conflicts and address differences without resorting to violence or oppression. The main challenge democracies face in gendered disinformation is thus not the increase of gendered hate speech or the targeting of individuals, but rather the manipulation of the existing state of gender inequality, which threatens to undermine essential democratic processes (Hedling, 2024). When pluralism is manipulated to drive excessive polarization, it can escalate into crises that threaten the stability and very existence of democratic systems (McCoy et al., 2018). A political climate where women and gender minorities face perceived heightened risks of harassment, abuse, or discrimination, can discourage their participation and silence the voices of marginalized communities, leading to reduced representation and systemic inequality.

Gendered disinformation adds to the list of how gender norms and their policy outcomes become situated in the interests of and interactions between states in the international system. Disinformation as a communicative practice contributes to reinforcing or manipulating public deliberations of facts and values that are productive of state identity (Wells & Friedland, 2023). In pluralistic democracies, the distortion of representation and the manipulation of democratic processes, such as discouraging women from pursuing political office due to concerns about the impact on their private lives, pose an existential threat. Gendered disinformation, therefore, operates as a catalyst that prompts states to identify their 'gendered vulnerabilities' (Hedling, 2024). By recognizing gendered

disinformation as a security issue, states must also confront their vulnerabilities, including the persistence of gender inequality and attitudes like misogyny and homophobia. The discourse surrounding gendered disinformation therefore also compels states to reevaluate and potentially fortify their stance on gender-related issues within the broader context of their foreign and security policies.

State responses to gendered disinformation must depart from robust analyses of such campaigns' scope, extent and effects in distinct domestic contexts. While research has confirmed a harsher climate for political debate in the digital age and the disproportionate targeting of women and sexual and ethnic minorities (Guerin & Maharasingam-Shah, 2020), the effects thereof are still unclear. Recent research in the US context of high polarization, suggests for instance, that in terms of electorate support, descriptive representation of gender and race may reinforce voter alignment (Weissman, 2024). This means that disinformation driving polarizing attitudes about gender may even benefit candidates and, by extension, disrupt systemic inequalities. At the same time, a major risk of gendered disinformation is that it silences targeted individuals, impacting the criteria for selecting societal leaders in the first place. These impacts are particularly challenging to study, especially in progressive democracies, where evolving social norms and changes over time can further complicate such analyses. This also means that communicative efforts to counter gendered disinformation may themselves become politically sensitive. Despite these complexities, safeguarding conditions for gender representation in societies, including protecting the free speech of targeted communities, is vital to psychological defence.

TARGETING GENDER HOW?

Gendered disinformation includes several tactics that use disinformation in ways that intersect with gender norms or gender identities to deliberately mobilize political polarization and inter-group threat perception. Political polarization refers to degrees of increasing ideological divide and fragmentation within a society, which can be influenced by how individuals' social identities shape their in-group perspectives and the social categorization of other social groups (Baldassarri & Bearman, 2007). Through processes of social categorization, individuals form emotional and psychological bonds with social groups, which are central to their self-identity. These attachments intensify when uncertainty arises, especially alongside perceived threats from other groups. This strengthens positive emotions toward one's group and negative feelings toward others, leading to heightened perceptions of intergroup threats and greater social distancing between groups (Renström et al., 2023). These processes can be set in motion when gender-based group identities such as "women" or "transgender people" are targeted in disinformation tactics. These tactics can thereby exploit the prevalence of gender-based violence, especially in digital settings (through bots and anonymous accounts) and therefore serve as a "wedge strategy" in terms of threat actors' interests in sowing division and discord in a targeted society (Wigell, 2019). There are essentially four broad categories of tactics involved in this form of disinformation: 1) misogynistic narratives, 2) homophobic or transphobic narratives, 3) sexually suggestive or sexualized content, and 4) targeted harassment.

Misogynistic narratives in disinformation perpetuate hatred, contempt, or prejudice against women and girls as a group. These narratives are a form of sexism that reinforces the idea that women are inherently inferior to men, supporting and sustaining patriarchal systems. By negatively portraying women and girls by reference to their gender identity, such narratives tap into deep-seated biases

(Banet-Weiser, 2021). For example, when female political candidates run for office, narratives reproducing "political misogyny" often suggest that women are less capable than men in handling positions of trust, responsibility, and high stress (Dovi, 2024). These narratives often depict women as weaker, less intelligent, and overly emotional while also questioning their ability to balance family life with demanding careers. These portrayals do not always stem from explicit misogyny or hatred but contribute to sustaining patriarchal norms. Such biases are prevalent, particularly in news media, which often reinforces these stereotypes (Bligh et al., 2012). Threat actors can exploit these harmful portrayals to drive social categorization along gendered lines to deepen social divides or to advance specific political agendas by sidelining or silencing certain voices.

Homophobic or transphobic narratives in disinformation focus on demeaning and demonizing LGBTQ+ communities, often in contrast to the so-called "natural order", the "traditional family", or children's rights (Edenborg, 2022). This tactic exploits polarization around political commitment to gender norms and human rights. By mobilizing conservative values, these narratives sow discord through the construction of LGBTQ+ communities, a threat to culture, tradition, religion and the nation-state. This form of gendered disinformation specifically targets LGBTQ+ communities, for instance, by depicting a so-called "gay lobby" as instrumental in European imperialism or trans people as violent and oppressive towards the rest of the population (Strand et al., 2021). Homophobia and transphobia are also mobilized to humiliate highly visible individuals. For instance, French President Emmanuel Macron has been targeted by disinformation campaigns about his alleged "extramarital gay relationship" and strong support from "the rich gay lobby" (Vilmer, 2019). These narratives may contribute to constructing LGBTQ+ communities as threatening in society; on the other hand, the effects of the threat construction itself can be used for political gain in ways that serve polarization differently. For instance, populist and far-right parties have also used pro-gender norms as a benchmark against multiculturalism and have claimed nationalist ideas of gender equality and as "a Western boundary" (Towns et al., 2014; Agius & Edenborg, 2019; Scrinzi & Blee, 2023).

Sexually suggestive or sexualized content is used to discredit or shame individuals through strategies of sexualization to deter women from public roles. Such content is often image- or video-based and depends on image manipulation and/or false identity attribution on social media platforms, increasingly often in the form of fake porn. This form of gendered disinformation serves to challenge women's competence and track records by framing them as sexually degenerate or immoral (Wilfore, 2022; Esposito, 2023). A common strategy is to attribute false identities as sex workers or to circulate sexualized rumours of past career paths or dating habits.

Finally, targeted harassment refers to acts deliberately aimed to discourage, silence or scare individuals in ways that intersect with the targeted gender identities. For instance, tactics involve privacy violations, leakage or doxxing by malicious actors. Doxxing refers to one or several person(s) (doxxer/doxxers) seeking private or personal identifying information about another individual (subject/target) and widely distributing it through undesired online mass media channels without the consent of that person, who would be made vulnerable by mass media disclosure. While both women and men experience doxxing, research has found that women, especially minority group women, are more likely to have their private information posted online and receive greater amounts of unwanted, vitriolic attacks and messages (Eckert & Metzger-Riftkin, 2020).

These tactics, recognised as gendered disinformation, often occur in tandem to target women (and other minority groups) in positions of power and visibility, such as politicians, journalists, or activists. When pursued as foreign information influence, they often pass below the radar of detection and attribution because gendered slurs and hate speech are normalised and often less regulated than other forms of discrimination (Weston-Scheuber, 2012). This makes gendered disinformation an effective tool for foreign information manipulation as it seamlessly blends with domestic contestation of gender norms and gender equality.

CONCLUSIONS

In conclusion, gendered disinformation and its consequences extend beyond individual harassment by silencing voices and discouraging representatives of marginalized communities from fully participating in democratic processes. It threatens pluralism and democratic integrity by deliberately targeting women, gender-nonconforming individuals and marginalized groups, thereby fueling polarizing attitudes in democratic societies.

The absence of conclusive research findings in how gender and disinformation intersect across domestic political contexts challenges attempts to track and counter manipulative strategies in this category. Addressing the forces that manipulate the status of equality and drive polarizing attitudes about gender requires psychological defence strategies to identify, analyze, and counteract the damaging effects of gendered disinformation. These strategies are essential to safeguarding democratic integrity and ensuring diverse perspectives in political debates.

DISCUSSION

- How do tactics like misogynistic narratives and targeted harassment used in gendered disinformation erode democratic processes and institutions?
- What specific risks does gendered disinformation pose to individuals in leadership and decision-making roles?
- Why is gender inequality a significant vulnerability in democracies and how does it increase the susceptibility of societies to gendered disinformation?
- What makes distinguishing between free speech and hate speech particularly challenging when addressing gendered disinformation?

ELSA HEDLING *is an Associate Professor and Associate Senior Lecturer in European Studies at Lund University. She is also affiliated with the Psychological Defence Research Institute. Her current research focuses on the politics of hybrid threats, post-digital propaganda, and disinformation targeting social identities, with a particular emphasis on gendered disinformation and its effects on democratic systems.*

MARTINA SMEDBERG *serves as a Programme Manager at the Psychological Defence Research Institute at Lund University. She is also pursuing a PhD at the Department of Communication at Lund University. Before joining the academic sector, she worked for over two decades as a diplomat with the Swedish Foreign Service.*

REFERENCES

Agius, C., & Edenborg, E. (2019). Gendered Bordering Practices in Swedish and Russian Foreign and Security Policy. *Political Geography,* 71, 56–66.

Baldassarri, D. and Bearman, P. (2007). Dynamics of Political Polarization. *American Sociological Review,* 72 (5), 784–811.

Banet-Weiser, S. (2021). Misogyny and the Politics of Misinformation. In H. Tumber & S. Waisbord (Eds.), *The Routledge Companion to Media Disinformation and Populism* (pp. 211–221. London: Routledge.

Bennett, L. W., & Livingston, S. (2018). The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions. *European Journal of Communication,* 33(2), 122–139.

Bligh, M. C., Schlehofer, M.M., Casad, B.J., & Gaffney, A.M. (2012). Competent Enough, but Would You Vote for Her? Gender Stereotypes and Media Influences on Perceptions of Women Politicians. *Journal of Applied Social Psychology,* 42(3), 560–597.

Dovi, S. (2024). On Political Misogyny. *American Political Science Review,* 1–14.

Eckert, S., & Metzger-Riftkin, J. (2020). 'Doxxing'. In K. Ross, Scarcelli, C.M., Moorti, S., & Cardo, V. (Eds.) *The International Encyclopedia of Gender, Media, and Communication* (pp. 1-5). Wiley Online Library.

Edenborg, E. (2022). Disinformation and Gendered Boundary Making: Nordic Media Audiences Making Sense of "Swedish Decline". *Cooperation and Conflict,* 57(4), 496–515.

Esposito, E. (2023). Online Gendered and Sexualised Disinformation Against Women in Politics. In Maci, S.M., Demata, M., McGlashan, M. & Seargeant, P. (Eds.), *The Routledge Handbook of Discourse and Disinformation*, (pp. 292–305). Routledge.

Guerin, C., &Maharasingam-Shah, E. (2020). *Public Figures, Public Rage: Candidate Abuse on Social Media.* Institute for Strategic Dialogue.

Hameleers, M. (2023). Disinformation as a Context-Bound Phenomenon: Toward a Conceptual Clarification Integrating Actors, Intentions and Techniques of Creation and Dissemination. *Communication Theory,* 33(1), 1–10.

Hedling, E. (2024). 'Gendered Disinformation'. In Aggestam, K. & True, J. (Eds.), *Feminist Foreign Policy Analysis* (pp. 137–153) Bristol, England: Bristol University Press.

Howard, J. W. (2019). Free Speech and Hate Speech. *Annual Review of Political Science* 22(1), 93–109.

McCoy, J., Rahman, T., &Somer, M. (2018). Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Polities. *American Behavioral Scientist,* 62(1), 16–42.

Reddi, M., Kuo, R., & Kreiss, D. (2023). Identity Propaganda: Racial Narratives and Disinformation. *New Media & Society,* 25(8), 2201–2218.

Renström, E. A., Bäck, H., & Carroll, R. (2023). Threats, Emotions, and Affective Polarization. *Political Psychology,* 44(6), 1337–1366.

Scrinzi, F. (2023). *The Racialization of Sexism: Men, Women and Gender in the Populist Radical Right.* 1st ed. New York: Routledge.

Strand, C., Svensson, J., Blomeyer, R., & Sanz, M. (2021). *Disinformation Campaigns about LGBTI+ People in the EU and Foreign Influence.* European Parliament, Policy Department for External Relations.

Towns, A., Karlsson, E., & Eyre, J. (2014). The Equality Conundrum: Gender and Nation in the Ideology of the Sweden Democrats. *Party Politics,* 20(2), 237–247.

U.S. Department of State. (2023). Summary of the UK-U.S. Roundtable on Countering Gendered Disinformation at the 67th United Nations Commission on the Status of Women. https://www.state.gov/summary-of-the-uk-u-s-roundtable-on-countering-gendered-disinformation-at-the-67th-united-nations-commission-on-the-status-of-women/.

Vilmer, J-B, J. (2019). The "Macron Leaks" Operation: A Post-Mortem. Atlantic Council. https://www.atlanticcouncil.org/wp-content/uploads/2019/06/The_Macron_Leaks_Operation-A_Post-Mortem.pdf.

Weissman, A. (2024). Descriptive Representation in an Era of Polarization. *The Journal of Politics*, https://doi.org/10.1086/732961

Wells, C., & Friedland, L.A. (2023). Recognition crisis: Coming to terms with identity, attention and political communication in the twenty-first century. *Political Communication*, 40(6). 681–699.

Weston-Scheuber, K. (2012). Gender and the Prohibition of Hate Speech. *QUT Law & Justice Journal,* 12(2), 132–150.

Wigell, M. (2019). Hybrid Interference as a Wedge Strategy: A Theory of External Interference in Liberal Democracy. *International Affairs,* 95(2), 255–275.

Wilfore, K. (2022). Security, Misogyny, and Disinformation Undermining Women's Leadership. In Haciyakupoglu, G., & Wong, Y. (Eds.) *Gender and Security in Digital Space* (pp. 124–142). London, UK: Routledge.

16. AI AND DISINFORMATION

CARL HEATH

SUMMARY

- AI has a dual impact on disinformation, as it both creates and combats it, reshaping the media landscape with significant societal implications.
- The use of AI in journalism and media raises concerns about bias, transparency, and the erosion of public trust, posing ethical challenges.
- Deepfake technology, powered by AI, threatens public trust in media by creating hyper-realistic but deceptive content.
- AI enables the scaling and personalization of misinformation, facilitating tailored disinformation campaigns and amplifying false narratives.
- Education and regulation play a critical role in mitigating AI-driven disinformation and ensuring its ethical deployment.

INTRODUCTION

The latest advent of artificial intelligence (AI) has significantly transformed the characteristics and evolution of digital media ecosystems, influencing various facets such as content creation, distribution, user engagement, and the overall dynamics of media consumption (Bontridder and Poullet, 2021). This transformation is underpinned by the integration of AI technologies, which have reshaped how digital media operates and interacts with its audience (Ziakis and Vlachopoulou, 2023). The capabilities of AI technologies to generate, manipulate, and disseminate content have raised critical concerns regarding the integrity of information and the potential for malicious exploitation. This transformation is not merely a technological shift; it represents a fundamental change in how information is produced, shared, and consumed, with profound implications for democracy and societal trust (Bontridder and Poullet, 2021). One of the most problematic aspects of AI's impact on digital media is its role in the generation of disinformation. In this chapter, we will explore the intricate relationship between digitalization, AI, and disinformation, highlighting the challenges and opportunities presented by this convergence.

THE DIGITAL MEDIA ECOSYSTEM AND DISINFORMATION

The impact of AI on traditional journalism and information dissemination is profound and multifaceted, reshaping the landscape of media production, distribution, and consumption. As AI technologies continue to evolve, they introduce both opportunities and challenges that significantly alter journalistic practices and the broader media ecosystem. This transformation is characterized by enhanced efficiency, the emergence of automated journalism, and the necessity for ethical considerations in the deployment of AI tools (Ali and Hassoun, 2019). AI's integration into journalism has led to significant advancements in the creation and distribution of news content. For instance, AI algorithms can analyze vast amounts of data to generate news articles, thereby expediting the news production

process. This capability is particularly evident in the realm of automated journalism, where AI systems can produce reports on routine events, such as sports scores or financial summaries, with minimal human intervention (Kotenidis and Veglis, 2021). Furthermore, AI technologies enhance the personalization of news delivery, tailoring content to individual preferences and consumption patterns, which can lead to increased audience engagement (Tejedor and Vilà, 2021). However, the rise of AI in journalism is not without its challenges. The reliance on algorithms raises concerns about the quality and credibility of news content. Automated systems may inadvertently propagate biases present in their training data, leading to skewed narratives and misinformation (Ali and Hassoun, 2019). This issue is compounded by the lack of transparency in AI decision-making processes, which can obscure the origins of news stories and diminish public trust in media institutions (Sánchez, 2022). The development within the field of journalism has many similarities to the development of AI in the context of disinformation.

ARTIFICIAL INTELLIGENCE AND DISINFORMATION

AI's influence on the generation, scaling, and mitigation of disinformation is a defining characteristic of the current information ecosystem. Advancements in artificial intelligence, particularly generative models, have significantly impacted the production and dissemination of disinformation, for example in the following areas:

TEXT GENERATION USING LARGE LANGUAGE MODELS

The advent of large language models (LLMs) has revolutionized text generation, offering unprecedented capabilities in producing coherent and contextually relevant content. However, this technological advancement is accompanied by significant risks, particularly concerning the dissemination of misinformation and disinformation. The potential for misuse by malicious actors is a pressing concern, as these models can generate text that is often indistinguishable from human-written content, thereby facilitating the spread of false information (Weidinger et al., 2021; Kreps, McCain and Brundage, 2020).

DATA POISONING

LLMs are susceptible to data poisoning attacks, which can increase the risk of spreading misinformation. Data poisoning involves the deliberate manipulation of training datasets to induce specific behaviors in the trained models, often leading to the generation of misleading or harmful outputs. This vulnerability arises from the reliance of LLMs on vast amounts of data, frequently sourced from the internet, where the integrity of information cannot always be guaranteed (Bender et al., 2021; Weidinger et al., 2021). One of the primary mechanisms through which data poisoning occurs is the injection of malicious samples into the training data. For instance, attackers can insert carefully crafted text that alters the model's predictions or biases its outputs towards specific narratives.

DEEPFAKES AND SYNTHETIC MEDIA

The advent of AI deepfakes has ushered in a new era of misinformation, posing risks to public trust and the integrity of information. Deepfakes, which are hyper-realistic synthetic media generated through advanced machine learning techniques, can manipulate audio and visual content to create misleading narratives (Vaccari and Chadwick, 2020). This capability raises concerns regarding their use in disinformation campaigns, particularly in political contexts, where they can distort reality and influence public opinion. Deepfakes can be weaponized to create false

narratives that mislead audiences, particularly during critical events such as elections or political debates (Fallis, 2020). For instance, the ability to fabricate videos of public figures saying or doing things they have not actually done can undermine trust in media and institutions (Vaccari and Chadwick, 2020). The ease of creating deepfakes, facilitated by widely available tools, means that even individuals with minimal technical skills can produce convincing deepfakes, further complicating the landscape of digital misinformation. Moreover, the implications of deepfakes extend beyond politics; they can also affect personal reputations and social dynamics (Nait-Ali et al., 2023). The technology can be exploited for malicious purposes, such as revenge porn or identity theft, leading to severe consequences for individuals targeted by such attacks. The psychological impact of encountering deepfakes can lead to a general erosion of trust in visual media, as audiences become increasingly sceptical of the authenticity of what they see. This scepticism can have broader societal implications, as it may hinder constructive discourse and the ability to discern factual information from fabricated content (Vaccari and Chadwick, 2020).

AI-POWERED PERSONA CREATION AND CHARACTER HACKING

The proliferation of AI-based characters in applications such as character.ai and Convai raises concerns regarding the potential for spreading disinformation through these characters. These platforms utilize advanced AI technologies to create interactive characters capable of engaging users in conversation through text, audio, pictures and video. However, the very capabilities that make these applications appealing also pose risks related to the dissemination of false or misleading information (Sebastian, 2023). One of the primary risks associated with AI-driven characters is their potential to spread msinformation. The advancements in these AI services can lead to inaccuracies that may misinform users, either through deliberate disinformation campaigns or through the inherent limitations of the models used. This is particularly concerning in contexts where users may not critically evaluate the information provided by these AI characters, leading to the acceptance of false narratives as truth.

GENDERED DISINFORMATION

As developed in the chapter on gendered disinformation, the issue of gender differences in the prevalence and severity of deepfakes is a critical area of research, particularly given the unique vulnerabilities that different genders face in the context of digital media manipulation. Deepfakes, which are often used to create non-consensual pornography or to misrepresent individuals, disproportionately affect women, raising significant concerns regarding gender-based violence and the erosion of autonomy. Research indicates that a significant majority of deepfake content is pornographic in nature and targets women. A report highlighted that approximately 96% of deepfake videos are non-consensual and involve female subjects, underscoring the gendered nature of this phenomenon (Laffier and Rehman, 2023). This disproportionate targeting of women not only reflects societal attitudes towards gender but also exacerbates existing issues of gender-based violence in digital spaces. The creation and dissemination of deepfakes can lead to severe psychological and reputational harm for women, making them particularly vulnerable to online harassment and abuse (Laffier and Rehman, 2023).

AI'S ROLE IN SCALING DISINFORMATION

While AI's capabilities in creating disinformation are significant, its ability to amplify and scale false narratives is equally concerning, for example using the following methods:

BREAKING LANGUAGE BARRIERS

AI has significantly transformed the landscape of language translation, breaking down barriers that previously hindered communication across diverse linguistic groups. For the main part, this development holds great values, as communication between people becomes easier. However, at the same time, this technological advancement also raises substantial risks concerning the proliferation of misinformation. As AI systems become more adept at translating languages, the potential for disseminating inaccurate or misleading information on a global scale increases, posing challenges to information integrity and public trust (Bontridder and Poullet, 2021). AI technologies, particularly LLMs, can generate and disseminate vast amounts of text at unprecedented speeds. This capability can lead to what has been termed an "AI-driven infodemic," where misinformation proliferates alongside genuine information. The inherent duality of AI technologies means that while they offer significant opportunities for enhancing communication, they also present risks of manipulating human behaviour and eroding trust in information sources (Angelis et al., 2023).

AUTOMATION OF CONTENT CREATION AND DISTRIBUTION

AI-powered automation significantly transforms the speed and scale of information propagation across digital platforms. This transformation is primarily driven by the integration of intelligent automation technologies, which enhance the efficiency and effectiveness of information dissemination processes. Intelligent automation (IA) utilizes digital technologies to innovate service processes, thereby facilitating rapid information flow across various platforms (Kowalkowski, Wirtz and Ehret, 2023). The ability of AI to analyze vast amounts of data in real-time allows for quicker responses to emerging trends and issues, which is critical in today's fast-paced digital environment. Moreover, the automation of content generation and distribution has led to an exponential increase in the volume of information shared across digital platforms. For instance, automated systems can curate and disseminate content tailored to specific audiences, thereby enhancing engagement and interaction. This capability is particularly evident in social media platforms, where algorithms determine the visibility of content based on user interactions, leading to a more dynamic and responsive information ecosystem (Kotenidis and Veglis, 2021). The proliferation of automated tools also means that information can be propagated across multiple channels simultaneously, reaching a broader audience at unprecedented speeds. The implications of this rapid information propagation are profound.

PERSONALIZATION AND TARGETING OF DISINFORMATION

The increasing sophistication of AI technologies, especially generative models, enables the creation and targeting of personalized disinformation campaigns. AI can enhance the personalization of disinformation through its ability to analyze vast amounts of data to tailor content to specific audiences. AI algorithms can sift through user data on social media platforms to identify preferences, beliefs, and vulnerabilities, allowing malicious actors to craft messages that resonate deeply with targeted individuals or groups. This capability can make

disinformation campaigns more effective and harder to detect (Bontridder and Poullet, 2021). These tools can generate persuasive narratives that align with the targeted audience's existing biases, making the misinformation more palatable and believable (Kreps, McCain and Brundage, 2020). Moreover, the deployment of AI in social media manipulation has been linked to the emergence of new tactics in malign influence operations. The efficiency and cost-effectiveness of these AI-driven campaigns lower the barriers for entry for malicious actors, enabling a wider range of individuals and organizations to engage in disinformation activities (Hartmann and Giles, 2020). The ease of use of disinformation tools poses a significant challenge for regulatory frameworks and societal resilience against misinformation. AI's role in enhancing the specificity of targeting also extends to the psychological manipulation of individuals (Hartmann and Giles, 2020). By leveraging insights from behavioural science, AI can craft messages that exploit cognitive biases, thereby increasing the likelihood of acceptance among the target audience. This manipulation is not limited to political contexts. It can also affect public health messaging, consumer behaviour, and more, leading to potentially harmful outcomes (Sebastian, 2023).

AI IN COMBATING DISINFORMATION

Despite its role in enabling disinformation, artificial intelligence also offers powerful tools to counteract its negative effects, for example:

AI-DRIVEN FACT-CHECKING TOOLS AND DETECTION

The rapid proliferation of information and misinformation in today's media landscape has propelled AI-based fact-checking into a crucial area of research. Scholars are exploring how to automate fact-checking using natural language processing (NLP), machine learning (ML), and extensive databases to assess the veracity of claims. Research demonstrates that NLP and ML techniques can automate various aspects of the fact-checking process, including document retrieval, evidence extraction, stance detection, and claim verification (Guo, Schlichtkrull and Vlachos, 2022). A mixed-initiative approach combining human expertise with AI's efficiency has shown improvements in accuracy when participants were exposed to correct model predictions (Nguyen et al., 2018). However, overreliance on AI models can lead to incorrect judgments when the models make mistakes, emphasizing the need for transparency in AI systems. A balanced approach combining human expertise with AI capabilities appears to be the most promising path forward in the field of automated fact-checking (Zeng, Abumansour and Zubiaga, 2021).

CHALLENGES IN DETECTION OF SYNTHETIC MEDIA AND MANIPULATED CONTENT

The detection of synthetic media and manipulated content is a central concern in digital security, as advances in artificial intelligence have facilitated the production of sophisticated deepfake videos, images, and audio (Nait-Ali et al., 2023). Deepfake technology, primarily driven by Generative Adversarial Networks (GANs), can convincingly replicate real people and events, undermining trust in digital information and blurring the line between fact and fiction. This proliferation of synthetic media and the increasing availability of tools to create deepfakes have made it crucial to develop effective detection techniques. Deepfake technology leverages GANs, in which two neural networks compete against each other to create increasingly realistic content. These tools are highly effective in fabricating

media that can deceive even trained observers, leading to the widespread use of deepfakes in misinformation and disinformation campaigns. The effectiveness of deepfakes in manipulating perceptions underscores the importance of developing robust detection frameworks capable of identifying and mitigating their impact on public discourse (Vaccari and Chadwick, 2020).

Detection of manipulated content is often carried out through advanced AI methods, particularly machine learning and deep learning frameworks. These methods analyse patterns and anomalies that differentiate synthetic media from genuine content (Gupta et al., 2023). A valuable means as to do this is through multi-modal fusion techniques that combine various types of analysis—such as audio, visual, and metadata—to improve the detection of synthetic content. Techniques such as fusion methods that blend multiple detection strategies have been particularly effective, as they provide a more comprehensive approach to identifying inconsistencies within the manipulated media (Gupta et al., 2023).

The challenges in AI-based detection of synthetic media are multifaceted, reflecting both the technical complexities and evolving strategies used in synthetic content creation. One of the most significant challenges is the ongoing improvement of GANs and other machine learning techniques, which makes deepfakes increasingly difficult to detect. As these generative models continue to advance, the line between real and fake content narrows, challenging existing detection frameworks. A further complication is the limitation of current detection methods to adapt dynamically to new threats (Köbis, Doležalová and Soraperra, 2021).

Another critical challenge lies in the limitations of human perception. Individuals often overestimate their ability to detect manipulated media, thereby increasing their vulnerability to misinformation. The general public lacks the skills to recognize the subtle cues that distinguish authentic content from sophisticated forgeries, and even experts struggle without advanced tools. This perceptual overconfidence suggests that detection cannot solely rely on human vigilance (Köbis, Doležalová and Soraperra, 2021). There is a need for widespread educational initiatives that aim to enhance media literacy and teach individuals to be more critical of the information they encounter. Improving media literacy could empower people to better identify manipulated content, reducing their susceptibility to deepfake-based misinformation (Köbis, Doležalová and Soraperra, 2021).

Ultimately, the detection of synthetic media and the challenges it presents are deeply tied to the role of AI in analyzing and mitigating foreign influence operations. The technical sophistication required to create deepfakes is paralleled by the complexity needed to detect and counter them (Hartmann and Giles, 2020). The use of AI must, therefore, encompass not only the identification of manipulated content but also an understanding of how this content fits into a broader strategy of malign influence. Addressing these challenges requires a comprehensive approach that integrates technological tools, regulatory measures, and public education.

EMERGING TRENDS AND CHALLENGES

The section explores phenomena such as AI-driven malinformation, the “liar’s dividend,” data voids, and the impact of AI-generated content on search engines.

AI SLOP AS MALINFORMATION

The term “AI slop” refers to a category of artificial intelligence systems that generate

content, often without rigorous oversight or ethical considerations, leading to the proliferation of malinformation. Malinformation, distinct from misinformation and disinformation, involves the dissemination of information that is harmful or misleading, often leveraging AI's capabilities to create and spread content at an unprecedented scale (Bontridder and Poullet, 2021). This phenomenon is particularly concerning in the context of social media, where AI-generated content can rapidly influence public opinion and behaviour, often without adequate checks and balances in place (Kreps, McCain and Brundage, 2020).

AI slop systems utilize generative models, such as large language models (LLMs), to produce text, images, or videos that can mimic authentic human-generated content. A risk with AI slop is the rapid dissemination of AI-generated content can overwhelm traditional fact-checking mechanisms, making it increasingly difficult for users to discern credible information from fabricated narratives (Kreps, McCain and Brundage, 2020).

THE LIAR'S DIVIDEND

The concept of the liar's dividend refers to a phenomenon where the proliferation of misinformation, particularly through advanced technologies like deepfakes, allows individuals—especially public figures—to deny the authenticity of incriminating evidence by claiming it is fabricated. The term has gained traction in discussions surrounding the implications of artificial intelligence (AI) in the dissemination of disinformation (Schiff, Schiff and Bueno, 2024). As AI technologies advance, they enable the creation of increasingly convincing fake content, which can be weaponized to undermine trust in legitimate information sources (Kreps, McCain and Brundage, 2020). The liar's dividend thus creates a paradox where the existence of deepfakes and other forms of manipulated media can lead to a general scepticism towards all media, including authentic content, thereby facilitating the spread of disinformation and eroding public trust in factual reporting (Schiff, Schiff and Bueno, 2024). In the context of political discourse, the liar's dividend serves as a strategic tool for politicians who may face damaging revelations. By labelling such revelations as "fake news" or deepfakes, they can deflect accountability and maintain their support base. This tactic has been observed in various political contexts, where leaders exploit the ambiguity surrounding the authenticity of media to dismiss legitimate criticisms (Schiff, Schiff and Bueno, 2022). The implications of this strategy are profound, as it not only allows for the evasion of accountability but also contributes to a broader culture of mistrust in media institutions. As individuals become more aware of the potential for deepfakes, they may develop a heightened scepticism towards all media, leading to what has been termed "reality apathy," where the distinction between truth and falsehood becomes increasingly blurred (Ahmed, 2021). The impact of the liar's dividend extends beyond individual politicians to societal implications, particularly in democratic contexts. The erosion of trust in media can have deleterious effects on public discourse and democratic processes. When citizens are unable to discern credible information from manipulated content, the very foundation of informed decision-making is compromised (Pawelec, 2022). This is particularly concerning in electoral contexts, where misinformation can influence voter perceptions and behaviours (Schiff, Schiff and Bueno, 2022). The liar's dividend thus poses a significant challenge to the integrity of democratic institutions, as it allows for the manipulation of public opinion through the strategic use of disinformation.

DATA VOIDS

Data void is a concept that refers to the absence or scarcity of reliable, high-quality information on specific topics or issues, which can lead to the proliferation of misinformation and disinformation. This phenomenon is of relevance in the context of artificial intelligence and its role in the dissemination of information. Data voids create fertile ground for malicious actors to exploit gaps in knowledge, often resulting in the spread of misleading narratives that can have significant social, political, and economic consequences. The COVID-19 pandemic has highlighted the critical importance of understanding data voids, as misinformation surged in the absence of credible information sources during this global crisis (Chong et al., 2022). In the realm of AI, the detection and analysis of data voids are of importance for developing effective strategies to combat misinformation (Purnat et al., 2021). AI technologies could be employed to identify these voids by analyzing online conversations and social media interactions, thereby revealing where misinformation is most likely to flourish.

AI BASED CONTENT AND THE QUALITY OF SEARCH

The emergence of AI-generated content on the internet has raised significant concerns regarding the quality of search engine results, the integrity of information, and the proliferation of disinformation. As AI technologies advance, they enable the rapid production of content that can mimic human writing, leading to an oversaturation of low-quality or misleading information online. This phenomenon complicates the task of discerning credible sources and exacerbates the spread of false narratives, thereby undermining public trust in digital information ecosystems (Kreps, McCain and Brundage, 2020).

FUTURE DIRECTIONS

To address the challenges posed by digitalization, AI, and disinformation, a nuanced and multifaceted approach is required that goes beyond technological solutions, considering the dynamics of the current digital ecosystem. AI plays a dual role in disinformation. It can be weaponized by malicious actors for large-scale manipulation, as well as being a key driver in amplifying false information, largely due to engagement-driven business models (Bontridder and Poullet, 2021). Current regulatory and technological measures such as AI-powered fact-checking tools are necessary but insufficient, as they fail to address the fundamental incentive structures that prioritize user engagement, which often promotes misleading content.

Regulatory frameworks must be carefully crafted to address disinformation while protecting fundamental freedoms like freedom of speech. Current initiatives, for example by the EU, reflect a complex balance between moderation and the protection of free expression.

Public awareness and media literacy are critical in combating AI-driven disinformation. As digital information flows grow, the need for media- and information literacy becomes more pressing. Education efforts must focus on raising awareness of biases in content prioritization and equipping the public to critically evaluate information.

The ethical deployment of AI systems is essential to ensure that technology does not undermine societal values. Developers must prioritize transparency and accountability, incorporating mechanisms to prevent AI from being used to deceive. AI's role in mitigating disinformation should extend to identifying social media bots, detecting malicious microtargeting, and disrupting harmful amplification practices.

Social media platforms must assume a proactive role in mitigating disinformation, which involves more than mere detection and removal of content. Providing users control over algorithmic recommendations can help curb the spread of disinformation by reducing reliance on engagement-centric systems.

Finally, sustained investment in research and innovation is critical to staying ahead of evolving AI-driven disinformation tactics. Research should not only advance technical capabilities for detecting disinformation but also deepen our understanding of the psychological mechanisms behind its success. Collaboration between technologists, social scientists, and policymakers can foster more comprehensive solutions that address both the technological and human elements of digital disinformation. A holistic response is necessary—one that extends beyond leveraging AI to fix AI-related issues, addressing the underlying ecosystem dynamics that contribute to the problem, and fostering a more democratic, transparent, and accountable digital environment.

DISCUSSION

- How can society balance the benefits of AI in media production with the ethical challenges it poses, particularly regarding transparency and public trust?
- What role do educational initiatives play in mitigating the risks associated with AI-driven disinformation, and how can they be effectively implemented?
- To what extent should regulatory frameworks intervene in the deployment of AI technologies to prevent misuse while preserving innovation and freedom of expression?
- How can AI technologies themselves be utilized to combat the spread of disinformation, and what are the potential limitations of these approaches?

CARL HEATH *is a senior researcher and focus area leader for digital resilience at RISE – Research Institutes of Sweden, as well as a researcher at the University of Gothenburg and part of the Swedish Centre for Digital Innovation (SCDI). He works in applied research relating to society's digital transformation, particularly concerning issues related to democracy, digital resilience, AI and innovation management. Heath has served as a Special Counsel for the protection of democratic dialogue for the Swedish government, examining democracy in the digital age, as it relates to disinformation, propaganda and hate speech.*

REFERENCES

Ahmed, S. (2021). Navigating the maze: Deepfakes, cognitive ability, and social media news skepticism. *SAGE Publishing,* 25(5), 1108-1129. Retrieved from https://doi.org/10.1177/14614448211019198.

Ali, W. and Hassoun, M. (2019). Artificial Intelligence and Automated Journalism: Contemporary Challenges and New Opportunities. 5(1). Retrieved from https://doi.org/10.20431/2454-9479.0501004.

Angelis, D. L. et al. (2023). ChatGPT and the rise of large language models: The new AI-driven infodemic threat in public health. *Frontiers Media,* 11. Retrieved from https://doi.org/10.3389/fpubh.2023.1166120.

Bender, M. E. et al. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT* '21, March 3–10, 610-623

Bontridder, N. and Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Cambridge University Press*, 3. Retrieved from https://doi.org/10.1017/dap.2021.20.

Caramancion, M. K. (2023). News Verifiers Showdown: A Comparative Performance Evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in News Fact-Checking. *arXiv:*2306.171776

Chong, M. et al. (2022). Delving into Data Science Methods in Response to the COVID-19 Infodemic. *undefined(undefined).*

Fallis, D. (2020). The Epistemic Threat of Deepfakes. *Springer Nature (Netherlands),* 34(4), 623-643. Retrieved from https://doi.org/10.1007/s13347-020-00419-2.

Gregory, S. (2021). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. *SAGE Publishing*, 23(3), 708-729. Retrieved from https://doi.org/10.1177/14648849211060644.

Guo, Z., Schlichtkrull, M. and Vlachos, A. (2022). A Survey on Automated Fact-Checking. *Association for Computational Linguistics,* 10, 178-206. Retrieved from https://doi.org/10.1162/tacl_a_00454.

Hanselowski, A. et al. (2019). A Richly Annotated Corpus for Different Tasks in Automated Fact-Checking. Retrieved from https://doi.org/10.18653/v1/k19-1046.

Hartmann, K. and Giles, K. (2020). The Next Generation of Cyber-Enabled Information Warfare. Retrieved from https://doi.org/10.23919/cycon49761.2020.9131716.

Kotenidis, E. and Veglis, A. (2021). Algorithmic Journalism—Current Applications and Future Perspectives," *Multidisciplinary Digital Publishing Institute*, 2(2), 244-257. Retrieved from https://doi.org/10.3390/journalmedia2020014.

Kowalkowski, C., Wirtz, J. and Ehret, M. (2023). Digital service innovation in B2B markets. *Emerald Publishing Limited*, 35(2), 280-305. Retrieved from https://doi.org/10.1108/josm-12-2022-0403.

Kreps, S., McCain, M. R. and Brundage, M. (2020). All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Cambridge University Press,* 9(1), 104-117. Retrieved from https://doi.org/10.1017/xps.2020.37.

Köbis, N., Doležalová, B. and Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *Cell Press,* 24(11), 103364. Retrieved from https://doi.org/10.1016/j.isci.2021.103364.

Laffier, J. and Rehman, A. (2023). Deepfakes and Harm to Women," *Journal of Digital Law & Legal Ethics,* 3(1), 1-21. Retrieved from https://doi.org/10.51357/jdll.v3i1.218.

Miller, S. et al. (2022). Integrating truth bias and elaboration likelihood to understand how political polarisation impacts disinformation engagement on social media. *Wiley.* Retrieved from https://doi.org/10.1111/isj.12418.

Nait-Ali, A. et al. (2023). Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions," *Multidisciplinary Digital Publishing Institute,* 12(10), 216-216. Retrieved from https://doi.org/10.3390/computers12100216.

Nguyen, T. A. et al. (2018). *Believe it or not: Designing a Human-AI Partnership for Mixed-Initiative Fact-Checking.*

Pawelec, M. (2022). Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions.

Purnat, D. T. et al. (2021). Infodemic Signal Detection During the COVID-19 Pandemic: Development of a Methodology for Identifying Potential Information Voids in Online Conversations. *Journal of Medical Internet Research,* 1(1). Retrieved from https://doi.org/10.2196/30971.

