MSB Swedish Civil Contingencies Agency

Lund University

RESEARCH REPORT

Countering Information Influence Activities

The State of the Art

Countering Information Influence Activities: The State of the Art, version 1.4 (1 July 2018)

James Pamment, Howard Nothhaft, Henrik Agardh-Twetman, Alicia Fjällhed
Department of Strategic Communication, Lund University

MSB's points of contact:
Fredrik Konnander

Publication number MSB1261 – July 2018
ISBN 978-91-7383-865-8

MSB has ordered and financed this report. The authors are solely responsible for the content.

Preface

This is version 1.4 of a report that aims to provide an overview of current thinking on how to counteract information influence activities. It was commissioned to support the Swedish Civil Contingencies Agency's (MSB) work in strengthening societal resilience against information influence activities. The report is intended to offer (1) a scientific overview to support the development of the MSB handbook Counter Influence Strategies for Communicators, (2) a guide and framework that can support the development of training and education on counter influence, and (3) a Swedish perspective on the knowledge currently available on information influence activities. The authors wish to thank MSB and the dozens of interviewees and reviewers without whom the study would not have been possible.

Foreword

The term “Fake News” has catapulted the topic of information operations into the centre of a heated and mostly ill-informed public debate. One common misconception is that the problem is new. Another mistake is to assume that information, as the most visible part of the problem, is necessarily the most important.

Both of these are at best only partly true. Political warfare—the disruption of another country’s public opinion and decision-making—dates back decades, if not centuries. Information operations form only part of the subversive arsenal. They are usually conducted along with the use of money, intimidation (legal and physical), cyber-attacks, and many other tactics.

But the information environment has indeed changed sharply, and mostly to the advantage of unscrupulous autocratic countries wishing to attack open societies. The internet enables ubiquity, immediacy and, most of all, anonymity of a kind undreamt of in past decades.

Technological change has also lowered barriers to entry in the media industries and disrupted the business models of incumbents. The era of media gatekeepers—in effect a cartel of trusted sources—has given way to a kaleidoscope of facts and opinions.

This diversity is welcome, but as with all technological change, we are still developing the norms and rules to manage it. In the meantime, our information landscape is ripe for misuse. The abuse of automated communication sources (bots) and paid-for disruptors (trolls) distorts public debate by promoting lies and swamping truth. Our attackers are also aided by weaker social and political immune systems. Amid a general (and largely welcome) decline in deference, respect for experts has diminished. For complex demographic, economic and social reasons, levels of social trust in many societies have ebbed.

At least in the short term, these problems are likely to worsen. We have seen attacks on the confidentiality of data (hacking) and its availability (swamping). The looming threat is attacks on data integrity. How will we react to “deepfakes”—seemingly authentic video and audio which purports to show our public figures saying words they never said, and doing things they never did?

As this excellent and timely report makes clear, the greatest strength of our societies—a open, robust public debate—now risks being its greatest vulnerability. Unconstrained by the requirements of honesty, truth or self-respect, attackers can use our information system to confuse us and disrupt our public life.

This report ably brings modern academic thinking on cognition and other topics to bear on the problem, notably in explaining how we form our individual and collective opinions. It offers no easy answers—indeed, it cautions against them. It notes, rightly, that whereas applications may be legitimate, illegitimate, or outright illegal, the techniques themselves are neutral. Parody and satire, for example, are the lifeblood of a modern democracy. If we cannot mock our rulers, we are not truly free. But the same techniques in the hands of an enemy state are no longer amusing and thought-provoking; they are aimed cynically at increasing polarisation and corroding the trust and respect that are essential for the proper functioning of our public life.

Grey areas abound and the choices we face are hard. An entirely passive response invites defeat. An overly zealous one is self-defeating. If we protect ourselves against attacks from autocratic closed societies by giving sweeping and arbitrary powers to our rulers, we may win the battle, but we lose the war.

The most important recommendation in this report is therefore to study information operations first, and to act cautiously in trying to mitigate or counter their effects. Crying “wolf!” (or “Fake News!”) at every news item we dislike is a sure way to erode credibility. Our adversaries’ biggest and most effective victories come when we do their work for them.