Rossetti, M. and Zaman, T. (2023). *Bots, disinformation, and the first impeachment of U.S. President Donald Trump.*

Sánchez, N. A. (2022). Addressing the Impact of Artificial Intelligence on Journalism: The perception of experts, journalists and academics," *University of Navarre,* 35(3), 105-121. Retrieved from https://doi.org/10.15581/003.35.3.105-121.

Schiff, J. K., Schiff, D. and Bueno, N. (2022). The Liar's Dividend: Can Politicians Use Deepfakes and Fake News to Evade Accountability?" Retrieved from https://doi.org/10.31235/osf.io/q6mwn.

Schiff, J. K., Schiff, D. and Bueno, S. N. (2024). The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?" *Cambridge University Press*, 1-20. Retrieved from https://doi.org/10.1017/s0003055423001454.

Sebastian, G. (2023). Exploring Ethical Implications of ChatGPT and Other AI Chatbots and Regulation of Disinformation Propagation. *RELX Group (Netherlands).* Retrieved from https://doi.org/10.2139/ssrn.4461801.

Shao, C. et al. (2018). The spread of low-credibility content by social bots. *Nature Portfolio,* 9(1). Retrieved from https://doi.org/10.1038/s41467-018-06930-7.

Školkay, A. and Filin, J. (2019). A Comparison of Fake News Detecting and Fact-Checking AI-Based Solutions. *20(4)*, 365-383. Retrieved from https://doi.org/10.33077/uw.24511617.ms.2019.4.187.

Tejedor, S. and Vilà, P. (2021). Exo Journalism: A Conceptual Approach to a Hybrid Formula between Journalism and Artificial Intelligence. *Multidisciplinary Digital Publishing Institute,* 2(4), 830-840. Retrieved from https://doi.org/10.3390/journalmedia2040048.

Vaccari, C. and Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News," *SAGE Publishing,* 6(1), 205630512090340. Retrieved from https://doi.org/10.1177/2056305120903408.

Wallace, E. et al. (2021). *Concealed Data Poisoning Attacks on NLP Models.* Retrieved from https://doi.org/10.18653/v1/2021.naacl-main.13.

Weidinger, L. et al. (2021). Ethical and social risks of harm from Language Models," *Cornell University.* Retrieved from https://doi.org/10.48550/arxiv.2112.04359.

Yadlin-Segal, A. and Oppenheim, L. Y. (2020). Whose dystopia is it anyway? Deepfakes and social media regulation. *SAGE Publishing,* 27(1), 36-51. Retrieved from https://doi.org/10.1177/1354856520923963.

Zeng, X., Abumansour, S. A. and Zubiaga, A. (2021). *Automated Fact-Checking: A Survey.*

Ziakis, C. and Vlachopoulou, M. (2023). Artificial Intelligence in Digital Marketing: Insights from a Comprehensive Review. *Multidisciplinary Digital Publishing Institute,* 14(12), 664-664. Retrieved from https://doi.org/10.3390/info14120664.

III. CASE STUDIES IN INFORMATION INFLUENCE

This section dives deeper into examples of malign information influence campaigns. The first three chapters focus on China and Russia, detailing some of the main trends in their information influence activities. Following this overview, the remaining chapters explore key issues including the outsourcing of government-led influence campaigns, election interference, the LVU (social services) campaign targeting Sweden, the role of arts and culture, and the role of video games. Together, these case studies demonstrate the breadth of malign information influence campaigns and offer concrete examples of how campaigns are constructed and implemented.

17. "TELLING CHINA'S STORY WELL": THE CHINESE COMMUNIST PARTY'S EFFORTS TO INFLUENCE THE GLOBAL INFORMATION SPHERE

BJÖRN JERDÉN, PERRY JOHANSSON VIG, ERIKA STAFFAS EDSTRÖM AND ALEXIS VON SYDOW

SUMMARY

- Chinese information operations are rooted in the People's Republic's Leninist party-state structure, which prioritizes propaganda and censorship to maintain its power and promote the party's goals, notably to project China's global rise as inevitable and beneficial to all.
- The Chinese Communist Party integrates domestic propaganda with foreign information operations, leveraging state-owned media, embassies and online platforms, as well as covert methods such as front organizations and disinformation campaigns.
- These operations target both global audiences and the ethnic Chinese diaspora, using controlled media, content-sharing agreements and state-sponsored narratives to influence perceptions and discourage dissent.
- China's military doctrine incorporates information warfare, which involves cyber operations, psychological tactics and influence campaigns, while its approach increasingly mirrors Russian strategies of amplifying societal divisions and sometimes interfering in foreign elections.
- While China's media influence is significant, particularly in the Global South, its global narrative efforts face resistance in high-income democracies. This demonstrates the need for enhanced media literacy, independent journalism and transparency to counteract its strategies.

To analyse Chinese information operations effectively, it is crucial to understand the nature of the People's Republic of China (PRC) as a Leninist party-state where all institutions that wield official power prioritize loyalty to the Chinese Communist Party (CCP). Since the founding of the PRC in 1949, its population has been subject to an extensive and efficient regime of censorship and propaganda.

In China, information manipulation and propaganda are primarily conducted to safeguard the party's survival, maintain its grip on power and enhance its status. The foreign-directed information operations discussed in this chapter are closely linked to internal propaganda, in terms of the content, methods and institutions involved. Like its internal propaganda, many of the information operations performed by the Chinese party-state are focused on protecting the party. Common themes in its messaging include highlighting the successes of China's socialist development and depicting the CCP's aim of incorporating Taiwan into the PRC as reasonable and justified.

In addition, the regime aims to project the idea that China's rise as a global power

is inevitable, non-threatening and beneficial to all. This objective has gained in importance with the rapid increase in China's power and growing global recognition of its influence. The campaign to "tell China's story well", launched by the country's most senior leader, Xi Jinping, in 2013, aims to mobilize the party-state around a proactive foreign propaganda agenda (CMP Staff, 2021).

However, China's foreign propaganda aims extend even further. The regime's ideology is fundamentally at odds with the liberal democratic values that have heavily influenced the norms and institutions of the current international order. At the same time, the Chinese regime believes that the world order is undergoing a fundamental shift, and that US-led western dominance will soon be a thing of the past, positioning China to take a central role in the emerging global order. To mitigate threats to the regime and legitimize China's ascent to superpower status, the normative foundation of the international order must be reshaped in such a way that it no longer favours liberal democratic values. Consequently, Chinese information operations ambitiously aim to define or redefine the global discourse on politics, human rights and international relations. In terms of messaging, this translates into support for China's various global policy initiatives, opposition to "US-led hegemony" and the construction of a "fairer" international order that accommodates a diversity of political systems and civilizations.

ACTORS AND METHODS IN CHINA'S INFORMATION OPERATIONS

The success of China's information operations relies heavily on a tightly regulated censorship regime, but also on an intricate and multifaceted political structure, in which all the actors involved are accountable to the CCP. The *Central Propaganda Department* of the CCP – under the direct supervision of the *Politburo*, the top policymaking body of the Party – is the primary institution responsible for propaganda (Harold et al., 2021). It plays a central role in crafting and disseminating the narratives that support the party's goals and policies. Official authorities, such as the *State Council Information Office* and the *Ministry of Foreign Affairs*, as well as embassies in foreign countries regularly use dedicated public outreach functions to communicate China's official position on matters within their area of responsibility to international audiences (d'Hooghe, 2015). In recent years, these authorities have been increasingly active in foreign media, complementing party-state media channels to influence local discourses on matters of interest. Since 2019, an increasing number of diplomats, officials and official media outlets have become active on western social media, in an attempt to shape the narrative on contested issues (Brandt & Schafer, 2020). Public diplomacy has shown a tendency towards increasingly combative and aggressive rhetoric in recent years, at least on prioritized political issues such as Taiwan. This trend is commonly referred to as "wolf-warrior" diplomacy (Dai & Luqiu, 2022).

The Central Propaganda Department controls all traditional media through the *State Administration of Press, Publication, Radio, Film and Television,* which reports to it. China has invested heavily in party- and state-owned media, resulting in extensive print, digital and social media activities in at least 12 languages for international propaganda purposes. The key official media outlets for foreign audiences are the China Global Television Network (CGTN), China Daily, China Radio International (CRI), Xinhua and the China News Service. Xinhua is the PRC's official state news agency. It maintains over 180 bureaus in more than 140 countries and regions.

Alongside official channels disseminating official propaganda, China also uses

covert methods in its information operations. Front organizations, support to foreign media and online disinformation are used to influence global discourse in discreet and sometimes deceptive ways. What these operations have in common is that non-CCP actors, such as international affairs commentators, academics, think tanks and opinion leaders, are used to serve as conduits for and amplifiers of the party's policies and perspectives, as well as for censorship and information manipulation. This is the modus operandi of the *United Front Work Department*, which coordinates efforts to align groups, both within and outside of China, with the interests of the CCP (Beauchamp-Mustafaga & Chase, 2019).

This practice extends to foreign "fellow travellers" who support these narratives, such as politicians, interest groups and civil society organizations. China's primary intelligence service, the *Ministry of State Security*, integrates intelligence collection with influence campaigns, and has established and overseen several think tanks and front organizations that seek to shape foreign perceptions of China (Joske, 2022). In addition, foreign influencers paid by China (Ryan et al., 2023) and fake websites (Fittarelli, 2024) play roles in Chinese information operations.

The ethnic Chinese diaspora, which by some definitions numbers over 60 million people worldwide, is of special importance to China's foreign information operations. The CCP regards the diaspora as an integral part of the Chinese nation, and actively seeks to align it with the party's goals and to prevent the emergence of anti-CCP political groups within it (Almén, 2020; To, 2014). For example, there are in many countries diaspora associations that support "the promotion of peaceful reunification", which endorse the CCP's view on the future incorporation of Taiwan into the PRC (Dotson, 2019). Foreign Chinese-language media play a central role in disseminating information to the diaspora. CCP influence over such media is established through financial support, journalistic guidelines and content-sharing agreements with Chinese party-state media. These efforts have resulted in widespread support for the CCP among diaspora media worldwide. In countries such as Australia, New Zealand and Sweden, almost all Chinese-language news media are loyal to the CCP, while regime-critical media have been marginalized (Kurlantzick, 2022; Staffas Edström, 2023a). Content sharing is also used to amplify the CCP's presence in established or traditional media outlets in foreign countries. Although successful in large parts of the Global South, this strategy faces more challenges in western countries, where such arrangements often pose reputational risks for the media outlets involved (Charon & Jeangène Vilmer, 2021).

China's domestic internet is heavily regulated. The *Cyberspace Administration of China* (CAC), China's national internet regulator, employs millions of people to monitor and censor online content (Fedasiuk, 2021). It is also involved in the regime's attempts to rewrite global internet standards to align them with its censorship and propaganda needs (Attrill & Fritz, 2021). CAC regulates Chinese social media platforms such as Wechat and Weibo, which play a significant role in the CCP's overseas public opinion strategy, particularly within overseas Chinese communities where these platforms are widely used. This contributes to self-censorship in the diaspora (Yang, 2021). The global success of the social media platform Tiktok, which is owned by the de facto Chinese company Bytedance, has brought discussion to the fore of the role of private sector companies in spreading CCP narratives and disinformation through platform content overseas (Staffas Edström, 2023b).

China's overseas online information operations prominently feature deception tactics.

Content farms produce significant amounts of low-quality and/or false information, while networks of fake accounts generate an impression of genuine engagement and launder information from state channels (Charon & Jeangène Vilmer, 2021).

Information operations are also an integral part of Chinese military doctrine (Mattis, 2018). The *People's Liberation Army* maintains departments dedicated to various related activities such as cyber operations, psychological operations and social media (Harold et al., 2021). A well-documented example is Base 311, a unit which appears to report to the *PLA Strategic Support Force*, which operates media companies and conducts influence operations targeted at Taiwan (Charon & Jeangène Vilmer, 2021).

The increasingly aggressive disinformation tactics that have been observed in this context are often described as reflecting a "Russification" of Chinese information operations, where the emphasis is on amplifying societal divisions rather than positive messaging about China. Convergence between Chinese and Russian narratives on social media has also been observed, as well as cooperative media activities between the two countries (Charon & Jeangène Vilmer, 2021). Like Russia, China has also engaged in online operations as a form of election interference, notably in Taiwan (Wilson, 2022; von Sydow, 2024).

ASSESSING THE EFFECTS OF CHINA'S INFORMATION OPERATIONS

As outlined above, China's foreign information operations are driven by two primary objectives. The first, to safeguard the party and the state, has been a notable success. By establishing a controlled media and internet environment, and exerting influence over Chinese-language press and media internationally, China has effectively regulated the information accessible to its extensive diaspora on both China and global affairs (To, 2014; Hamilton & Ohlberg, 2020). The second major objective, however, which involves promoting China's narratives to the international community, has yielded more varied outcomes.

Favourable views of China have been in decline in many high-income countries since 2015. In general, people in the "Global South" tend to have a more positive view of China, though negative perceptions have risen in some countries and regions in recent years (Asia Society, 2025; Pew Research Center, 2023). However, measures of China's popularity with the global public do not fully capture the effectiveness of China's information operations. Self-censorship or belief of disinformation campaigns, even among a small minority of people, can be equally significant results of such operations, albeit much harder to quantify (Cook, 2023).

A 2022 report by the US think tank Freedom House rated the intensity of Beijing's media influence efforts as "high" or "very high" in 16 of the 30 countries examined (Cook, 2022). The study analysed variables such as content sharing agreements with Chinese state-owned media, disinformation campaigns and reporting on topics that the Chinese government deems sensitive. The study also highlighted a significant disparity in the ability of democracies to counter CCP influence. Only half the countries were assessed as "resilient", while the other half were deemed "vulnerable". Other regional studies mapping CCP influence indicate high levels of CCP media interference, especially in the Global South (Dubow et al., 2022; Thibaut et al., 2022). The "China Index" data base, hosted by the Taiwanese organization Doublethink Lab (2024), provides details of the vehicles for and patterns of this influence in close to 100 countries.

TACKLING HARMFUL CHINESE INTERFERENCE IN THE GLOBAL INFORMATION SPHERE

Through its information operations, the CCP aims to reshape the global discourse on politics, human rights and international relations. From a democracy and human rights perspective, its methods raise significant concerns. The CCP's domestic propaganda relies on strict censorship to establish information control within China. At the same time, the party exploits the inherent openness of democratic societies to information operations. Increasingly, it is also exporting its censorship regime abroad, resulting, among other things, in a diminishing number of CCP-critical Chinese-language media outlets and increased efforts to confine the Chinese diaspora within a controlled social media and news environment dominated by the Party's narratives. Understanding the CCP's strategies and tools is essential in order to create effective policies to counter information manipulation efforts and reinforce media ecosystems at the local, national and global levels. To achieve increased resilience, efforts should involve enhancing media literacy, robust support for independent journalism, transparency concerning media ownership and investments, and the dissemination of best practices among allies and beyond institutional silos.

DISCUSSION

- How does the structure of the CCP as a Leninist party-state influence its approach to both domestic and international information operations?
- Which strategies and methods does the CCP employ to shape global narratives and influence international perceptions of China?
- Do China's information operations challenge liberal and democratic values?
- What challenges do democracies face in countering the CCP's information operations and which strategies can be employed to enhance resilience?

BJÖRN JERDÉN *is Director of the Swedish National China Centre at the Swedish Institute of International Affairs (UI). He has a PhD in Political Science from Stockholm University and a master's degree in international relations from Malmö University. He has been a visiting fellow at National Chengchi University, National Taiwan University, National Chengkung University and Harvard University. Jerdén has studied Chinese in Shanghai, Lund and Taipei.*

PERRY JOHANSSON VIG *is a Stockholm University associate professor and doctor of philosophy. His research has mainly dealt with China's actual and imaginary relations with the West. Still at the intersection between culture, politics and psychology, he is now finishing two major projects on communist influence campaigns against Sweden during the Cold War. Johansson has worked at some of the world's leading universities and lived for 15 years in East Asia.*

ERIKA STAFFAS EDSTRÖM *is a former analyst at the Swedish National China Centre at UI. She completed the Swedish Foreign Ministry's Diplomatic Training Programme in 2024 and is currently working at the Embassy of Sweden in Hanoi as Head of Section for Political Affairs and Public Diplomacy. She has previously spent more than seven years in China and worked at the Embassy of Sweden in Beijing and the Swedish Chamber of Commerce in China. Staffas Edström has a degree in Politics and Economics from Lund University and has also studied at the Beijing Language and Culture University and Yunnan Normal University.*

ALEXIS VON SYDOW *is an analyst at the Swedish National China Centre at UI. He has a master's degree in Chinese from Stockholm University and has studied political science at the Swedish Defence University. He has a background as a translator and lived in China for eight years. His areas of expertise include China's foreign policy as well as issues of security and influence.*

REFERENCES

Almén, O. (2020). *The Chinese Communist Party and the Diaspora.* FOI.

Asia Society. (2025). Global Public Opinion on China. https://asiasociety.org/policy-institute/global-public-opinion-china/about.

Attrill, N. & Fritz, A. (2021). *China's cyber vision How the Cyberspace Administration of China is building a new consensus on global internet governance.* ASPI.

Beauchamp-Mustafaga, N. & Chase, M. S. (2019). *Borrowing a Boat Out to Sea: The Chinese Military's Use of Social Media for Influence Operations.* Johns Hopkins University School of Advanced International Studies.

Brandt, J. & Schafer, B. (2020). *Five Things to Know About Beijing's Disinformation Approach.* GMF. https://securingdemocracy.gmfus.org/five-things-to-know-about-beijings-disinformation-approach/.

Charon, P. & Jeangène Vilmer, J. (2021). *Chinese Influence Operations: A Machiavellian Moment.* IRSEM.

CMP Staff (2021). *The CMP Dictionary – Telling China's Story Well.* China Media Project. https://chinamediaproject.org/the_ccp_dictionary/telling-chinas-story-well/

Cook, S. (2022). *Beijing's Global Media Influence 2022.* Freedom House. https://freedomhouse.org/report/beijing-global-media-influence/2022/authoritarian-expansion-power-democratic-resilience.

Cook, S. (2023). *Countering China's Malign Influence Operations in the United States.* Freedom House. https://www.intelligence.senate.gov/sites/default/files/documents/os-scook-092723.pdf.

Dai, Y. & Luqiu, L. R. (2022). Wolf Warriors and Diplomacy in the New Era. *China Review*, 22(2), 253–283. https://www.jstor.org/stable/10.2307/48671506

Dotson, J. (2019). *The United Front Work Department Goes Global: The Worldwide Expansion of the Council for the Promotion of the Peaceful Reunification of China.* The Jamestown Foundation. https://jamestown.org/program/the-united-front-work-department-goes-global-the-worldwide-expansion-of-the-council-for-the-promotion-of-the-peaceful-reunification-of-china/

Doublethink Lab. (2024). China Index. https://china-index.io/.

Dubow, B., Greene, S. & Rzegocki, S. J. (2022). *Tracking Chinese Online Influence in Central and Eastern Europe.* CEPA. https://cepa.org/comprehensive-reports/tracking-chinese-online-influence-in-central-and-eastern-europe/.

Fedasiuk, R. (2021). *A Different Kind of Army: The Militarization of China's Internet Trolls.* The Jamestown Foundation. https://jamestown.org/program/a-different-kind-of-army-the-militarization-of-chinas-internet-trolls/

Fittarelli, A. (2024). *Paperwall: Chinese Websites Posing as Local News Outlets Target Global Audiences with Pro-Beijing Content.* The Citizen Lab. https://citizenlab.ca/2024/02/paperwall-chinese-websites-posing-as-local-news-outlets-with-pro-beijing-content/

Hamilton, C. & Ohlberg, M. (2020). *Hidden Hand: Exposing how the Chinese Communist Party is Reshaping the World.* London: OneWorld Publications.

Harold, S. W. et al. (2021). *Chinese Disinformation Efforts on Social Media.* RAND.

d'Hooghe, I. (2015). *China's Public Diplomacy.* Brill.

Joske, A. (2022). *Spies and Lies: How China's Greatest Covert Operations Fooled the World.* Hardie Grant Books.

Kurlantzick, J. (2022). *Beijing's Global Media Offensive: China's Uneven Campaign to Influence Asia and the World.* Oxford University Press.

Mattis, P. (2018). *China's 'Three Warfares' in Perspective.* War on the Rocks. https://warontherocks.com/2018/01/chinas-three-warfares-perspective/

Pew Research Centre. (2023). *China's Approach to Foreign Policy Gets Largely Negative Reviews in 24-Country Survey.* https://www.pewresearch.org/global/2023/07/27/chinas-approach-to-foreign-policy-gets-largely-negative-reviews-in-24-country-survey/.

Ryan, F., Knight, M. & Impiombato, D. (2023). Singing from the CCP's Songsheet: The Role of Foreign Influencers in China's Propaganda System. ASPI.

Staffas Edström, E. (2023a). *Ro med kommunistpartiets åror: en granskning av den kinesiska medienärvaron i Sverige.* NKK.

Staffas Edström, E. (2023b). *Kan vi lita på Tiktok?* NKK.

von Sydow, A. (2024). *China's Foreign Election Interference: An Overview of its Global Impact.* NKK

Thibaut, K. et al. (2022). *China's Discourse Power Operations in the Global South. An Overview of Chinese Activities in Sub-Saharan Africa, Latin America and the Middle East.* Atlantic Council. https://www.atlanticcouncil.org/wp-content/uploads/2022/04/Chinas_Discourse_Power_in_the_Global_South.pdf.

To, J. (2014). *Qiaowu: Extraterritorial Policies for the Overseas Chinese.* Leiden: Brill Academic.

Wilson, K. L. (2022). Strategic Responses to Chinese Election Interference in Taiwan's Presidential Elections. *Asian Perspective,* 46(2), 255–277. 10.1353/apr.2022.0011

Yang, F. (2021). *Translating tension: Chinese-language media in Australia.* Lowy Institute.

18. RUSSIA'S INFORMATION INFLUENCE

ANNELI AHONEN

SUMMARY

- There are five pillars that can be seen as creating Russia's information influence ecosystem: Official government communications, state-funded global messaging, cultivation of proxy sources, weaponisation of social media and cyber-enabled disinformation.
- The connection to Russia is sometimes openly acknowledged, and in other cases completely hidden.
- Distinctive features of Russia's information influence include high numbers of channels and messages, and dissemination of manipulative content. It is also rapid, continuous, and repetitive, and lacks consistency.
- The effect is not to persuade or appear credible but to sow confusion, disarray and distrust.
- During the past years, the Kremlin has continued to invest vast resources in its information influence capacities.

The practice of spreading disinformation is often referred to as originating in early Russian and Soviet intelligence services practices and deriving from the Russian word "dezinformatsiya", though there is no absolute certainty of its etymology.[25] As part of the Soviet deception game and active measures, Thomas Rid (2020) explains how disinformation was and is used in "covert operations designed to achieve overt influence" – something that the CIA on its side used to call "political warfare".[26]

Rid divides the development of the modern era in disinformation into four waves, first one beginning in the 1920s-1930s, when radio started shaping journalism and forgeries were used sometimes targeting both Soviet Union and the United States at the same time. The second wave after the World War II professionalised use of "political warfare" and "disinformation", which were used to deepen existing tension and contradictions within the adversary's policies. The third wave starts in the late 1970s, when the term "active measures" was already widely at use in the Soviet intelligence agencies. The operations were taking a more active form with the Soviet Union gaining an upper hand, Rid describes. The fourth wave of disinformation came long after the Soviet Union collapsed, in the mid-2010, when use of deceptive campaigns was reshaped by new technologies and internet platforms.[27]

Historically, active measures and disinformation were largely understood as

[25] Definition of disinformation. 2 Dec. 2024, https://www.merriam-webster.com/dictionary/disinformation.
[26] Active measures: the secret history of disinformation and political warfare. By Thomas Rid. New York. : Farrar, Straus and Giroux . 2020.
[27] Ibid.

the domain of intelligence services. The current practical use of the term "disinformation" has expanded so that many different actors, both state and non-state, have the ability of running malign influence operations, but at the core of the concept remains that disinformation is exploited with the intent to cause harm.

RUSSIAN STATE'S INFLUENCE STRATEGY

Russia's approach to disinformation is perhaps best understood with the concept strategic deception. "The Russian approach to the conflict can be described as a combination of tools perfected during the Soviet period and reactivated, first in the context of domestic power struggle and later in that of Russian foreign and security politics", Pynnöniemi (2016) writes. Ranging from political, informational, economic, financial and military spheres, different means are used to put the adversary into a defensive posture. Conditions are this way created for a surprise, often of military nature.[28]

Weiss and Pomerantsev (2014) write that the Kremlin exploits the idea of freedom of information to inject disinformation into society. The effect is not to persuade (as in classic public diplomacy) or earn credibility but to sow confusion via conspiracy theories and falsehoods. The tools stretch across media, elite influencers, party politics, finance, NGOs, the expert community, and cultural activities. [29]

During the past years, the Kremlin has continued to invest in its information influence capacities. Only according to the official budget, state media in its various forms receives more than 1,3 billion euros yearly.[30] Several players engage in the activities, from the Kremlin and Presidential Administration to different ministries, armed forces, intelligence agencies, companies, state-run NGOs and proxies, as well as state media. While still in the 2016 US election the Kremlin kept influence operations at an arm's length and involved companies such as Internet Research Agency in the campaign, after Russia's invasion of Ukraine these practices have been more openly used by the state apparatus.

[28] Pynnöniemi, K., & Rácz, A. (2016). Fog of falsehood: Russian strategy of deception and the conflict in Ukraine. The Finnish Institute of International Affairs. https://www.fiia.fi/wp-content/uploads/2017/01/fiiareport45_fogoffalsehood.pdf
[29] Pomerantsev, P., & Weiss, M. (2014). The menace of unreality : how the Kremlin weaponizes information, culture and money. Institute of Modern Russia.
[30] Debunk.Org (2023). Kremlin spent 1.9 billion USD on propaganda last year, the budget exceeded by a quarter. Debunk.Org. https://www.debunk.org/kremlin-spent-1-9-billion-usd-on-propaganda-last-year-the-budget-exceeded-by-a-quarter

US Global Engagement Center has a useful illustration of the Kremlin's strategic tools.[31]

It describes five pillars that create Russia's disinformation and propaganda ecosystem: Official government communications, state-funded global messaging, cultivation of proxy sources, weaponisation of social media and cyber-enabled disinformation. The connection to Russia is sometimes openly acknowledged, and in other cases completely hidden.

The Kremlin's disinformation strategy is not always succeeding, and it has blind spots. The authoritarian governance was, for example, notably late in preparing the Russian audiences to the full-scale invasion of Ukraine. This failure was likely impacted by how the decision to invade was made – relatively late, within a very small group of people and with little coordination between Putin and the government administration.[32]

THE CASE OF RT AND SPUTNIK IN SWEDEN

When Sputnik was founded in late 2014 from the remnants of the dismantled Russian state news agency Ria Novosti and radio broadcasting service Voice of Russia, new language versions of the websites were rapidly created across Europe. Sputnik's goal was to cover over 30 mainly smaller language audiences, while the Russia Today (RT) TV network would focus on reaching large, global audiences in English, Arabic, Spanish, and later German and French. Sputnik's initial aim was to enter the Nordic markets, and Swedish Sputnik had a short presence running from spring 2015 to spring 2016, at which point all of Sputnik's Nordic operations

[31] U.S. Department of State, Global Engagement Centre. (2022). Kremlin-funded media: RT and Sputnik's role in Russia's disinformation and propaganda ecosystem. GEC special report. https://www.state.gov/wp-content/uploads/2022/01/Kremlin-Funded-Media_January_update-19.pdf
[32] Dr Neville Bolt et al. (2023). Kremlin communication strategy for Russian audiences before and after the full-scale invasion of Ukraine, NATO Strategic Communications Centre of Excellence. https://stratcomcoe.org/publications/kremlin-communication-strategy-for-russian-audiences-before-and-after-the-full-scale-invasion-of-ukraine/293

were eventually closed. Sputnik assumed the characteristics of tabloid media with a strong anti-establishment angle. RT's tagline would later become, Question More, indicating its chief method of casting doubt on mainstream news sources by introducing often spurious narratives based around conspiracies, speculation, and disinformation.

During its 12 months of operation, Sputnik Sverige performed a role in Russia's information influence efforts within the Swedish information environment. From nearly 4,000 articles published by Sputnik between April and December 2015, the most common themes about Sweden were 'Crisis in the West', 'Positive image of Russia', and 'Western aggressiveness'. Ukraine, the EU, NATO, and the United States were mentioned often and depicted negatively in the articles. Sweden itself received significantly less coverage. Outright faked content appeared on a limited scope, with disinformation about the shooting down of MH17 an important example demonstrating the value of the platform to the Kremlin[33]. Rather than providing tailored, hyper local news to Sweden, the Sputnik experiment did not seem to click with target audiences.

Another study identified during years 2014–2019 the following narratives in Sputnik's English-language coverage about Sweden[34]:

- Sweden is an ultraliberal state in decay
- A conflict-torn space
- An invaded and unsafe space by migrants and criminals
- A country that exaggerates liberal values

The Russian leadership offered no explanation for shutting down the Sputnik operation in Nordics. Potential reasons may include that the websites were not economically feasible, or because the poor quality of content and translations did not appeal to the audiences. Another possibility is that the English language RT/Sputnik served as a channel to reach the Nordic audiences with general international news equally well. Perhaps more likely is that existing Swedish fringe news platforms performed a similar role to Sputnik, but with better localisation. Given the low levels of exposure to Russian state media, it seems likely that resources were quickly diverted to more promising markets. A survey from 2020 found that 7% of Swedes consumed the international RT or Sputnik channels, with young, men, and supporters of fringe or right-wing Sweden Democrats party overrepresented[35].

There is insufficient data about how Sputnik Sverige was resourced. However, it does seem that other Russian state media were actively pursuing scandalous stories in Sweden during this period. An illustrative example is a Sputnik article in English and French about Sweden wanting to shoot missiles from Gotland at Russian troops. This is an example of more subtle distortion, where Sputnik used

[33] Kragh, M., & Åsberg, S. (2017). Russia's strategy for influence through public diplomacy and active measures: The Swedish case. Journal of Strategic Studies, 40(6), 773–816. https://doi.org/10.1080/01402390.2016.1273830

[34] Deverell, E., Wagnsson, C., & Olsson, E.-K. (2021). Destruct, direct and suppress: Sputnik narratives on the Nordic countries. The Journal of International Communication, 27(1), 15–37. https://doi.org/10.1080/13216597.2020.1817122

[35] Wagnsson, C., Blad, T., & Hoyle, A. (2024). 'Keeping an eye on the other side': RT, Sputnik, and their peculiar appeal in democratic societies. The International Journal of Press/Politics, 29(4), 1109–1133; Wagnsson, C. (2023). The paperboys of Russian messaging: RT/Sputnik audiences as vehicles for malign information influence. Information, Communication & Society, 26(9), 1849–1867. https://doi.org/10.1080/1369118X.2022.2041700

quotes from the Governor of Gotland and a military commentator, but omitted the context and mistranslated their statements to make it seem like they were plotting an attack on St Petersberg.[36] In a second example, Russia state media channel NTV visited Gotland's County Administrative Board and filmed a Swedish civil servant pointing at several locations on a map, adding a voiceover that explained how Gotland had been at the centre of wars in the 20th century. The civil servant had, however, been asked to show the nature reserves in Gotland.[37] In a third example targeting Sweden during this period, NTV offered money to local teenagers in Rinkeby in 2017 "to do some action in front of the camera" in an attempt to capture footage of gang violence.[38]

Sputnik Sverige did not last for long, but the information influence tactics used were aligned with other Russian state media operating abroad. Faked content appeared early in the news cycle in situations where the Russian official position was based on creating uncertainty, mistrust, or generating multiple conspiracies. This was evidenced consistently in UK and French media regulators' impartiality conclusions on RT stories about the poisoning of Russian ex-agent Sergey Skripal, chemical attacks in Syria, as well as RT Germany's coverage of COVID-19 in its YouTube channels[39].

Impartiality breaches by Russian state media would only increase following the Russian invasion of Ukraine[40]. Post-invasion, apart from spreading disinformation and propaganda, RT has also acquired cyber capabilities and engaged in military procurement in support of Russia's war efforts. As of 2024, RT and Sputnik have been sanctioned both by the EU and the US.[41]

TACTICS

The current Russian model for propaganda has been characterized as "the firehose of falsehood". Its distinctive features are high numbers of channels and messages, and dissemination of manipulative content. It is also rapid, continuous, and repetitive, and not consistent.[42]

Based on whistleblowers' accounts and leaked documents, it is known that Russia's presidential administration distributes orders to main state media in weekly meetings on what and how to report on specific issues[43]. This tool is called "temnik", instruction and guidelines from the authorities which are shared with the journalists. These instructions have played an important role in establishing control

[36] NATO StratCom CoE (2015). Disinformation in Sweden.
[37] EUvsDisinfo. (2018, July 16). In Sweden, resilience is key to combatting disinformation. EUvsDisinfo. https://euvsdisinfo.eu/in-sweden-resilience-is-key-to-combatting-disinformation/
[38] Ibid., EUvsDisinfo. (2017, March 8). Russian TV offers money for staged 'action' in Sweden? EUvsDisinfo. https://euvsdisinfo.eu/russian-tv-offers-money-for-staged-action-in-sweden/
[39] Deutsche Welle (2021.) YouTube deletes RT's German YouTube channels. DW.Com. https://www.dw.com/en/youtube-deletes-rts-german-youtube-channels-after-covid-misinformation-strike/a-59343394, CSA (2018). Manquements à l'honnêteté, à la rigueur de l'information et à la diversité des points de vue: Mise en demeure de RT France. http://web.archive.org/web/20180709163542/http://www.csa.fr/Espace-Presse/Communiques-de-presse/Manquements-a-l-honnetete-a-la-rigueur-de-l-information-et-a-la-diversite-des-points-de-vue-mise-en-demeure-de-RT-France, BBC (2019) Ofcom fines Russian news service £200,000 over impartiality. https://www.bbc.com/news/entertainment-arts-49126466
[40] Ofcom (2022). Ofcom finds RT in breach of due impartiality rules. https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-standards/ofcom-finds-rt-in-breach-of-due-impartiality-rules/
[41] U.S. Department of State (2024). Alerting the world to RT's global covert activities. United States Department of State. https://www.state.gov/alerting-the-world-to-rts-global-covert-activities/, Council of the EU (2022). EU imposes sanctions on state-owned outlets RT/Russia Today and Sputnik's broadcasting in the EU. https://www.consilium.europa.eu/en/press/press-releases/2022/03/02/eu-imposes-sanctions-on-state-owned-outlets-rtrussia-today-and-sputnik-s-broadcasting-in-the-eu/
[42] Paul, C., & Matthews, M. (2016). The Russian 'firehose of falsehood' propaganda model: Why it might work and options to counter it. RAND Corporation. https://www.rand.org/pubs/perspectives/PE198.html
[43] EUvsDisinfo. (2019, February 4). The weekly meetings. EUvsDisinfo. https://euvsdisinfo.eu/the-weekly-meetings/

over Russian media's coverage.[44]

Russia's sophisticated use of online campaigns making use of bots, trolls and most recently AI-generated content is a continuous attempt to influence global audiences and important political events such as elections. One of the main aims has been to erode the West's support to Ukraine. These campaigns involve extensive monitoring of target audience's opinions and crafting of messages accordingly. Some of these are funded and coordinated by the Russian presidential administration.[45] One of the contractors, Social Design Agency, claimed to have published 34 million comments from January to April 2024 alone.[46] Many of the managers of such campaigns have been sanctioned after Russia's full scale invasion of Ukraine in 2022.[47]

WELL-KNOWN RUSSIAN DISINFORMATION STORIES

Several disinformation campaigns have been attributed to Russia and have been well-documented. They serve as examples providing insight to Russia's methods and aims in spreading disinformation, and some stories trace back to Soviet times.

US DEVELOPED HIV AND DELIBERATELY SPREAD IT

The campaign initiated by the KGB started with an obscure newspaper in India called the Patriot in 1983. A campaign also described in the chapter about conspiracy theories published an anonymous letter headlined "AIDS may invade India: Mystery disease caused by US experiments." Allegedly the letter was written by a "well-known American scientist and anthropologist" in New York. It claimed AIDS is a result of the Pentagon's experiments to develop biological weapons.[48] Later, the story spread globally to hundreds of newspapers, magazines, radio and TV and received coverage in over 80 countries in more than 30 languages.[49] Similar stories are exploited by Russian disinformation still today.[50]

UKRAINIANS CRUCIFIED A LITTLE BOY

The story of Ukrainians having crucified a three-year-old boy was first published by Eurasianist political philosopher and propagandist Aleksandr Dugin and later

[44] EUvsDisinfo. (2017, March 30). Temnik—The Kremlin's route to media control. EUvsDisinfo. https://euvsdisinfo.eu/temnik-the-kremlins-route-to-media-control/
[45] Meduza (2024). 'Latching onto successful projects' Leaked documents suggest Kremlin spin doctors are presenting popular movies and TV to Putin as propaganda wins. https://meduza.io/en/feature/2024/02/27/latching-onto-successful-projects
[46] Savchuk, ., & Mironyuk, . (2024, September 23). Temniki, memy, "reyestr feikiv ta doprealnosti". Jak pratsyue agenciya-pidryadnik Kremlya iz dezinformatsii. (Temniki, memes, "registry of fakes and pre-realities". How the Kremlin's disinformation contractor agency works). Radio Svoboda. https://www.radiosvoboda.org/a/skhemy-kreml-dezinformatsiya-asd-hambashydze/33121898.html
[47] U.S. Department of Justice. (2024, September 4). Justice Department disrupts covert Russian government-sponsored foreign malign influence operation targeting audiences in the United States and elsewhere. https://www.justice.gov/opa/pr/justice-department-disrupts-covert-russian-government-sponsored-foreign-malign-influence; The Council of the European Union (2023). Council Decision (CFSP) 2023/1566. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A32023D1566
[48] Boghardt, T (2009). Soviet Bloc Intelligence and Its AIDS Disinformation Campaign. Studies in Intelligence Vol. 53, No. 4 (December 2009). https://web.archive.org/web/20100324175917/https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol53no4/pdf/U-%20Boghardt-AIDS-Made%20in%20the%20USA-17Dec.pdf
[49] United States Department of State (1987). Soviet Influence Activities: A Report on Active Measures and Propaganda, 1986 – 87. https://www.globalsecurity.org/intell/library/reports/1987/soviet-influence-activities-1987.pdf
[50] EuvsDisinfo. Database query "hiv". https://euvsdisinfo.eu/disinformation-cases/?text=hiv

reported by Russian state-controlled Channel One in June 2014[51] after Russia had started its military aggression in Ukraine. According to the story, the boy had been publicly crucified by Ukrainian soldiers at "Lenin square" in Sloviansk. Journalists debunked the story – there was no video or audio recordings of the incident, and there is even no Lenin square in Sloviansk.[52]

LISA-CASE

In 2016, a 13-year-old Russian German girl "Lisa" had gone missing for 30 hours. She was reported by Russian state-controlled media to have been raped my migrants. The story was fake, and the German police confirmed that she had been with a friend during that time. The story was reported by Russian media extensively, including its foreign arms like RT and Sputnik. Social media channels amplified the message. The reporting was followed by demonstrations by thousands of people from German Russian minority, and in the end Russian Foreign Minister Sergei Lavrov commented on the issue, saying that he is concerned about the German police ability to work on such cases.[53] The so called Lisa-case was an early wake-up call especially in Germany about Russian information influence tactics.

IMPACT

One of the most difficult questions is what kind of impact Russia's information influence has. There are many ways to approach the question – social media engagement of specific campaigns can be measured, it is possible to look at the consequences of a campaign and their impact on real lives of their targets and victims, or we can rely on surveys following the change of people's opinions in certain topics Russia's influence is spreading.

One of the illustrative examples is Russia's interference attempts in the US election in 2016. Russia's interference attempts took the US by surprise and the government, election administration, big tech platforms and the media were unprepared for what followed. There was limited amount of warnings or exposure of the interference attempts ahead of the vote, and media largely amplified the hack-and-leaks of Democratic National Committee networks aimed at discrediting Hillary Clinton. In 2020, the Senate Select Committee on Intelligence found that the US administration struggled to determine the appropriate response, frozen by 'paralysis of analysis'.[54]

In the end, it was confirmed that 126 million people saw Russian disinformation ahead of the 2016 election.[55]

[51] Bykov. D. (2014). Zachem TV, Aleksandr Dugin and Galina Pyshnyak raspyali malchika (Why did TV, Aleksandr Dugin and Galina Pyshnyak crucify a boy). https://sobesednik.ru/dmitriy-bykov/20140715-dmitriy-bykov-zachem-tv-aleksandr-dugin-i-galina-pyshnyak-ra; Pervyi Kanal (2014). Bezhenka iz Slavyanska vspominaet, kak pri nei kaznili malenkogo syna i zhenu opolchentsa. (A refugee from Slavyansk recalls how her young son and the wife of a militiaman were executed in front of her) https://web.archive.org/web/20240121000806/https://www.1tv.ru/news/2014-07-12/37175-bezhenka_iz_slavyanska_vspominaet_kak_pri_ney_kaznili_malenkogo_syna_i_zhenu_opolchentsa

[52] Tyschuk, O (2014). Istoriu o "raspyatom malchike" dlya Pervogo kanala pridumala zhena boevika DNR. (The story about the "crucified boy" for Channel One was invented by the wife of a DPR militant). Fakty.com.ua. https://fakty.com.ua/ru/ukraine/polituka/20140714-1521089/, Epifanova, M. (2014). I eto – ne predel? (And this is not the limit?) Novaya Gazeta. https://web.archive.org/web/20140719021523/https://novayagazeta.ru/politics/64440.html

[53] Meister, S. (2016). The "Lisa case". Germany as a target of Russian disinformation. NATO Review. https://www.nato.int/docu/review/articles/2016/07/25/the-lisa-case-germany-as-a-target-of-russian-disinformation/index.html

[54] Senate Select Committee on Intelligence (2020). U.S. Government response to Russian activities. https://www.intelligence.senate.gov/press/senate-intel-releases-bipartisan-report-obama-admin-response-russian-election-interference

[55] Solon, O., & Siddiqui, S. (2017, October 31). Russia-backed Facebook posts 'reached 126m Americans' during US election. The Guardian. https://www.theguardian.com/technology/2017/oct/30/facebook-russia-fake-accounts-126-million

Another example from Central and Eastern Europe is how the results of pro-Russian narratives spreading can be seen in surveys. Notably in 2024, in Slovakia more people blame the West or Ukraine itself for the war in Ukraine (51% combined) than Russia (41%).[56]

Researcher Keir Giles recommends seeing subversion and Russia's influence as a process rather than an event and looking at long-term trends. He compares what was normal in the information space in English-speaking countries in 2020 compared to 2015: "This comparison reveals spectacular change in an astonishingly short time. Assisted by the policies and algorithms of social media platforms, Russia has ridden and accelerated trends of fragmentation, distrust, and the spawning of alternative realities – and is now joined by a wide range of foreign and domestic imitators who choose to emulate Russian tactics for their own political ends, amplifying the damage done."[57]

RESPONSE: RESOURCES TO FOLLOW

Several governmental, media, fact-checking and civil society initiatives have been set up to counter foreign disinformation and information influence. Following resources offer up-to-date information and debunk Russia's disinformation:

EUvsDisinfo
EU's initiative to counter pro-Kremlin disinformation.

Psychological Defence Agency
Swedish governmental agency that counters foreign information influence.

NATO Strategic Communications CoE
NATO Center of Excellence which publishes regular analysis on Russian disinformation narratives, tactics and impact on social media.

Hybrid CoE
Helsinki Center of Excellence is an NGO whose member states are EU or NATO members. Focuses on countering hybrid threats.

DFRlab
Atlantic Council's initiative Digital Forensic Research Lab publishes analysis on foreign influence campaigns on social media.

EUDisinfoLab
An NGO specialising in exposing online influence campaigns from various sources.

DISCUSSION

- What are the main aims of Russia's information influence?
- What are some of the distinct features of Russian disinformation and propaganda?
- How would you describe the role of Russian state-controlled media, like RT and Sputnik, in the disinformation ecosystem?
- Mention an example of Russia's information influence and what does it tell us about techniques used to disinform.
- Does Russian disinformation have any impact?

[56] Globsec (2024). Globsec Trends 2024 Slovakia. https://www.globsec.org/what-we-do/publications/globsec-trends-2024-slovakia
[57] Keir Giles (2021). Assessing Russian success and failure. Baltic Defence College Russia Conference, Tartu. https://www.researchgate.net/publication/349776264_Assessing_Russian_Success_and_Failure

ANNELI AHONEN *has a long experience in developing democratic responses to disinformation on an international level. She works currently as a Research Associate at Cardiff University. Recently she published a report on cyber-enabled influence campaign "Ghostwriter" and the responses to it, and she follows closely especially Russia's disinformation. Prior to that, she worked at the Institute for Strategic Dialogue as a senior fellow during the German Federal election. As Head of East Stratcom Task Force at the European External Action Service, she led the team's work in strategic communications in the Eastern Partnership countries and the EU's response to Russia's disinformation, including running the EUvsDisinfo campaign. Earlier she worked as a journalist for ten years, out of which 2009–2016 in St. Petersburg for the biggest Finnish newspaper Helsingin Sanomat.*

19. THE COMMERCIAL DISINFORMATION ECOSYSTEM

DAREJAN TSURTSUMIA & JAMES PAMMENT

SUMMARY

The chapter introduces current research on commercially motivated actors involved in production of disinformation. Such industries thrive in contexts marked by precarious labour conditions and war economies, as exemplified by Russia and the Kremlin's commercial disinformation infrastructure. The chapter unveils the range of actors and their roles within this outsourced ecosystem, while also raising questions about ethical responsibilities of the communication industry and the evolving relationships between states and technology companies in the Western context. The empirical material is based on the analysis of the leaked documents of Social Design Agency of Russia, detailed in a report published by the Swedish Psychological Defence Agency in 2025.

THE DIMENSIONS OF COMMERCIAL DISINFORMATION RESEARCH

Disinformation research is increasingly urged to move beyond event-driven case studies and to instead examine the global and systemic production of false information. The recurrent spread of intentionally misleading and manipulative content has sparked academic debates about the role of for-profit, non-state actors who produce disinformation for financial gain (Ong & Cabañes, 2018; Hughes & Waismel-Manor, 2021; Grohmann & Ong, 2024; Diaz Ruiz, 2025). Within the broader and loosely defined field of disinformation studies, the notion of "disinformation-for-hire" has emerged, drawing attention to the organisational and labour dimensions of disinformation as a service (Grohmann & Ong, 2024). Marketing scholarship has also begun to address disinformation's commercial aspects, emphasizing that the circulation of false information is deeply embedded in contemporary media infrastructures and communication practices, rather than merely an unintended by-product of these systems (Diaz Ruiz, 2023). Similarly, scholars studying tech platforms have addressed problems of monetisation of disinformation (Hua et al., 2022), and the consequences of de-platforming malicious actors (Horta Ribeiro et al., 2024; Innes & and Innes, 2023).

Investigative reporting has further revealed the global scope of the disinformation industry. In an article for BuzzFeed News, produced in partnership with the Taiwanese outlet The Reporter, Silverman et al. (2020) documented firms in Malaysia, Georgia, Egypt, Israel, the United Arab Emirates, Ukraine, Brazil, Indonesia, and Poland that specialise in delivering political campaigns. Structured primarily as PR or marketing companies, these firms range from "bot farms" employing multiple social media operators, to organisations offering semi-automated content creation and dissemination services. Many allegedly draw inspiration from Russian and Iranian influence playbooks, but their primary

motivation is financial gain (Silverman et al., 2020). In an interview for the same investigation, Jonathan Corpus Ong noted that 'dark PR' companies in the Philippines actively seek to destigmatise their activities by adopting neutral terminology, for example, describing fake news sites as "supplemental pages", thereby normalising the practice. Similarly, members of a small Macedonian agency, made infamous for profiting from digital advertising by producing false content during the 2016 U.S. elections, openly rationalised their activities as purely market driven. Some even argued that responsibility lay with Americans who failed to distinguish truth from falsehood, rather than with those producing the content (Hughes & Waismel-Manor, 2021).

The commercial side of disinformation does not, however, solely include the organisational and labour dimensions. Scholars have also examined the responsibilities of the big tech platforms in profiting from the circulation of disinformation. Programmatic advertising infrastructure such as AdTech, often profits from the dissemination of harmful or misleading content (Stevens, 2017; Braun & Eklund, 2019; Diaz Ruiz, 2025). According to these studies, advertising technology firms benefit financially from spreading disinformation, while brands inadvertently fund these activities by having their ads placed on such websites. In addition, the engagement-driven logic of social media platforms, including disclosures from whistleblower Frances Haugen, reveal Facebook's systematic prioritisation of user engagement over societal wellbeing ('The Facebook Files', 2021).

Grohmann and Ong (2024) caution, however, against viewing tech platforms solely as negative forces, arguing that such a stance risks reinforcing a techno-deterministic perspective. Furthermore, it is important to acknowledge that many tech platforms have cooperated with investigative researchers and state authorities to identify and dismantle accounts engaged in coordinated inauthentic behaviour (CIB), 'typosquatting', and other activities that violate platform policies *(Affidavit in Support of Seizure Warrant,* 2024). Recent decisions by Meta and X to scale back their content moderation efforts have, however, created a more disinformation-friendly environment, a move that Mark Zuckerberg has controversially framed as "reducing censorship" (Duffy, 2025).

In conclusion, research on commercial disinformation highlights the systematic production of false information by examining the underlying factors that enable and sustain it. According to current research, these factors include state governance, which fosters precarious labour conditions and permits a lack of ethical oversight in sectors such as media and communications. They also encompass digital platforms including social media and programmatic advertising infrastructure, and their affordances, which, whether intentionally or unintentionally, facilitate the large-scale and increasingly personalised dissemination of disinformation and its monetisation, as well as produce profit for the platforms themselves.

RUSSIAN COMMERCIAL DISINFORMATION AGENCIES – SDA AND STRUCTURA

Due to its recent notoriety in connection with the war in Ukraine, the Social Design Agency SDA, Агентство Социального Проектирования) and its sister company Structura (Группа Компаний Структура) became some of the most scrutinised for-profit disinformation agencies in the world (a position previously held by Mikhail

Prigozhin's Internet Research Agency). Social Design Agency and Structura are based in Moscow, and support the Kremlin's influence operations domestically and abroad, offering services such as media monitoring, content production, content delivery, and dissemination. Founded in December 2017 by political consultant Ilya Gambashidze, both SDA and Structura, together with the Russian nonprofit Dialog (АНО Диалог), were implicated in conducting influence operations after Russia's full-scale invasion of Ukraine in 2022. Their campaigns targeted audiences in Germany, France, the United States, the United Kingdom, Italy, Israel, and Ukraine, among others. One of their most prominent tactics was the creation of "mirror websites" that mimicked established Western news outlets to mislead Western audiences into consuming Kremlin-aligned narratives. The tactic was publicly uncovered in the summer of 2022 under the label *Operation Doppelgänger*, following investigations by journalists (Von Wienand et al., 2022; Brühl et al., 2022) as well as technical assessments conducted by Meta (Ben Nimmo & David Agranovich, 2022) and organizations within the Foreign Information Manipulation and Interference (FIMI) community (Qurium, 2022; Alaphilippe et al., 2022; Viginum, 2023). However, most of knowledge about SDA and Structura stems from files released from the SDA servers in the spring and summer of 2024, containing over 3,100 internal documents, which were distributed to a small number of journalists and scholars (Pamment & Tsurtsumia, 2025).

The leaked document titled "The Employees of SDA" (Сотрудники АСП), shows that the organisation employs at least 107 individuals, operating both from Moscow and various regions across Russia managing several campaigns simultaneously. The human resources were mainly dispersed across Russia's big cities and regions, although some were found in Crimea and Bulgaria. Little is known about the working conditions or motivations of employees from the leaked files, making it difficult to assess the labour dimension of the agency's activities. However, it is plausible that the same precarious conditions in Russia that drive individuals to enlist in the army also make them more susceptible to employment as bot-operators or content creators within disinformation agencies.

Based on its outputs, ranging from textual and visual materials to content delivery techniques and social media monitoring/listening tools, SDA is most plausibly characterised as an advertising, communications, or public relations firm. At the same time, its services extend to media monitoring, and its long texts often mimic the style of newspaper articles, giving it the appearance of a news organisation. Yet, unlike genuine journalism, SDA's "news" is unconstrained by ethics or reality. In addition, SDA strategically borrows the language of security and intelligence, branding itself as the "informational forces of Russia." They employ marketing terminology such as "market penetration," but also frame their operations in military and intelligence terms, coining phrases like "infiltrating enemy territory" or "fostering subversion" in Western countries.

THE RUSSIAN DIGITAL COMMERCIAL DISINFORMATION ECOSYSTEM

In the aftermath of Russia's full-scale invasion in Ukraine in 2022, Sergei Kiriyenko, the First Deputy Chief of Staff of the Presidential Administration of Russia, and one of President Putin's closest aides, was tasked with consolidating political consent in Russian-occupied Ukraine. Among other extensive duties, he was entrusted with developing the Russian digital public sphere and online

communication infrastructure (Holger Roonemaa et al., 2024; Cardiff University, 2024). In 2023, he secured €600 million from the state budget to support his network of digital propaganda companies and autonomous non-profit organizations, ANOs, such as ANO Dialog (АНО Диалог) and ANO IRI (Internet Development Institute - АНО Институт Развития Интернета). Kiriyenko personally approves and oversees all strategic activities of the ANOs and smaller companies within the network, making sure they are repeatedly validated and expanding their budgets and operations. SDA and Structura operate within this broader outsourced disinformation infrastructure. At times they appear to function as a single entity in relation to the information influence campaigns, but despite their operational overlap, the two organisations formally bid separately for government contracts, thereby creating the illusion of fair competition.

The figure below illustrates a segment of the Kremlin's outsourced propaganda ecosystem, based on the leaked files from SDA. The seats around the table represent the individuals and organisations directly engaged in the planning and strategic management of SDA's campaigns against the Western countries and Ukraine.

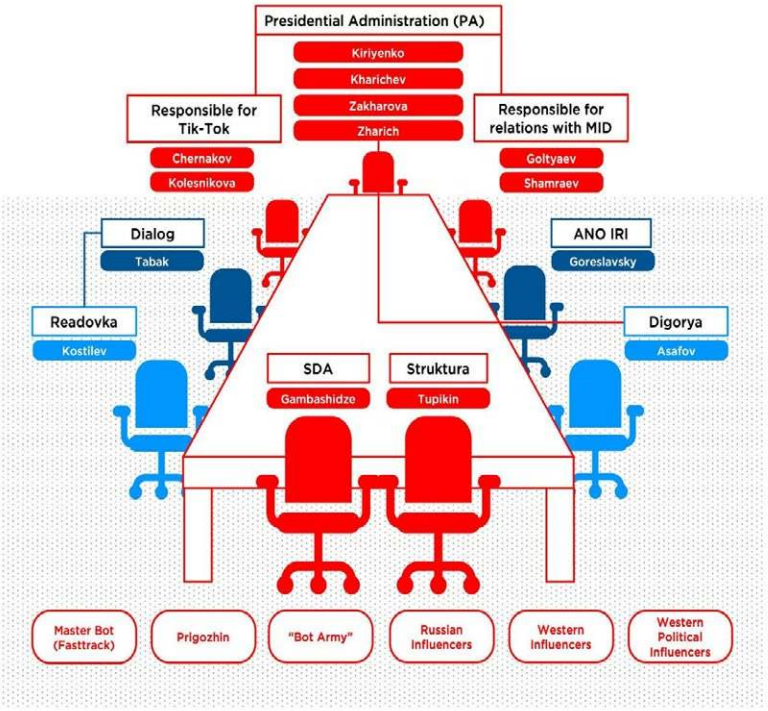


Figure 1. A representation of regular participants in Presidential Administration meetings at which strategic decisions about SDA operations were made from 'Beyond Operation Doppelgänger – Capability Assessment of Social Design Agency' by James Pamment and Darejan Tsurtsumia.

NON-PROFITS

Dialog (АНО Диалог) has played an active role in recent influence operations against Ukraine and the West but is mainly targeting the Russian population. Designed to establish a unified system for monitoring public sentiment across

Russia's regions, Dialog has technical capabilities to operationalise stealthy modes of disinformation delivery (*Affidavit in Support of Seizure Warrant*, 2024). Dialog envisions itself as a creator of the Russian 'state internet' or Gosdigital (Госдиджитал), completely saturated with pro-Putin content (Holger Roonemaa et al., 2024). Dialog imagines that nationalistic PR content that promotes Putin and his regime should accompany every Russian citizen from kindergarten and continue through school, museums, libraries, sports activities, and even interactions with municipal and central government institutions, ultimately extending to hospitals. Gosdigital in its most ambitious imaginary is a "cradle-to-grave" propaganda service.

Another important platform under Kiriyenko's oversight is the Institute of Internet Development (ANO IRI), tasked with producing propagandistic content for the Russian internet. In 2023, ANO IRI received nearly €220 million from the state budget, followed by €180 million in 2024 (Holger Roonemaa et al., 2024). Meanwhile, Dialog was allocated approximately €68 million. ANO IRI is responsible for producing films, blogs, games and TV series that promote Kremlin narratives, while Dialog and Dialog Regions provide the necessary online infrastructure to disseminate IRI's content. IRI focuses on creating 'cultural' content designed to evoke national pride and heroism, to persuade residents of occupied territories that they are better off with Russia, and to convince Russian residents that their quality of life has improved. Dialog, in turn, guarantees that these messages reach every Russian with internet access.

COMMERCIAL ACTORS

Some of the most notable Russian commercial actors who contribute to the digital disinformation ecosystem are social listening software Brand Analytics and Medialogia, chatbot services Fasttrack by Masterbot, and Dolphin Cloud and Anti-Detect Browser software by Zhitniakov Software Solutions. Brand Analytics is a Russian IT company specialising in digital and social media data analysis. It is a key technology behind SDA's media monitoring, and social listening functions. Many of the automated monitoring outputs appear to be sourced from this dashboard. Their technology is similar to Western commercial social listening tools such as Brandwatch and Talkwalker. On its website, Brand Analytics portrays itself as a leading IT company and developer of social media monitoring and analysis systems for the Russian and CIS markets, claims to have worked with more than 1,300 companies, and to process an average of 3 billion messages per month. As an industry standard, Brand Analytics also provides training courses to help clients master the software: to define the scope of platforms, target audiences, and specific objectives, as well as interpret the software's outputs.

Medialogia, a competitor to Brand Analytics, has specialised in social media monitoring and analysis since 2003. The company reportedly employs around 650 people and processes roughly 100 million social media posts alongside one million digital media messages each day, amounting to a monthly volume comparable to Brand Analytics' claimed 3 billion messages. According to its own figures, Medialogia automatically tracks around 250,000 platforms and over 2.5 billion social media accounts. Like Brand Analytics, it promotes proprietary tools such as the Media Index and SM Index, while placing particular emphasis on sentiment analysis in the development of its text analysis technologies. The company claims to employ deep learning models, neural networks capable of assessing not only

the overall meaning of texts but also the attributes of the entities they reference. Medialogia maintains a relatively streamlined product portfolio, focusing primarily on mass media and social media monitoring, and in addition to its core services, offers a range of webinars and training courses.

Fasttrack is a Russian company offering chatbot services to commercial clients. It appears to be developed by Masterbot (ID: 9721010347), a computer technology company first registered in 2016.[58] The services are managed through a dashboard called dashboard.fstrk.io. Fasttrack chatbots are compatible with not only Russian chat services, but international platforms, such as WhatsApp, Telegram and Viber. It is likely that these services were employed by SDA and Structura to flood the chat groups, creating an appearance of a lively discussion and harass Ukrainian citizens by sending them personal messages and fake survey questionnaires.

In order to stay undetected influence campaigns orchestrated by SDA were delivered through Dolphin Anti-detect browser. An anti-detect browser is a specialised web browser designed to conceal digital fingerprints - such as cookies, IP addresses, and other tracking data - thereby preventing websites from identifying or linking user accounts. These browsers enable operators to manage multiple accounts simultaneously and perform various activities without leaving detectable traces.[59] While Dolphin might not be the best anti-detect browser on the market, it is a Russia-based software, and therefore safer to use, can be modified to meet clients' needs, and is part of the Russian business infrastructure.

There is a general trend towards nationalisation of digital services and platforms in Russia, as well as towards creating Russian analogies of the Western platforms. Much of it seems to be connected to Sergey Kiriyenko and his authority over the Russian digital sphere. Kiriyenko has since 2021 exerted control over the popular Russian social media network VKontakte (VK), installing his son as a CEO. In March 2025, VK released a new 'national' messaging app which is intended to reduce reliance on Western messaging apps (Anton Troianovski, 2025). Russia seems to be determined to build its own digital infrastructure, by creating and supporting the national analogies of the Western platforms. It is not evident if Russia is trying to copy Chinese products or if there is a knowledge exchange between the Chinese and Russian digital sectors.

Among Russia's national digital platforms is Sreda, a software suite developed specifically for the government sector. It provides services such as email, calendars, cloud storage, task management, organizational structure management, and e-conferencing, functionally resembling Microsoft's offerings. The platform is accessible across multiple devices and claims to deliver a high level of security. Its integrated chat function is intended to replace foreign messaging apps as the primary communication tool for civil servants.[60] Originally launched as Arm GS (арм гс), short for "Automated Workplace for a Civil Servant" (автоматизированное рабочее место государственного служащего), the system has since been rebranded as Sreda (Среда), a term that translates both as "environment" and "Wednesday." Today, more than 350,000 civil servants reportedly use the platform for data management[61].

[58] https://fstrk.io; http://fasttrackbot.com; https://www.rusprofile.ru/id/10584994
[59] https://multilogin.com/glossary/antidetect-browser/
[60] https://web.archive.org/web/20241001145821/https://www.tadviser.ru/index.php/
[61] https://sreda.digital.gov.ru/

CONCLUSION

Commercial disinformation is blooming in Russia, and it has been successfully rationalised under the patriotic guise of "national interests." The Kremlin's disinformation ecosystem encompasses digital, business, and educational platforms, all of which receive financial benefits from the exchange. The war in Ukraine has transformed Russia into a war economy in which businesses and institutions are mobilised and financially incentivised toward a single overarching objective. Those who do not directly contribute to the physical war effort instead provide informational and intellectual support for the Kremlin's hybrid warfare, targeting both the West and Russia's own citizens. Among non-military businesses that are part of the Kremlin's commercial disinformation ecosystem are news agencies, social media platforms, digital software companies, advertising and PR firms, non-profits tasked with creating "cradle-to-grave" propaganda, and educational institutions that produce new cohorts of political technologists, software engineers, AI specialists, and communication professionals.

Beyond domestic efforts, the Kremlin actively seeks to penetrate Western social media platforms such as Facebook, Instagram, and YouTube, as well as messaging services like WhatsApp and Telegram, which serve as key channels for information distribution. This strategy is often described as "waging war on the enemy's turf." Digital platforms are increasingly perceived as extensions of physical territory. SDA and Structura make a considerable effort to go undetected while infiltrating those territories at scale, and when exposed, they often frame the resulting publicity as a marker of success. Western governments have responded to such breaches by imposing sanctions on the organisations and their leadership, but this does not seem to slow the efforts down.

A trend can be observed in Russia's efforts to develop domestic analogues of Western digital platforms and services, as well as their subsequent nationalisation under Kremlin control. Russia has long maintained its own platforms, such as Vkontakte and Yandex, which are now firmly influenced by state authorities. Unlike China, Russia's digital ecosystem is not as insulated from the global information environment, nor does it attempt to replicate the Chinese model. Instead, it mirrors Western platforms while simultaneously exhibiting a gradual tendency toward increased isolation and the restriction of Russian citizens' access to the global digital sphere. These trends are likely to continue.

It is unclear, however, if entities such as the ANOs, SDA and Structura could function within a financially competitive environment. ANO Dialog and ANO IRI are not expected to produce financial profit, and are subsidised as part of the war effort, and there is also limited evidence of SDA's financial stability. Their success is measured by KPIs that produce numbers reflecting the penetration of information and awareness at best, but do not show the reception of that information or its influence on actual behaviour. This does not mean that their efforts should be dismissed; in fact, they often use industry standard evaluation methods. However, it means that their fate is highly contingent on events unfolding at the front, and on the geopolitical context more broadly. Put simply, it may be questioned whether these products and instruments of a war economy could survive in a peace economy.

Learning the Russian context helps to address certain trends in platform-state relationships within democratic societies, especially in times of heightened security concerns. The Russian case demonstrates the wide range of actors that can

participate in and profit from commercialised disinformation infrastructures. These are not limited to news outlets, advertising agencies or shady communication consultants, but extend across a diverse array of digital services, including social listening tools, anti-detect browsers, chatbot providers, and others that both sustain and benefit from such exchanges. The ecosystem of organisations responsible for producing disinformation-as-a-service therefore becomes fully integrated and dependent on the continued degradation of global affairs, contributing to a negative spiral. If totalitarian regimes and wars necessitate the growth of organisations like SDA, ANO IRI and ANO Dialog to ensure their survival, these organisations are at least as dependent on the continuation of the regime and the war to survive.

DISCUSSION

- How does the commercialisation of disinformation complicate traditional understandings of hybrid warfare as primarily a state-driven activity?
- What ethical responsibilities do advertising, and content creating industries bear in sustaining democratic values in the West?
- In what ways does Russia's integration of PR firms, non-profits, and digital services into its disinformation ecosystem differ from Western platform-state relationships?
- How might the creation of domestic Russian digital platforms (e.g., Sreda, VK, national messaging apps) affect the future of global information warfare?

DAREJAN *Tsurtsumia is a PhD student at Karlstad University studying the commercialization of disinformation, particularly in the Russian context.*

JAMES PAMMENT *is Director of the Lund University Psychological Defence Research Institute. His main research interest is in the role of strategic influence in international relations, both its legitimate sides (e.g., public diplomacy and aid) and illegitimate (e.g., propaganda and hostile foreign interference). Previous affiliations include the Carnegie Endowment for International Peace, Swedish Defence University, the EU-NATO Hybrid Threats Center of Excellence, and the University of Texas at Austin.*

REFERENCES

Affidavit in Support of Seizure Warrant. (2024). https://www.justice.gov/d9/2024-09/doppelganger_affidavit_9.4.24.pdf

Alaphilippe, A., Machado, G., Miguel, R., & Poldi, F. (2022). *Doppelganger – Media clones serving Russian propaganda.* EU DisinfoLab. https://web.archive.org/web/20250902091359/https://www.disinfo.eu/doppelganger/

Anton Troianovski. (2025, August 10). The Quiet Technocrat Who Enacts Putin's Ruthless Agenda. *The New York Times.* https://web.archive.org/web/20250902092613/https://www.nytimes.com/2025/08/10/world/europe/putin-russia-ukraine-war-sergei-kiriyenko.html

Nimmo, B. & Agranovich, D. (2022). *Removing Coordinated Inauthentic Behavior from China and Russia* (Threat Report). Meta. https://web.archive.org/web/20250902090144/https://about.fb.com/news/2022/09/removing-coordinated-inauthentic-behavior-from-china-and-russia/amp/

Braun, J. A., & Eklund, J. L. (2019). Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism. *Digital Journalism,* 7(1), 1–21. https://doi.org/10.1080/21670811.2018.1556314

Brühl, J., Heubl, B., & Hurtz, S. (2022, September 7). Propaganda mit gefälschten SZ-Videos*. Süddeutsche Zeitung.* https://web.archive.org/web/20250903134440/https://www.sueddeutsche.de/politik/desinformation-russische-propaganda-sz-fake-news-1.5651531

Cardiff University. (2024). *Putin's 'Little Grey Men': Russia's Political Technologists and Their Methods.* Security, Crime and Intelligence Innovation Institute. https://www.cardiff.ac.uk/__data/assets/pdf_file/0008/2875274/Little-Grey-Men-Report-Final-compressed-1.pdf

Diaz Ruiz, C. (2023). Disinformation on digital media platforms: A market-shaping approach. *New Media & Society,* 1–24. https://doi.org/10.1177/14614448231207644

Diaz Ruiz, C. (2025). *Market-Oriented Disinformation Research: Digital Advertising, Disinformation and Fake News on Social Media*. Taylor & Francis. https://doi.org/10.4324/9781003506676

Duffy, C. (2025, January 7). Meta gets rid of fact checkers and says it will reduce 'censorship'. *CNN.* https://web.archive.org/web/20250613151441/https://edition.cnn.com/2025/01/07/tech/meta-censorship-moderation

Grohmann, R., & Ong, J. C. (2024). Disinformation-for-Hire as Everyday Digital Labor: Introduction to the Special Issue. *Social Media + Society,* 10(1), 20563051231224723. https://doi.org/10.1177/20563051231224723

Holger Roonemaa, Marta Vunsh, Anastasiia Morozova, Carina Huppertz, & Mattias Carlsson. (2024). Kremlin Leaks: Secret Files Reveal How Putin Pre-Rigged His Reelection. *VSquare.* https://vsquare.org/kremlin-leaks-putin-elections-russia-propaganda-ukraine/

Horta Ribeiro, M., Jhaver, S., Martinell, J. C. i, Reignier-Tayar, M., & West, R. (2024). *Deplatforming Norm-Violating Influencers on Social Media Reduces Overall Online Attention Toward Them* (No. arXiv:2401.01253). arXiv. https://doi.org/10.48550/arXiv.2401.01253

Hua, Y., Horta Ribeiro, M., Ristenpart, T., West, R., & Naaman, M. (2022). Characterizing Alternative Monetization Strategies on YouTube. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1–30. https://doi.org/10.1145/3555174

Hughes, H. C., & Waismel-Manor, I. (2021). The Macedonian Fake News Industry and the 2016 US Election. *PS: Political Science & Politics*, 54(1), 19–23. https://doi.org/10.1017/S1049096520000992

Innes, H., & and Innes, M. (2023). De-platforming disinformation: Conspiracy theories and their control. *Information, Communication & Society,* 26(6), 1262–1280. https://doi.org/10.1080/1369118X.2021.1994631

Ong, J., & Cabañes, J. V. (2018). Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines. *Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines*. https://doi.org/10.7275/2cq4-5396

Pamment, J., & Tsurtsumia, D. (2025). *Beyond Operation Doppelgänger: A Capability Assessment of the Social Design Agency* (No. 8). Psychological Defence

Agency. ISBN: 978-91-989646-3-9

Qurium. (2022). *Under the hood of a Doppelgäng*er. Qurium Media Foundation. https://web.archive.org/web/20250902091132/https://www.qurium.org/alerts/under-the-hood-of-a-doppelganger/

Silverman, C., Lytvynenko, J., & Kung, W. (2020, January 7). Disinformation For Hire: How A New Breed Of PR Firms Is Selling Lies Online. *Buzzfeed News.* https://web.archive.org/web/20250413131844/https://www.buzzfeednews.com/article/craigsilverman/disinformation-for-hire-black-pr-firms

Stevens, D. (2017). *Google Makes Millions off of Fake News* (Google Transperancy Project. Campaign for Accountability). https://web.archive.org/web/20250513030208/https://campaignforaccountability.org/report-google-makes-millions-from-fake-news/

The Facebook Files. (2021, September 15). *Wall Street Journal.* https://web.archive.org/web/20250613104214/https://www.wsj.com/articles/the-facebook-files-11631713039?gaa_at=eafs&gaa_n=ASWzDAh_IryYNt98jECLhggnT_aI4sHoOzB1xd2nXtRyy76XXPx0oGhmqjRnZs5n-4M%3D&gaa_ts=684c00a8&gaa_sig=IIBMaanmsAn2cOEgFFSiHxNkDFviLjXn9ylzLA_Pvf8dYQTvfd9vCd5RYiY0I5q9Pqra9P9CET9RkoatthMAdg%3D%3D

Viginum. (2023). *RRN: une campagne numérique de la manipulation de l'information complexe et persistante.* https://web.archive.org/web/20250902091606/https://www.sgdsn.gouv.fr/files/files/20230619_NP_VIGINUM_RAPPORT-CAMPAGNE-RRN_VF.pdf

Von Wienand, L., Steurenthaler, S., & Loelke, S. (2022, August 29). Infokrieg – Wie Putin-Trolle die Deutsche Öffentlichkeit Manipulieren. *T-online.* https://web.archive.org/web/20250902085226/https://www.t-online.de/nachrichten/deutschland/gesellschaft/id_100042596/ukraine-krieg-prorussische-kampagne-das-steckt-hinter-den-fake-artikeln.html

20. ELECTION INTERFERENCE: A PSYCHOLOGICAL DEFENCE PERSPECTIVE

SEBASTIAN BAY

SUMMARY

- Hybrid threats to elections blend cyberattacks, subversion (e.g., illicit political finance), and coordinated information influence activities – often via social media – to disrupt electoral processes and erode trust in democratic institutions.
- Sophisticated information influence activities pose a growing threat to the integrity of elections.
- These operations exploit psychological vulnerabilities to manipulate public opinion, influence the outcome of elections, and undermine trust in the conduct of elections.
- To safeguard electoral integrity, establish a robust and adaptive legal framework; implement proactive risk assessment and prevention; build resilience to stressors and shocks; and maintain effective systems for crisis response and recovery. This work depends on coordinated multi-stakeholder collaboration and strengthened psychological defence.

Sophisticated hybrid threats increasingly threaten the integrity of democratic elections. These operations often exploit psychological vulnerabilities to manipulate public opinion, spread disinformation and undermine trust in elections. This chapter examines election protection, highlighting the role of information influence activities in election interference.

THE BEGINNING

State-sponsored election interference has a long history. Various nations have employed diverse tactics, techniques, and procedures (TTPs) to influence electoral outcomes in other countries (Bay et al., 2022), and a range of cases from the 20th century detail foreign interference in national elections. Levin (2020) identified 117 cases of overt or covert partisan electoral interventions by the United States and the USSR/Russia between 1946 and 2000. Levin (2020) argues that electoral interventions will occur when two conditions are met: when a great power sees its interests threatened by a candidate or party during democratic elections, and then there is a significant other domestic actor who wants or can be supported (Levin, 2020).

Ohlin (2020) argues that the 2016 US presidential election was a watershed moment in our contemporary understanding of threats to elections, marking a shift in our understanding of how foreign actors interact with democratic processes in the 21st century (Ohlin, 2020). The current model of election interference,

where cyber-attacks and information influence activities in the digital domain are the centre of gravity, is a shift from previous methods. The 2016 interference wasn't the first time these methods were used, but it is probably the most studied, investigated, and researched case (Ibid.). This shift predates the 2016 US presidential election and can be traced back to, at least, the Russian interference in the Ukrainian presidential and parliamentary elections in 2014, where Russian state actors deployed similar tactics, techniques, and procedures (Bay et al., 2022). Elements of modern election interference techniques can be traced back to the advent of social media. In 2008, during the post-election violence in Kenya, social media, in combination with traditional media, played a significant role as a platform and vector for disinformation and hate speech. In the 2013 Kenyan election, Cambridge Analytica was hired by the ruling Jubilee Party to conduct a targeted advertising and disinformation campaign. Their tactics included exploiting ethnic tensions and spreading false information about opposing candidates (Wambui & Oka, 2022).

While drawing a sharp line between old and new methods is difficult, historically, foreign election interference has often focused on traditional diplomacy, propaganda, and espionage to sway public opinion and policy decisions. However, from the early 2010s, election interference has transformed into something more overt and sophisticated, leveraging advanced cyberattacks, subversive techniques, digital platforms and social media to disseminate disinformation, sow discord, undermine public trust in the electoral process, and influence the outcome of elections (Bay et al., 2022; Bay, 2024).

UNDERSTANDING THE CURRENT THREAT LANDSCAPE

Today, foreign interference targeting elections often takes the form of hybrid threats to elections. Hybrid threats are defined by the European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE) as:

> *"[...] harmful activities that are planned and carried out with malign intent. They aim to undermine a target, such as a state or an institution, through a variety of means, often combined. Such means include information manipulation, cyberattacks, economic influence or coercion, covert political manoeuvring, coercive diplomacy, or threats of military force. [...]" (Hybrid CoE, 2025).*

Recently observed foreign election interference can often be categorised as a hybrid threat because antagonistic actors combine a variety of means to undermine the integrity of democratic processes. Hybrid threats in the context of elections are diverse and sophisticated, encompassing cyberattacks, information influence activities, physical attacks and other subversive means that seek to disrupt the electoral process. These threats aim to influence public opinion, erode trust in democratic institutions, undermine the legitimacy of elections, and manipulate the outcome. Information influence activities are a core component of hybrid threats to elections. They leverage disinformation and other information influence activities to influence public discourse, manipulate voters, and undermine trust in the election process (Bay, 2024).

Hybrid threats to elections can be categorised into four main areas (Bay, 2024):

- **Threats to the Conduct of Elections:** These threats seek to disrupt the electoral process, targeting election officials, infrastructure, or logistics. Examples include physical attacks, cyberattacks, and attempts to manipulate or sabotage voting equipment or systems.

- **Threats to Trust in Elections:** These threats aim to undermine public confidence in the legitimacy of elections by spreading disinformation, promoting conspiracy theories, and creating a perception of widespread voter fraud or manipulation.
- **Threats to the Will and Ability to Vote:** These threats aim to influence voter behaviour by discouraging participation or undermining voters' voting ability. This can involve voter intimidation, spreading false information about voting procedures, or making it difficult for specific groups to register or cast their votes.
- **Threats to Political Decision-Making:** These threats aim to influence voters' voting preferences using illegitimate or illegal means. These can include inauthentic coordinated activity on digital platforms or subversive attempts to influence political decision-making, including illicit financing and subversion of candidates.

Information influence activities can manifest in several forms as part of the broader hybrid threat portfolio targeting elections. These operations frequently employ disinformation, social media manipulation, and cyberattacks to form a central aspect of the overall influence operation targeting the election. Information influence activities seek to influence the perception and behaviour of target audiences (Bay et al., 2022; Bay, 2024).

Various actors engage in information influence activities targeting elections, each with distinct motives and strategies. A range of state actors have used information influence activities to interfere in other nations' elections. These activities may be conducted directly by government agencies or through proxies such as private companies or individuals tasked with spreading disinformation or influencing public opinion. Russia, China, and Iran are notable examples of state actors involved in such operations (Ibid.).

Non-state actors, including extremist groups, political activists, and hacktivists, also partake in information influence activities. Their motives can vary but often include promoting specific political agendas, disrupting the electoral process, or undermining trust in institutions. These actors frequently manipulate social media and disinformation campaigns to further their objectives (Ibid.).

Domestic actors, such as political parties, political organisations, and individuals, are increasingly involved in information influence activities either directly or as proxies for foreign actors. Their motivations may include gaining political advantage, suppressing voter turnout, or undermining trust in the electoral system (Ibid.).

RECENT EXAMPLES OF ELECTION INTERFERENCE

Recent examples of election interference highlight the diverse tactics employed by various actors to influence electoral outcomes and public opinion. There is an ever-growing list of elections targeted by election interference. Election interference is a global phenomenon and a global challenge to democracy. This section will highlight some notable cases since 2014.

THE RUSSIAN INTERFERENCE IN UKRAINIAN ELECTIONS IN 2014

Russia's 2014 military intervention in Ukraine was the most overt form of interference in the country's democratic processes. Still, beyond the direct military engagement, Russia also carried out a broader campaign of political interference aimed at destabilising Ukraine's government and shaping its political direction.

Part of this campaign involved undermining the legitimacy of the 2014 presidential election through elaborate disinformation and cyberattacks against the election infrastructure (AFP, 2014; Clayton, 2014; Greenberg, 2017).