Edward Lucas
Senior Vice President at the Center for European Policy Analysis (CEPA)
July, 2018

Table of Contents

1. Introduction

Sweden is exposed to attempts by foreign actors and their proxies to sway policy and undermine institutions, including democratic elections and free and open debate, by diplomatic, informational, military and economic means.[1] The application of persuasive, coercive and erosive power in the so-called grey zone between peace and war and the existence of so-called hybrid threats has been documented and confirmed by journalists, researchers and government agencies around the world and acknowledged by Swedish Prime Minister Stefan Löfven.[2] The targeting of elections in the United States, France, Germany and the United Kingdom are high-stakes examples that have attracted worldwide attention and continue to be investigated.

Overview of the report

This report covers one aspect of hybrid threats and the application of power in the grey zone: the *informational dimension* manifested in *information influence activities* (*informationspåverkan*). Information influence activities are here understood as *the targeting of opinion-formation in illegitimate, though not necessarily illegal ways, by foreign actors or their proxies*. This targeting is used to support and amplify diplomatic, economic and military pressure; hence information is considered a crucial multiplier of the hybrid influence spectrum. We focus this report on the role of public sector communicators on the basis that communication officers may be among the first to come into contact with information influence efforts, for example in their media monitoring, social media, customer relationship and citizen engagement roles. Although public sector communicators are the suggested audience for this report, we believe that it may also be of interest to politicians, journalists, decision-makers, researchers, students and the general public.

The process by which an organisation can systematically counteract information influence activities is reflected in the report's structure. Understanding what information influence is, and what it is not, constitutes the first step (chapter 2). The second step is identifying the means that are typically employed (chapter 3). Equipped with a thorough understanding of information influence activities in context, communication officers can then begin to prepare their own organisations, take action in specific cases and engage in a continuous learning-cycle (chapter 4). Fig. 1 illustrates the process.

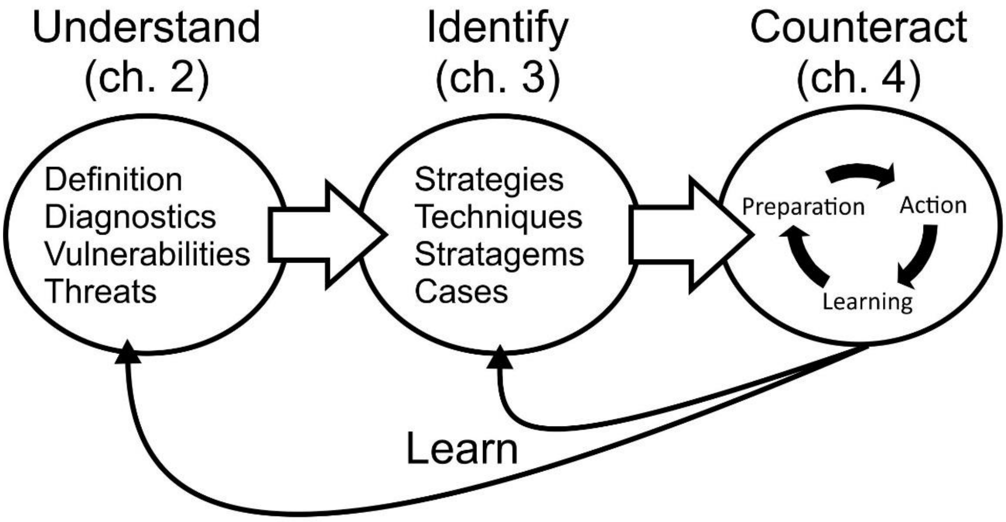


Figure 1: The process of systematically counteracting information influence activities and structure of the report

So, what are information influence activities? In contrast to public diplomacy – which constitutes legitimate informational power exerted across borders to influence policy outcomes – the defining characteristic of information influence activities is that they not only utilize but *exploit* the open system of opinion formation in Western democracies, turning its greatest asset, free and open debate, into a vulnerability. Information influence activities exploit open and free opinion-formation by mimicking legitimate behaviour to gain access to and influence the public sphere. What the pretence conceals is a typically four-fold violation of the minimum rules of civilized debate: (1) Information influence activities contain *deceptive* elements, i.e. the techniques of information influence obscure, mislead and disinform; (2) information influence activities are not interested in a constructive solution to a problem, but *intend* to do harm (as evidenced, for example, by support for both sides in a divisive issue to invoke confrontation); (3) information influence activities are *disruptive*, i.e. they not only intend to do harm, but really do (as evidenced, for example, by destruction of property); (4) information influence activities constitute *interference*, i.e. foreign information influence activities, sometimes via domestic proxies, interfere in domestic democratic processes and the sovereignty of states.

The four dimensions, encapsulated in the acronym DIDI, can be used as diagnostic criteria to differentiate genuine cases of information influence activities. Such cases are probably rarer than current media coverage suggests. Many other forms of public communication also exploit the relative generosity of the ground rules of democratic debate, albeit to a lesser degree, for political, economic, cultural or other ends. Political campaigning might contain deceptive elements, for example, but it is difficult to make the case that a political party interferes with issues in its own home country, or that it intends to undermine national sovereignty. If a public relations agency, for example, were hired to conduct a campaign that was deliberately deceptive, intended to do harm, disrupted democratic processes, and interfered in an issue without legitimacy, it would be reasonable to question the source of the funding of the campaign and to consider the application of some of the counter influence techniques suggested in this report.

Our understanding of the DIDI-criteria as *diagnostic* indicates a mindset centred on remedial action: if there is a fire, our focus is being able to identify the fire early and to extinguish it, not on apprehending the arsonist, since that is the task of other agencies. The DIDI-criteria are therefore complemented by a description of numerous *techniques* typically employed in information influence campaigns as well as several arrangements of these techniques in typical *stratagems*. Techniques that information influence campaigns employ include relatively benign approaches such as use of humour and rhetoric, all the way through to the forging of documents or the use of false identities. Once again, the emphasis is on diagnostics: the utilization of these techniques should not be considered proof of an information influence campaign. The same holds for the arrangement of several techniques in stratagems, although the case now becomes stronger: not only indications of influence techniques, but of coordinated activities. The consideration of techniques as well as stratagems reflects our understanding that information influence campaigns are best detected on the level of chain-of-events, in tandem with an analysis of the context and assessment of the likelihood of hostile intentions.

The focus on judgment, containment and correction derives from the way this report was commissioned as preparatory work. During 2018, this report will be supported by one or more specialised products aimed at communication professionals in the public sector, and a series of workshops aimed at civil servants. In contrast to journalists, communicators in organisations are not so much faced with the question: *'Is this definitely and irrefutably a case of information influence and should it be exposed?'* Communicators are faced with the question: *'Do the events, as far as we can ascertain them, warrant a reaction from us, and what should a reasonable reaction be?'* It must not be forgotten, here, that cutting-edge information influence activities are *designed* to be difficult to detect and purposively at odds with familiar categories of right and wrong. What is far more important than establishing a case to the satisfaction of scientific criteria, therefore, is a measured and balanced response that reflects not only the severity of the threat and the certainty about it, but also the potential harm that might derive from a heavy-handed response.

Unfortunately, this consideration does not always seem to take place. At present, the impact of information influence campaigns on liberal society appears to be doubly detrimental: where deception remains undetected or unchallenged, illegitimate influence is gained; where it is revealed and exposed, trust in media and confidence in institutions is further undermined, especially when the cries of 'fake news' or 'interference' turn out to be overblown. Political solutions are thus caught between a rock and a hard place: on the one hand, a *laissez-faire*-attitude might lead to ever more blatant attempts to exert unfair influence; on the other hand, demands to crack down endanger open and free debate as one of the greatest strengths of liberal society. Suspicions entertained by both ends of the legitimate political spectrum, for example that the fake news debate will be exploited by the centre to discredit everything outside the mainstream-corridor, further polarizes society. We advocate vigilance, not paranoia.