In late May 2014, just days before Ukraine's presidential election, a wave of cyberattacks threatened to undermine the country's voting process. A pro-Russian hacktivist group calling itself CyberBerkut, later suspected to have ties to Russia's GRU military intelligence, targeted Ukraine's Central Election Commission in a coordinated cyberattack designed to sabotage and discredit the election. The first phase of the attack began four days before Ukrainians went to the polls. Hackers infiltrated the Central Election Commission's computer infrastructure and deleted crucial files, rendering the vote-tallying system inoperable. Soon afterwards, CyberBerkut publicised its success by leaking stolen documents and emails, claiming to have "destroyed the computer network infrastructure" used for the election. Ukrainian authorities, however, managed to restore the deleted files and rebuild the system from backups within a day. CyberBerkut had also embedded malicious software that would have skewed reported results in favor of a far-right candidate. Government cybersecurity teams removed the malware 40 minutes before vote totals were scheduled to go live. Russian state media reportedly aired the fake results anyway (Ibid.).

Shortly after the polls closed, the attackers launched a large-scale distributed denial-of-service (DDoS) attack. By flooding the central election servers with massive amounts of illegitimate data requests, the hackers temporarily blocked the legitimate flow of actual voter data. This tactic reportedly delayed the final tally by about two hours (Ibid.).

Russia also actively supported pro-Russian politicians and parties within Ukraine, using these figures to advance policies aligned with Moscow's interests and counter Ukraine's pro-European aspirations. In addition, Russian oligarchs with substantial business interests in Ukraine were leveraged to strengthen Moscow's influence and interference (Lange-Ionatamišvili, 2014; Sæther, 2023).

Another element of Russia's interference was its exploitation of existing political and social cleavages in Ukraine. By amplifying divisions across different regions and ethnic groups, Russia exacerbated internal tensions. A significant component of Russia's efforts was an elaborate information influence campaign designed to sway public opinion, undermine faith in democratic institutions, and promote pro-Russian narratives. Disinformation was circulated widely via state-controlled media outlets such as RT and Sputnik, social media platforms, and online proxies. These operations sought to paint Ukraine as a failed state beset by extremism and to foster discord within its society (Ibid.).

THE RUSSIAN INTERFERENCE IN THE US PRESIDENTIAL ELECTION IN 2016

Russia's Interference in the 2016 US Presidential Election serves as a prominent example of state-sponsored election interference. Russian intelligence agencies orchestrated a multi-faceted campaign involving disinformation, hacking, and social media manipulation (DNI, 2017). These efforts aimed to influence public opinion, sow discord, and undermine trust in the electoral process. The 2016 US election interference is notable for its scale and sophistication. It employed various tactics, including cyberattacks, social media manipulation, and disinformation (Bennett &

Livingston, 2018; DNI, 2017; Howard et al., 2018; Rid, 2020).

Russian hackers targeted the Democratic National Committee and Clinton campaign emails, leaking information to damage their reputations and undermine public confidence. Russian hackers also targeted the US election administration infrastructure in a sophisticated and targeted way to undermine the credibility of the system. Russian actors also used bots, trolls, and fake accounts to spread disinformation, amplify inflammatory content, and polarise public opinion. Lastly, Russia also disseminated disinformation designed to undermine the legitimacy of the election, sow distrust in the democratic process, and both promote and demote political candidates.

The 2016 US election interference served as a wake-up call for the international community, highlighting the vulnerability of democratic processes to digital manipulation and the need for more robust defences. It catalysed research into hybrid threats and information influence activities and prompted a global effort to develop strategies for countering disinformation and election interference (Ohlin, 2020). The 2016 US election interference showcased the potential of digital platforms to manipulate public opinion and influence elections, even if the techniques had been developed and used previously. Since 2016, the number of threat actors has increased, and the tactics and techniques have evolved (Bay et al., 2022; Bay, 2024).

THE 2020 US PRESIDENTIAL ELECTION

In the aftermath of the 2020 US Presidential Election an Intelligence Community Assessment by the U.S. National Intelligence Council assessed that while no foreign actor attempted to alter or interfere with any technical aspects of the election process, a range of foreign actors, but primarily Russia and Iran, attempted to interfere with the political process (NIC, 2021). In the year leading up to the election, Meta countered a dozen influence operations targeting US audiences. According to Meta, there was an equal distribution of networks from Russia, Iran and the United States, underscoring the foreign and domestic nature of election interference (Gleicher et al., 2021). The #StopTheCount Campaign during the 2020 US Presidential Election represents another significant case of election interference. This primarily domestic campaign spread disinformation about voter fraud and systemic irregularities, aiming to undermine confidence in the election results. The disinformation campaign culminated in the storming of the US Congress on January 6, 2021, in an attempt to stop the certification of the election results (Bay, 2024; Center for an Informed Public et al., 2021).

THE 2024 EUROPEAN PARLIAMENT ELECTION

The European Digital Media Observatory's (EDMO) Task Force on the 2024 European Parliament Election reported widespread dissemination of false narratives in at least 11 European countries during 2023. These narratives primarily focused on accusations of voter fraud, foreign influence, and alleged unfair practices. The intent behind these falsehoods appeared to be to undermine the legitimacy of democratically elected representatives and discredit the election system. 2023 also saw the emergence of AI-generated evidence fabricating election fraud, further complicating efforts to maintain public trust in the electoral process (Panizio, 2023).

According to the Task Force's final report, disinformation surrounding the elections had become increasingly sophisticated and geographically widespread by late 2023. In particular, the share of EU-focused disinformation detected by fact-checkers rose sharply from 5% in January 2024 to 15% by May 2024. Many of these falsehoods repurposed known conspiracy tropes for the electoral context—alleging "rigged" voting processes, fanning climate-change denial, or portraying migrants as malicious forces "seizing power" in Europe. According to EDMO, Russia emerged as a major state actor behind these efforts, using multiple disinformation campaigns to undermine trust in the EU's institutions and to weaken public support for Ukraine. These operations ranged from "Operation Matrioska," which manipulated fact-checkers and media organisations with forged leads, to "Pravda," a growing network of copycat websites impersonating legitimate media outlets uncovered by Viginum, the French agency for countering information influence activities (EDMO, 2024).

EU member states were also targeted by the Russian-backed Doppelgänger campaign, aimed at disrupting the electoral process across France, Germany, Italy, Poland, and Spain. This operation involved creating fake domains that mimicked legitimate news sources and operating a network of 47 inauthentic websites to spread disinformation. By impersonating well-known outlets, the Doppelgänger campaign injected confusion into the election discourse, casting doubt on genuine polling data and stoking fears of foreign meddling. In several instances, threat actors combined these deceptive websites with AI tools to fabricate "evidence" of vote tampering, compounding challenges for fact-checkers and election administrators. Overall, the Task Force's findings underscore how evolving disinformation techniques, especially the strategic use of AI, can pose a substantial risk to voter confidence and electoral integrity (EU DisinfoLab 2024; EDMO 2024).

THE RUSSIAN INTERFERENCE IN THE 2024 PRESIDENTIAL ELECTION IN MOLDOVA

The 2024 presidential election and concurrent EU membership referendum in Moldova were marked by a complex spectrum of interference tactics originating both from abroad, chiefly Russia, and domestic actors. Russia's concerted efforts combined financial and cyber-based subversion with more traditional forms of voter manipulation. Investigations by Moldovan authorities revealed illicit external funding funnelled to pro-Russian candidates, as well as major vote-buying schemes (Bryjka, 2024; RFL/RL, 2024b; Olari et al., 2024; OSCE, 2024; Ntousas & Pleșca, 2024).

In tandem, cyber intrusions targeted state agencies and Moldova's Central Electoral Commission, with repeated attempts to breach voter databases, disrupt government websites, and spread false information via phishing and malware campaigns. These intrusions were partly attributed to Russia-linked groups using "hack-and-leak" operations, wherein stolen or fabricated data were strategically published to discredit pro-European candidates. At the same time, large-scale disinformation campaigns spread using social media and contained narratives aiming, among other things, to stoke fears around EU accession (Antoniuk, 2024; Bryjka, 2024; Ntousas & Pleșca, 2024).

Domestically, illegitimate practices included "carousel voting," where groups were allegedly bused to multiple polling stations to cast repeated votes. Additionally, intimidation tactics, a surge of bomb threats at polling stations, AI-generated

death threats, and an attempted arson at the Central Election Commission, were possibly used to discourage election turnout and undermine the election (ibid.).

Analysts have noted, however, that the sheer breadth of the interference, spanning covert financing, cyberattacks, social media manipulation, and on-the-ground voter intimidation, represents a disturbing template for hybrid threats to elections. Thus, Moldova's experience highlights the need for robust defences against sophisticated election interference (Ibid.).

A GLOBAL PROBLEM

Even if the cases above single out recent interference in Western elections, election interference is a global problem and information influence activities target elections on all major continents. Scholars have documented how electoral processes worldwide—in Asia, Africa, or Latin America—face diverse and evolving manipulation tactics. Research highlights that such tactics range from orchestrated disinformation campaigns and malicious cyberattacks to covert financial operations and misleading advertising strategies. They often draw on extensive influence networks spanning domestic and foreign actors, sometimes using paid advertising or clandestine "troll armies" to sway citizens' perspectives (cf. Corpus Ong & Cabañes, 2019; Corpus Ong & Tapsell 2021; Gleicher et al., 2021).

INFORMATION INFLUENCE TACTICS AND TECHNIQUES

The tactics and techniques employed in information influence activities targeting elections are multifaceted, sophisticated, and constantly evolving. One key strategy involves the fabrication of content that mimics reputable media sources. Malicious actors craft fake news stories, manipulate images, and generate tailored disinformation to mislead the public and influence voter perceptions. This is further amplified by the use of advanced AI technologies, which enable the creation of realistic fake videos, audio recordings, and other synthetic media, known as deepfakes. These deepfakes can be used to spread false narratives, manipulate public opinion, and damage reputations, often by presenting fabricated evidence as genuine (EU DisinfoLab, 2024; Gorman & Livine, 2024).

Amplification and manipulation of this content are achieved through various methods. Malicious actors utilise vast networks of fake accounts, bots, troll farms, and cyborg accounts to spread disinformation widely and rapidly on social media platforms. This strategy aims to manipulate public discourse and sway opinions by exploiting algorithms and user networks. Data-driven strategies also segment voters into specific groups for personalised political advertisements. This technique, often fuelled by the collection and analysis of personal data, targets vulnerable populations with tailored disinformation, exacerbating divisions and influencing voting behaviour. Another tactic, astroturfing, involves creating the illusion of grassroots support or opposition for a candidate or issue by orchestrating fake social movements. This deceptive tactic manipulates public perception by presenting artificial public sentiment as organic and authentic, influencing electoral outcomes (Bay et al. 2022; Bay, 2024; cf. Gleicher et al., 2021).

Cyber operations are increasingly used as an active component in information influence activities targeting elections. Malicious actors engage in hack-and-leak operations, stealing sensitive information and strategically leaking it to damage reputations and influence voter perceptions. This tactic combines hacking with disseminating leaked data, often timed to maximise impact on public opinion

during election cycles. Cyberattacks targeting the technological components of the electoral process, such as voting systems, voter registration databases, and election-related communication systems, aim to disrupt the electoral process, create confusion, undermine trust in the system, and potentially manipulate election results (Bay et al. 2022; Bay, 2024).

Distributed Denial of Service (DDoS) attacks are launched against election authority websites and electronic voting systems to overwhelm them and disrupt services, creating confusion and delays in the electoral process and potentially hindering voter access and confidence in the system. Spear phishing attacks, personalised phishing attacks against election officials and politicians, aim to gain unauthorised access to sensitive information, compromising the integrity of election-related data and communications and potentially influencing elections or undermining public trust (Ibid.).

Manipulative strategies further complicate the landscape. Influence-for-hire services employ third-party firms to conduct hacking and disinformation operations. They provide specialised services that make it difficult to trace the source of influence campaigns and protect the primary actors from direct attribution and accountability. These actors also fuel societal polarisation by spreading real or manipulated information that heightens partisanship and distrust in democratic governance. This strategy creates and deepens conflicts within the electorate, making it harder for society to reach a consensus on critical issues and potentially undermining the legitimacy of electoral outcomes (Bay et al. 2022; Bay, 2024; Corpus Ong & Cabañes, 2019; cf. Gleicher et al., 2021).

WHY ARE ELECTIONS VULNERABLE?

Elections are inherently susceptible to information influence activities due to a confluence of factors. The high stakes involved in elections, the compressed timeframe within which campaigns occur, and the inherent trust placed in the electoral process by the public all contribute to the vulnerability (cf. Alihodžić, 2023; Bay, 2022)

This vulnerability is exacerbated by the often low general understanding of the election system among the public, the media, and politicians. Misunderstandings or a lack of knowledge about election procedures can be exploited to sow confusion and doubt about the legitimacy of election outcomes (Bay et al., 2022; Bay, 2024; cf. Pamment et al., 2018).

Furthermore, the election infrastructure itself can present numerous vulnerabilities. Voting machines, voter registration databases, election logistics and online platforms can all be targeted by malicious actors seeking to disrupt or manipulate the electoral process. If inadequately protected, these systems provide entry points for cyberattacks and other forms of interference. Any lack of robust cybersecurity measures further exacerbates these vulnerabilities, leaving election systems and political parties open to attacks that can compromise the integrity of elections and the associated data (cf. Alihodžić, 2023; Bay et al., 2022; Bay, 2024, cf. Van der Staak & Wolf, 2019).

These vulnerabilities highlight the complex and multifaceted nature of protecting elections from information influence activities. Addressing these challenges requires comprehensive strategies to bolster public trust, secure election infrastructure, and enhance media literacy among the electorate.

MITIGATION STRATEGIES

The increasing vulnerability of elections to information influence operations necessitates a multifaceted approach to safeguarding democratic processes. To counter these threats, nation-states, international organisations, civil society, and researchers have identified a range of mitigation strategies, including e.g. (cf. Bateman & Jackson 2025; Bay, 2024; Bay. 2025; Pamment et al., 2018; Pamment & Isaksson, 2024; Wilson, Johnson & Levine, 2023).

1. Improving Election Protection
   Assessing and preventing electoral risks, building resilience against stress and shocks, and establishing efficient crisis response and recovery mechanisms to ensure the integrity of its electoral processes. This includes conducting regular audits of election systems and post-election evaluations to identify areas to improve for future elections.
2. Enhancing Transparency and Public Awareness
   Improving transparency in the conduct of elections to enable source criticism and informed decision making. Educating the public about the tactics and techniques used by antagonists. This can involve teaching citizens critical thinking skills, helping them identify and evaluate sources of information, and promoting media literacy.
3. Enhancing Media Resilience
   Supporting media outlets to combat disinformation by enhancing their ability to identify election-related disinformation, promoting fact-checking initiatives, and encouraging ethical reporting standards.
4. Collaboration and Coordination
   Fostering collaboration and coordination between government agencies, law enforcement, cybersecurity experts, and election authorities to share information and coordinate responses to contribute to more resilient elections.
5. Conducting Exercises
   Regularly testing and evaluating election systems and procedures for countering potential threats to elections through tabletop exercises and simulations.
6. Legislative Action
   Strengthening legislation to address disinformation and manipulation, criminalise interference in elections, and protect election officials from harassment.
7. Dedicated Resources for Law Enforcement
   Allocating dedicated resources and establishing specialised units within police and prosecution services to combat illegal election interference. This involves training law enforcement officers and prosecutors, dedicating resources to enable effective investigations, and ensuring coordination between these units and other government agencies for timely and effective responses to threats against the electoral process.
8. Cybersecurity Measures
   Enhancing the cybersecurity of election infrastructure, election systems, media and political parties to protect against cyberattacks and data breaches. This includes investing in robust security systems, training election officials and staff on cybersecurity best practices, and developing incident response plans.

9. Fact-checking and Debunking
   Developing and supporting fact-checking organisations that can identify and debunk false and misleading information. This is closely related to providing citizens with credible sources of information, highlighting the tactics used by information influencers, and promoting a culture of critical thinking, as highlighted in point number 2.
10. Strategic Communication Planning
    Developing coordinated messaging strategies to preempt and respond to disinformation, particularly by anticipating evolving narratives that foreign or malicious actors may push.
11. Removing Inauthentic Networks
    Encouraging rapid detection and takedown of accounts or pages engaged in coordinated inauthentic behavior, thereby reducing the reach and impact of disinformation campaigns.
12. Threat Intelligence
    Centralising or coordinating the expertise of relevant agencies and stakeholders (including intelligence, military, and civilian bodies) to identify, analyse, and trackforeign and domestic information influence activities.
13. Deterrence
    Combining public disclosures (such as naming-and-shaming), legal actions, and asset freezes and other offensive actions against malicious foreign actors to increase the costs of hybrid threats to elections.

While far from exhaustive, these mitigation strategies can contribute to a more secure and trustworthy electoral process, enhancing the resilience of democratic systems against information influence activities.

MITIGATION STRATEGIES DEPLOYED BY ELECTION MANAGEMENT BODIES TO COUNTER ELECTION-RELATED DISINFORMATION

Research by Asplund & Casentini (2024) shows that ahead of recent elections and referendums in Argentina (2023), Australia (2022 and 2023), Canada (2019 and 2021), Germany (2021), and South Africa (2021) electoral management bodies have adopted a wide range of comprehensive approaches to counter election-related disinformation through real-time monitoring, transparent information dissemination, and collaboration with social media platforms (Asplund & Casentini, 2024).

In Australia, the Australian Electoral Commission launched a “Stop and Consider” campaign, urging voters to question the reliability and timeliness of electoral information, and introduced a disinformation register detailing instances of false content and corresponding remedial actions. In Argentina, the National Electoral Chamber partnered with Meta to improve transparency in political advertising. It also introduced a WhatsApp chatbot to supply voters with accurate information about polling stations, voter eligibility, and complaint procedures in an attempt to mitigate the spread of misinformation. Canada’s Chief Electoral Officer established a Social Media Monitoring Unit to detect and remove misleading narratives regarding voting procedures, a function reinstated in 2021 and extended to multiple online platforms, aligning with the mandate of its Electoral Integrity Office to strengthen public confidence in elections (Ibid.).

In Germany, the Federal Returning Officer identified and corrected widespread false claims circulating about the electoral process. At the same time, Agence France-Presse collaborated with Facebook to create a chatbot enabling users to verify digital materials with editorial support. The South African Electoral Commission worked with Media Monitoring Africa and various social media platforms during the 2021 municipal elections to detect misleading advertisements and swiftly address reported cases through a cooperative system that allowed stakeholders to report suspected disinformation (Ibid.).

In Sweden, authorities have established dedicated national, regional, and local election protection networks to coordinate efforts among key agencies and stakeholders. These networks actively identified and addressed security risks such as disinformation, cyberattacks, and physical threats. They also aim to strengthen resilience through targeted training programmes, specialised learning resources, public awareness initiatives, and enhanced cybersecurity. Furthermore, they have developed crisis response and recovery mechanisms that focus on rapid information sharing and coordinated action across various levels of government. Central to Sweden's strategy is a collaborative, whole-of-society approach, seeking to build broad societal resilience to disinformation and other forms of election interference (Bay, 2025).

In Moldova, authorities and allied institutions adopted a coordinated, multi-level strategy to safeguard election integrity amid an onslaught of hybrid threats in 2024. Government agencies launched investigations into vote-buying and illicit campaign financing, clamped down on cyberattacks aimed at critical infrastructure, and ramped up public awareness campaigns to help voters recognise and resist disinformation. Specialised units within law enforcement and intelligence services worked with election officials to address emerging risks, foster swift information exchange, and ensure a rapid response to online and offline threats. The government fortified its cybersecurity posture by partnering with international experts, reinforcing legal frameworks, and classifying essential electoral systems for prioritised protection. Through these measures, supported by legal action, greater resources for independent oversight, and targeted crisis management planning, Moldova strove to bolster overall societal resilience and uphold trust in democratic processes despite persistent foreign interference. These efforts were further amplified by international partners, who offered training, cybersecurity assistance, and sanctions on malign actors, underlining a shared commitment to protecting Moldova's elections (Bryjka, 2024; RFL/RL, 2024a; OECD, 2024; McGrath, 2024).

At the European Union level, the Commission has introduced a series of measures aimed at reinforcing democracy, ensuring free and fair elections, and safeguarding electoral rights for EU citizens. This includes enacting the new Regulation on the transparency and targeting of political advertising in April 2024, strengthening rules to protect the integrity of elections and bolstering public trust in political campaigns. Underpinning these efforts are the European Democracy Action Plan and the Defence of Democracy Package, which focus on enhancing media freedom, combating disinformation, and engaging citizens in public life. The European cooperation network on elections and the updated Code of Conduct for the 2024 European Parliament elections further reinforced these initiatives. By promoting transparency, supporting civic participation, and encouraging robust national election networks, the Commission seeks to nurture a more resilient democratic environment throughout the Union (European Commission, 2024).

Collectively, these measures underscore how election authorities and national and regional bodies have leveraged a broad toolset - from advanced monitoring to transparent communication strategies and institutional partnerships - to try to safeguard the integrity of democratic processes from a broad range of election interference.

CONCLUSION

Information influence activities are a significant threat to the integrity of elections. These activities can undermine public trust, distort the electoral process, and influence election outcomes. To safeguard democratic processes and ensure that elections are free and fair, it is essential to understand the nature of information influence activities, identify the threat actors involved, and develop comprehensive strategies for countering these threats. The approaches outlined in this chapter provide an overview of methods to build more resilient and secure elections.

Election protection requires a multi-faceted approach that combines technical security measures, legislative action, public awareness campaigns, and psychological defence strategies. Proactively addressing vulnerabilities, fostering collaboration, and promoting media literacy can mitigate the impact of information influence activities and safeguard the integrity of democratic elections.

The psychological dimension of election protection is crucial. By understanding the vulnerabilities exploited by information influence activities and implementing appropriate psychological defence strategies, we can safeguard the integrity of democratic elections and ensure a fair and informed electorate. This requires ongoing vigilance, robust legislative and institutional measures, a collective effort to strengthen investigative media, and promoting media literacy, transparency and critical thinking. By embracing a multi-faceted approach, it is possible to strengthen the defences against the growing challenge of hybrid threats to elections to safeguard future elections and democracy.

DISCUSSION

- How can we strengthen public trust in elections, and what role do public awareness campaigns play in this effort?
- Which strategies effectively counter information influence activities and other hybrid threats targeting elections?
- In what ways do psychological vulnerabilities contribute to the spread of election-related disinformation, and how can we address them?
- What practical steps can governments, election authorities, and citizens take to safeguard elections?
- Is it possible to build an election system that is less vulnerable to information influence activities and other hybrid threats? How can it be done?

SEBASTIAN BAY *is a researcher specialising in election security, hybrid threats, and disinformation. His expertise in elections stems from managing various projects aimed at reducing electoral risks, enhancing resilience, and establishing effective response and recovery mechanisms for Sweden's general elections in 2018 and 2022. Sebastian has also authored numerous reports on election security and hybrid threats for the European Centre of Excellence for Countering Hybrid Threats, the Swedish Defence Research Agency (FOI), the NATO Strategic Communications Centre of Excellence, and International IDEA. He holds a master's degree in Political Science, a bachelor's degree in Politics and Economics, and a bachelor's degree in Intelligence Analysis, all from Lund University, Sweden.*

REFERENCES

Alihodžić, S. (2023). *Protecting Election: Risk Management, Resilience-Building and Crisis Management in Elections.* Stockholm, Sweden: International IDEA.

Antoniuk, D., (2024). *Moldova's government hit by flood of phishing attacks*. Somerville, U.S.: The Record. Recorded Future News. https://therecord.media/moldovas-government-hit-by-flood-of-phishing-attacks

Bastos, M. T., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review,* 37(1), 38–54.

Bay, S., Appelgren, J., Isaksson, E., Lindgren, J., & Thunholm, P. (2022). *Hot mot svenska allmänna val - Exempel och scenarier för valadministrationen.* Stockholm: FOI.

Bay, S. (2024). *Countering hybrid threats to elections: From updating legislation to establishing collaboration networks.* Hybrid CoE Research Report 12. Helsinki: European Centre of Excellence for Countering Hybrid Threats.

Bay, S. (2025). *Protecting Electoral Integrity: The Case of Sweden.* Stockholm: International IDEA.

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication,* 33(2), 122–139.

Bryjka, F. (2024). *Russian interference nearly overwhelmed Moldovan presidential election-referendum vote.* Warsaw, Poland: Polish Institute of International Affairs. https://pism.pl/publications/russian-interference-nearly-overwhelmed-moldovan-presidential-election-referendum-vote

Corpus Ong, J., & Cabañes, J. (2019). *Four Work Models of Political Trolling in the Philippines.* Riga: NATO Strategic Communications Centre of Excellence.

Corpus Ong, J., & Tapsell, R. (2021). *Mitigating Disinformation in Southeast Asian Elections.* Riga: NATO Strategic Communications Centre of Excellence.

European Commission. (2024). *Democracy and electoral rights.* https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/democracy-eu-citizenship-anti-corruption/democracy-and-electoral-rights_en

European Digital Media Observatory (EDMO). (2024). *Final report – Outputs and outcomes of a community-wide effort.* Florence, Italy: European Digital Media Observatory's Task Force on the 2024 European Parliament Elections.

EU DisinfoLab. (2024). *What is the Doppelganger operation? List of resources. Brussels, Belgium: EU DisinfoLab.* https://www.disinfo.eu/doppelganger-operation/

Director of National Intelligence (DNI). (2017). *Assessing Russian Activities and Intentions in Recent US Elections.* Washington D.C., U.S.: Office of the Director of National Intelligence.

Hybrid CoE (2025). Hybrid threats as a phenomenon. Helsinki: European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE). https://www.hybridcoe.fi/hybrid-threats-as-a-phenomenon/

Fjällhed, A., Pamment, J., & Bay, S. (2021). A Swedish perspective on foreign election interference. In J. D. Ohlin & D. B. Hollis (Eds.), *Defending democracies: Combating foreign election interference in a digital age.* New York, U.S.: Oxford Academic.

Gleicher, N., Franklin, M., Agranovich, D., Nimmo, B., Belogolova, O., & Torrey, M. (2021). *Threat report: The state of influence operations 2017–2020*. Menlo Park, U.S.: Meta.

Gorman, L. & Levine, D. (2024). *The ASD AI Election Security Handbook*. Washington D.C.: The Alliance for Securing Democracy (ASD) at the German Marshall Fund of the United States (GMF).

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & Francois, C. (2018). *The IRA, social media and political polarization in the United States, 2012–2018.* Oxford, U.K.: Computational Propaganda Research Project.

Lange-Ionatamišvili, E., 2014. *Analysis of Russia's information campaign against Ukraine.* Riga, Latvia: NATO Strategic Communications Centre of Excellence.

Levin, D. H. (2020). *Meddling in the ballot box: The causes and effects of partisan electoral interventions.* Oxford, U.K.: Oxford university press.

McGrath, S. (2024). *Moldova narrowly votes to secure path toward EU membership after accusing Russia of interference.* Chisinau, Moldova: Associated Press. https://apnews.com/article/moldova-elections-eu-referendum-russia-325cb2c13beb1d76565a6e2aadef971a

National Intelligence Council (NIC). (2021). *Intelligence Community Assessment. Foreign Threats to the 2020 US Federal Elections.* Washington D.C., U.S.: National Intelligence Council.

Ntousas, V., Pleșca, L. (2024). *Russian meddling in Moldova.* Bucharest, Romania: German Marshall Fund of the United States.https://www.gmfus.org/news/russian-meddling-moldova

Ohlin, J. D. (2020). *Election Interference: International Law and the Future of Democracy.* Cambridge, U.K.: Cambridge University Press.

Olari, V., Calmis, D., Gigitashvili, G. (2024). *Malign interference in Moldova ahead of presidential election and European referendum.* Chicinau, Moldova: Atlantic Council DFRLab. https://dfrlab.org/2024/10/18/malign-interference-moldova/

Organisation for Security and Co-operation in Europe (OSCE). (2024). *Moldova, Presidential Election and Constitutional Referendum, 20 October 2024: Statement of Preliminary Findings and Conclusions.* Chisinau, Moldova: Organization for Security and Co-operation in Europe (OSCE). https://www.osce.org/odihr/elections/moldova/578815

Pamment, J., Nothhaft, H., Agardh-Twetman, H., & Fjällhed, A. (2018). *Countering information influence activities: The state of the art.* Lund, Sweden: Lund University.

Pamment, J., & Isaksson, E. (2024). *Psychological defence: Concepts and principles for the 2020s.* Karlstad, Sweden: Myndigheten för psykologiskt försvar.

Panizio, Enzo. *Disinformation Narratives during the 2023 Elections in Europe.* Florence, Italy: European Digital Media Observatory, 2023.

Radio Free Europe/Radio Liberty (RFL/RL). (2024a). *Moldovan Authorities Tell Voters To Ignore 'Fake' Messages Ahead Of Runoff.* Chisinau, Moldova: Radio Free Europe/Radio Liberty. https://www.rferl.org/a/moldova-election-sandu-stoianoglo-corruption-russia/33184310.html

Radio Free Europe/Radio Liberty (RFL/RL). (2024b). *Moldova says it has uncovered a Russian-funded voter-rigging plot.* Chisinau, Moldova: Radio Free Europe/Radio Liberty. https://www.rferl.org/a/moldova-election-sandu-shor-russia-eu-vote-rigging/33145103.html

Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare.* New York, U.S.: Farrar, Straus and Giroux.

Sæther, T. (2023). War of Broken Fraternity: Competing Explanations for the Outbreak of War in Ukraine in 2014. *The Journal of Slavic Military Studies,* 36(1), 28–56.

Center for an Informed Public, Digital Forensic Research Lab, Graphika, & Stanford Internet Observatory (2021). *The Long Fuse: Misinformation and the 2020 Election.* Stanford, U.S.: Election Integrity Partnership.

Van der Staak, S., & Wolf, P. (2019). *Cybersecurity in Elections. Models of Interagency Collaboration.* Stockholm, Sweden: International IDEA.

Wambui, W., & Oka, J. (2022). *Disinformation in Kenya's Political Sphere: Actors, Pathways and Effects.* Nairobi, Kenya: KICTANet & CIPESA.

Wilson, R. D., Johnson, K., & Levine, D. (2023). *Lessons from other democracies: Ideas for combatting mistrust and polarization in US elections.* Washington DC, U.S.: Alliance for Securing Democracy at the German Marshall Fund & Election Reformers Network.

21. THE LVU DISINFORMATION CAMPAIGN AGAINST THE SOCIAL SERVICES IN SWEDEN

MAGNUS RANSTORP

SUMMARY

- Disinformation as a Strategic Tool: The LVU campaign was a deliberate disinformation effort that falsely claimed Swedish authorities systematically abduct and assimilate Muslim children. This narrative quickly spread from domestic social media groups to international platforms, amplifying distrust in Swedish social services.
- Collaboration Between Domestic and International Actors: The campaign gained traction through a combination of Swedish activists, radical Islamist influencers, and foreign media outlets like Al Jazeera. Figures such as Shuounislamiya played a crucial role in framing LVU as part of a global conspiracy against Muslims, significantly damaging Sweden's international reputation.
- Exploitation of Emotional and Religious Narratives: By using emotionally charged content—such as videos of crying children and distressed parents—the campaign leveraged fear and anger, making it difficult for authorities to counter with factual explanations. The messaging fueled an "us vs. them" mentality, portraying Swedish institutions as oppressive forces targeting Muslims.
- Security and Societal Consequences: The campaign has deepened mistrust between immigrant communities and Swedish authorities while increasing threats and hate against social workers. Additionally, Sweden's security risks have heightened, as radical Islamist groups have used the campaign, along with Quran burnings, as justification for potential attacks.
- The Need for Stronger Counter-Disinformation Strategies: The campaign revealed Sweden's vulnerabilities in handling international disinformation. Authorities must improve communication strategies, particularly with local Arabic-language media and international outlets, to challenge false narratives and rebuild trust with affected communities.

The LVU campaign, which targeted Swedish social services and their work under the law on the care of young people (LVU), is a clear example of how disinformation campaigns can take root nationally and then spread globally. The campaign quickly mobilized forces in Sweden and attracted attention from international actors who, through social media, spread false narratives claiming that Swedish authorities systematically abduct and assimilate Muslim children. A crucial factor in the early

spread of the campaign was a Swedish Facebook group that organized opinion and protests in Sweden, while social media figures like Shuounislamiya and other international profiles and foreign media amplified the campaign globally. This dynamic collaboration and co-creation between domestic and international actors gave the campaign significant global traction.

This chapter is a summary of a 2023 study[62] conducted by the Swedish Defence University on behalf of the Psychological Defence Agency, focusing on the disinformation campaign against Swedish social services, which began to spread in early 2022. The study analyzes the origins, structure, and spread of the campaign, as well as the role both domestic and international actors played in its success. Special attention is given to how Swedish public opinion was mobilized on social media and how this mobilization connected to international forces, as well as the long-term consequences for Swedish societal security.

THE RISE OF THE LVU CAMPAIGN AND SWEDISH MOBILIZATION

The domestic protest movement surrounding the LVU campaign began as an expression of dissatisfaction and distrust toward Swedish social services. Particularly among certain immigrant families, primarily those of Muslim background, there was growing concern that children were being removed without valid reasons. This concern stemmed from longstanding distrust of Swedish authorities in socially vulnerable areas, where residents often felt marginalized or unfairly treated.

A Swedish Facebook group, “Children’s Rights, My Rights,” was founded in this context during the summer of 2021 and quickly became a central platform for people wanting to share their stories about social services and seek support from others in similar situations. The group rapidly grew, attracting 30,000 members united in highlighting what they perceived as unjust child removals from their families under LVU. By using Facebook’s algorithms and features to spread emotionally charged posts, videos, and images, the group helped create a powerful narrative that Swedish authorities were “kidnapping” children from immigrant families. The discussion in the Facebook group was often marked by a hostile tone, with social workers and foster parents being targeted, which reinforced distrust and fuelled the spread of disinformation.

This group also became a platform for organizing protests and demonstrations across Sweden, where people could publicly express their dissatisfaction. These demonstrations attracted media attention, creating a platform where affected individuals could voice their frustrations and share their stories. The Swedish public opinion against social services was thus built through a combination of social media and physical gatherings. The presence of foreign media also meant that the coverage spread from the demonstrations outwards in the Muslim world and came back to Sweden like a communication ’boomerang’ to residents of socially disadvantaged areas which amplified the conspiratorial narrative.

[62] Magnus Ranstorp & Linda Ahlerup (2023), The LVU Campaign: The disinformation campaign against Swedish social services: Disinformation, conspiracy theories and domestic-international links related to malign influence activities by non-state actors.Swedish Defence University. https://fhs.diva-portal.org/smash/get/diva2:1874819/FULLTEXT01.pdf

The political party Nyans, led by Mikail Yüksel, also made the LVU issue a central part of its 2022 election campaign. The party was an early participant in the demonstrations and promised to raise the LVU issue politically, giving the campaign increased legitimacy. Yüksel and other party members were regular speakers at demonstrations and garnered significant support from the Facebook group.

Another key figure is George Touma, a barber and social media personality with over 1.4 million followers. In January 2022, he published a viral video about a Syrian family in northern Sweden whose children had been taken by social services. The video spread widely, both nationally and internationally, and added emotional intensity to the campaign. Although Touma later distanced himself from the hate campaign, he continued to post content about the LVU issue.

Several actors from the radical Islamist environment, such as Abu Raad and Raad al-Duhan, contributed to spreading disinformation by claiming that Swedish social services were selling children to paedophiles and deliberately trying to destroy Muslim families. Their content, often shared by international influencers like Moustafa El-Sharqawy (known as Shuounislamiya), circulated widely on social media, contributing to the global spread of the campaign.

The radical Islamist group Hizb ut-Tahrir also became involved, organizing demonstrations and spreading narratives that LVU is part of a Swedish secularization and assimilation policy targeting Muslims.

STRATEGIES AND NARRATIVES IN THE CAMPAIGN

The campaign relied on several interwoven narratives that played on feelings of fear, anger, and frustration. One of the most prominent claims was that social services deliberately targeted children from immigrant families, particularly Muslims, to assimilate them into a secular lifestyle. These false narratives, alleging that “Swedish authorities kidnap Muslim children,” fuelled fears that the government had a hidden agenda to erase the Muslim identity of children, portraying the campaign as a state-sanctioned attack on Islam.

Emotionally charged posts and videos, often featuring crying parents and children, spread through social media groups and stirred strong reactions among their audiences. These stories, which depicted social services as oppressive, made it difficult for Swedish authorities to counter the campaign with factual explanations of the LVU process. The disinformation exploited a “us versus them” rhetoric, portraying Swedish authorities as oppressors and immigrant families as victims, which effectively mobilized sympathy and anger.

The campaign also tapped into fears of discrimination, Islamophobia, and the oppression of minorities, exploiting the existing tensions between religious and cultural expectations and Swedish legal principles. By portraying the message that “your children could be next,” the campaign created strong mobilization, not only amplifying distrust of social services but also connecting to broader radical Islamist narratives about the West being in conflict with Islam.

The narratives were also adapted to Swedish media coverage and the actions of authorities, where claims that this was a disinformation campaign were interpreted as attempts to suppress debate. This led to the creation of a strategic narrative that “the state is evil,” further deepening distrust and polarization within society.

INTERNATIONAL ACTORS' INVOLVEMENT

A crucial factor in the campaign's international spread was the involvement of foreign actors who amplified the domestic voices. Moustafa El-Sharqawy, known as Shuounislamiya, became a central figure in spreading the campaign globally. With a large following on platforms like Telegram, YouTube, and others, he focused on specific LVU cases and framed them as part of a broader conspiracy against Muslims in Sweden. El-Sharqawy, who warns Muslim women and youths against traveling to the West due to its perceived moral decay, scanned Western media for misconduct, which he highlighted and discussed on his social platforms.

In January 2022, he posted a viral video in which a Syrian man claimed that his children had been "kidnapped" by Swedish social services. The video spread widely in the Arab world and quickly became a cornerstone of the campaign. What began as a national protest in Sweden gained international traction through El-Sharqawy's platform, where he used his influence to bolster false claims that Sweden was oppressing Muslims.

Other international actors, including religious leaders and both radical and moderate Islamist preachers, quickly picked up on the campaign to push their own agendas. By framing the LVU removals as part of a global conspiracy against Muslims, they strengthened the religious dimension of the campaign and gave it additional legitimacy in the Muslim world. Foreign media outlets like Al Jazeera and TRT Arabic reported uncritially on the campaign, and hundreds of millions of viewers across the Middle East were exposed to the narrative that "Muslim children are being kidnapped" by Swedish authorities — a narrative that often went unchallenged.

The petition "In Defence of Children," signed by 142 religious leaders and prominent individuals in the Arab world, called for children taken by social services to be returned to their families and suggested forming committees to act as intermediaries between authorities and immigrant communities. It also supported the Nordic Committee for Human Rights (NCHR) in its work to protect family units and limit the influence of social services. The petition further called for an end to social services' involvement in how families raise their children in religion and morality and proposed that immigrant communities be allowed to run their own boarding schools and orphanages under their supervision and with their educational plans.

COLLABORATION BETWEEN DOMESTIC AND INTERNATIONAL ACTORS

The interaction between Swedish protest groups and international actors illustrates how social media can serve as a bridge between domestic movements and international influence campaigns. Swedish activists used social media to share their personal stories and experiences, and these quickly gained international attention when shared by foreign influencers and media figures. This dynamic led to a situation where a primarily local issue soon took on massive international dimensions.

The connection to international actors, particularly through Shuounislamiya, enabled the campaign to reach a global audience, further damaging Sweden's image in many Muslim-majority countries. By leveraging the power of social media,

these actors were able to spread a narrative that became increasingly difficult to refute due to its emotional appeal and rapid spread.

IMPACT OF DISINFORMATION

The disinformation in the LVU campaign was particularly effective because it was rooted in deeply emotional narratives that were difficult to counter with traditional information campaigns. By focusing on crying children and distressed parents, the campaign succeeded in creating an image of Sweden as a state that systematically oppressed Muslims by breaking up their families.

The combination of domestic protests, organized through social media, and international channels like Shuounislamiya and Al Jazeera, created a feedback loop where interviews with activists and stories about LVU removals were amplified and legitimized by international actors. This helped the campaign expand beyond Sweden's borders and become an international issue, affecting Sweden's global reputation.