Communicators in public organisations are well-advised to keep the purpose of counter-influence activities in mind. The purpose is not to outwit and expose the source of an information influence activity. The mandate of communicators is first and foremost to *protect democratic values*, by strengthening the capacity of audiences to make up their own mind free from illegitimate influences. For communicators, two conflicting values appear to be of relevance, namely:

- Free and open debate, as far as legitimate
- Confidence in public authorities, as far as justified

It is not the purpose of this report to offer political solutions or to prescribe ways of reconciling conflicting values. On a societal level, the report simply describes, without political recommendation, a spectrum of response strategies found in the literature. This spectrum ranges from the most relaxed approach of trusting in the self-correcting powers of the public sphere ('ignoring') to a hard-line approach involving heavy regulation and emergency powers ('crack-down'). While eschewing recommendations on societal level, the report does suggest a basic framework for communicators at the organisational level. The framework involves three very basic steps, namely *preparation*, *action* and *learning*. The report then synthesizes, from the literature, a variety of approaches communicators in organisations may utilize to prepare, act and learn in systematic ways. Special consideration is given to the element of action and the mandate of the communicator. For communicators, four levels of response are suggested: *assess*, *inform*, *advocate* and *defend*. Assessment and information are always considered legitimate reactions, whereas the escalation to advocacy and defence depends on the gravity of the situation and the mandate of the communicator.

The role of MSB: Preparation for a handbook

This report is part of a collaboration between MSB and Lund University in preparation for the September 2018 election. The Swedish Civil Contingencies Agency (MSB) is the government agency with overarching responsibility for civil protection, public safety, emergency management and civil defence.[3] Within these broad areas, the agency is also responsible for parts of what is commonly referred to as psychological defence.[4] Sweden has a long history of including psychological resilience in its total defence doctrine, for the purpose of safeguarding democratic and open society against external threats such as hybrid warfare.[5] Although commonly associated with the Cold War, the principles of psychological and total defence remain relevant in the current geopolitical climate. In its recent report, the US Senate's Committee on Foreign Relations concluded that Sweden, among the other Nordic States, displayed extraordinary immunity against malign influence, presumably because of excellent education systems that emphasize critical thinking, low levels of corruption and high interpersonal trust.[6]

Identifying and countering information influence activities from external actors has become increasingly prioritized by MSB. The National Security Strategy of 2017 points to a deteriorating security situation in northern Europe and the Baltic sea region.[7] This is further emphasised in the Defence Policy Enforcement Decision for 2015-2020, where information influence campaigns are specifically highlighted as a challenge.[8] Under these circumstances, the Swedish parliament has stated that public agencies and institutions should develop the capacity to identify, analyse and counter information influence campaigns.[9]

Since 2016, MSB has been tasked to (1) develop its own capacities related to information influence campaigns; and (2) contribute to the preparedness and capacities of other agencies, institutions and relevant stakeholders through knowledge development and dissemination, and support for cooperation and coordination.[10] Subsequently, MSB has identified the need for an accessible and communicative counter-influence 'tool kit' to strengthen agencies' and institutions' capacity to counter information influence campaigns. This report is a step toward achieving that goal.

In 2017, MSB commissioned the Department of Strategic Communication at Lund University to develop tangible advice on how information influence campaigns can be countered. The aim of the project, of which this report is one part, is to provide communication staff at public agencies and institutions with available and easy-to-use training materials for identifying and countering information influence campaigns. Under the working title of "Counter-influence Strategies for Communicators" the project runs over a full year, with the end goal of producing a handbook-style product that can educate and empower communicators to strengthen societal resilience against hostile influence efforts.

Purpose, methodology, research questions

The report draws upon the insights of practitioners, researchers and other experts, and synthesises these into a single resource where knowledge is presented in a systematic, accessible and actionable way. Methodologically, the research rests on a database of literature initially provided by MSB but developed and expanded by the researchers.[11] At the current stage, the database contains over 1,000 academic texts, reports, and other relevant sources from fields such as strategic communication, law, strategy, international relations and social psychology. Journalistic accounts of information influence campaigns have been utilised to contextualize and situate academic findings, as scholarly work naturally lags behind rapidly unfolding events. For the same reason, complementary interviews, workshops and consultations have been conducted with experts and practitioners to improve our understanding of this complex phenomenon.

The questions that guided our work are the following:

What is the current state of scientific knowledge concerning information influence activities and how can information influence campaigns be countered?

- *What strategies, techniques and stratagems can be identified?*
- *What forms of preparation, analysis and activities can best support the counteracting of information influence activities?*

Structure of the report

Chapter 2 offers a brief introduction to information influence activities, the systemic vulnerabilities that information influence seek to exploit, and the types of threat that societies may face.