CONSEQUENCES AND LESSONS LEARNED

The spread of the LVU campaign has had serious consequences for Sweden, both domestically and internationally. It has deepened the mistrust between immigrant communities and Swedish authorities while damaging Sweden's international reputation, particularly in Muslim-majority countries. The campaign has also led to threats and hate directed at social workers and other officials within social services, worsening the security situation for Swedish authorities.

A key lesson from the campaign is that disinformation spread through social media is extremely difficult to combat, especially when it is based on emotional and religious narratives. It requires not only swift and accurate information from authorities but also long-term efforts to build trust between authorities and residents in the most affected areas. Social workers need to work more directly with communities, building trust and explaining how the LVU process works.

At the same time, Sweden must improve its strategies for countering international disinformation and protecting its global reputation. The LVU campaign, combined with a series of Quran burnings in 2023, has heightened security risks, drawing attention from several terrorist groups focused on targeting Sweden.

Another important takeaway is that foreign media, such as Al Jazeera, played a crucial role in the spread of the campaign through uncritical reporting and the use of hashtags aligned with key actors. Foreign media outlets played a significant role in amplifying the campaign, but there has been a clear lack of nuanced voices, allowing much of the disinformation to go unchallenged. This highlights the need for better relations between Swedish authorities and international media to combat disinformation and provide a balanced perspective. The LVU campaign also exposed a deep mistrust of Swedish authorities in certain vulnerable areas and the existence of parallel societal structures, which require long-term trust-building efforts. A potential communication channel is local Arabic-language media in Sweden, which could be more effectively used to reach different population groups. The Swedish government also needs to rethink its communication strategies during these types of crisis – how to calibrate an official response with foreign media, through which channels communication should occur and what messages to convey. The lesson learned from the LVU-campaign is that communication is necessary in multiple languages to multiple audiences to manage down the impact of negative narrative campaigns.

The campaign has experienced periods of increased activity, particularly between January and February 2022, but it has never entirely disappeared. It resurfaced in the fall of 2022 and is expected to continue influencing Sweden. Combined with events such as the Quran burnings in 2022–2023, the LVU campaign has contributed to making Sweden a priority target for radical Islamist groups.

DISCUSSION

- How do social media platforms contribute to the spread of disinformation campaigns like the LVU case, and what strategies can authorities use to counteract them effectively?
- What role did international actors, such as foreign media and radical influencers, play in amplifying the LVU disinformation campaign, and how did their involvement impact Sweden's global reputation?
- Why was the LVU campaign particularly effective in mobilizing public opinion, and how did the use of emotional narratives influence perceptions of Swedish social services?
- What are the potential long-term security and societal consequences of disinformation campaigns targeting government institutions, and how can Sweden strengthen its psychological defence against such threats?

DR. MAGNUS RANSTORP *is a leading authority on terrorism and violent extremism, with over three decades of global research experience. He is Strategic Adviser at the Center for Societal Security at the Swedish Defence University and Associate Professor of Political Science. Between 2011–2024, he played a key role in the EU Radicalisation Awareness Network, contributing to European counter-radicalization efforts. He now serves on the Research Committee at the EU Knowledge Hub on the Prevention of Radicalisation, shaping strategies to combat extremist threats. The study underlying this chapter was awarded "The Golden Magnifying Glass" on Source Criticism Day (March 13, 2024), recognizing its significant contribution to countering disinformation.*

22. ARTS, CULTURE AND PSYCHOLOGICAL DEFENCE

ANNA MCWILLIAMS

SUMMARY

- History show that the arts have been used in war for many different reasons such as spreading propaganda, to entertain or to heal trauma.
- The kinds of art that has been used includes everything from music, theatre, literature, to digital arts and performance and varies through time and between different conflicts.
- Today we see how many artists take an active role in supporting the Ukrainian war effort through their art both within Ukraine as well as internationally.
- The subjective nature of the arts means people can be influenced on an emotional level. There needs to be an understanding of who is using arts and for what purpose as the use of art can result in harm such as becoming an infringement on freedom of expression.

The arts, such as theatre, literature, music or visual arts gives us the opportunity to take part in other people's feelings, thoughts and perspectives. The arts can entertain, make us forget our own lives for a while, but also help us make sense of the world around us or heal trauma. The subjective nature of the arts can be an influence that reach us on an emotional level and can therefore be a way to influence people's behaviour, thoughts or actions as part of a psychological defence or as information influence. The following text will demonstrate how the arts has been used in conflicts, with an emphasis on 20th and 21st Century wars, and how the arts is still used to influence people in conflicts to this day.

FROM ANTIQUITY TO WORLD WAR I

History demonstrate uncountable examples of how the arts have had many different roles in conflict and war. Going back to antiquity, civilisations such as the ancient Greeks and Romans spread propaganda through architecture, monuments or different types of art. The aim was to glorify war and the soldiers that fought it. This is a way to instrumentalise art that has been used again and again throughout history (Anghel and Zbuchea, 2023). The iconography used during antiquity has reappeared as symbols also for later powers, such as the roman eagle used by Napoleon and by the Nazis during World War II. During the 17th and 18th century, many generals commissioned art that portrayed themselves at the forefront as brave warriors with the battle in the background. There is no coincidence that if death was portrayed in art it was solely of the enemy and not of an army's own troops. Instead, paintings often focussed on the bravery of one's own men (Anghel and Zbuchea, 2023, p.10).

Looking at how art has been used during the conflicts of the 20th century does, however, give us a more complex picture. Although the glorification of war still

carries on, there are many other ways that art is used in attempts to influence different audiences. During World War I knowledge about the horrific conditions on the battlefields were starting to spread through art, in poetry as well as in songs about the war. In the United States singing became a way to deal with the trauma following the war, especially for the families of soldiers who never returned home (Gier, 2008, p. 19).

WORLD WAR II

One of the most famous examples of the arts being used for state propaganda can be found in Germany during World War II. Soon after the Nazi party took over government in 1933, Joseph Goebbles became propaganda minister and controlled most of the German art through the government agency Reich Chamber of Culture (Reichskulturkammer). No artist was allowed to work in Germany without being affiliated with the Reich Chamber. In this way, the state kept control over arts and performances (Carlberg, 2016, p. 18), implementing the Nazi party's line of only allowing "Aryan culture". Music was often used to mobilise people and resources for the Nazi "cause" such as at the Nürenberg rallies, a series of events starting in 1923 where large number of people were gathered to celebrate the Nazi party's success (Bergh and Sloboda, 2010, p. 4).

In the United Kingdom, the state funded theatre companies to tour the country and play in everything from local theatres to staff canteens at factories and mines (Heinrich, 2010). There were several reasons behind this; it was a way to entertain and lift the moral of the population in difficult times, but it was also a way to remind people what they were fighting for. Even if the British state never controlled the arts in the same way as was done in Germany, the British government funded these productions, which therefore led to political influence over the arts. Productions that were uplifting, educational and national were encouraged and often Shakespeare plays were chosen, as it was considered the ultimate example of British culture (Heinrich, 2010, p. 62).

In the United States, the entertainment industry, especially Hollywood, became increasingly involved in the production of films that helped to paint a "correct" view of the war. One of the first films to take a stance in the war was Charlie Chaplin's film *The Great Dictator* from 1940 (Koppes Clayton and Black, 1990, p. 31). Even if these films may not have been direct state propaganda, they glamorised the war, often trying to make war appear more human by telling personal stories, for example Abbot and Castello's film Caught in the Draft. Hollywood films, which were still to some extent exported to Europe, were also used to counter anti-American propaganda and to portray the enemy as a monster, for example *Menace of the Rising Sun* or *Beast from the East* (Koppes Clayton and Black, 1990, p. 60).

VIETNAM AND THE IRAQ WAR

Although elements of anti-war protests had existed amongst art previously the Vietnam war became a change in the amount of anti-war art produced (Kinsella, 2005). Through music, artists hoped to humanise the war and make it less abstract for an American audience. They hoped that through music's ability to evoke an emotional response, they could get people to identify with the victims of war, both soldiers and civilians, and overcome a distinction of us and them (Kinsella, 2009).

The war in Iraq between 2003-2011 also demonstrates another change in how the arts were used in war as technological development made music portable

to a greater extent, which meant that it could be used by the soldiers in several ways. Studies of American soldiers in Iraq have shown that music, often metal or rap, was used 'to "pump them up" for possible combat situations' (Pieslak, 2007, p. 137). Music was also used as a military tactic against the enemy as the U.S. military prepared to retake control over Falluja on 31 March 2004 (Pieslak, 2007, p. 130). Music was then played from Humvees that surrounded the town to disturb the enemy and stop them from sleeping, which in turn was aimed at weakening their defence. Interviews with former American soldiers in the war has shown that the use of metal, rap and rock was chosen as it was culturally offensive (Pieslak, 2007, p. 130). Similar tactics had been used in Panama City in 1989 as part of the American Operation "Just Cause" as they tried to arrest the Panama dictator Manuel Noriega when he was applying for asylum at the Papal Nunciatura (the embassy of the Vatican state). On this occasion rock and metal such as AC/DC, Mötley Crüe, Metallica and Led Zeppelin was played to unsettle Noriega and at the same time drown out the sound of negotiations for the journalists nearby (Pieslak, 2007, p. 129).

UKRAINE

In Ukraine today we see many artists take an active role in supporting the Ukrainian war effort. It appears that in 2022 there was a sharp change from the arts more generally being non-political to actively engaging with the war (Tukova, 2023, p. 194). Tukova describes the objectives behind musicians' involvement as foremost a will to support the Ukrainian Armed Forces as well as reinforcing cultural diplomacy, especially through spreading Ukrainian music world wide for their cause (Tukova, 2023, p. 194). This music is described by Tukova as 'transferring many sensations and emotions of witnesses of the war into musical sounds' (Tukova, 2023, p. 196).

The reasons behind the art and performances that take place in Ukraine today appear to vary and includes raising hope, documenting what is happening and dealing with the trauma of war. The artist Sasha Anisimova has photographed bombed buildings and drawn everyday situations on top of them of what people would be doing right now if there were no war. This includes working, reading a book, watering a plant or doing yoga (Walfisz, 2022). Her art has multiple meanings, expressing both loss and hope for normal life to return (Walfisz, 2022).

Street art has become a way to protest against the war and a platform to spread messages to a wide audience. Themes often portrayed are connected to freedom of Ukraine such as the mural located on Independence Square in Kyiv portraying "Berehynia", a goddess from Slavic mythology that after 1991 became a protecting mother figure of the Ukrainian nation (Leahy, 2024, p. 21). Children are often used to show both the vulnerability of war but also as a sign of resilience, hope for the future and humour (Källström, 2023). Artists include both Ukrainian artists and international artists such as British artist and activist Banksey, Italian artist Salvatore Benintende alias TVBOY and French street artist Christian Guemy (Källström, 2023).

Several artists' initiatives also sell Ukrainian art such as the auction site *Sunseed Art* that sells Ukrainian art, where part of the profit goes to the Armed forces of Ukraine and charitable organisations (Sunseed Art Website, 2024). Another example of financial aid to the war through art is the band Kalush Orchestra who won Eurovision in 2022 and sold the trophy to raise money for the war (Welslau and Selck, 2024, p. 18).

The examples of arts in Ukraine today described above are only a few examples of the kinds of activities that artists and performers engage in. The war has inspired a new interest in Ukrainian culture including art, both within the country as well as internationally. Within Ukraine it is often connected with an interest in connecting to a Ukrainian national identity and to better understand Ukrainian culture (Lytvynenko, 2023).

DISCUSSION

Although the examples above are based on literature published in English with the result of a focus on western culture, we can see how the arts can play many different roles in wars. The arts has no requirement to stay objective, quite the opposite. Through the arts ability to portray feelings and to invite its' audience to be subjected to other people's experiences and feelings, it has a power to reach us on an emotional level (Kinsella, 2005). It is this emotional level that can become a powerful tool in psychological warfare to cause harm, or in psychological defence to create resilience, to heal or rebuild.

Although discussions of art during war is often referred to as propaganda as for example music during the war in Ukraine (Tukova, 2023, p. 203), using the term propaganda suggests that art is always produced for communication, to convey a specific message. This is not the case. Although art can be a way to convey a message it can also be used for other means; as part of healing after a conflict or to try to understand the complexities of a war. The way art becomes entangled with conflict and war demonstrate the complexities of how people react and act in war and that art can be public, communal or private.

The arts can humanise, something that is especially important during and following a war as conflict tends to dehumanise the enemy (Premratna and Bleiker, 2016). Arts can also be used as part of this dehumanising process, to create an "other." Researchers such as Bergh and Sloboda have therefore suggested that not using music as part of peaceful purposes could leave it open for abuse by those who want to create or maintain conflicts (Bergh and Sloboda, 2010, p. 4). It is, however, not clear who should be using the arts and for what purpose. Many states, including Sweden, apply a principle that is generally known as "arm length principle" which means that the arts should not be directed by governments but should function autonomously. Although funding will always to some extent steer art, the aim of the principle is implemented to stop governments from directly influencing the arts in any direction. This is to protect the freedom of expression within art, an important part of a democratic society (see further discussion in Dahlberg, 2019). Historical examples, however, have shown that this principle can be hard to uphold during a war and that states have used art to influence their own population as well as an external enemy.

There is also a tendency towards assuming that the arts, such as music, always has positive powers, often described as 'the power of music' or 'music as a universal language' (Howell, 2023, p. 153). As historic and current examples show, however, this is far from the case as arts can also be used to cause damage or influence people in a negative way (Bergh and Sloboda, 2010, p. 5). Research has shown that information influence in social media is particularly effective if affective content is shared, that which plays on our emotions (Nilsson, Olsson and Ekman, 2022, p. 50). If we do not understand the mechanisms that guide these processes and how it can affect us, we risk leaving ourselves open to harm.

DISCUSSION

- Give some examples of the kinds of art we see in war throughout history?
- Give some examples of the reasons have art been used in war throughout history?
- Can the use of the arts in war affect the freedom of expression and if so how?

ANNA MCWILLIAMS, *PhD, is a senior researcher and associate professor at the Swedish Defence Research Agency. Her research focusses of heritage and culture during war and conflict. She also acts as an advisor on questions of heritage and culture in connection with the Swedish total defence.*

REFERENCES

Anghel, S. and Zbuchea, A. (2023). Artists at war, *Europolity*, vol. 17, no. 2.

Bergh, Arild and Sloboda John. (2010). Music and Art in Conflict Transformation: A Review. *Music and Arts in Action*. Volume 2, Issue 2. 2010.

Carlberg, A. (2016). *Hitlers lojala musiker. Hur musiken blev ett vapen i tredje rikets propaganda.* Stockholm: Santérius Förlag.

Dahlberg, S. (2019). *What is the length of an arm? How "arm's lengths distance" is used in art and cultural politics in Sweden today.* Master Thesis, University of Gothenburg. https://scenochfilm.se/wp-content/uploads/2019/11/Susanna-Dahlberg-What-is-the-length-of-an-arm-190506.pdf

Gier, C. (2008). Gender, Politics, and the Fighting Soldier's Song in America during World War I. *Music & Politics,* Number 1, Winter 2008.

Howell, G (2023) Peaces of music: understanding the varieties of peace that music-making can foster. Peacebuilding, 11:2, 152-168.

Heinrich, A. (2010). Theatre in Britain during the Second World War. NTQ. 26:1, Cambridge University Press.

Kinsella, T. P. (2009). A Season in Hell: Art Song and the American War in Vietnam. In Heberle, M. (Ed.), *Thirty Years After: New Essays on Vietnam War Literature, Film and Art.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Kinsella, T. P. (2005). *A world of hurt: Art music and the American War in Vietnam.* University of Washington ProQuest Dissertations & Theses. https://www.proquest.com/docview/305423230?%20Theses&fromopenview=true&fromunauthdoc=true&pq-origsite=gscholar&sourcetype=Dissertations%20

Koppes, C. R. and Black, G. D. (1990). *Hollywood Goes To War – How Politics, Profits and Propaganda Shaped World War II Movies.* Berkeley: The Free Press.

Källström, L. (2023). Street Art Against War: With Stencil Marks and Paint Cans in Ukraine. *Baltic Worlds* BW: 2023.

Leahy, E. L. (2024). State Murals, Protest Murals, Conflict Murals: Evolving Politics of Public Art in Ukraine. *Arts*, 13:1.

Lytvynenko, A. (2023). Ukrainian Cultural and Art TV Programs During Russia's Invasion of Ukraine. *Scientific Journal of Polonia University,* 58:3.

Nilsson, P-E, Olsson, S., and Ekman, I. (2022). *Den nya informationsmiljöns*

topografi - Teknik, människa och strategi i osäkerhetens tidevarv. FOI-R--5342--SE

Pieslak, J. R. (2007). Sound Targets: Music and the War in Iraq, *Journal of Musicological Research,* 26:2-3, 123-149.

Premaratna, N. and Bleikerin, R. (2016). Arts and Theatre for Peacebuilding. In Oliver P. Richmond, Sandra Pogodda and Jasmin Ramovic (eds). *The Palgrave Handbook of Disciplinary and Regional Approaches to Peace.* Basingstoke: Palgrave MacMillan.

Sunseed Art Website. (2024). https://sunseed-art.com/en/ Accessed 29-11-24]

Tukova, I. (2023). Art music and war: Ukrainian case 2022. *Musicologica Brunensia.* Vol. 58, iss. 2.

Walfisz, J. (2022). Meet the Ukrainian artist creating images of hope amongst the rubble of Kharkiv. *Euronews.* https://www.euronews.com/culture/2022/04/01/meet-the-ukrainian-artist-creating-images-of-hope-amongst-the-rubble-of-kharkiv

Welslau, L. M. and Selck, T. J. (2024). Geopolitics in the ESC: Comparing Russia's and Ukraine's use of cultural diplomacy in the Eurovision Song Contest. *New Perspectives,* Vol. 32(1) 5–29.

23. INFORMATION INFLUENCE AND VIDEO GAMES

ELSA ISAKSSON

SUMMARY

- Video games are dynamic, politicized arenas where various actors compete to influence audiences, making them powerful tools for strategic communication and propaganda.
- Gaming platforms' immersive nature, social connections, unmoderated forums, and data collection mechanisms make them vulnerable for being exploited.
- They are increasingly being utilized as tools for malign information influence.
- Despite their extensive reach and impact, video games remain an under-examined area in the study of information influence and disinformation.

In April 2023, Brad Smith, President of Microsoft, raised alarms about Russian intelligence agencies and The Wagner Group's efforts to infiltrate gaming communities. Following a leak of top-secret Pentagon documents related to the Ukraine war through Discord, Smith disclosed that Microsoft's threat analysis team had identified ongoing Russian attempts to penetrate these online communities (Plunkett, 2023). This is just one of many recent examples of how video games and adjunct platforms has been utilized for malign information influence.

Video games has been recognized as powerful tools for strategic communication and propaganda (Pamment, Falkheimer & Isaksson 2023; Foust 2021; Schulzke 2013). As argued by Foust (2021), they are far from being politically neutral, apolitical spaces where people merely interact and play cooperatively. Instead, they are "... vibrant, contested, growing, lucrative, politicised spaces, where actors of all sizes and ideologies compete to influence the minds of their audiences. Video games are where politics happen". These attributes make games attractive tools for strategic communication and propaganda, as they offer a relatively inexpensive means to reach and engage audiences while monitoring their reactions (Schulzke, 2013). The interactive nature of video games, coupled with their ability to deeply immerse players in alternate realities, creates an ideal environment for influencing perceptions and spreading disinformation. As argued by Delwiche (2007, p. 92), "video games have the potential to shape attitudes and behavior in ways that Goebbels could never have dreamed." Similarly, Foust (2021) notes that games are a contested space for political dialogue, where "governments and corporations, journalists and activists, and players of every stripe are competing to tell stories and shape perceptions about the world." The gaming domain is ripe with vulnerabilities. However, while social media platforms have been extensively studied for their roles in information influence and disinformation, video games remain an underexplored domain despite their massive reach and potential for impact (Pamment, Falkheimer & Isaksson 2023).

THE SIGNIFICANCE OF VIDEO GAMES

The practice of playing video games dates back to the early 1960s, with the first games developed in the United States and Japan. By 1982, the gaming industry had surpassed the combined turnover of the Hollywood film and pop music industries (Rogers & Larsen, 1984). Today, it is one of the largest global cultural industries with 3 billion people playing (Turner 2024). In Sweden, half of all men and over a third of women aged 16–84 played video games in 2021 (Folkhälsomyndigheten, 2022).

VIDEO GAMES AS A TOOLS FOR INFLUENCE

As part of Lund University Psychological Defence Research Institute's ongoing exploration into potential malign foreign influence in under-researched domains, the recent report "Malign foreign interference and information influence on video game platforms: understanding the adversarial playbook" written by James Pamment, Jesper Falkheimer & Elsa Isaksson and published by the Psychological Defence Agency has shed light on how gaming platforms can be exploited by foreign powers for information interference. The study identified more than 40 influence techniques deployed in the gaming domain, categorized into six overarching tactics. Here, these tactics are detailed:

TACTIC 1: REFRAMING REALITY

This tactic involves using video game content to reshape how people see reality. Threat actors use video game elements like aesthetics, storytelling, mechanics, and realism to blur the line between games and reality. They might substitute game footage for real events, turn real-life actions into game-like scenarios, and incorporate video game themes into propaganda efforts. This tactic harnesses the immersive nature of games to influence perceptions and behaviors in the real world. For instance, in 2017, the Russian Ministry of Defence posted images on Facebook and Twitter claiming to show evidence of US and Da'esh cooperation. These photos were later debunked by Bellingcat, revealing them to be cropped screenshots from the mobile game AC-130 Gunship Simulator (Higgins, 2017). Similar, ARMA 3, a first-person shooter game initially developed by Czech company Bohemia Interactive in 2013 and subsequently updated in various versions, has been involved in multiple instances of misinformation and disinformation. Examples include a video that purported to show the Israel-Hamas conflict actually used footage from the game (AFP Sri Lanka, 2021) and a video circulated on Facebook in 2020, claiming to depict the shooting down of a US military plane by Taliban militants in Afghanistan and viewed nearly 9 million times, that was also sourced from ARMA 3 (AFP Pakistan, 2021).

Other examples include the adopting of gaming tropes in real-life situations, as in the livestreamed terrorist attacks in Christchurch, Pittsburgh, El Paso, and Halle which all made references to games and used a gamified language as well as adopted the visual style and perspective typical of first-person shooter when livestreaming the attacks (Schlegel, 2020). Other examples of the tactic include how Jihadist organizations, including Da'esh, utilize elements from video games for recruitment and propaganda purposes where they have incorporated footage from games such as Call of Duty (Schlegel, 2021).

TACTIC 2: PROJECTING AUTHORITY

This tactic includes using video games to assert authority to further geopolitical or political-economic competition by, for example, censor or encourage self-censorship in line with authoritarian norms and values, harvest data, or to conduct espionage. This involves the gaming industry being targeted as a sphere of influence or hybrid domain where threat actors aim to exert influence by either directly owning or indirectly controlling one or several aspects of the gaming industry. Ownership of gaming studios by authoritarian states or organizations can thus drive the production of games that propagate specific ideologies. In the Russian propaganda game "The Best in Hell", for example, players take on the role of Wagner Group fighters in the "Special Military Operation," where their mission involves capturing Ukrainian cities and following commands issued by a computer-generated Yevgeny Prigozhin (Röpcke 2024). Similar, in February 2023, the Ukrainian Minister of Digital Transformation officially urged Sony, Microsoft, and Valve to cease the sale of Atomic Heart, a video game prominently featuring Soviet Union, KGB, and Russian military themes. The minister expressed concerns on Twitter[63], stating, "I do believe neither of these businesses support bloody regime, murders or romanticizing communism. Brand new level of Russian digital propaganda – using gaming industry". He linked his tweet to the letter sent to the companies, highlighting that the game was developed by the Russian studio Mundfish. The minister's letter raised alarms about potential financial support to Russia's budget through game purchases, potentially funding the conflict against Ukraine.

The tactic also involves censoring of democratic norms and values. Video game producers can also self-censor due to fear of exclusion from authoritarian markets. There are many examples of how both Russia and China prohibit the positive mention of homosexuality in video games. A law against "propaganda of non-traditional sexual relations" was adopted by the Russian parliament in 2013 (Elder, 2013), effectively outlawing the distribution of information about gay relationships or gay rights through any media, as well as the holding of gay pride marches and rallies (MacDonald, 2016).

TACTIC 3: HACKING SYSTEMS

This tactic includes the use of video game and game adjacent data to gain access or leverage over computer systems. It involves threat actors using cyber skills to break into gaming systems, targeting gamers, companies, and the gaming industry itself. It can include Phishing through gaming platforms and associated services, exploiting DRM or anti-cheat software vulnerabilities with malware, breaching gaming companies for stealing player and proprietary data, using stolen accounts for money laundering, extortion, or leaking classified information, exploiting unmoderated gaming forums to leak harmful or illegal content, or laundering money through gaming economies or prepaid game vouchers.

For example, in April 2023, the US Justice Department and Pentagon launched an investigation following the leak of top-secret documents on a Discord channel associated with the Minecraft computer game. The leaked documents, which later circulated on platforms like 4Chan, Telegram, Twitter, and major media outlets worldwide, contained intelligence concerning Russia's invasion of Ukraine (Toler, 2023). Allegedly leaked by a member of the National Guard, the

[63] https://twitter.com/FedorovMykhailo/status/

documents were shared with an international community to provide them with current affairs updates (Alexander, 2023). Concerns arose that these leaks might include manipulated information, referred to as tainted leaks, aimed at deliberately misleading readers (Adams, 2023). Another example includes the events of March 2023, where GSC Game World, the Ukrainian developer behind STALKER 2, found themselves targeted in a cybersecurity breach where hackers gained access to their game data and attempted to extort concessions. An anonymous user on VKontakte, a Russian social media platform, claimed responsibility for the breach and issued demands that included Russian-language localization, an apology to Russian and Belarusian players, and the restoration of a banned Discord account (Hall, 2023).

TACTIC 4: INTERACTIVE PROPAGANDA

This tactic includes the use of video game design to promote interactive forms of propaganda. It involves the spreading of traditional and interactive propaganda through games, fostering gaming communities within the game environment, and potentially radicalizing and mobilizing players towards specific ideologies or causes. Threat actors focus on video games, creating their own games that mimic popular franchises, developing mods for existing games to reach established audiences quickly, or using in-game advertising to promote their agendas.

Mods, short for modifications, are user-created changes to video games, ranging from minor tweaks to major overhauls. While some mods are officially approved by game developers, others are created and distributed without permission. In 2022, Roblox removed two unauthorized mods depicting the Russia–Ukraine conflict. These mods allowed players to engage in combat scenarios, including bombings of cities like Mariupol. Despite being unauthorized, one of these mods, “War on Larkiv: Ukraine,” gained popularity with a 71% user review score and over 90,000 plays in less than two weeks, fueled by attention on platforms like TikTok (Tidy, 2022).

TACTIC 5: SOCIAL PROPAGANDA

In this tactic, video game-adjacent platforms are used for propaganda through social connections. It involves leveraging game-adjacent platforms, such as Discord, Steam, or Twitch to create communities where ideologies and propaganda can thrive. By employing techniques similar to those used on social media, threat actors can build relationships, spread their messages, and intensify divisions. The goal is to forge community bonds that extend beyond the gaming environment, influencing interactions across various tech platforms and even into real life. Threat actors leverage video game chat functions to spread propaganda, to harass individuals, and distribute illegal content. They can create false accounts to manipulate platforms and use advertisements to promote ideologies. Misuse of content moderation tools to ban or restrict user accounts, and exploiting areas with lax enforcement, are also common tactics. These actions exploit gaps in moderation to manipulate and often aim to polarize communities within and outside gaming.

For example, the use of video games and particularly the platforms and forums connected to games has been a growing trend for spreading right-wing extremist propaganda. Right-wing terrorists are increasingly using gaming platforms to spread their propaganda to younger audiences. These efforts not only disseminate extremist ideology but also provide recruitment opportunities and foster relationships among members of the right-wing extremist movement (Europol, 2021).

TACTIC 6: PSYCHOGRAPHIC TARGETING

In this tactic, data attained from video game and adjunct platforms is utilized for psychographic targeting and profiling which enable threat actors to gain deeper insights into individual users and market segments. This information can then be used to enhance profiling and targeting of specific user groups, to develop complex datasets for various purposes, and to improve advertising strategies. This tactic can help threat actors to understand individuals and groups better, supporting goals such as identifying susceptible populations, aiding espionage, and refining machine learning and AI. Data collected from user interactions, purchases, social behaviours, and VR/AR experiences are merged with other datasets to further these objectives.

CONCLUSION

The six tactics described - reframing reality, projecting authority, hacking systems, interactive propaganda, social propaganda, and psychographic targeting - illustrate diverse and sophisticated methods that threat actors have and can employ to manipulate gaming environments for their own purposes and how gaming platforms have become battlegrounds for influence. While the tactics outlined underscore the vulnerabilities inherent to gaming platforms, they also highlight opportunities for stakeholders to address these challenges through collaboration, innovation, and education. By fostering resilience among users, promoting transparency in the industry, and integrating safeguards, it is possible to mitigate the risks posed by malign information influence and ensure that gaming spaces remain safe and inclusive.

DISCUSSION

- How can policymakers and the gaming industry collaborate to mitigate the exploitation of gaming platforms for malign information influence purposes?
- In what ways do malign information influence efforts within video games differ from those seen on social media platforms?
- What lessons can be learned from how social media companies work with malign information influence?
- How can video games also be used as a tool to increase resilience against malign information influence and disinformation?

ELSA ISAKSSON *is a PhD candidate at Lund University in strategic communication and psychological defence and part of the Psychological Defence Research Institute. Her research focuses on underexplored domains and techniques of malign information influence, one of them being gaming environments. She has a background in political science and terrorism and political violence studies and is a contributing author to the report "Malign foreign interference and information influence on video game platforms".*

REFERENCES

Adams, P. (2023, April 10). Ukraine war: Who leaked top secret US documents – and why? *BBC*. https://www.bbc.com/news/world-europe-65225985

AFP Pakistan. (2021, April 28). This clip actually shows computer-generated imagery from a video game. *AFP Fact Check*. Retrieved from https://factcheck.afp.com/clip-actually-shows-computer-generated-imagery-video-game-0

AFP Sri Lanka. (2021, May 26). This footage does not show Israel's air defense system -- it was mainly created from a video game. *AFP Fact Check*. Retrieved from https://factcheck.afp.com/footage-does-not-show-israels-air-defense-system-it-was-mainly-created-video-game

Alexander, H. (2023, April 13). REVEALED: Leaker who posted top secret Pentagon documents in a Discord group 'works at a military base and is in his 20s' - as close friend describes him as 'fit, strong, armed and trained... like something out of a crazy movie'. *Daily Mail*. https://www.dailymail.co.uk/news/article-11967397/Leaker-posted-secret-Pentagon-documents-works-military-base-20s.html

Delwiche, A. (2007). From The Green Berets to America's Army: Video games as a vehicle for political propaganda. In J. P. Williams & J. H. Smith (Eds.), *The player's realm: Studies on the culture of video games and gaming*. McFarland & Co.

Elder, M. (2013, June 11). Russia passes law banning gay "propaganda". *The Guardian*. Retrieved from https://www.theguardian.com/world/2013/jun/11/russia-law-banning-gay-propaganda

Europol. (2021). *EU Terrorism Situation & Trend Report*. Publications Office of the European Union, Luxembourg. https://www.europol.europa.eu/publication-events/main-reports/european-union-terrorism-situation-and-trend-report-2022-te-sat

Folkhälsomyndigheten. (2022). Resultat från regeringsuppdrag att genomföra en befolkningsstudie om spel om pengar. https://www.folkhalsomyndigheten.se/contentassets/58e8320a52944ef8bd2b013e068e2c5e/resultat-franregeringsuppdrag-att-genomfora-en-befolkningsstudie-om-spel.pdf

Foust, J. (2021, March 25). Video games are the new contested space for public policy. *Brookings*. Retrieved from https://www.brookings.edu/techstream/video-games-are-the-new-contested-space-for-public-policy/

Hall, C. (2023, March 13). STALKER 2 team says it was hacked by pro-Russian group, is being blackmailed. *Polygon*. https://www.polygon.com/23637543/stalker-2-hack-russian-war-in-ukraine-threats

Higgins, E. (2017, November 14). The Russian Ministry of Defence publishes screenshots of computer games as evidence of US collusion with Da'esh. *Bellingcat*. Retrieved from https://www.bellingcat.com/news/mena/2017/11/14/russian-ministry-defence-publishes-screenshots-computer-games-evidence-us-collusion-daesh/

MacDonald, K. (2016, December 7). Russian MPs are not the first to try to write LGBT people out of video games. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2016/dec/07/russian-mps-lgbt-out-video-games

Pamment, J. Falkheimer, J. & Isaksson, E. (2023). Malign foreign interference and information influence on video game platforms: Understanding the adversarial playbook. *The Psychological Defence Agency*. Retrieved from file:///Users/el2105is/Downloads/mpf-skriftserie-23-03-malign-foreign-interference-and-information-influence-on-video-game-platforms-understanding-the-adversarial-playbook%20(3).pdf

Rogers, E. M., & Larsen, J. L. (1984). *Silicon Valley fever: Growth of high-technology culture*. Basic Books.

Röpcke, J. (2024, April 19). Russen entwickeln brutales Mord-Spiel. *Bild*. https://www.bild.de/politik/ausland-und-internationales/patriotisches-spielprojekt-russen-entwickeln-brutales-mord-spiel-661fd8cb56b4320eb3766641

Schlegel, L. (2020). Jumanji extremism? How games and gamification could facilitate radicalization processes. *Journal for Deradicalization, 23*. https://journals.sfu.ca/jd/index.php/jd/article/view/359

Schlegel, L. (2021). The gamification of violent extremism & lessons for P/CVE. *Radicalization Awareness Network*. Retrieved from https://home-affairs.ec.europa.eu/networks/radicalisation-awareness-network-ran/publications/gamification-violent-extremism-lessons-pcve-2021_en

Schulzke, M. (2013). Rethinking military gaming: America's Army and its critics. *Games and Culture, 8*(2), 59-76. https://doi.org/10.1177/1555412013478686

Tidy, J. (2022, September 30). Roblox removes 'meat grinder' Ukraine v Russia game. *BBC News*. Retrieved from https://www.bbc.com/news/technology-63078950

IV. COUNTERMEASURES

This section explores the measures that can be taken to mitigate, counteract, or push back on malign information influence campaigns. The firs chapter returns to questions of cognition, this time in the context of drawing upon psychological mechanisms as techniques for minimising the negative impact of disinformation. Examples then follow of education initiatives, crisis communication as a countermeasure, and lessons learned from Ukraine's experiences of dealing with Russian information influence. The next two chapters explore the concepts of deterrence and attribution as methods of dissuading threat actors from continuing with their activities. Finally, the section concludes with discussion on the ethics of countermeasures.

24. PSYCHOLOGICAL MECHANISMS RELATED TO INFLUENCE AND PERSUASION

BJÖRN PALMERTZ

SUMMARY

- Interdisciplinary Foundations: The chapter integrates insights from cognitive psychology, social psychology, and evolutionary psychology to explore mental processes, social influences on behaviour, and the adaptive nature of psychological mechanisms.
- Evolutionary Perspectives: Evolutionary psychology emphasizes the role of adaptations, by-products, and random effects in shaping behaviour, with evolved psychological mechanisms providing flexible yet historically rooted responses to survival and reproduction challenges.
- Cognitive Processes and Heuristics: Human reasoning involves deductive and inductive approaches, while heuristics like availability, representativeness, and confirmation biases simplify decisions but can lead to errors in judgment.
- Attitudes, Emotions, and Decision-Making: Attitudes and emotions influence decision-making through mechanisms like risk aversion, framing effects, and cognitive shortcuts, with strategies varying based on goals, social context, and emotional states.
- Influence and Persuasion: Persuasion occurs through central or peripheral routes depending on motivation and cognitive engagement, with factors like the communicator's credibility, message framing, and audience characteristics shaping effectiveness.

During the last two decades, the use of influence operations by both state and non-state actors, such as terrorist and extremist organizations, has become increasingly evident to decision-makers and populations alike. The digital revolution, which has significantly affected information dissemination and social exchange, alongside the increased interconnectedness of key societal systems and infrastructure, has introduced new opportunities as well as vulnerabilities. Additionally, changes in labour markets and the demographic composition of many societies have expanded opportunities and horizons for some segments of the public, while leaving others uncertain about their place or representation in traditional media or political spheres. This dynamic has had a profound impact on both national policies and international relations, particularly in relation to various referendums and elections.

Understanding the multifaceted nature of this challenge is crucial for developing societal resilience and effective countermeasures. Central to this understanding

is a grasp of the human psychological mechanisms that underpin how we interpret and communicate with the world around us. The research field of psychology encompasses several sub-fields that, while overlapping and interrelated, each offer distinct perspectives.

In this chapter, we will draw on three of these sub-fields: 1) Cognitive Psychology focuses on higher mental processes, including attention, language use, memory, perception, problem-solving, and thinking. 2) Social Psychology studies how social variables affect individual behaviour, attitudes, perceptions, and motives, as well as group and intergroup phenomena. It thereby seeks to identify and understand the factors that lead to certain behaviours and emotional responses in social contexts. 3) Evolutionary Psychology, which emphasizes the adaptiveness of behaviour and mental processes, based on the premise that mental capabilities have evolved over millions of years to serve specific adaptive purposes (APA, 2013).

EVOLUTION

Evolution by natural selection remains the only scientifically confirmed theory connecting broad psychological research (Mayr, 1982). This research identifies three evolutionary products: adaptations which are characteristics inherited or reliably developed through natural selection to solve survival or reproduction challenges. By-products, traits linked to adaptations but not directly solving those challenges, and random effects that arise from chance mutations or environmental changes (Tooby & Cosmides, 1992).

One way of structuring and describing the core units of human nature is *Evolved Psychological Mechanisms.* These are fundamental to human nature, specifically designed to address survival or reproduction challenges that have arisen throughout history. They are structured to process only pertinent information, making organisms aware of the adaptive problems they encounter. The input received is transformed into output through if-then decision rules, dictating responses such as attacking, fleeing, or remaining still, and can manifest as physiological activities like fear, informational cues that prompt other psychological mechanisms, or overt behaviours such as running away. These outputs aim to resolve certain adaptive problems, such as adopting an aggressive stance when faced with a threat, but they do not always lead to beneficial choices or reactions in every given situation, since they have been developed for success over very long periods of time. Due to their specificity, complexity, and variety of these mechanisms they provide a wide range of behavioural options. Rather than acting solely on rigid instincts, we often have multiple courses of action to choose from and can sometimes design our environment to reduce exposure to challenges (Buss, 1999).

REASONING

Two central concepts to illustrate information processing. *Deductive reasoning* refers to logically reaching conclusions from certain statements, a challenge is that a conclusion can have high validity, meaning it follows the correct form of reasoning, but still be untrue and not consistent with the content of the statements. *Inductive reasoning,* on the other hand, involves forming conclusions with varying degrees of certainty that do not necessarily follow from the premises. The strength of an inductive argument depends on factors such as the representativeness of observations, the number of observations, and the quality of evidence.

Due to the vast number of decisions humans make, we often rely on *heuristics*—assumptions based on past experiences—to make quick, automatic choices conserving our time and energy. One aspect of this is the *availability heuristic*, which proposes that humans judge occurrences easily remembered as more probable (Tversky & Kahneman, 1973; McKelvie, 1997). Another process that can mislead us is when making correlations between events. We often expect things to be related and treat them as such, even if there is no correlation or the connection is weaker than assumed. *Stereotypes* are an example, where generalizations about groups, often negative, are reinforced since we pay more attention to behaviours that fit the stereotype (Hamilton, 1981; Sears, 1983).

Another connection that doesn't always serve us well is the *representativeness heuristic* which explains judgments based on how much events or objects resemble each other. For instance, when judging if someone is a software developer, participants might ignore important facts and statistics if given personality descriptions. This can lead to overestimating a likelihood based on the presented traits, causing errors in reasoning (Kahneman & Tversky, 1972). Humans are also prone to *confirmation bias*, also described in the chapter about cognitive biases, where we focus on information that supports our existing beliefs and ignore what does not (Nickerson, 1998). A historical example is the British intelligence services' early assessment of the German V2-rocket program during WWII. Despite indications of its existence already in 1939, tensions across the scientific intelligence community led to the program being underestimated, resulting in grave misjudgements and a late response (Hunt, 2021).