Chapter 3 gives an overview of how to identify information influence activities. It covers some influence strategies, offers a detailed taxonomy of the main influence techniques currently in use, and provides examples of the some of the ruses (or stratagems) that combine several techniques to achieve specific goals. It concludes with a case study demonstrating how different techniques fit together into a hostile campaign.

Chapter 4 focuses on how to counter information influence activities, including some societal-level approaches to countering, a discussion of the role of public sector communicators, and a three-step overview of preparatory, actionable and learning activities. The chapter concludes with a short assessment of the limits of counter-influence strategies.

2. Understanding information influence activities

Opponents in public debate, be it about political, economic or social issues, have always known the strategy of accusing the other side of employing illegitimate techniques to gain undeserved advantages: the slur ‘populism’ is one example, the postulation of a media bias another. In the current climate, this move has been made easier and more attractive by the introduction of an ominous ‘factor X’: *covert foreign machinations*. Information influence activities are a real threat, but an atmosphere where everyone accuses everyone else of being someone else’s puppet, and in the vaguest of terms, is not conducive to open and free debate. It is important, therefore, to begin with a clear, yet workable understanding of what information influence activities are – and even more importantly, what they are not. Accordingly, the report begins with a discussion of the crucial distinction between legitimate democratic debate on one side and information influence activities on the other, with due consideration given to the ambiguous zone in between. In essence, we offer a simple model of opinion formation in liberal democracies. This model serves as an organising principle, helping to map different influence techniques depending on the vulnerabilities they seek to exploit.

2.1 Definition of information influence activities

MSB bases its work within the field of information influence on two overarching definitions which broadly captures the way information influence activities work and the ways in which they can be deployed as part of larger influence campaigns. Its definitions are broadly as follows:

- *Information influence activities (informationspåverkan)*, also known as cognitive influence activities, are activities conducted by foreign powers to influence the perceptions, behaviour and decisions of target groups to the benefit of foreign powers. Information influence activities can be conducted as a single activity or as part of a larger information influence operation combining various and multiple activities.

- *Influence campaigns (påverkanskampanjer)* are the coordinated efforts of a foreign power comprised of several influence activities and/or influence operations where each activity (or operation) has one or several ends of their own intended to help achieve the ends of the influence campaign as a whole. This could include influencing (1) the decisions of politicians and other decision-makers in the public sector; (2) parts or the whole of Swedish public opinion; (3) political decisions or public opinion in other countries where Swedish sovereignty, the goals of Swedish security, or other Swedish interests can be negatively affected.[12]

These definitions clearly capture the phenomena from MSB's perspective, which is firmly rooted in the security goals for Swedish society (*målen för samhällets säkerhet*).[13] However, the definitions adopt a national perspective. As such, they may not offer a sufficiently pragmatic understanding for the purpose of this report which seeks to consider the organisational level of the communicator. Therefore, this report understands information influence activities in a slightly different way:

- Information influence activities are the *illegitimate* attempt to influence opinion-formation in liberal democracies (legitimacy);

- They are conducted to benefit foreign powers, whether state, non-state or proxies (intention);

- They are conducted in the context of peace, war and hybrid threat- or grey zone-situations, i.e. situations of tension that are neither peace nor war (ambiguity).

The fulcrum of this definition is the concept of legitimacy. Contrary to public diplomacy – which is the application of legitimate information power – information influence campaigns are illegitimate, although not necessarily illegal, for three interrelated reasons, which are ultimately *moral* in nature:

- Because they *deceive* people.[14] Information influence activities try to look legitimate when they are not. During the 20th century, industries emerged with the purpose of managing – and influencing – public opinion. Public relations, public affairs, public diplomacy and lobbying are examples of the legitimate efforts of organisations to influence public opinion in support of their interests. Information influence activities mimic established forms of media and engagement in the public sphere to leverage the system and the trust that people bestow upon it.

- Because they *exploit vulnerabilities*. By gaining access under false pretences, information influence activities not only utilize but exploit the system of opinion-formation. Liberal democracies, based on the assumption of good will, have open systems for public debate. Information influence activities betray this generosity and turn one of democracy’s greatest assets, free and open debate, into a vulnerability.