Another heuristic is the *fundamental attribution error*, also termed correspondence bias, a partial explanation for why we often are surprised when ordinary people commit terrible acts. It leads us to attribute behaviours to personal dispositions rather than situational factors. For instance, in a car accident, we are more likely to blame a driver than consider external factors like road conditions (Barjonet, 1980). When evaluating extreme behaviours, we expect evil actions from inherently bad people and good actions from kind individuals (*Ross, 1977)*. However, very ordinary people can, as history and psychological experiments have shown, engage in highly destructive behaviour under certain conditions (Milgram, 1974).

ATTITUDES AND EMOTIONS

Attitude research is a contested field, with ongoing debates over definitions and the role of attitudes. Consequently, the predictive connection between attitudes and behaviour is also disputed. Early studies using questionnaires revealed a weak correlation between respondents' stated attitudes and their actual behaviour (LaPiere, 1934). Modern studies propose that this weak correlation may be due to the complex nature of the relationship between attitudes and behaviour. Various factors influence this relationship, including the accessibility of the attitude, whether it is expressed publicly or privately, and the strength of the individual's identification with a group that holds the normative attitude (Doll & Ajzen, 1992).

Even under ideal conditions, people have a limited capacity to process information, often relying on mental shortcuts. Nisbett and Ross have described this as people being *cognitive misers* (Nisbett & Ross, 1980). Subsequent research on social thinking expanded on this by introducing the concept of the *motivated tactician,* where individuals select from various cognitive strategies based on their goals, motives, and needs. These choices can prioritize adaptability and accuracy or speed and self-esteem, leading to diverse outcomes (Fiske & Taylor, 1991).

What factors influence attitude formation? Research offers several explanations. One is the *mere exposure effect* (Zajonc, 1968), which posits that repeated exposure to an object enhances positive or negative responses, though this effect diminishes over time (Bornstein, 1989). *Classical conditioning* also plays a role; attitudes toward a stimulus can become more positive or negative when consistently paired with another stimulus that elicits corresponding emotions. One study, for example, found that messages were more persuasive when presented together with music for which the individual had a positive predisposition (Galizio & Hendrick, 1972). *Instrumental conditioning* also influences attitudes. Behaviours followed by positive consequences are more likely to be repeated, while those followed by negative consequences are less likely. This type of learning depends on the frequency, timing, and strength of the reinforcement (Kimble, 1961).

Another facet which influences decision-making are emotions. They can manifest in various ways: *Expected emotions* are those sensed in anticipation of an expected emotional response, such as the elation in winning a game. *Immediate emotions* occur during the decision-making process. These can be *integral*, linked to the decision itself, like anxiety, or *incidental*, unrelated to the decision but still impactful, like irritation from ambient noise (Schlösser et al., 2013).

A phenomenon related to these emotional types is *risk aversion* — the common belief that potential losses weigh more heavily than equivalent gains (Tversky & Kahneman, 1991). Experiments reveal that people overestimate the impact of a loss before experiencing it compared to their actual happiness after the result (Kermer, 2006; see also the discussion on prospect theory in the chapter about cognitive biases). This misjudgement may stem from underestimating their own coping mechanisms when faced with unfavourable outcomes. Incidental emotions such as mood and environment also affect decisions. For instance, a study found that university admissions decisions prioritized academic attributes on cloudy days and non-academic attributes on sunny days (Simonsohn, 2007). The *framing effect* of how choices are presented is another influencing factor. When options are framed as gains, people prefer risk-averse strategies; but when framed as losses, they tend to take risks. In one study, participants consistently followed this pattern, even though the number of deaths from a disease outbreak was identical across two test groups. The only difference was the framing of the choices (Tversky & Kahneman, 1981).

INFLUENCE AND PERSUASION

Petty and Cacioppo theorize that people take one of two routes when faced with persuasive messages. If individuals are motivated and able to reason about an issue, they are likely to follow the central route to persuasion, which involves evaluating the arguments. Strong and compelling arguments increase the likelihood of persuasion. However, in situations where individuals lack motivation or are distracted, the weight of the arguments matters less. They may then follow the peripheral route to persuasion, where easily comprehensible and familiar statements have an advantage (Petty & Cacioppo, 1986). This is often used in advertising, where products or companies are associated with clear, positive images or symbols. Given that the goal of persuasion is behavioural change, researchers have also studied whether the outcomes of these two routes lead to the same type of change. It was found that the central route tended to lead to more durable change (Petty & Krosnick, 1995).

Looking at the communication process itself there are several interesting findings. A communicator's identity significantly influences how the message is received. The perceived expertise of the source can prompt action (Olson & Cal, 1984), while trustworthiness is enhanced by confident delivery (Erickson et al., 1978). Additionally, a message is seen as more reliable if it appears to lack persuasive intent (Walster & Festinger, 1962). For instance, a businessman criticizing pollution is seen as more credible than a pro-environment politician delivering the same message (Eagly et al., 1978). In addition, likability plays a role; we are more responsive to messages from attractive or similar individuals (Dion & Stein, 1978; Krisberg, 2004). Those who share our appearance or mannerisms have a greater impact, specifically regarding subjective preferences such as values and taste. However, for objective decisions based on facts, dissimilar communicators can boost confidence, as their judgment is seen as more independent and credible (Goethals & Nelson, 1973).

The message's persuasive appeal also depends on the audience. Emotionally charged messages are more effective for those uninvolved or guided by emotions (Chaiken, 1980), while well-educated, analytical individuals respond better to reasoned arguments (Cacioppo et al., 1983). Consistency also matters; people are more likely to be influenced if the appeal aligns with their prior public commitments (Sherman, 1980). Research has also found that one sided arguments can be effective if the audience is already in agreement. However, for people that already disagree or are exposed to opposing views, messages acknowledging such perspectives are more advantageous (Jones & Brehm, 1970).

An important aspect of influence is the *primacy effect*, which highlights the advantage of presenting information early. For example, candidates listed first on a ballot have a higher chance of being elected. Conversely, the limitation of human memory encourages the recollection of relatively new experiences – the *recency effect* (Miller & Campbell, 1959).

RELEVANCE FOR DEFENCE AGAINST MALIGN INFLUENCE OPERATIONS

Understanding psychological and social processes is a key aspect for improving the development of capabilities regarding identification, analysis and countermeasures of malign influence operations. Expanding our knowledge about the mental processes and social circumstances that shape our attitudes, decisions, and actions as humans is always important, but especially so in times of conflict, exposing societal and technical vulnerabilities as well as introducing an array of security threats affecting democratic states. To understand and track how antagonistic actors combine capabilities to exert influence against other societies, ranging from public diplomacy and social media to military demonstrations and subversion is one key aspect. Gaining continuous insight into existing and approaching vulnerabilities is another.

Malign influence operations often exploit cognitive shortcuts like the availability heuristic, confirmation bias, and representativeness heuristic to manipulate perceptions and decision-making. For instance, disinformation is often designed to be emotionally salient and memorable, creating the illusion of truth through repetition. Countermeasures need to address these biases by promoting media literacy, encouraging critical thinking, and deploying strategies that reasonably prepare the public for potential false narratives before they are deployed by an adversary.

Adding to this understanding, social psychology highlights the power of group identification, social norms, and intergroup relations in shaping attitudes and behaviours. Malign influence operations frequently amplify divisive narratives to exploit societal fault lines and sow discord. To address such tactics, countermeasures can include fostering inclusive narratives that emphasize shared values, deploying counter-messaging that targets specific groups with culturally resonant appeals, and leveraging trusted community leaders to disseminate accurate information.

In addition to these social dimensions, evolutionary psychology emphasizes the role of adaptive mechanisms in decision-making, particularly in response to threats. Influence campaigns often rely on fear-based messaging to elicit heightened emotional responses and incite certain behaviours. To counter these approaches, balanced framing techniques can mitigate fear-based messaging while simultaneously leveraging positive emotions and trust-building efforts to enhance public confidence in credible sources.

Another perspective can be found in the dual-route model of persuasion (central and peripheral routes) which offers a deeper understanding of how individuals process influence attempts. Adversaries may rely on peripheral cues, such as attractive visuals or repetition, to sway less-engaged audiences. Counter-influence activities, however, can benefit from employing central-route strategies, presenting evidence-based arguments to engaged audiences while also using peripheral cues to effectively reach broader demographics.

The effectiveness of any counter-influence effort is also dependent on the credibility of the communicator. Research demonstrates that messages from trusted, confident, and non-partisan sources are more persuasive. This underscores the importance of ensuring transparency in communication, carefully selecting spokespersons, and aligning messaging with the values and priorities of the target audience (Eagly et al., 1978; Clark & Evans, 2014).

By integrating insights from the field of psychology into long- and short-term counter-influence activities, the freedom of action and potential effect of adversaries who rely on exploiting cognitive and emotional vulnerabilities can be limited. The goal is not merely to respond to individual malign activities but to cultivate an informed and resilient public capable of resisting manipulation and offering a strong societal platform for more short-term operationally specific countermeasures, ensuring the long-term security and integrity of democratic institutions and the population they serve.

DISCUSSION

- Cognitive Psychology: How do heuristics like the availability and representativeness biases influence decision-making in everyday scenarios, and what strategies can mitigate their effects?
- Social Psychology: In what ways do social variables, such as group dynamics or cultural norms, shape individual attitudes and behaviours, and how can this understanding be applied to address social issues?
- Evolutionary Psychology: What are Evolved Psychological Mechanisms, and how do their historical adaptive purposes sometimes result in maladaptive outcomes in modern environments?
- Influence and Persuasion: How do the central and peripheral routes to persuasion differ in their impact on long-term behaviour change, and what practical examples demonstrate their effectiveness in communication?

BJÖRN PALMERTZ *is a senior advisor at the Psychological Defence Research Institute, Lund University. With over 25 years in defence, security, and communication, he has worked for among others the Swedish Armed Forces, the Swedish Psychological Defence Agency, RISE Center for Cybersecurity, Turner Broadcasting and DreamWorks SKG. As an expert on hybrid threats, influence operations and foreign interference his research and advisory centres on how these challenges can be identified, analysed and countered from the perspective of democratic societies. He has contributed to books such as Hybrid Warfare (2021) and the Routledge Handbook of Disinformation and National Security (2023).*

REFERENCES

American Psychological Association. (2013). Glossary of psychological terms.

Barjonet, P. E. (1980). L'influence sociale et des representations des causes de l'accident de la route. Le Travail Humain, 43, 243-253.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. Psychological Bulletin, 106, 265-289.

Buss, D. M. (1999). Evolutionary psychology – the new science of the mind. Needham Heights: Allyn & Bacon.

Cacioppo, J. T., et al. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. Journal of Personality and Social Psychology, 45, 805-818.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. Journal of Personality and Social Psychology, 39, 752-766.

Clark, J., & Evans, A. (2014). Source credibility and persuasion: The role of message position in self-validation. Personality & Social Psychology Bulletin, 40(8).

Dion, & Stein. (1978). Physical attractiveness and interpersonal influence. Journal of Experimental Social Psychology, 14, 97-109.

Doll, J., & Ajzen, I. (1992). Accessibility and stability of predictors in the theory of planned behavior. Journal of Personality and Social Psychology, 63, 754-765.

Eagly, A. H., et al. (1978). Casual inferences about communicators and their effect on opinion change. Journal of Personality and Social Psychology, 36, 424-435.

Erickson, et al. (1978). Speech style and impression formation in a court setting: The effects of powerful and powerless speech. Journal of Experimental Social Psychology, 14, 266-279.

Fiske, S. T., & Taylor, S. E. (1991). Social cognition (2nd ed.). New York, NY: McGraw Hill.

Galizio, M., & Hendrick, C. (1972). Effect of musical accompaniment on attitude: The guitar as a prop for persuasion. Journal of Applied Social Psychology, 2, 350-359.

Goethals, G. R., & Nelson, E. R. (1973). Similarity in the influence process: The belief- value distinction. Journal of Personality and Social Psychology, 25, 117-122.

Hamilton, D. L. (1981). Illusory correlation as a basis for stereotyping. In D. L. Hamilton (Ed.), Cognitive processes in stereotyping and intergroup behavior. Hillsdale, NJ: Erlbaum.

Hunt, B. (2021). Lost in space: The defeat of the V–2 and post-war British exploitation of German long-range rocket technology. Air Power History, 68(1), 17–36.

Jones, R. A., & Brehm, A. W. (1970). Persuasiveness of one- and two-sided communications as a function of awareness there are two sides. Journal of Experimental Social Psychology, 6, 47-56.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3(3), 430–454.

Kermer, D. A., et al. (2006). Loss aversion is an affective forecasting error. Psychological Science, 17, 649-653.

Kimble, G. A. (1961). Hilgard and Marquis' conditioning and learning (2nd ed.). New York, NY: Appleton-Century-Crofts.

Krisberg, K. (2004). Successful truth – anti-smoking campaign in funding jeopardy: New commission works to save campaign. Medscape.

LaPiere, R. T. (1934). Attitudes vs. actions. Social Forces, 13, 230-237.

Mayr, E. (1982). The growth of biological thought. Cambridge, MA: Cambridge University Press.

McKelvie, S. L. (1997). The availability heuristic: Effects of fame and gender on the estimated frequency of male and female names. Journal of Social Psychology, 137, 63-78.

Milgram, S. (1974). Obedience to authority. New York, NY: Harper and Row.

Miller, N., & Campbell, D. T. (1959). Recency and primacy in persuasion as a function of the timing of speeches and measurements. Journal of Abnormal and Social Psychology, 59, 1-9.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2(2), 175–220.

Nisbett, R. E., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgement. Englewood Cliffs, NJ: Prentice Hall.

Olson, J. M., & Cal, A. V. (1984). Source credibility, attitudes, and the recall of past behaviours. European Journal of Social Psychology, 14, 203-210.

Petty, R. E., & Cacioppo, J. T. (1986). Communication and persuasion: Central and peripheral routes to attitude change. New York, NY: Springer-Verlag.

Petty, R. E., & Krosnick, J. A. (1995). Attitude strength: Antecedents and consequences. Hillsdale, NJ: Erlbaum.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 10, pp. 173–220). New York, NY: Academic Press.

Schlösser, T., Dunning, D., & Fetchenhauer, D. (2013). What a feeling: The role of immediate and anticipated emotions in risky decisions. Journal of Behavioral Decision Making, 26(1), 13–30.

Sears, D. O. (1983). The person-positivity bias. Journal of Personality and Social Psychology, 44, 233-250.

Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. Journal of Personality and Social Psychology, 39, 211-221.

Simonsohn, U. (2007). Clouds make nerds look good. Journal of Behavioral Decision Making, 20, 143-152.

Tooby, J., & Cosmides, L. (1992). Psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), The adapted mind (pp. 19-136). New York: Oxford University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice. Quarterly Journal of Economics, 106, 1039-1061.

Walster, E., & Festinger, L. (1962). The effectiveness of ”overheard” persuasive communications. Journal of Abnormal and Social Psychology, 65, 395-402.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. Journal of Personality and Social Psychology, 9, 1-27.

25. EDUCATION AGAINST DISINFORMATION

THOMAS NYGREN & ULRICH K. H. ECKER

SUMMARY

- Education plays a crucial role in strengthening resilience against disinformation by fostering critical thinking and encouraging responsible news consumption. Developing the ability to assess information critically enables individuals to navigate complex digital environments more effectively.
- Addressing the challenge of disinformation through education requires sustained and systematic efforts rather than simple or immediate solutions. Given the complexity of misinformation and its evolving nature, long-term, engaging strategies are essential for meaningful impact.
- Effective training approaches should go beyond theoretical instruction. Programs that combine demonstration, explanation, critical engagement, reflections, games, social interaction in structured ways and hands-on practice offer greater opportunities for learners to internalize and apply their knowledge.

Identifying and countering disinformation has become essential in an age of pervasive digital information. Disinformation—intentionally misleading information—spreads effectively on social media and digital platforms, impacting public trust, elections, and public health, posing a substantial threat to democracy (European Commission, 2022a; Juul & Ugander, 2021). Advanced technologies, including AI, amplify these challenges, making it increasingly difficult for individuals to discern truth from manipulation (Goldstein et al., 2024; Köbis et al., 2021). Education systems now play a central role in developing citizens' critical literacy based on general cognitive skills and subject-specific knowledge, strengthening individuals' ability to accurately evaluate domain-specific content (Osborne & Pimentel, 2022) we all depend on the expertise of others. For example, most readers of Science accept the anthropogenic origin of climate change. Yet far fewer have actually read a report of the Intergovernmental Panel on Climate Change (IPCC). By fostering deep subject knowledge, educational efforts enable people to critically analyze information in areas such as science, health, and history, providing a foundation for identifying biased narratives and recognizing trustworthy sources. To address the challenge of disinformation, education and training must provide people with good subject knowledge and up-to-date competencies adapted to the digital realities we all face today, acknowledging that disinformation is a moving target. Thus, developing digital civic literacy—combining factual knowledge, critical thinking, and reflective attitudes—supports a democratic society by preparing citizens to engage with information responsibly (Nygren & Guath, 2022).

EDUCATIONAL INTERVENTIONS THAT WORK

Short interventions can support individuals' ability to distinguish credible news from misleading information through short tips on thinking about where the information comes from, questions about whether they really believe the information is accurate, and short videos and games about disinformation (Ecker et al., 2022; Roozenbeek et al., 2023). However, such interventions have often been tested in online panels with participants who know they are in a research study, meaning that interventions have rarely been assessed in real-world complex situations (Roozenbeek et al., 2024). But what works in lab studies or when people are paid to participate in an experiment does not necessarily work as well in everyday life.

GOOD NEWS THAT EDUCATES

From a broader perspective, we can see that the consumption of news can have a positive impact on people's knowledge of what is happening in the world—knowledge that can protect against disinformation on current affairs.

Studies of the relationship between news habits and the ability to determine the credibility of news have previously shown unclear relationships, where good news habits are not necessarily related to the ability to identify disinformation (Damstra et al., 2021). However, current research suggests that there may indeed be positive associations; for example, efforts to foster regular engagement with news content can be beneficial for knowledge formation, truth discernment, and trust in news, especially for consumers with low baseline interest in news (Altay et al., 2023, 2024), even in times when the news focuses on heated issues such as war and extremism (Altay et al., 2024). News that highlights misleading narratives can also raise awareness of circulating misinformation and boost trust in legacy media by drawing attention to the importance of editorial standards (Altay et al., 2023; Thorson, 2024). At the same time, it can be concluded that good news habits alone are not sufficient, and researchers call for a combination of measures to help people deal with disinformation.

MEDIA AND INFORMATION LITERACY (MIL) TRAINING

Education has long been recognized as central to countering the spread and impact of disinformation. Bateman and Jackson (2024) note that there is considerable research evidence that MIL can be effective in countering disinformation. However, they emphasize that to be effective, such training needs to be carried out accurately using research-based principles. This type of training often requires the presence of teachers and content adapted to the target audience. Good training includes a mix of teacher-led and student-active exercises (Martella et al., 2024). The teacher in such a practice shows, explains, challenges and creates conditions for practice and reflection (Nygren, 2019). Thus, MIL training faces significant challenges in terms of resourcing, scalability, time, and reach. It can take many years to implement comprehensive education programs on a broad scale, making it both costly and time-consuming to have far-reaching impact and, more specifically, to engage those most vulnerable to disinformation (Bateman & Jackson, 2024).

If implemented well, MIL training consistently shows promising results. Providing knowledge and awareness of journalistic practices, media manipulation and disinformation techniques, as well as promoting an understanding of the media

industry, the internet and digital technologies is key (Bateman & Jackson, 2024; European Commission, 2022a). MIL has long been part of public education and humanities education in developed democracies, especially in subjects that emphasize critical reading and thinking, such as language, essay writing, civics and rhetoric. Public libraries have also historically promoted media literacy. MIL education for adults can take place in libraries, senior centers, at leisure events or in professional contexts. There is currently research showing how MIL training can work for both very young people (e.g. Kohnen et al., 2020) and an older population (Moore & Hancock, 2022).

Teaching disinformation resilience has shown promising results by using games that let individuals practice manipulation techniques, such as Bad News, where players simulate the role of a disinformation creator to understand persuasive tactics better and learn to identify misinformation techniques. This experiential approach enhances users' ability to recognize and resist misinformation, with effectiveness seen across cultural settings, including real-world classrooms (Axelsson et al., 2024; Roozenbeek et al., 2020). However, such interventions should be reinforced over time for sustained impact, as initial enthusiasm and learning effects can fade (Leder et al., 2024; Maertens et al., 2023). Beyond gamification, skills like lateral reading—checking information against multiple sources—are foundational in teaching critical evaluation, resembling fact-checking methods and involving digital tools like reverse image searches (Nygren et al., 2021). This approach, often described as "technocognition" (Lewandowsky et al., 2017), emphasizes understanding and using digital tools to verify media authenticity, particularly useful in deepfakes and AI-generated content contexts. To learn to fact-check like a pro, it is (a) good to see how experts do it, (b) actively practice the techniques, and (c) get feedback, which can stimulate people to use digital tools effectively. Authentic and tailored learning environments, where people can practice real-life situations, can strengthen their ability to deal with disinformation (Axelsson & Nygren, 2024). Training in these skills not only equips users to detect manipulations but can also foster awareness of how algorithms shape content exposure. Such knowledge is essential for citizens' navigation in today's information landscape.

ANALYTICAL THINKING AND THOUGHTFULNESS

Another important skill for detecting and evaluating misleading information is being able to apply a slow, analytical and reflective mindset (Roozenbeek et al., 2021). This contrasts with the more intuitive, automatic and emotional thinking we use when reacting quickly to information, which digital environments often stimulate (Martel et al., 2020). Encouraging thoughtfulness and analytical thinking helps people to step back, question the credibility of information and its source, and analyze the arguments. It can be valuable to create conditions for this through desirable difficulties with friction, providing participants with instructive challenges that force them to stop, think and rethink (Nygren, 2023, 2025). Training aimed at developing this skill should focus on giving participants time to reflect, question and discuss the information they encounter.

INTEGRATION OF SUBJECT KNOWLEDGE FOR SPECIFIC AREA

Integrating subject knowledge into teaching efforts relating to disinformation is crucial, especially in areas that are often affected by disinformation, such as scientific issues related to climate change or vaccines. Science education can

serve as an important tool to provide students with knowledge about how scientific processes work and how to interpret data and research results. People with a basic understanding of the methods of science are likely to be better equipped to recognise pseudoscience and false claims (Osborne & Pimentel, 2022). It is important for people to understand the scientific consensus in areas such as climate change, which can help them distinguish between accurate scientific information and misleading propaganda (Bayes et al., 2023; Osborne & Allchin, 2024).

Social studies and history are also important subjects to understand, for instance, political propaganda when recognizing how authoritarian regimes and extreme political groups use disinformation to influence public opinion. People who understand historical examples of propaganda and have better political knowledge are better prepared to recognise modern disinformation campaigns (Nygren, 2019; Vegetti & Mancosu, 2020). It needs to be pointed out, however, that some motivated political misperceptions might be more common among knowledgeable people (Flynn et al., 2017; Miller et al., 2016).

REPETITION AND LONG-TERM EFFECTS

Effectively teaching against disinformation cannot be a one-off effort but requires a continuous process where people are allowed to practice and improve their skills regularly. One of the primary functions of human memory is to build stable representations; updating one's memory and revising one's knowledge are, therefore, cognitively challenging tasks that require effort and persistence (Ecker et al., 2022). Retrieval practice, for example, through repeated (self-)testing, is crucial for new knowledge and skills to become consolidated in people's long-term memory and behavioural patterns (Brown et al., 2014). Skills that are not practised regularly may be lost (Maertens et al., 2021, 2023), making it important to create learning environments where, for example, students in schools are continuously given the opportunity to review and analyse information. One possible solution is to integrate source evaluation, media analysis and general information literacy lessons across different subjects, so that students practice these skills in different contexts (European Commission, 2022a), and combine such efforts with ongoing monitoring of impacts, to assess improvements and long-term effects of systematically repeated interventions.

POTENTIAL SIDE EFFECTS

People often accept information that aligns with their beliefs due to confirmation bias and motivated reasoning, which can make it challenging to correct ideological or political misperceptions, especially when these views are reinforced by partisan media or political leaders (Ecker et al., 2022). Educational interventions designed to address disinformation can inadvertently heighten general scepticism, leading to cynicism or overconfidence in one's ability to discern falsehoods (Altay, 2022; Haider & Sundin, 2020; Nygren et al., 2025). To mitigate these issues, disinformation education should emphasize intellectual humility, encouraging students to remain reflective about their abilities and receptive to different perspectives. Building a learning environment that promotes self-reflection and critical analysis can support balanced critical thinking while avoiding the potential side effects of cynicism or undue confidence (Bowes & Fazio, 2024; Prike et al., 2024). Educators should create an environment where learners can self-reflect on their misconceptions through discussion and critical analysis. A special focus on those who have difficulty understanding is needed to avoid unwanted side effects (Nygren et al., 2025).

CHANGING ATTITUDES AND DEVELOPING DIGITAL CIVIC LITERACY

Changing attitudes is often more challenging than filling knowledge gaps. While knowledge can be imparted through fact-based teaching, changing attitudes requires a deeper and longer-term process involving reflection, discussion and social interaction. Teaching should not only focus on filling knowledge gaps but also on basic attitudes and approaches to information. (European Commission, 2022a).

A key component of this work is developing digital civic literacy – the ability to analyze and understand information and act as responsible and aware digital citizens. Children and adults need to learn to understand how their own information consumption and sharing affects themselves and society. Developing this type of literacy means that people learn to be critical, reflective and responsible when interacting with information in the digital world (Kozyreva et al., 2020, 2022; Nygren & Guath, 2022).

Teaching may also need to engage participants on a personal level, allowing them to reflect on their own information habits and discuss them with others. Collaboration and group discussions can be effective tools to help participants see things from different perspectives and question their own preconceptions. This can be particularly useful for people in privileged societal positions (Galinsky et al., 2006). Research shows that students who are given the opportunity to practice source evaluation and discuss and reflect on their opinions in groups can change their attitudes and become more critical of disinformation (Axelsson et al., 2024). In education, it can be helpful to see and practice different perspectives on an issue to become more reflective about difficult questions (Lo & Adams, 2018). There is potential for polarization, stereotypes, and misconceptions to be reduced by bringing people together and exchanging perspectives through systematic exercises (Bruneau & Saxe, 2012; Paluck et al., 2021). Seeing how political opponents can be sympathetic and have great similarities can also be helpful, and recalibrating potentially distorted, polarized perceptions of the world can support democratic values (Voelkel et al., 2023).

EVALUATE THE IMPACT, LACK THEREOF AND SIDE EFFECTS

Evaluating the effectiveness of disinformation interventions requires objective and holistic measurement methods, as self-reported assessments can be unreliable (Jones-Jang et al., 2021); for instance, participants who feel confident in their fact-checking skills may actually lack the necessary expertise. Beyond knowledge acquisition, evaluations should verify if people can apply learned concepts to analyze new information critically. To accurately gauge the impact on attitudes and digital literacy, it is essential to track changes in skills and attitudes over time and compare these to self-assessed confidence levels (European Commission, 2022b). Effective assessment must also account for unintended effects, such as cynicism or overconfidence, to enable educators to mitigate these through adapted teaching approaches (Nygren et al., 2024). A comprehensive evaluation approach, including long-term follow-ups, is critical for understanding the benefits and potential drawbacks of education against disinformation (Tay et al., 2023).

SUMMARY OF CHALLENGES AND OPPORTUNITIES OF EDUCATION AGAINST DISINFORMATION:

OPPORTUNITIES:

- Education can help build resilience to disinformation by promoting good news habits and critical thinking.
- Training that demonstrates, explains, challenges and creates opportunities for practice and reflection is recommended.
- Retrieval practice, repeated training, and integrating subject knowledge with source evaluation can yield long-term results.
- Gamification and simulations can give people a practical understanding of how disinformation is spread and how it can be managed.
- Collaboration and social interaction in structured ways can help people shift their perspectives and reflect on their own views.

CHALLENGES:

- There are no quick fixes to the complex challenge of disinformation education; systematic, engaging, long-term efforts are needed.
- The effects of training can be short-lived if repetition and long-term interventions are not used.
- Psychological factors, such as the cognitive difficulty of belief updating and motivated reasoning, can make it difficult to correct misperceptions.
- Teaching can potentially have side effects, such as excessive scepticism or overconfidence.
- Teaching technical skills to identify disinformation is difficult, especially when technology is evolving rapidly.

Teaching against disinformation thus requires a careful balance between efforts to provide people with factual knowledge and develop their ability to think critically and analyze information in today's (and tomorrow's) digital world.

DISCUSSION

- How can professional communicators balance the need for engaging and accessible content while ensuring accuracy and fostering critical thinking in audiences?
- Reflect on the challenges of simplifying complex topics without oversimplifying or distorting information and consider strategies to encourage audiences to evaluate sources critically.
- What role can professional communicators play in shaping news habits and digital civic literacy?
- Consider how media professionals can contribute to promoting responsible information consumption, encouraging lateral reading, and countering the spread of disinformation through their reporting, writing, or media production.
- How can communicators address the risk of unintended side effects, such as increased skepticism or overconfidence, when educating audiences about disinformation?
- Reflect on how messaging strategies can be designed to promote critical thinking, intellectual humility and trust in credible sources and avoid unwanted side-effects.

THOMAS NYGREN *is professor of history and civics education, Uppsala University. His research focuses on how educational interventions can help individuals navigate the complexities of disinformation and develop critical thinking skills. He is the PI of multiple projects about education against misinformation (the News Evaluator, Digital Literacy Across Disciplines, and SEGA:D) and was an expert for the European Commission on education against disinformation.*

ULLRICH K. H. ECKER*, is professor, School of Psychological Science, University of Western Australia, and a cognitive scientist studying human cognition, in particular the impact of misinformation on memory, reasoning, and behaviour. He has led several large research projects and is currently working on a project titled "The misinformation future – Confronting emerging threats". Together with his students and collaborators, he has published 110+ peer-reviewed papers, holds editorial roles at two academic journals, and is a lead author of the Debunking Handbook 2020.*

REFERENCES

Altay, S. (2025, March 20). How Effective Are Interventions Against Misinformation? https://doi.org/10.31234/osf.io/sm3vk_v2

Altay, S., Hoes, E., & Wojcieszak, M. (2024). *News on Social Media Boosts Knowledge, Belief Accuracy, and Trust: A Field Experiment on Instagram and WhatsApp.* OSF. https://doi.org/10.31234/osf.io/hq5ru

Altay, S., Nielsen, R. K., & Fletcher, R. (2023). News can help! The impact of news media and digital platforms on awareness of and belief in misinformation. *The International Journal of Press/Politics,* 19401612221148981.

Axelsson, C.-A., Nygren, T., Roozenbeek, J., & van der Linden, S. (2024). Bad News in the civics classroom: How serious gameplay fosters teenagers' ability to discern misinformation techniques. *Journal of Research on Technology in Education,* 0(0), 1–27. https://doi.org/10.1080/15391523.2024.2338451

Bateman, J., & Jackson, D. (2024). *Countering Disinformation Effectively: An Evidence-Based Policy Guide.* https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en

Bayes, R., Bolsen, T., & Druckman, J. N. (2023). A Research Agenda for Climate Change Communication and Public Opinion: The Role of Scientific Consensus Messaging and Beyond. *Environmental Communication,* 17(1), 16–34. https://doi.org/10.1080/17524032.2020.1805343

Bowes, S. M., & Fazio, L. K. (2024). Intellectual humility and misinformation receptivity: A meta-analytic review. *Advances.in/Psychology, 2,* e940422. https://doi.org/10.56296/aip00026

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Harvard University Press.

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of 'perspective-giving' in the context of intergroup conflict. *Journal of Experimental Social Psychology*, 48(4), 855–866. https://doi.org/10.1016/j.jesp.2012.02.017

European Commission. (2022a). *Final report of the Commission expert group on tackling disinformation and promoting digital literacy through education and*

training: Final report. Publications Office of the European Union. https://doi.org/10.2766/283100

European Commission. (2022b). *Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training.* Publications Office of the European Union. https://doi.org/10.2766/28248

Damstra, A., Boomgaarden, H. G., Broda, E., Lindgren, E., Strömbäck, J., Tsfati, Y., & Vliegenthart, R. (2021). What Does Fake Look Like? A Review of the Literature on Intentional Deception in the News and on Social Media. *Journalism Studies,* 1–17. https://doi.org/10.1080/1461670X.2021.1979423

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology,* 1(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Fazio, L., Rand, D., Lewandowsky, S., Susmann, M., Berinsky, A. J., Guess, A., Kendeou, P., Lyons, B., Miller, J., Newman, E., Pennycook, G., & Swire-Thompson, B. (2024). *Combating misinformation: A megastudy of nine interventions designed to reduce the sharing of and belief in false and misleading headlines.* OSF. https://doi.org/10.31234/osf.io/uyjha

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics. *Political Psychology,* 38, 127–150. https://doi.org/10.1111/pops.12394

Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and Perspectives Not Taken. *Psychological Science,* 17(12), 1068–1074. https://doi.org/10.1111/j.1467-9280.2006.01824.x

Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2024). How persuasive is AI-generated propaganda? *PNAS Nexus, 3*(2), pgae034. https://doi.org/10.1093/pnasnexus/pgae034

Haider, J., & Sundin, O. (2020). Information literacy challenges in digital culture: Conflicting engagements of trust and doubt. *Information, Communication & Society*, 1–16. https://doi.org/10.1080/1369118X.2020.1851389

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't. *American Behavioral Scientist,* 65(2), 371–388. https://doi.org/10.1177/0002764219869406

Juul, J. L., & Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences,* 118(46), e2100786118. https://doi.org/10.1073/pnas.2100786118

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience, 24*(11), 103364. https://doi.org/10.1016/j.isci.2021.103364

Kohnen, A. M., Mertens, G. E., & Boehm, S. M. (2020). Can Middle Schoolers Learn to Read the Web Like Experts? Possibilities and Limits of a Strategy-Based Intervention. *Journal of Media Literacy Education,* 12(2), 64–79.

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens Versus the Internet: Confronting Digital Challenges With Cognitive Tools. *Psychological Science in the*

Public Interest, 21(3), 103–156. https://doi.org/10.1177/1529100620946707

Kozyreva, A., Wineburg, S., Lewandowsky, S., & Hertwig, R. (n.d.). Critical Ignoring as a Core Competence for Digital Citizens. *Current Directions in Psychological Science,* 2022(0), 09637214221121570. https://doi.org/10.1177/09637214221121570

Leder, J., Schellinger, L. V., Maertens, R., van der Linden, S., Chryst, B., & Roozenbeek, J. (2024). Feedback exercises boost discernment of misinformation for gamified inoculation interventions. *Journal of Experimental Psychology: General, 153*(8), 2068–2087. https://doi.org/10.1037/xge0001603

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Lo, J. C., & Adams, C. I. (2018). Civic literacy through literacy instruction: Using Structured Academic Controversy in a government classroom. *Citizenship Teaching Learning and Instruction, 13*(1), 83–104.

Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied, 27*(1), 1–16. https://doi.org/10.1037/xap0000315

Maertens, R., Roozenbeek, J., Simons, J., Lewandowsky, S., Maturo, V., Goldberg, B., Xu, R., & Linden, D. S. van der. (2023). *Psychological Booster Shots Targeting Memory Increase Long-Term Resistance Against Misinformation.* OSF. https://doi.org/10.31234/osf.io/6r9as

Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications,* 5(1), 1–20.

Martella, A. M., Schneider, D. W., O'Day, G. M., & Karpicke, J. D. (2024). Investigating the intensity and integration of active learning and lecture. *Journal of Applied Research in Memory and Cognition.*

Miller, J. M., Saunders, K. L., & Farhart, C. E. (2016). Conspiracy Endorsement as Motivated Reasoning: The Moderating Roles of Political Knowledge and Trust. *American Journal of Political Science,* 60(4), 824–844. https://doi.org/10.1111/ajps.12234

Moore, R. C., & Hancock, J. T. (2022). A digital media literacy intervention for older adults improves resilience to fake news. *Scientific Reports, 12*(1), 6008. https://doi.org/10.1038/s41598-022-08437-0

Nygren, T. (2019). *Fakta, fejk och fiktion: Ämnesdidaktisk digital kompetens för lärare.* Natur & Kultur Akademisk.

Nygren, T. (2023). *AI i skolan: Möjligheter och utmaningar i undervisningen.* Natur och kultur.

Nygren, T. (2025). Artificial Intelligence in Schools: Educational Challenges and Opportunities. Cambridge University Press.

Nygren, T., Al-Afifi, M., & Axelsson, C.-A. W. (2025). *Boosting Fact-checking in the Classroom: Verifying War Photos and the Pitfalls of Overconfidence in Education against Disinformation, Education Inquiry. https://doi.org/10.1080/20004508.2025.2558407*

Nygren, T., & Guath, M. (2022). Students Evaluating and Corroborating Digital News. *Scandinavian Journal of Educational Research*, 66(4), 549–565. https://doi.org/10.1080/00313831.2021.1897876

Nygren, T., Guath, M., Axelsson, C.-A. W., & Frau-Meigs, D. (2021). Combatting visual fake news with a professional fact-checking tool in education in France, Romania, Spain and Sweden. *Information, 12*(5), 201.

Osborne, J., & Allchin, D. (2024). Science literacy in the twenty-first century: Informed trust and the competent outsider. *International Journal of Science Education,* 1–22. https://doi.org/10.1080/09500693.2024.2331980

Osborne, J., & Pimentel, D. (2022). Science, misinformation, and the role of education. *Science, 378*(6617), 246–248. https://doi.org/10.1126/science.abq8093

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice Reduction: Progress and Challenges. *Annual Review of Psychology, 72*(Volume 72, 2021), 533–560. https://doi.org/10.1146/annurev-psych-071620-030619

Prike, T., Holloway, J., & Ecker, U. K. H. (2024). Intellectual humility is associated with greater misinformation discernment and metacognitive insight but not response bias. *Advances.in/Psychology, 2,* e020433. https://doi.org/10.56296/aip00025

Roozenbeek, J., Culloty, E., & Suiter, J. (2023). Countering Misinformation. *European Psychologist, 28*(3), 189–205. https://doi.org/10.1027/1016-9040/a000492

Roozenbeek, J., Linden, S. van der, & Nygren, T. (2020). Prebunking interventions based on "inoculation" theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review, 1*(2). https://doi.org/10.37016//mr-2020-008

Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R. H., Sultan, M., & van der Linden, S. (2021). Susceptibility to misinformation is consistent across question framings and response modes and better explained by open-mindedness and partisanship than analytical thinking. *Judgment and Decision Making.*

Roozenbeek, J., Remshard, M., & Kyrychenko, Y. (2024). Beyond the headlines: On the efficacy and effectiveness of misinformation interventions. *Advances.in/Psychology, 2,* e24569. https://doi.org/10.56296/aip00019

Tay, L. Q., Lewandowsky, S., Hurlstone, M. J., Kurz, T., & Ecker, U. K. H. (2023). A focus shift in the evaluation of misinformation interventions. *Harvard Kennedy School Misinformation Review.* https://doi.org/10.37016/mr-2020-124

Thorson, E. (2024). *How News Coverage of Misinformation Shapes Perceptions and Trust.* Cambridge University Press.

Vegetti, F., & Mancosu, M. (2020). The Impact of Political Sophistication and Motivated Reasoning on Misinformation. *Political Communication, 37*(5), 678–695. https://doi.org/10.1080/10584609.2020.1744778

Voelkel, J. G., Stagnaro, M., Chu, J., Pink, S. L., Mernyk, J. S., Redekopp, C., Ghezae, I., Cashman, M., Adjodah, D., & Allen, L. (2023). Megastudy identifying effective interventions to strengthen Americans' democratic attitudes. *OSF Preprints. March, 20.*

26. CRISIS COMMUNICATION FOR PSYCHOLOGICAL DEFENCE

JESPER FALKHEIMER

SUMMARY

- Crisis communication is goal-oriented communication before, during and after crises in relation to different groups, stakeholders and society as a whole
- Four principles should guide crisis communication operations: preparedness, credibility, openness and speed
- Like practice in psychological defence, crisis communication combines proactive approaches (prevention, preparation, monitoring) with reactive approaches (strategies and tactics for coping with the crisis, responding, attribution, repairing as well as recovery and return to normalcy).
- There are three fundamental crisis communication approaches defined by different overall objectives: instructing, adjusting and reputation management

Psychological defence is a multidisciplinary field of knowledge, and the practice is multifaceted. But among all connections to different disciplines, one stands out as inseparable to psychological defence – crisis communication. This chapter aims to summarize and show how crisis communication knowledge is beneficial for countering malign foreign interference and information influence campaigns.[64] The goal for actors spreading disinformation is to create some form of crisis in societies, groups and for individuals – a collective perception of uncertainty, chaos, mistrust, anger and loss of common meaning and direction. Like practice in psychological defence, crisis communication combines proactive approaches (prevention, preparation, monitoring) with reactive approaches (strategies and tactics for coping with the crisis, responding, repairing as well as recovery and return to normalcy).

To manage a disaster or physical breakdown obviously a broad task, but the management of the crisis, as defined above, largely means communicating. Crisis communication naturally takes place when a crisis event has occurred, during the acute phase and the recovery phase. But as is shown in crisis management theory it is also important for organizations to communicate during other phases, during the preparation phase (if it exists) and during the learning phase. During the preparation phase communication can be decisive for the development of the crisis. But also, to prepare for a crisis, under regular operating conditions, risk communication is essential. Legitimacy and credibility are central aspects both in crisis communication and psychological defence – and it these social constructs are built over time, not in acute situations.

[64] This chapter is based in a part of a crisis communication handbook written by same author published in Georgia by Center for Development and Strategy (with Tinatin Aghniashvili), within the framework of the project "Strategic Communications for Better Future", funded by the U.S. Department of State.

Crises happen at different levels. Sometimes a crisis is limited and mostly of concern for a small group or an individual, sometimes crises are having organizational and societal consequences. In this chapter the focus is focus on crises which may be serious threats to the basic structures or the fundamental values and norms of society, which under time pressure and highly uncertain circumstances necessitates making critical and fast decisions (Rosenthal et al 1989)[65]. Even if the focus are crises which threaten societal structures it also includes advice about organizational crises, since these two different crisis perspectives both add to each other and may lead to conflicting strategies – defending public or organizational interest. But crisis communication always involves organizations since they act as "the entity responsible for managing the crisis" and have a legal responsibility for managing many crises (Coombs & Holladay 2022)[66].

While there are physical events and incidents such as accidents or disasters that create crises, there are also crises that amplify due to moral scandals or experienced misbehaviour. Malign information influence activity may take its starting point in either or. It is important to understand that a crisis is always about how people *create meaning* about what is or has happened, not necessarily about what is or has happened. That said, for professional crisis communicators the goal must always be that people get a correct and fair perception of what is happening or what has happened and may act and respond based on facts rather than misperceptions. Obviously, some people do have a direct experience of an event or incident, but the overall majority in a society is totally dependent on secondary sources such as news media, social media platforms, governmental or corporate information sources. Crises are mediated for most of us. This is why crisis communication is such an interesting but complex practice, and why fact-based, fast and transparent crisis communication is so important for the functioning of society and its institutions, and a prerequisite for avoiding disinformation and misinformation.

CRISIS AND COMMUNICATION AND FUNDAMENTAL STRATEGIES

Crisis communication is goal-oriented communication before, during and after crises in relation to different groups, stakeholders and society. In practice crisis communication includes different operations such as:

- preparing and implementing crisis communication organizations
- developing action plans, training and exercises
- creating crisis communication scenarios
- building collaborative relationships with important stakeholders, collaborators and opinion leaders
- production and dissemination of information to the public during a crisis through various media channels
- give instructional warnings, communicating with journalists through press conferences and other activities

[65] U. Rosenthal, M. T. Charles, and P. 't Hart, eds., Coping with Crisis: The Management of Disasters, Riots and Terrorism (Springfield: Charles C. Thomas, 1989), 10.
[66] Timothy W. Coombs and Sherry J. Holladay, eds., The Handbook of Crisis Communication, 2nd ed. (Hoboken, NJ: Wiley-Blackwell, 2022).

- supporting top management with advice and information including the decision-making
- supporting co-workers with crisis information during and after the crisis
- auditing responses (attitudes and behavior) related to the crisis and crisis communication activities and establishing systems for learning after the crisis.

Crisis communication is a given combination of *crisis* and *communication*. The word crisis originally comes from the Greek *krisis* and does not necessarily mean something negative. In fact, *krisis* as a word stands for a deciding moment or a critical turning point which may lead to recovery or death. Even in societal crises this definition may have some value since some societal crises indeed may lead to something better, if managed in a beneficial way. This is obviously not the case with crises related to disasters or accidents but may be relevant when it comes to some political, financial or organizational crises. A crisis may be called an extreme event, an *extraordinary event* or a serious social disturbance. The word crisis is also used by some researchers as the condition when an organization cannot handle an extreme event in a satisfactory manner. It means that if a difficult situation is managed in a good way, the crisis will not occur. On the other hand, a crisis which is badly managed and communicated may lead to a *double crisis* – a second crisis related to the fact that the first crisis was mismanaged. This occurs very often in the so-called post-crisis phase when news media and other actors scrutinize the management of the crises and may find that information was uncoordinated, rare and sometimes even false.

The word *communication* comes from the Latin word *communicare* and can, simplified, be defined in two ways. Either one defines communication as the transmission and exchange of information between a sender and a receiver or mass of receivers, or one defines communication as sharing common meaning and sense between participants in a symmetric process. The first view is called a transmission perspective, and the second view is called a sensemaking perspective. Both perspectives have their merits even if the sensemaking view is more relevant if one aim to create deep human understanding and a common meaning, since it builds on dialogue as a foundation (Falkheimer & Heide 2023). But one of them does not necessarily exclude the other, since they focus on different aspects and phases of the communication process. The transmission perspective is still a primary approach in crisis situations when an organization needs to inform a large group of people as fast as possible. One example of this is during an evacuation situation, where instructive information needs to be transmitted to all concerned actors at once through public information systems and news media. But there is often also a need of adding a sensemaking communication approach in these situations, when people that are resistant to evacuate need to have interpersonal dialogues with emergency personnel (that still have a persuasive aim, so these are not true dialogues of course).

The crisis communication ideal is based on an idea that one should be open and inform as completely as possible about what happens in a crisis and what consequences it has for the public to protect fundamental values such as the life and health of citizens. There are three fundamental crisis communication approaches defined by different overall objectives (Sturges 1994)[67].

[67] David L. Sturges, “Communication through Crisis: A Strategy for Organizational Survival”, Management Communication Quarterly 7, no. 3 (1994): 297-316.

- **Instructing information:** Crisis information that tell the public, stakeholder and individuals how to react to the crisis with public safety as the directing principle. A typical crisis message following this approach would be a simple, clear and mass communicated instruction telling people what to do.
- **Adjusting information:** Crisis information that helps people psychologically cope with the magnitude of the crisis. This involves communication efforts concerning the psychological state of the public or groups during a crisis, aiming to decrease uncertainty, avoid panic and amplification of crises. A typical adjusting crisis message focus on reassuring the public that there is a plan and leadership and that things are taken care of.
- **Reputation management communication:** Crisis information that will affect the legitimacy or an image about the organization among the public and stakeholders. This includes image repair strategies (from denials to excuses) and other communication efforts intended to improve an organization's reputation before, during or after a crisis.

Instructional and adjusting crisis communication needs to be prioritized in all crises, putting the public in centre, while reputational crisis communication mainly serves the self-interest of organizations. It is also a fact that the reputational damage of organizations in a crisis becomes severe of instructing and adjusting communication is not working well.

DIFFERENT TYPES OF CRISES

Aside from the division between organizational and societal crisis, which are tied together in crisis communication practice, it is important to understand different types of crises – since the type also affects which strategy to use.

Firstly, you can differentiate between crises caused by nature and those caused by humans. But many crises obviously involve a combination of these.

Second, you can make a division between when crises are unintentional or intentional. Among the unintentional we find accidental crises such as natural disasters, epidemics and (most) technical accidents. Among the intentional crises we find terrorism, fraud, riots and sabotage, as well as cyber-attacks against companies or government.

Third, a common typology divides crises into different characters: natural crises, technological crises, social crises, leadership crises and economic crises. These crises are also linked to each other in the sense that one may cause the other.

A fourth division of crises is based on the course or passage of events, i.e. how fast the crisis event develops and ceases. This division leads to four types of crises (Hart & Boin 2001)[68]:

- the fast burning
- the cathartic
- the slow burning
- the long shadow crisis

[68] P.'t Hart and A. Boin, "Between Crisis and Normalcy: The Long Shadow of Post-Crisis Politics," in Managing Crises: Threats, Dilemmas, Opportunities, ed. Uriel Rosenthal (Springfield: Charles C. Thomas, 2001).

The fast-burning crisis occurs and ends quickly, the cathartic crisis ends quickly after a slow emergence, the slow burning crisis creeps forward slowly and slowly fades away, and the long shadow crisis may occur quickly but creates continued further problems, often of a political nature. Many accidents, such as major flight accidents, are fast-burning crises. Fraud in an organization leading to bankruptcy, may be an example of the cathartic crisis. Flooding is an example of the slow-burning crisis. The earthquake in the Indian Ocean in 2004 that caused several tsunamis on December 26 leading to a human disaster in many countries may be an example of a long shadow crisis.

Finally, one of the more well-known and used definition of crisis types is *situational crisis communication theory,* SCCT (Coombs 2022)[69]. The latter is based on an understanding of how people attribute the causes of crises to other people's behaviors and events – in other words who is to blame. Most of us tend to project responsibility on others, also in crisis situations. In SCCT, the crisis analysis takes its starting point in the current situation in that an assessment is made of how people are supposed to relate to and perceive a crisis and its causes based on both emotional and cognitive reasons. According to SCCT we can typically find *three different crisis types*.

- When the organization is also a *victim* (attributed weak responsibility which means that legitimacy or trust is hardly threatened at all). Typically, this includes natural disasters, false rumors, terrorist attacks or sabotage.
- When the organization experience an unintended crisis due to *unfortunate circumstances* (the organization is attributed minimal responsibility, which means moderate threat to reputation or trust). Typically, this includes technical errors and accidents.
- When the organization is *responsible* and could have prevented the crisis (attributed great responsibility that leads to a serious threat to legitimacy or trust). Typically, this includes accidents caused by people (managers, employees) e.g. in the form of industrial accidents or products exploding, and misdemeanors (e.g. cheating, breaking the law or rules and taking risks that lead to accidents).

After concluding crisis type – which may not be totally clear in the beginning of the process – it must be established whether there are aggravating circumstances from previous crises and what trust capital the organization has from before. After that, a communication strategy is chosen that is suitable regarding the situation, type of crisis, trust capital and previous crisis history. Based on the SCCT, there is some general advice for crisis communication strategy. Among other things, to:

- Only engage in instructive (e.g. factual warnings) and adaptive crisis communication (focus on psychological aspects) crisis communication during crises when the organization is a victim, has trust capital and lacks previous aggravating crises in its history.
- Act with mitigating crisis communication (don't take responsibility) during crises where the organization is a victim or in connection with unintended accidents.
- Make efforts with restorative crisis communication (take responsibility) during crises that are accidents or that could have been avoided where the organization has a previously weak trust capital.

[69] Timothy W. Coombs, Ongoing Crisis Communications: Planning, Managing and Responding, 6th ed. (Thousand Oaks, CA: Sage, 2022).

- Make efforts with denial crisis communication when rumors, which are not true, are spread.
- Do not mix denial with diminishing or restorative crisis communication strategies.

FOUR CRISIS COMMUNICATION PRINCIPLES

Based in experience and research it is possible to define four crisis communication principles that may be guiding lights for every crisis communicator, independent of national context[70], and which may also be relevant for the practice of psychological defense. These principles combine and balance two considerations – efficiency and ethical practice.

PREPAREDNESS

Every organization must be prepared for a crisis event to occur. This is also an important assignment for crisis communicators which are supposed to be the ears and eyes of institutions and authorities. Many organizations, especially public ones, must also be prepared to deal with crisis events out in society even though they themselves are not directly affected by them and responsible for their own part. For the organization, preparedness as a communication principle means to:

- auditing crisis signs and signals.
- analyse risks where vulnerability and possible crises are listed.
- state consequences of these risks.
- set guidelines and take measures to prevent crises.
- prepare a crisis management plan and a crisis communication plan.
- outline a crisis organization/institution.
- identify everyone who may be affected by various crises.

It is necessary to prevent crisis events by implementing measures that prevent such or mitigate the consequences of them. For crisis communicators this means preparing staff, organization, channels, platforms as well as auditing opinions and so forth.

CREDIBILITY

Legitimacy is built over time and may be seen as a capital to spend when a crisis occurs. Legitimacy is usually defined as an individual or collective perception about which actors and institutions have the right to rule, regulate, and decide[71]. In other words, it is about who and if you trust an institution, authority, organization or even an individual. Credibility is necessary for building legitimacy and what you can work with in an operational way. The principle of legitimacy means that you and your organization need to focus not only on plans, strategies and actions during crises, but on strategies that enhance legitimacy during "normal" times. We know that *honesty, openness and competency* are important dimensions of being experienced as credible and in the long run legitimate. Therefore, there is a need of focusing these dimensions on your organization all the time and communicating them to all possible stakeholders and to the wider public.

[70] These principles are inspired by different sources but especially SEMA (Swedish Emergency Management Agency), "Crisis Communications Handbook" (Huskvarna: NRS Tryckeri, 2008:3).
[71] J. Falkheimer. Legitimacy Strategies and Crisis Communication. in Oxford Research Encyclopedia of Politics (Oxford: Oxford University Press, 2021).

OPENNESS

Openness towards the public, news media and other authorities is a crucial communication principle, based on ethical and moral standards. Hiding what has happened or not telling the truth about possible consequences is never a good idea. Openness is also recommendable from an efficiency perspective since crisis communication depends on, as mentioned, legitimacy. The public or concerned groups in a crisis follow authorities' recommendations and instructions only when they trust the same authorities. Openness may sound as a clear principle, but in practice it is surrounded by complexity and not always so easy to follow. This is due to two factors. First, you may not have access to correct information and the whole picture of what has happened. Information about what has happened may be unverified and unreliable. Second, there may be judicial constraints that make it impossible to be totally transparent. These constraints may be linked to the protection of national safety as well as individual privacy. The principle of openness can guide us as a principle but does not mean that all information can be instantly free. You as a crisis communicator are responsible for not spreading false or unverified information and following legal standards. You can be open anyway by being open about uncertainties and explaining the situation as much as possible and explain considerations for privacy that needs to be fulfilled. It is important to avoid so called an information vacuum, a situation where you remain silent towards the public (and news media) due to different reasons. In an information vacuum rumours, misinformation (unintended), disinformation (intended) and conspiracy theories emerge and amplify.

SPEED

A crisis usually happens fast. Even if there may be signals about what is happening, a crisis means that you as a crisis communicator need to respond instantly. This is, as discussed above regarding openness, not always easy since you do not have the information you need. But the public as well as news media demand instant crisis information. In practice authorities cannot compete with the speed of news media, and they shall not compete either. Still, one may need to learn from some of the logics of how professional news organizations work, since they have efficient working methods. Professional news organizations are:

- constantly auditing signals in the outside world and are quick to identify upcoming crises.
- having preparedness to quickly staff the newsroom in the event of a crisis.
- are constantly trained in processing and producing information for a wide audience.
- working in the field as close to events and crises as possible.
- can, thanks to digital publishing, correct inaccuracies quickly.

As emphasized above one may learn something from news organizations when organizing and preparing crisis communication, but one should avoid competing with the speed that journalists have. Compared to news journalists, crisis communicators can never spread information that is not certain or verified – this may also be a principle for news journalism, but in practice the standard of verification is not as high as for authorities.

DISCUSSION

- What competence and resources are needed to cope with larger societal crises?
- What is the role of news media, social media and other platforms in contemporary crises?
- Speed and openness are two principles of effective crisis communication. But whichs risks do these principles also create when appled in practice?
- What can actors in the psychological defence learn from crisis communication theory and practice?

JESPER FALKHEIMER, *PhD, is Professor of Strategic Communication, Department of Communication, Lund University. He is national coordinator for the research network Communication and Media in Crisis and War (Campus Totalförsvar) and a researcher at the Lund University Psychological Defence Research Institute. He is also Professor II at Kristiana University of Applied Sciences, Norway, Honorary Professor at Hong Kong Polytechnic University, Visiting Professor at University of Johannesburg and Editor-in-Chief for Journal of Communication Management. His research interests are strategic communication in general, and crisis communication, disinformation and communication management.*

REFERENCES

Boin, A., 't Hart, P., Stern, E. K. & Sundelius, B. (2005). *The politics of crisis management: Public leadership under pressure.* Cambridge: Cambridge University Press.

Coombs, T. W. (2022). *Ongoing crisis communications: Planning, managing and responding.* 6th Edition. Thousand Oaks, CA: Sage.

Coombs, T.W. and Holladay, S. (2022). *The Handbook of Crisis Communication,* 2nd ed. Hoboken, NJ: Wiley-Blackwell.

Falkheimer, J. (2021). Legitimacy Strategies and Crisis Communication, in *Oxford Research Encyclopedia of Politics*. Oxford: Oxford University Press.

Hart P., Boin A. (2001). Between crisis and normalcy: the long shadow of post-crisis politics. In: Rosenthal ABU (ed) *Managing crises: threats, dilemmas, opportunities.* Charles C. Thomas.

Mitroff, I., Shrivastava, P. and Udwadia Firdaus E. (1987). Effective Crisis Management. *The Academy of Management Executive* (1987-1989), Vol. 1, No. 4 (Nov. 1987), pp. 283-292.

Rosenthal, U., Charles, M. T. & 't Hart, P. (red.). (1989). Coping with crisis: The management of disasters, riots and terrorism. Springfield: Charles C. Thomas.

Sturges, D. L. (1994). Communication through crisis: A strategy for organizational survival. *Management Communication Quarterly.* Vol. 7(3), 297-316.

Swedish Emergency Management Agency, KBM (2008). *Crisis Communications Handbook.* Huskvarna: NRS Tryckeri, Rapport 2008:3.

27. LESSONS LEARNED FROM THE WAR IN UKRAINE

IVAR EKMAN AND PER-ERIK NILSSON

SUMMARY

- The war in Ukraine shows clearly that communications writ large is a matter of life and death in modern warfare.
- The fact that Ukraine was prepared for this when the Russian full-scale invasion happened in 2022 made a big difference for the trajectory of the information struggle.
- The Ukrainian experience is that civil society holds vast informational fire power, and can play a big and important role.
- Informational control and restrictions on the enemy's access to the informational battlefield also were important parts of Ukrainian success in the first two years of the full-scale war.
- Adaptability, speed, and emotional resonance are key in modern warfare's informational domain.

Communication for the sake of life and death. That is what Ukraine's struggle on the information front since the full-scale invasion of February 2022 has been about. Not simply a fight to gain the greatest number of "likes" on social media, or to be hailed as making the wittiest meme mocking the Russian leadership. But a constant battle for the unity of the nation, where failures in communication infrastructure can cost the lives of civilians; where soldiers' haphazard usage of cell phones makes them visible targets for strikes; where the battle of the perceived reality abroad is essential for political, economic, and military support; where information manipulation and psychological operations are constantly trying to put a wedge into the morale of the Ukrainian people, their will to fight, and their relations with their supporting nations (Nilsson and Ekman, 2024; Ekman and Nilsson, 2023 – most of the conclusions of this chapter are based on these two texts).

When the full-scale invasion took place, Ukrainian society and its state institutions were, in some fundamental ways, uniquely prepared for an across-the-board information battle with a unique enemy, Russia. It is worth keeping in mind that Russia in the run-up to 2022 generally was viewed as preeminent among great powers for mastering the informational aspect of warfare. Russia had also showed the world its apparent skills in a number of instances, from the largely bloodless annexation of Crimea by "little green men" in 2014, to the destabilizing interventions during the 2016 US presidential elections. Russia seemed to be particularly adapted in mastering the mediatisation of war compared with its antagonists, gaining a much better grasp of the dynamics of the new digital information environment (Hoskins and O'Laughlin's, 2015; Nilsson and Ekman, 2024, p. 33-38). In other words, Russia, armed with its bot farms and a century-

long experience of exerting influence through information manipulation (Horbyk, Prymachenko, and Orlova, 2023), was expected to roll over the Ukrainians as easily on the information battlefield, as many predicted they would in the icy forests and fields of Ukraine.

Ukraine, however, came to this battlefield prepared. The war with Russia did not begin in 2022 – it began in 2014. Events such as the occupation and illegal annexation of Crimea, and morale-sapping battles like the one in Debaltseve in the Donbass in 2015, had taught Ukraine bitter but crucial lessons on both the importance of managing information and communications in a strategic way, but also the strengths and potential weaknesses of Russian capabilities.

This realisation was shared across Ukrainian society. Ukraine had gone through a major political upheaval dubbed" the Revolution of Dignity" in 2013-14, where civil society had played a crucial role. Thus, many in Ukraine – both inside and outside the state apparatus – saw the need of creating not only their own media ecology free from Russian influence but also measures to counter Russian information manipulation and influence, and the ability to formulate a strong message of their own. Crucially, this development was not only seen as necessary to resist and dispel Russian interference, but it was also considered to be essential that it was carried out in line with the country's need for reliable information, as a country at war, and for safeguarding democracy and a rights-based society.

Many of these efforts – such as the fact-checking organisation StopFake, and the media outreach organisation Ukraine Crisis Media Center – were sprung out of civil society, laying the foundation for a true whole-of-society approach to strategic communication. At the same time, in this period the state began to reform its own handling of issues relating to communication and information security. In 2014, a Ministry of Information Policy was launched. Several governmental institutions adjusted and adapted their communications strategies, policies, and doctrines in the following years. For example, the Ministries of Culture, Defence, and Foreign Affairs developed capabilities and know-how in their communication with domestic and international audiences. At the time, the Armed Forces of Ukraine realised that domestic public relations and offensive communications against the adversary required modernisation. The ensuing work by government agencies was carried out by, among other things, studying how other countries approached the issue of strategic communications and through the direct support of NATO and EU countries.

During this period, several measures were also implemented to restrict Russian influence in the Ukrainian information environment. In October 2014, the government banned 14 Russian television channels from the Ukrainian cable networks. In 2017, the government issued an executive order mandating internet service providers to restrict access to prominent Russian websites and social media platforms, including the second most popular, VKontakte. This move was prompted by concerns about the Kremlin's influence over Russian social media and the potential for collecting data on Ukrainian citizens. In early 2021, the Government shut down three domestic TV channels with Kremlin affiliations. All these actions spurred debate on freedom of speech but were defended by the authorities as necessary for the safeguarding of Ukraine. President Volodymyr Zelenskyy, as an example, motivated the 2021 shutdown by citing the urgency to "fight against the danger of Russian aggression in the information arena" (Dickinson, 2021).

As governmental institutions worked with capability development, extensive competence on matters related to strategic communication and information security that had emerged outside of government was incorporated into state structures. The interplay of government, corporations, and civil society informed the capability development and eventually led to several institutionalised government functions. For example, in 2021, the Centre for Strategic Communication and Information Security was established under the Ministry of Culture and Information Policy, while the Centre for Countering Disinformation was set up under the National Security and Defense Council.

In sum, when Russian forces crossed the border in the early hours of February 24, 2022, Ukraine had been laying the groundwork for a functional information warfare and strategic communication machinery for several years. As one veteran "information warrior" put it: "The eight years taught us a lot. We learnt the Russian playbook, their narratives, the main actors, their main tricks. In February, when they attacked us, we were prepared" (Ekman & Nilsson, 2023, p. 66).

However, this did not mean that success on the information front was by any means guaranteed. Indeed, a number of key aspects of how Ukraine handled strategic communication in the early phase of the war was practically impossible to prepare for, but turned out crucial for the battle that came. One of, if not the, most important of these aspects was the Ukrainian president Volodomyr Zelenskyy. He came to the war as an actor, comedian, and satirical commentator of Ukrainian and Russian politics turned postmodern leader who ran a presidential campaign with a vague anti-establishment and populist political program. He struggled in his first years as president, but when the full-scale invasion took place, Zelenskyy rose to the occasion, not the least by staying in Kyiv. "Churchill with an iPhone" as a British journalist baptised him (Freedland, 2022), Zelenskyy became the foremost symbol of the Ukrainian one-voice policy. The vision is of a policy that can be understood as a communications pyramid where strategic messages trickle down and are amplified by other actors. As one senior Ukrainian communicator puts it: "The communication pyramid is a very simple communication model. Important messages are delivered by important people – the president should speak first, then the respective ministers and subordinated structures should take it further" (Ekman & Nilsson, 2023, p. 28 -29). These communications efforts, however, have not been organised hierarchically in the bureaucratic sense. Instead, the endeavour has been polyphonic. In terms of strategic communication, this translates into a dynamic and reciprocal communications process, involving top-down as well as bottom-up channels, where information flows and are amplified bi-directionally between the parties.

How, then, has this been achieved? The (likely) most fundamental aspect of this process is a potent "rallying around the flag" effect, manifested in the Ukrainian people's morale and will to defend their country. Notably, the concept of willingness to defend extends beyond the mere protection of a country's current borders and institutions. In the case of Ukraine, as examined by Jānis Bērziņš and Victoria Vdovychenko (2022), the "rally around the flag" effect represents also a societal commitment to the country's future, specifically towards increased democratisation and integration with NATO and the EU. According to a senior Ukrainian communicator, Zelenskyy embodied this broader societal response: "Zelenskyy feels it very well, that is why he and his team are fast and creative. They feel the mood of the Ukrainians. He is not avant-garde; the society is avant-garde. As president, you cannot betray these people. You cannot be weak when they are

so good and strong. It is a mistake to say that Zelenskyy gives this push, it is the Ukrainian society doing this. It comes from the bottom to the top, not the other way around" (Ekman & Nilsson, 2023, p. 73).

Ukrainian strategic communications since 2022 have by all accounts been both efficient and innovative and have been forged into a distinct brand radically different from its Russian counterpart. The Ukrainian case underlines the importance of both reactive measures (analysis and debunking) and proactive steps (anticipatory communication and attacking the sources of disinformation) to foresee and neutralise potential disinformation (Kalenský and Osadchuk, 2024). From this perspective, speed and proactive communications are pivotal, as they shape the narratives of events, ideally before the adversaries can develop their own line of attack. In Ukraine, the proactive generation of rapid responses is due to formal and informational agility, which is essential in today's fast-paced information environment.

A less innovative, but also crucial part of the Ukrainian management of the information environment, has been information control. In times of war, nations employ various strategies to manage the information disseminated through the news media. This was certainly relevant on 24 February 2022, when President Zelenskyy declared a state of emergency in Ukraine, implementing measures that included a prohibition on creating and spreading information that could cause destabilisation. Shortly thereafter, martial law was enforced and, in early March, the Commander in Chief, Valerii Zaluzhnyi, ordered restrictions on conveying information that could disclose military actions. In March 2022, as an outcome of the situation, the National Security and Defense Council of Ukraine formalised an existing initiative that sought to consolidate the resources of major commercial TV networks and public service, enabling a unified broadcast across the channels, thereby establishing Ukraine's "United News" format (Ukr. "Єдині новини," often also referred to as the TV "marathon," or "telethon"). When formalising this arrangement, Russian military aggression and disinformation were cited as driving factors. The decision empowered the Ukrainian broadcasting regulator to integrate national TV channels under the "United News" platform, which, until further notice by the government, ensured a unified broadcast. Additionally, in April 2022 the move led to the disconnection of three TV channels linked to former president Petro Poroshenko.

While it is too early to assess the Marathon's long-term impact, critical voices in Ukraine express concerns about the risk of promoting uniformity and even partial censorship, potentially undermining democratic principles. This raises the crucial question of whether implementing tools intended to combat disinformation and foreign influence—tools that may seem effective in the short term—could ultimately erode democracy and free speech, leading to unexpected and adverse outcomes in the long run.

In summary, Ukrainian strategic communication in wartime has adapted to difficult challenges through a number of skilful measures: developing a unified institutional voice, employing innovative and proactive counter-disinformation tactics based on a deep understanding of the enemy, tailoring messages to diverse audiences, and exercising control over the national narrative. This holistic approach underscores the importance of adaptability, speed, and emotional resonance in modern warfare's informational domain. As such, Ukrainian strategic communication is a case in point that illuminates a nation that is adjusting, adapting, and innovating

capabilities to manage meaning in the third phase of the mediatisation of war.

However, it is worth remembering that the war is still raging. Just as Ukraine adapted its capabilities in the years leading up to 2022, there are signs Russia has adapted its communications and information operations to Ukrainian achievements in the early phases of the war, and that Russian methods and messages are now having greater impact. In addition, communications always have a close relationship to reality on the ground, and Ukrainian messaging was buffeted by battlefield successes during the war's first year. Now that the outcome of the kinetic war is much more uncertain, the same uncertainty is also felt in the unpredictable terrains of the modern information environment.

DISCUSSION

- What is the best strategy for governments to harness the power of civil society and engaged citizens for strategic communications purposes?
- How can governments, organizations, and civil society effectively collaborate to combat disinformation while maintaining trust, independence, and ethical integrity?
- What strategies can organizations implement to ensure communication continuity and resilience, preventing over-reliance on individual figures?
- How can organizations balance proactive and reactive communication strategies to effectively counter disinformation without amplifying false narratives?

IVAR EKMAN, *MIA, is an analyst at the Swedish Defence Research Agency (FOI), specializing in strategic communications and intelligence methodology. He is currently the program director of Glimt, an open crowd forecasting project supporting Ukraine's war efforts.*

PER-ERIK NILSSON, *Ph.D., is a senior researcher at the Swedish Defence Research Agency (FOI), specializing in information and cyber warfare. His recent work explores Ukrainian strategic communications, Russian cyber operations, the evolution of cognitive warfare, and China's use of information technology for global influence.*

REFERENCES

Bērziņš, J. and Vdovychenko.V. "Willingness to Fight for Ukraine: Lessons for the Baltic States." *BSR Policy Briefing Series*, no. 9. Centrum Balticum Foundation, Turku, 2022: https://www.naa.mil.lv/sites/naa/files/document/J.Berzins_Victoria%20Vdovychenko_%20Willingness%20to%20fight%20for%20UkraineLe ssons%20for%20the%20Baltic%20states.pdf.

Dickinson, P. "Analysis: Ukraine Bans Kremlin-Linked TV Channels." Blog post, Atlantic Council, February 5, 2021: https://www.atlanticcouncil.org/blogs/ukrainealert/analysis-ukraine-bans-kremlin-linked-tv-channels/.

Ekman, I. and Nilsson, P-E. "Ukraine's Information Front; Strategic Communication During Russia's Full-Scale Invasion of Ukraine." *FOI-R--5451--SE*, Totalförsvarets Forskningsinstitut, Stockholm, 2023: https://www.foi.se/rest-api/report/FOI-R--5451--SE.

Freedland, J. "A Key Reason Putin's Bloody Invasion is Faltering? He's no Match for Zelenskiy's iPhone." *The Guardian,* March 25, 2022: https://www.theguardian.com/commentisfree/2022/mar/25/churchill - iphone -volodymyr.

Horbyk, R., Prymachenko, R., and Orlova, D. (2023). The Transformation of Propaganda: The Continuities and Discontinuities of Information Operations, from Soviet to Russian Active Measures. *Nordic Journal of Media Studies,* 5(1), 68-94: https://doi.org/10.2478/njms-2023-0005.

Hoskins, A. and O'Loughlin, B. (2015). Arrested War: The Third Phase of Mediatization. *Information, Communication & Society,* 18(11), 1320-1338: https://doi.org/10.1080/1369118X.2015.1068350.

Kalenský, J. and Osadchuk, R. (2024). How Ukraine fights Russian disinformation; Beehive vs mammoth. *Hybrid CoE Research Report* 11, Helsinki, https://www.hybridcoe.fi/wp-content/uploads/2024/01/20240124-Hybrid-CoE-Research-Report-11-How-UKR-fights-RUS-disinfo-WEB.pdf.

Nilsson, P-E. and Ekman, I. (2024). Be Brave Like Ukraine': Strategic Communication and the Mediatization of War. *National Security and the Future* 25(1), 21-65: https://doi.org/10.37458/nstf.25.1.2.

28. DETERRENCE IN THE INFORMATION ENVIRONMENT

HEDVIG ÖRDÉN

SUMMARY

- Deterrence strategies are common in security and military policy. The strategy aims to influence an adversary's cost-benefit analysis.
- Traditional deterrence theory was developed in a Cold War context and include 'deterrence by denial' and 'deterrence by punishment'. These strategies have the purpose of signalling to an adversary that an attack is either unlikely to succeed or too costly.
- Deterrence theory was developed with the purpose of managing nuclear threats. In subsequent 'waves' of deterrence theory, the thinking on deterrence has evolved to address a range of novel threats and a diverse set of threat environments.
- Deterrence in the information environment requires strategies beyond traditional approaches and draws on insights from cyber deterrence and thinking about deterrence in relation to non-state actors.
- Deterrence strategies for the information environment include 'deterrence by denial', 'deterrence by detection' and 'deterrence by de-legitimization'.

The aim of this chapter is to introduce deterrence theory and describe how it can be applied in the information environment. To this end, the chapter provides a short history of thinking about deterrence and introduces three key concepts: *deterrence by denial, deterrence by detection* and *deterrence by de-legitimization.*

Deterrence plays a key role in security and military policy. In short, deterrence refers to a set of practices whereby a defender manipulates an adversary's preliminary cost-benefit calculation related to a future attack. In deterrence theory, such manipulation traditionally comes in two forms. Deterrence may involve signalling to an adversary that any aggression will inevitably be met with retaliation, thereby raising his perception of *costs* associated with an attack. This is called *deterrence by punishment*. Alternatively, a defender can make it clear that the aggressor's envisaged attack is unlikely to succeed, thereby lowering the expected *gains*. This latter form of deterrence is commonly described as *deterrence by denial*. All deterrence strategies rely on *credible signalling* to the adversary. Any cost-raising measures must be clearly communicated to the presumed attacker. Both threats of retaliation and defence capabilities must be perceived as credible.

Deterrence is a Cold War concept. Reflecting the power distribution during the Cold War, and the threat of nuclear annihilation, traditional deterrence theory was highly state-centric. There is much to learn from traditional thinking on deterrence. Nevertheless, deterrence practice must also be adapted to the environment where the aggression plays out. As this chapter will show, deterrence theory has evolved

over time in response to the emergence of non-conventional threats and in response to non-state threat actors. The thinking on how to adapt traditional deterrence theory for the information environment is still evolving. A purpose of this chapter is therefore to give a brief introduction to deterrence in relation to malign information influence and interference and relate emerging discussions on the topic to the broader development of deterrence theory.

The chapter unfolds in the following way. First, it outlines a short history of deterrence theory. Second, the chapter briefly describes the challenges of applying deterrence to the information environment. Third, it introduces how the traditional *deterrence by denial* can be adapted for an information age through societal resilience. Fourth, it introduces the novel *deterrence by detection* inspired by the cybersecurity and cyber deterrence literature. Fifth, the chapter outlines *deterrence by de-legitimization*, drawing on discussions about deterrence in the context of non-state actors. Lastly, the chapter summarizes the findings and suggests a set of questions to facilitate further discussion.

A SHORT HISTORY OF DETERRENCE THEORY

The practice of trying to dissuade an adversary from attacking has a long history. However, a systematic literature on deterrence first emerged during the Cold War. This early deterrence literature primarily focused on how to manage the nuclear threat. Taking as a core premise, the idea of a rational actor operating under conditions of uncertainty, the first 'waves' of deterrence theory introduced well-known concepts in international security such as 'Mutually Assured Destruction' (MAD). Seen through the lens of deterrence, a rational actor would choose to refrain from attack if he was certain of massive retaliation. Security could then – seemingly paradoxically – be upheld through a mutual threat of nuclear annihilation.

The early thinking about deterrence came to influence policymakers and shape international relations during the Cold War (Jervis, 1979). As Lupovici notes (2010: 708), the theory was appealing to decision makers because it provided a 'simple, and even simplistic' solution to highly complicated foreign policy issues. Over time, the thinking on deterrence also evolved with the changing threat environment. Retaining its place as a central tool for statecraft beyond the nuclear age, deterrence theory came to be applied in response to emerging non-kinetic as well as asymmetric threats (Knopf, 2010). The theory also underwent development through internal scholarly debates pointing out the inherent limitations of early frameworks.

A 'third wave' of deterrence theory in the 1970s questioned the premise of the rational actor. This literature points to the central role played by policymakers' beliefs and interpretations of the world as well as the cognitive limitations associated with decision-making under stress (Jervis, 1979). During the subsequent 'fourth' wave of deterrence theory, scholars furthermore highlighted how deterrence is also a *learned* response. Rather than being a direct and rational response to a threat, functioning deterrence is a system of thinking anchored in institutions and part of strategic culture (Lupovici, 2010: 714). Seen this way, any deterrence theory developed in response to changes in the threat landscape would also require an adaptation of institutionalized practices on the international level.

Contemporary approaches to deterrence commonly draw on key traditional constructs while also adapting these to the particularities of the threat. The below section introduces some current thinking about deterrence as a strategy for managing non-conventional threats in the information environment, drawing on the most up to date literature.

DETERRENCE IN THE INFORMATION ENVIRONMENT

The question of how deterrence can be enacted in the information environment has emerged as an important question with a growing focus on non-kinetic and non-conventional threats on the policy level.

The literature on deterrence in the information environment seeks inspiration from traditional deterrence theory as well as novel perspectives on deterrence related to non-conventional threats and threat actors. As we will see below, cyber-attacks and terrorism share some similarities with foreign influence and interference activities. Like cyber-attacks, information influence and interference is often enacted in the digital environment where establishing the identity of threat actors is a challenge. Information influence can also be pursued in combination with cyber-attacks through so-called 'hack and leak' operations. Moreover, like terrorist actors, influence actors operate on a cognitive level, seeking to manipulate the perceptions of a broader population.

Despite these similarities to other non-conventional threats, foreign influence and interference remains a distinct kind of practice. The uniqueness of this threat requires careful considerations about the limitations of existing approaches to deterrence as well as the development of novel deterrence practices.

DETERRENCE BY DENIAL

To convince an adversary that aggression is unlikely to lead to significant gains is a fundamental principle of deterrence. Adapted to the information environment, this form of *deterrence by denial* involves building societal resilience.

The proposed turn to deterrence by denial through a focus on strengthening social resilience is motivated by the uniqueness of the perceived threat in a democratic context. Liberal democracies depend on contestation and open debate for public opinion formation. Threat actors make use of this openness for destructive purposes (Farrell & Schneier 2018), mobilizing existing societal cleavages to their own advantage (Wigell, 2021: 52). However, freedom of speech and freedom of opinion remain core liberal democratic values to be defended, and the strengthening of democracy can be seen as a viable path to deny an attacker any gains without compromising core democratic ideals (Wigell, 2021).

Deterrence by denial relies on strengthening the resilience to malign information influence within broad strands of the population. Pamment and Palmertz (2023: 27-28) for instance call for a strengthening of cybersecurity together with media literacy. To be successful, they argue that this strategy must involve the pinpointing and addressing of *societal* and *technical* vulnerabilities (Pamment & Palmertz, 2023). Cybersecurity measures can address adversary practices such as phishing and cyber-attacks whereby threat actors gain information. Media literacy campaigns, source criticism and fact checking, in turn, work to strengthen the broader population against malign information influence. Broadening the scope for resilience-building to include a more substantial societal transformation, scholars like Wigell (2021) propose a turn to 'democratic deterrence'. Through this focus on strengthening democracy, democratic deterrence involves creating a favourable domestic media environment for citizen activism, investigative journalism and enhanced transparency in political processes (Wigell, 2021). While these strategies differ in scale, domestic resilience measures all come with the potential to deny threat actors any gains from destructive activities in the information domain. Through deterrence by denial, it is possible to push the cost-benefit calculation to the defender's advantage.

Resilience strategies based on media literacy and fact-checking are widely embraced by policymakers. Part of their appeal is that such initiatives, by focusing on individual resilience-building, can easily generate support across (party) political boundaries. Still, resilience through media literacy rarely addresses systemic

vulnerabilities (Monsees, 2023: 162). Comprehensive strategies for democratic deterrence – such as strengthening the domestic media ecosystem and enacting novel transparency legislation related to foreign influence and interference – demand substantial political commitment and potentially broad agreement on resource allocation. As a result, 'democratic deterrence' is more difficult to implement in practice.

Another obstacle to societal resilience is coordination. An effective resilience strategy, integrating both technical and societal elements, necessitates collaboration among diverse stakeholders, including governments, private technology firms, and civil society organizations (Pamment and Palmertz, 2023). These actors often pursue conflicting interests, leading to divergent problem definitions and disagreements over appropriate countermeasures (Ördén, 2019). Consequently, while a 'whole-of-society' approach to deterrence by denial is promising, this strategy also introduces complex challenges.

DETERRENCE BY DETECTION

A novel strategy for deterrence in the information environment is *deterrence by detection.* This form of deterrence leverages the fact that threat actors in the information domain often employ obfuscation to the advantage of the defender.

The attribution of foreign information influence and interference poses a challenge in comparison to many conventional threats. Threat actors furthermore utilize *deniability* as a core strategy. Uncertainty about the identity of an attacker allows actors to evade responsibility. Deliberate obfuscation can also be employed to instill uncertainty among a population. In this context, deterrence scholars have demonstrated how defenders can use the threat of detection, or public attribution of attacks, to manipulate an actor's behaviour. The bulk of scholarship on deterrence by detection seeks inspiration from the literature on cyber deterrence. Like cyber-attacks, contemporary information influence activities tend to operate in the digital environment.

A viable defensive strategy in the digital environment is therefore to develop appropriate technical capabilities for detection. The fundamental idea underlying deterrence by detection is that 'our adversaries are less likely to commit opportunistic acts of aggression if they know they are being watched' (Mahnken & Kim, 2021: 4). Through enhanced threat intelligence and digital forensics, states can for instance develop tools for identifying threat actors and mapping their activities (Mueller et al., 2019: 108). To this end, defending actors strive to enhance capabilities for assessing 'adversary capabilities, intentions and opportunities' (Pamment & Agardh-Twetman, 2019: 127). Drawing on important insight from cybersecurity, models have been introduced for mapping specific tactics, techniques and procedures (TTPs) used by both state and non-state threat actors in the information environment (Innes & Ahonen, 2025: 20). The existence of such advanced capabilities using past data allows defenders to better attribute future threat activities, thereby contributing to enhanced deterrence by detection.

In addition to such technical or forensic solutions, deterrence by detection also depends on defender's *credible signalling* of existing surveillance capabilities. For deterrence to work, a threat actor must believe that such capabilities *exist* (Mahnken & Kim, 2021: 1). They also need to know what activities a defender regards as red lines. To this end, attribution frameworks have been developed and adapted for the information environment (See for instance Palmertz et al., 2025). As highlighted by cyber deterrence scholars, transnational agreements can further

enhance deterrence by detection (Mueller et al., 2019). Such shared agreements can create transnational standards that are perceived as *'unbiased, legitimate and valid,* even among parties who might be antagonistic' (Mueller et al., 2019: 115. Orig. italics.).

The work on deterrence by detection in the information environment has made great strides in the recent decade. At the same time, attribution using TTPs builds on past data and the approach remains a challenge when applied to new threat actors (Innes & Ahonen, 2025: 21). What is more, agreements on standards between allied states can enhance capabilities and allow civil society actors – such as Open-Source Intelligence (OSINT) groups and activists – to contribute to detection and attribution. Still, transnational standards which also include antagonistic parties remains a challenge. In a Cold War context of potential nuclear annihilation, the mutual desire to avoid confrontation had the potential to generate shared institutional practices. In the information context, however, deterrence by detection often needs to be combined with other deterrence practices to be effective in manipulating the cost-benefit calculation of a potential attacker. An example includes deterrence by detection followed by deterrence by punishment in the form of a threat of sanctions.

DETERRENCE BY DE-LEGITIMIZATION

If deterrence through detection emphasises the development of technical capabilities, *deterrence by de-legitimization* focuses on cost-raising through the mobilization of societal norms. This entails an expansion of deterrence beyond physical punishment (Wilner, 2011: 10). Instead, this form of deterrence operates through the threat of *social punishment (*Knopf, 2010: 18).

Deterrence by de-legitimization was initially developed in response to the terrorist threat. In the early discussions on terrorist deterrence, sub-state actors were often seen as 'undeterrable' due to seemingly 'irrational' motivations[72]. While deterrence scholars refute such ideas, they nevertheless suggest that effective deterrence requires rethinking when it comes to asymmetric threats and ideologically motivated sub-state actors. Wilner (2011: 14) for instance argues that deterrence should target the *social influence* of terrorist groups. In this pursuit, defenders can furthermore employ 'discourse as a source of leverage' (Knopf, 2010: 19). Thus, the aim of deterrence by de-legitimization is to 'reduce the challenger's probability of achieving his goals by attacking the legitimacy of the beliefs that inform his behavior' (Wilner, 2011: 26). Through the employment of strategic communication defenders can work to undermine a threat actor's credibility in the public realm.

In the information environment, deterrence by de-legitimization can be viewed as an extension of deterrence by detection. Having first identified a threat actor, a defender can publicly call attention to this actor's undesirable behaviour. This way, defending actors can mobilize the existing social norms to their own advantage and public attribution turns into a form of social punishment. Deterrence by de-legitimization can be achieved through public statements by officials attributing threat activities to a specific actor. Examples include the public attribution of Russian election interference in Germany or Covid-19 disinformation disseminated by Russia, China and Iran by the European Union (Hedling & Ördén, 2025). On

[72] This assumption of irrationality is linked to the hard-line ideological, religious or political motivations of terrorist actors. Deterrence scholars refute such ideas of irrationality but also suggest other paths forward for deterrence since the cost-benefit calculation of terrorist actors might look quite different. For more discussions on this, see for instance Wilner (2015).

a lower level, non-state actors and civil society can furthermore contribute to de-legitimization through the employment of messaging targeting the reputation of adversaries (Pamment & Agardh-Twetman, 2019: 131).

Effective deterrence by de-legitimization in the information environment also comes with challenges. The strategy depends on the existence of shared societal norms for acceptable behaviour (Wilner, 2024: 70). To a certain extent, such norms exist on a domestic level and among like-minded states on the international level. An example is norms around election interference among certain liberal democratic states. Nevertheless, a challenge with de-legitimization is the political risk which springs from the employment of such practices in contexts where shared social norms are under question. Threat actors operating in the information environment often exploit existing domestic vulnerabilities associated with polarization, targeting highly politicized issues. Enacted in a polarized environment, de-legitimization practices might backfire and contribute to further domestic polarization (Hedling & Ördén, 2025). Political actors may then opt for other forms of deterrence when the level of political risk is deemed too high (Hedling & Ördén, 2025).

CONCLUSION

Deterrence theory has evolved significantly since the Cold War, adapting to a novel threat environment. Traditional concepts such as 'deterrence by denial' and 'deterrence by punishment' have been reconsidered in light of contemporary challenges, and expanded on, in the context of non-conventional threats and non-state actors. This has led to novel strategies such as 'deterrence by detection' and 'deterrence by de-legitimization'.

In the information environment, deterrence by denial has been adapted to focus on societal resilience. Strengthening democratic institutions, enhancing media literacy, and bolstering cybersecurity are key components of this approach to deterrence. Similarly, deterrence by detection aims to deter threat actors by enhancing detection and attribution capabilities and producing shared frameworks for attribution. Deterrence by de-legitimization further expands deterrence theory by leveraging societal norms and strategic communication to undermine the credibility of threat actors.

This chapter highlights the necessity of integrating multiple deterrence strategies to effectively counter contemporary threats in the information environment. Given the unique characteristics of the information environment – such as a potential for anonymity, deniability, and societal impact – a one-size-fits-all deterrence model remains insufficient. Instead, a combination of deterrence by denial, detection, and de-legitimization is required for defence against influence and interference activities.

DISCUSSION

- What are the basic principles of deterrence, and how can they contribute to enhancing security?
- What are the challenges to successful deterrence in the information environment?
- What are the benefits and challenges of deterrence by denial, deterrence by detection and deterrence by de-legitimization respectively?
- How can such difficulties be overcome?

HEDVIG ÖRDÉN *is a researcher at the Psychological Defence Research Institute, Lund University, and a postdoc at the Department of Political Science and Public Management, the University of Southern Denmark. She is also an affiliated researcher at the Europe Programme at the Swedish Institute for International Affairs. Her work is situated within critical security studies and critical intelligence studies. She publishes on topics related to security and foreign information influence, intelligence and security expertise, and intelligence and liberal democracy.*

REFERENCES

Farrell, H., & Schneier, B. (2018). Common-knowledge attacks on democracy. *Berkman Klein Center Research Publication,* (2018-7).

Hedling, E. & Ördén, H. (online first, March 31). 'Disinformation, Deterrence and the Politics of Attribution', *International Affairs.*

Innes, M. & Ahonen, A. (2025). Attribution and Information Influence Operations. A 'Field Guide' for Open-Source Investigators and Researchers. ADAC.io EU Horizon Project Deliverable 1.1. Psychological Defence Research Institute; Lund University.

Jervis, R. (1979). Deterrence theory revisited. *World Politics,* 31(2), 289-324.

Knopf, J. W. (2010). The Fourth Wave in Deterrence Research. *Contemporary Security Policy,* 31(1), 1–33.

Kuerbis, B., Badiei, F., Grindal, K., & Mueller, M. (2022). Understanding transnational cyber attribution. *Cyber Security Politics,* 220.

Lupovici, A. (2023). Deterrence through inflicting costs: Between deterrence by punishment and deterrence by denial. *International Studies Review,* 25(3), viad036.

Lupovici, A. (2010). The emerging fourth wave of deterrence theory—Toward a new research agenda. *International Studies Quarterly,* 54(3), 705-732.

Mahnken, T. G., & Sharp, T. (2020). Deterrence by detection: Using preemptive surveillance to prevent opportunistic aggression in the Asia-Pacific Region. *Australian Naval Review,* (1), 65-72.

Monsees, L. (2023). Information disorder, fake news and the future of democracy. *Globalizations,* 20(1), 153-168.

Mueller, M., Grindal, K., Kuerbis, B., & Badiei, F. (2019). Cyber attribution. *The Cyber Defense Review,* 4(1), 107-122.

Palmertz, B., Isaksson, E. & Pamment, J. (2025). A Framework for Attribution of Information Influence Operations. ADAC.io EU Horizon Project Deliverable 1.1. Psychological Defence Research Institute; Lund University.

Pamment, J., & Agardh-Twetman, H. (2019). Can there be a deterrence strategy for influence operations?. *Journal of information warfare,* 18(3), 123-135.

Pamment, J., & Palmertz, B. (2023). Deterrence by denial and resilience building. In *Routledge Handbook of Disinformation and National Security* (pp. 20-30). Routledge.

Wigell, M. (2021). Democratic deterrence: How to dissuade hybrid interference. *The Washington Quarterly,* 44(1), 49-67.

Wilner, A. (2024). 'Deterrence by De-legitimization in the Information Environment: Concept, Theory, and Practice.' *Deterrence in the 21st Century: Statecraft in the Information Age.*

Wilner, A. S. (2015). *Deterring Rational Fanatics.* University of Pennsylvania Press.

Wilner, A. S. (2011). Deterring the undeterrable: Coercion, denial, and delegitimization in counterterrorism. *The Journal of Strategic Studies, 34*(1), 3-37.

Ördén, H. (2019) 'Deferring substance: EU policy and the information threat', *Intelligence and National Security,* 34(3), 421-437.

29. ATTRIBUTION

HEDVIG ÖRDÉN & JAMES PAMMENT

SUMMARY

- Attribution in foreign information influence and interference is complex, as digital environments enable both tactical obfuscation and the amplification of content by unwitting actors, including media and individuals.
- A central difficulty lies in distinguishing legitimate democratic influence from illegitimate manipulation.
- Scholars identify two dimensions of attribution: technical attribution, focused on identifying unique tactics and techniques, and political attribution, which involves assigning responsibility in public to signal, deter, or blame.
- While technical attribution adapts cybersecurity methods to build cases against threat actors, political attribution requires navigating risks tied to domestic vulnerabilities and international tensions.
- Ultimately, attribution is not only about knowing who is behind an operation but also about weighing the consequences of making that knowledge public.

The question of *who* is behind influence campaigns as well as *how* to establish a threat actor's identity lies at the heart of policy discussions pertaining to foreign information influence and interference. Information influence operations are often enacted in the digital environment. This facilitates obfuscation, both in terms of the tactics used and the actors involved. What is more, influence operations enacted in the digital environment feed off the structural features and affordances of traditional and social media. A foreign threat actor might initiate the spread of certain content, but a successful influence campaign involves the – often unwitting – sharing of this content by a range of other actors, including individual users and traditional media outlets. A final element complicating attribution in a liberal democratic context is how to distinguish between legitimate and malign forms of influence. Democratic politics is essentially about individuals exercising their legitimate right to influence and publicly voice their opinions.

This chapter introduces scholarly insights on attribution in the context of foreign information influence and interference. A key argument is that attribution has two dimensions which, in turn, come with different challenges. On the one hand, attribution is the practice of establishing a threat actor's identity. This involves a focus on identifying a threat actors' unique tactics and techniques. This can be termed *technical attribution*. On the other hand, attribution is also a distinct social practice for distributing responsibility and blame, enacted in a specific political setting. When enacted in public, this form of attribution can be employed to

send signals to a threat actor, for example for deterrence. This is termed *political attribution*. This chapter explores the distinct challenges involved in the two forms of attribution.

The difficulties associated with establishing threat actors' identities have been referred to as the 'attribution problem'. Solutions often draw on insights from the cybersecurity literature, modifying the approaches to fit the threat of foreign information influence (Pamment & Smith, 2022; Palmertz et al., 2025). The focus here is to leverage technical, behavioural and contextual evidence to build a case against a threat actor (Palmertz et al., 2025). However, the challenges arising with practices of public attribution cannot be solved by technical means alone. In public attribution, decision-makers also need to navigate political risk (Hedling & Ördén, 2025). The political challenges spring from the fact that threat actors often exploit existing societal vulnerabilities. An act of public attribution directed at a foreign actor can thereby become entangled with domestic politics, as well as more complex international geopolitical controversies. Put simply, just because you know who did it, doesn't mean that you necessarily want to go public with that information.

ATTRIBUTION AS IDENTIFICATION

Pamment & Smith (2022) and Palmertz et al (2025) argue that attribution consists of several considerations that can be placed in a matrix. The first two types of evidence belong to technical attribution:

- **Technical evidence,** which consists of the observable traces that an adversary leaves behind at the level of digital signals;
- **Behavioural evidence,** which builds an understanding of the methods by which different threat actors carry out their work (this is often termed Tactics, Techniques and Procedures or TTPs).

These two types of evidence offer a solid picture of what happened at a technical level, by drawing together digital signals with known patterns of behaviour associated with threat actors. The next two types of evidence are more complex assessments that form the baseline for political attribution.

- **Contextual evidence,** which consists of an assessment of the content and context of the influence campaign, with the goal of trying to understand the motivations of the adversary;
- **A legal & ethical assessment** of whether assigning blame is desirable and proportionate based on one's own strategic concerns. This might for example include whether an attribution is escalatory, or may involve political or commercial fallout, diplomatic concerns, or litigation.

The identification stage takes these four categories and explores them across three types of sources. The source informs not just what information is collected and how it is assessed, but also whether and to what extent it may be made public. Evidence derived from open sources can be analysed and discussed by any actor. Evidence derived from proprietary data is rich in technical and behavioural information which is only released at the discretion of the data owners, which are often in the private sector. Evidence derived from classified intelligence is either shared purely as a sanitized contextual or legal assessment, or selected parts are declassified and shared. The three information sources shaping attribution are:

- **Open source** relies on publicly available information and OSINT. NGOs, media, researchers, and intelligence agencies use it when they lack access to classified data. Attribution often comes from investigative journalism, crowdsourced research, and analysis of open datasets, building circumstantial cases by examining tactics and narratives. It can be strengthened by linking activities to web domains, IPs, financial records, and company ownership, creating evidence of responsibility.
- **Proprietary sources** use privileged data from digital platforms, private intelligence firms, data brokers, and cybersecurity companies. This technical and behavioural data helps to identify the traces of the operations in the "backend" of infrastructure such as a social media platform. Attribution often comes from platform takedowns or commercial intelligence reports, sometimes supplemented with open-source evidence. Legal assessments are tied to platform policies and host country laws, but decisions can be influenced by commercial or geopolitical concerns. While proprietary data supports strong attribution, its scope is limited by which systems the data owner has privileged access to, and cross-platform insights rely on discreet data-sharing partnerships.
- **Classified sources** rely on secret information, sometimes combined with open or proprietary data, and are mainly the province of governments and the military. They address specific intelligence needs, usually for internal use, but can be shared with allies or the public through statements and reports. Attribution seeks to link the information influence operations to a strategic view of a threat actor's broader hostile activities.

An analyst's visibility of, and perspective on, an information influence operation will depend on their objectives, priorities, resources, and access to information. Few, if any, analysts have both access to the spectrum of potentially available data and the resources and freedom to fully explore them. Having an awareness of the strengths and weaknesses of, and gaps in specific types of access is therefore essential to, for example, identifying potential bias and assessing probability. There is an assumption, however, that government attributions have the opportunity to consider all types of evidence, and therefore they are often seen as authoritative. However, it is worth bearing in mind that intelligence assessments are often probability or likelihood based, and hence that many investigations contain a degree of complexity best offset by shared or combined analyses from multiple organisations.

THE POLITICS OF ATTRIBUTION

The act of public attribution can serve purposes beyond simply exposing the identity of a threat actor. Viewed on a social level, attribution is also the practice of determining responsibility for an undesirable act. Attribution is essentially an act of assigning and distributing *blame* (Shaver, 2012). Mobilizing this modality, the public attribution of foreign information influence functions as a form of social punishment whereby an actor is delegitimized by reference to their norm-transgressive behaviour.

When used as a foreign policy tool, the attribution of foreign information influence can serve as a form of *deterrence*. Using public attribution for deterrence draws on the logic of *deterrence by de-legitimization*. Public acts of attribution leverage societal norms to target the social influence of actors (Wilner, 2011: 14). The power of social punishment inherent in attribution however gives rise to complications for

decision makers operating in a liberal democratic setting.

The challenges associated with public attribution spring from the unique nature of the threat as well as the nature of attribution as a form of social punishment. Democracy is essentially about a public contestation of opinions (Farrell & Schneier, 2018: 6). Foreign threat actors tend to exploit this feature to amplify existing societal cleavages (Karlsen, 2019), making information influence a threat where external elements intersect with fundamental processes of democratic deliberation. The practice of attribution – the use of social punishment to delegitimize foreign actors – in a democratic setting thereby generates domestic political risk (Hedling & Ördén, 2025: 9). When acts of attribution are enacted in a polarized environment, or in relation to highly politicized issues, they have the potential to deepen societal and political divides. Decision-makers could be perceived as meddling in democratic deliberation, seeking to attack political rivals or to discredit citizens' legitimate views and opinions by reference to foreign adversaries[73].

This set of unique challenges highlight the politics involved in attribution. Decisions on attribution are political in the sense that they are often 'related to a perceived political need for attribution (or for non-attribution)' (Hedling & Ördén, 2025: 7). When navigating political risk in the context of attribution, decision-makers can employ different strategies. They can decide for *public attribution*, refrain from attribution and instead pursue a strategy of *non-attribution* or they can opt for the middle way of *diffused attribution* (Hedling & Ördén, 2025: 11-12).

Public attribution is the practice whereby state actors publicly establish responsibility for foreign information influence. This can be done through official statements establishing responsibility for a campaign, which may be followed by other foreign policy decisions such as sanctions or bans. The practice of public attribution operates by the logic of *deterrence by delegitimization* in that it establishes responsibility and distributes social punishment (Wilner, 2011). Low levels of domestic political risk would make this strategy appealing to actors. Public attribution can, for instance, be employed in reaction to undesirable acts committed by recognized longtime adversaries in a domestic context where foreign policy sentiments are widely shared among the population.

Non-attribution refers to the act of refraining from public attribution despite having sufficient technical evidence about the identity of a threat actor. Decision makers can decide against attribution following a cost-benefit calculation which takes into account existing levels of domestic resilience as well as the domestic political risks associated with a public statement. Timing is a crucial component in decisions on non-attribution (Hedling & Ördén, 2025: 11). In addition, policymakers often consider the risk of escalation (in essence, the fear of making the situation worse by antagonising an adversary), or of surrendering the technical advantages (or "tradecraft") that would enable them to keep tracking interference campaigns. In the case of Russian interference in the 2016 US elections, the Obama campaign for instance decided to refrain from attribution during the ongoing election

[73] Research demonstrates how attribution of foreign information influence and disinformation more generally can be used by political actors as a strategic tool, also in a domestic context. Studies for instance show how right-wing populist actors employ disinformation attribution as a strategy to delegitimize media institutions and political rivals (Hameleers, 2020; Egelhofer et al., 2022). What is more, citizens often engage in selective attribution. In other words, they attribute the dissemination of false information along ideological lines, to political actors with whom they disagree (Hameleers & Brosius, 2022; Tong et al., 2020).

campaign. Documents show how the decision to attribute the activities to Russia, and issue sanctions, only after the elections was linked to fears about a potential domestic political backlash following public attribution (Hedling & Ördén, 2025: 11).

Diffused attribution offers a way of balancing the political risks associated with public attribution with transparency. In diffused attribution, responsibility for attribution is shared and distributed between different actors (Hedling & Ördén, 2025: 11-12). Decision makers operating in highly polarized domestic environments can for instance choose to attribute foreign information influence through other actors. Examples are international organizations or other credible non-state actors, such as investigative journalists or open-source intelligence organizations. Diffused attribution essentially mobilizes the logic of *deterrence by detection*. As an attributional practice, it publicly demonstrates a technical and forensic capability for detection of foreign information influence campaigns and, thereby, sends a signal to the adversary. Threat actors are made aware that their campaigns have been detected, and the public is kept informed of ongoing threat activities. At the same time, diffused deterrence allows decision makers to navigate domestic political risk by leaving the act of social punishment to other actors.

Engaging with the politics of attribution demonstrates how attribution seen as simply an act of determining the identity of a threat actor is insufficient for understanding the full chain of attributional practices regarding foreign information influence. Actors might have sufficient technical evidence to determine the responsibility for an attack but still decide to abstain from public attribution. The absence of public attribution statements should therefore not be viewed as an absence of knowledge. What is more, the development of technical and forensic frameworks for attribution – while essential for deterrence – will not solve the political risks associated with using attribution for purposes of deterrence.

CONCLUSION

Attribution is not simply a technical problem but is also inherently political. The decision to attribute is based on many factors, only some of which are tied to the likelihood of a specific threat actor being responsible for a deed. While the processes may be more developed in for example cybersecurity and financial crime, the analysis of information influence operations has also developed practices of attribution that help to shape perceptions of the field in the mind of the public.

DISCUSSION

- What purposes can attribution serve in relation to foreign information influence?
- What are the key challenges of attribution in each case?
- How can these challenges be overcome?

HEDVIG ÖRDÉN *is a researcher at the Psychological Defence Research Institute, Lund University, and a postdoc at the Department of Political Science and Public Management, the University of Southern Denmark. She is also an affiliated researcher at the Europe Programme at the Swedish Institute for International Affairs. Her work is situated within critical security studies and critical intelligence studies. She publishes on topics related to security and foreign information influence, intelligence and security expertise, and intelligence and liberal democracy.*

JAMES PAMMENT *is Director of the Lund University Psychological Defence Research Institute. His main research interest is in the role of strategic influence in international relations, both its legitimate sides (e.g., public diplomacy and aid) and illegitimate (e.g., propaganda and hostile foreign interference). Previous affiliations include the Carnegie Endowment for International Peace, Swedish Defence University, the EU-NATO Hybrid Threats Center of Excellence, and the University of Texas at Austin.*

REFERENCES

Farrell, H., & Schneier, B. (2018). Common-knowledge attacks on democracy. Berkman Klein Center Research Publication, (2018-7).

Egelhofer, J. L., Boyer, M., Lecheler, S., & Aaldering, L. (2022). Populist attitudes and politicians' disinformation accusations: effects on perceptions of media and p oliticians. *Journal of Communication, 72*(6), 619-632.

Hameleers, M. (2020). My reality is more truthful than yours: Radical right-wing politicians' and citizens' construction of "fake" and "truthfulness" on social media—Evidence from the United States and the Netherlands. *International Journal of Communication, 14,* 18.

Hameleers, M., & Brosius, A. (2022). You are wrong because I am right! The perceived causes and ideological biases of misinformation beliefs. *International Journal of Public Opinion Research, 34*(1), edab028.

Hedling, E., & Ördén, H. (2025). Disinformation, deterrence and the politics of attribution. *International Affairs,* iiaf012.

Innes, M. & Ahonen, A. (2025). Attribution and Information Influence Operations. A 'Field Guide' for Open-Source Investigators and Researchers. ADAC.io EU Horizon Project Deliverable 1.1. Psychological Defence Research Institute; Lund University.

Karlsen, G. H. (2019). Divide and rule: ten lessons about Russian political influence activities in Europe. *Palgrave Communications, 5*(1), 1-14.

Pamment, J. & Smith, V. (2022) *Attributing Influence Operations: toward a community framework.* Riga: NATO Strategic Communication Centre of Excellence & EU-NATO Hybrid Centre of Excellence

Palmertz, B., Isaksson, E. & Pamment, J. (2025). A Framework for Attribution of Information Influence Operations. ADAC.io EU Horizon Project Deliverable 1.1. Psychological Defence Research Institute; Lund University.

Shaver, K. G. (2012). *The attribution of blame: Causality, responsibility, and blameworthiness.* Springer Science & Business Media.

Spitzberg, B. H., & Manusov, V. (2021). Attribution theory: Finding good cause in the search for theory. In *Engaging theories in interpersonal communication* (pp. 39-51). Routledge.

Tong, C., Gill, H., Li, J., Valenzuela, S., & Rojas, H. (2020). "Fake news is anything they say!"—Conceptualization and weaponization of fake news among the American public. Mass Communication and Society, 23(5), 755-778.

Wilner, A. S. (2011). Deterring the undeterrable: Coercion, denial, and delegitimization in counterterrorism. The Journal of Strategic Studies, 34(1), 3-37.

30. THE ETHICS OF COUNTERING DISINFORMATION

ALICIA FJÄLLHED

SUMMARY

- Beyond stating that disinformation constitutes false and intentionally misleading messages, we also understand it as a morally condemnable behaviour.
- To assess a message to be false or intentionally misleading is, however, not easy for actors only holding that message and no other information.
- The assessment of messages leads to different ideas of what constitutes an appropriate and justified response.
- Whether to respond or not, and in what way, is also bound to a moral judgment engaging in a balancing act between removing unethical content while not infringing on freedom of speech.
- In this process of defining, assessing, and responding to disinformation, it is important that we use the same moral system to both condemn behaviour and make sure that we stay within the same moral boundaries in the response.

You work for a public organization in a democratic country which is about to hold general elections. One day, you find the below post on your organisation's social media channel. How would you respond?

> *Just great! The election organization is such a smooth-running machine… You only have to stand in line for HOURS to cast your vote. Democracy at its finest!*

To counter disinformation is to engage in moral reasoning. We often describe disinformation as false or misleading communication spread with the intent to cause harm. Such a description sets up criteria for how to assess whether a message can be defined as disinformation. However, we also describe disinformation as bad, wrong, or use other value-judgements to make a normative statement. This ties disinformation to ethics and morals, as we are taking a stance on how we think actors *ought* to behave, and where disinformation becomes phenomena that *ought* to be countered. Therefore, ethics and morals are found at the centre of the process of countering disinformation, moving from the idea of *what* disinformation is to an understanding of *why* we assess it as normatively wrong. The moral dimension is present from the theoretical definitions, to practices of assessing and determining cases of disinformation, and ultimately finding morals in the very boundaries constituting if and how we ought to respond. Returning to the above-presented social media post, whether you consider removing, ignoring, or responding to the post, each choice is grounded in a moral evaluation of the post itself where you need to morally justify your response.

As this chapter will show, the theoretical definition of disinformation may be simple enough, but it also creates moral dilemmas when used as a tool to identify, assess, and counter disinformation. Ethics is present throughout this process, boiling down to a final question: Would you respond to the message, on what grounds, and how do you make sure you stay within the moral lines you used to pass this moral judgement?

IMMORAL DISINFORMATION

Strategic communication refers to goal-oriented communication (Zerfaß, Verčič, Nothhaft, & Werder, 2018). As disinformation is false information spread with the intent of causing harm, it is goal-oriented and therefore a form of strategic communication. Measures to counter disinformation are also goal-oriented, and thus also a form of strategic communication. The difference lies in a moral evaluation, presenting disinformation as immoral while strategic countering of disinformation aspires to use moral forms of communication. This moral line is drawn based on a moral framework found in the ideal of democratic societies. The ideal is theoretically reflected in Jürgen Habermas' (1990) theory of discourse ethics, which is formed upon two principles for the nature of the conversation and three criteria for the nature of statements in such conversation.

First, discourse ethics states that everyone affected by a decision should be free to participate in a common open conversation about the issue. Through this discourse, the participants can arrive at an agreement on how their society, its organisations and citizens, ought to act—that is, leading to a collective moral agreement. This is why we condemn foreign actors interfering in a domestic debate, and why we condemn actors creating multiple accounts on a mass scale to manipulate the impression of a public opinion on social media.

Furthermore, the moral framework dictates that only some claims ought to carry weight in the discourse, ideally derived from three moral values—they ought to convey the truth, they ought to be sincere, and they ought to be driven by constructive intentions. From this, disinformation manifests as an immoral form of communication, as it breaks with all three rules; it is not truthful but presents lies, it is not sincere but aims to be deceptive and does not aim to result in constructive intentions but is driven by a disruptive aim.

ARE YOU CERTAIN THIS IS DISINFORMATION?

Disinformation is defined as false or misleading messages spread with the intent to do harm. When responding to the post, actors thus face two central dilemmas tied to the two criteria defining disinformation. The first moral dilemma arises when we try to assess whether the post contains false information. Say that election workers on site confirm to you that there have been no hour-long cues. The post, then, would contain false information. But what if the message was not meant as a factual statement? The intent could have been to express an opinion, or a subjective experience. Maybe the person was there, and experienced the cue as being very, very long, posting the message as a sarcastic statement, an expression of this person's dissatisfaction with the voting system. How does that affect your evaluation of whether the post is a case of disinformation, and does it effect your idea of an appropriate response? The line between false and true information is not as easily determined as we might think, as "parody and satire do not easily fall into a binary analysis of truth and falsity" (Kahn, 2021, p. 3, see also Fallis, 2014; 2015).

While disinformation clearly refers to false statements, one could ask whether we are more interested in the statement as it was meant by the one communicating it, the statement as literally written, or the statement as interpreted by its audience?

The second dilemma derives from disinformation defined as intending to cause harm. Human rights organisations argue it impossible to establish "the absence and presence of intent to cause harm" (Kahn, 2021, p. 2). And if we return to the scenario presented in this chapter, it is difficult to imagine how actors could make such an assessment in a practical sense. For, again, how do you determine that this is a case of actors intending to do harm, and not a well-intended misunderstanding or an expression of one's opinion.

As this section has shown, the definition of disinformation may be seen as clear and simple, but when put to the test as a tool to assess real messages this does not necessarily mean that it is easy to determine the presence of the phenomenon.

THE MORAL DILEMMA IN COUNTERING

Habermas' (1990) moral framework also justifies societies' countermeasures against disinformation, as the process is intended to keep the discourse as close to the ideal as possible. All countermeasures face moral questions, where even the decision not to act can be morally justified or condemnable. Efforts to counter disinformation is not moral per se, but rests on actors' ability to "maintain moral authority by making the case that it has been harmed, that it has normative standing to engage in counter-interventions, and that it does so in an appropriate manner" (Bjola, 2018, p. 306). In short, the response has to be justified, resting on an assessment of the situation, an evaluation of options, and rests on arguments supporting one's course of action.

To determine the right cause of action is presented as a "balancing act", balancing between the moral expectation to counter disinformation whilst also respecting freedom of speech (see for example Bontcheva, Posetti, Teyssou, Meyer, Gregory, Hanot & Maynard, 2020). This means that even false statements can be protected by law,[74] and furthermore that...

> *...the human right to impart information and ideas is not limited to "correct" statements, that the right also protects information and ideas that may shock, offend and disturb, and that prohibitions on disinformation may violate international human rights standards, while, at the same time, this does not justify the dissemination of knowingly or recklessly false statements by official or State actors (OSCE, 2017, p. 1).*

The balance is also arguably shifting alongside the changing security landscape. In the shift from times of peace to the presence of disinformation used as a hybrid threat by foreign powers, the balance has shifted towards public calls to action for harsher measures against disinformation (Bjola, 2018). This leads to a new set of concepts, alongside hybrid threats finding government concepts such as *information influence* in Sweden (MPF, n.d.) or *Foreign Government Interference* as used by social media platforms (Meta, 2019), creating potentially more restrictive moral rules for foreign actors than domestic ones (see Ördén and Pamment (2021) for a discussion). However, most actors encountering the post on social media may not have the tools to determine who spread the message. These actors may have to work under the assumption that it might be a member of the public, leading to

[74] See chapter titled Freedom of expression and countering of malign influence activities in Sweden

a much wider moral leeway for the one communicating the message and a more restricted set of possible actions to counter it.

To counter disinformation means a wide range of actions. For example, platforms filtering out social media entries through algorithmic solutions. Returning to the balancing act, this, however, simultaneously risks removing legitimate expressions of freedom of speech. Similarly, new legal frameworks have been the subject of the same criticism (OSCE, 2017), with some legal frameworks being celebrated as examples where states take responsibility for the integrity of the information environment, while other initiatives are criticized as government tools used to label criticism toward oneself as cases of disinformation to enable legal actions against that criticism.

This balancing act also arises in everyday countering of disinformation, such as when deciding which action to take against the initially presented social media post. Some would suggest that the appropriate response is to delete it, if so probably on the grounds that it is assessed as containing false claims intended to cause harm. Another person, however, would perhaps assess the same entry as a sarcastic statement, and thus decide to do nothing—to protect every person's right to express their dissatisfaction, even if this "may shock, offend and disturb" (OSCE, 2017, p. 1). Both measures carry risk. In the first case, one risks being criticized for silencing a voice in the open public debate. In the second case, one risks being criticized for allowing a message to sow distrust against the democratic system and discourage people from voting.

STRATEGICALLY COUNTERING DISINFORMATION

To act morally in countermeasures against disinformation requires a strategic approach. For the risk at stake is essentially not only that we silence democratic voices, but by doing so we also drive the public's distrust towards democratic institutions. Imagine, for example, that the post was spread by someone that did not mean to do harm. Perhaps it was meant as an expression of their experience or based on a rumour that genuinely worried them. Thinking strategically, how do you think they would react to your response? How would they interpret your silence, how would they respond if you chose to remove their post, or if you phrased your response in a certain way? What would the effect of your response be, either to that person or among the others who also read the conversation between the two of you online?

By removing content, you might run the risk of being perceived as censuring criticism or hiding something, adding fuel to the debate that something's off, confirming the suspicion embedded in the post. Perhaps you even decide to put restrictions on the account spreading the message, leading them to migrate to new platforms where their conversation continues. While countering should act to protect the discourse, the very act of pushing some actors away may create the very problem we sought to address. Just as we assess disinformation based on the risk of causing harm, it is important that we make such a risk assessment in relation to our countermeasures.

One way to ensure that we are living our own moral values is to follow the very principles proposed by Habermas (1990), to engage in an open conversation. By responding to the post, you have an opportunity to spread your own organisations' sense of the facts in the case to the author of the original post and to others that read this entry. Beyond that, the very fact *that* you are responding as well as

how is a way to *show* the values you are protecting—the inclusive, informed, and respectful debate—where the very tone in that message becomes an opportunity to embody the role of an actor that can rise to the occasion and deliver a response that builds trust towards the public organisation you are representing.

Aside from the response to the message, we also need to ensure that the moral framework itself is a product of that open conversation. Whether it is a conversation between friends and family, with colleagues at work, in conversations between organisations exchanging experiences, or in the public debate—it is here that we can expect to find the answer as to what constitutes ethical or unethical forms of communication. Through that conversation, the above-presented boundaries may change as we continue to develop our definitions of disinformation and other forms of immoral communication, as well as develop our sense of the ethics of countering disinformation.

DISCUSSION

- You work for a public organization in a democratic country about to hold a general election. One day you find the entry below on your social media platform. What would be an ethical response?
- Just great! The election organization is such a smooth-running machine... You have to stand in line for HOURS to cast your vote. Democracy at its finest!
- What do you think is most appropriate – to remove, ignore, or respond to this message?
- If you decide to respond, what would be the response?
- Based on your response, what do you think will be the author's reaction—and how do you think other readers will react?

ALICIA FJÄLLHED *(PhD) is a researcher at the Swedish Defence Research Agency, focusing on communicative dimensions in crisis and war. Her research departs from the dynamics between strategic and moral communication, exploring the process of defining, assessing and countering immoral communication. The chapter draws upon her doctoral dissertation Strategic moral communication (Lund University, 2023).*

REFERENCES

Bjola, C. (2018). The ethics of countering digital propaganda. *Ethics & International Affairs,* 32(3), 305–315. doi:10.1017/So892679418000436.

Bontcheva, K., Posetti, J., Teyssou, D., Meyer, T., Gregory, S., Hanot, C., & Maynard, D. (2020). *Balancing act: Countering digital disinformation while respecting freedom of expression, Broadband Comission research report on 'freedom of expression and addressing disinformation on the Internet'*. ITU, UNESCO & Broadband Commission for Sustainable Development. https://en.unesco.org/publications/balanceact.

Fallis, D. (2014). The varieties of disinformation. In L. Floridi & P. Illari (Eds.), *The philosophy of information quality,* 135 Synthese Library 358, (pp. 135–161). Springer International Publishing Switzerland.

Fallis, D. (2015). What is disinformation?. *Library Trends,* 63(3), 401–426.

Fjällhed, A. (2023). *Strategic moral communication: A metatheoretical and methodological response to the normative perspective on strategic communication* [Doctoral dissertation, Lund University].

Habermas, J. (1990). *Moral consciousness and communicative action.* MIT Press, Cambridge, Massachusetts.

Khan, I. (2021). *Disinformation and freedom of opinion and expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan.* (Report A/HRC/47/25; Report of the Special Procedure of the Human Rights Council). United Nations Human Rights Council. https://digitallibrary.un.org/record/3925306?ln=e

Meta. (2019, October 21). *How we respond to inauthentic behaviour on our platforms: Policy Update,* https://about.fb.com/news/2019/10/inauthentic-behavior-policy-update/

MPF (n.d.). *Psychological Defence Agency*. https://mpf.se/psychological-defence-agency

Ördén, H., & Pamment, J. (2021). *What is so foreign about foreign influence operations?* Carnegie Endowment for International Peace. https://carnegie-production-assets.s3.amazonaws.com/static/files/Orden_Pamment_ForeignInfluenceOps2.pdf

OSCE. (2017). *Joint declaration on freedom of expression and "fake news", disinformation and propaganda.* https://www.osce.org/fom/302796

Zerfaß, A., Verčič, D., Nothhaft, H., & Werder, K. P. (2018). Strategic communication: Defining the field and its contribution to research and practice. *International Journal of Strategic Communication,* 12(4), 487–505. doi:10.1080/1553118X.2018.1493485

Psychological Defence Agency

LUNDS UNIVERSITET

PSYCHOLOGICAL DEFENCE AGENCY
VÅXNÄSGATAN 10
SE-653 40 KARLSTAD
REGISTRATOR@MPF.SE
010 - 183 70 00
WWW.MPF.SE