- Because they *break the rules* that govern constructive open and free debate; organized trolling,[15] for example, is not conducive to constructive dialogue.

2.2 Diagnosing illegitimate influence

What are the rules of open and free debate, and how can infractions be detected? There are many rules, of course, some spoken, some unspoken, some generally agreed on, some contested. For communicators tasked with making a judgment, four diagnostic criteria may be helpful in determining what requires a response:

- *Deception*: Legitimate influence is open and transparent about its source, origins and its purpose.

- *Intention*: Legitimate influence intends to contribute toward a constructive solution, even if the nature of that solution is contested. Although one should assume good will, in cases of information influence activities there is reason to believe that the intent is merely to do harm.

- *Disruption*: Legitimate influence ends where the disruption to society is disproportionate to or outweighs the potential benefits of that disruption. Strikes and protests for a specific social purpose, for example, constitute legitimate disruption.

- *Interference*: The legitimacy of engagement in open and free debate rests, at least partly, on being personally affected by an issue. The clandestine involvement of a foreign power in an election, for example, constitutes interference.

Actors tasked with countering hostile influence will not always be able to ascertain the presence of all factors. The characteristics will rarely be fully evident in isolated events. The proper level to detect and counteract attempts to influence is therefore the *chain-of-events* that they are part of. While single events may provide a useful starting point for identification, it is the chains-of-events – or *stratagems*, i.e. typical manoeuvres that play out over time and in varying fora (see 3.3) – that reveal an information influence campaign. The more factors present in the DIDI diagnosis, the more justified is the diagnosis that information influence campaigns, i.e. a coordinated effort, is taking place:

- *Deception*: The chain of events involves attempts to influence opinion formation by deceptive means such as e.g. factually incorrect news reporting or the use of false experts.

- *Intention*: The chain of events is, according to the best available evidence, conducted, controlled or instigated by an actor with perceived hostile intent, i.e. to undermine or otherwise harm society to further own goals and objectives.[16] Motivations for information influence activities can be commercial, political, criminal, personal, or in support of military operations.

- *Disruption*: The chain of events undermines, harms society and/or otherwise hinders the normal functioning of societal institutions or shows the potential to do so.

- *Interference*. The chain of events involves actors, especially foreign actors or their proxies, that have little or no business in interfering with the issue at hand; the involvement in the issue encroaches on the sovereignty of the state.

It is important that the diagnostic criteria are considered holistically, as four intersecting features of a campaign. The application of a single criterion will lead to the erroneous conclusion that perfectly legitimate political, activist or even business practices count as information influence activities. It is not a coincidence that techniques employed in information influence activities overlap with journalism, public affairs, public diplomacy, lobbying and public relations; mimicry of these techniques is part of the modus operandi. Furthermore, as the rules of conduct in the public sphere are often contested, open and free debate will always be characterized by an 'edge' of controversial practices. Fig. 2.1 illustrates that not everything that goes on in free and open debate is 'white'. The figure also illustrates, however, that information influence activities take place on the edge of open and free debate; they only pretend to conform. Communicators should note, moreover, that illegal influence attempts, such as blackmail or bribery, are a matter for the security forces.

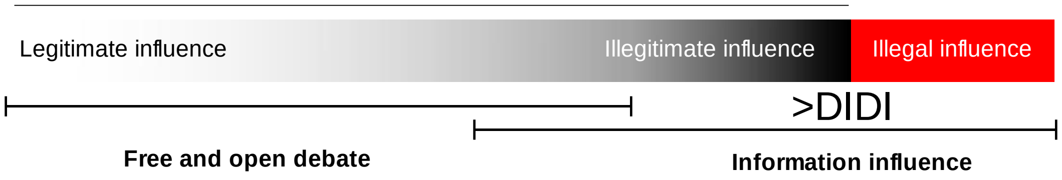


Figure 2.2: The continuum between legitimate, illegitimate and illegal influence.

2.3 Exploitation of vulnerabilities

Information influence activities *exploit the system of opinion formation*. To understand how, it is important to consider why citizens trust in the system, i.e. how the system works and how information influence activities exploit it. To do so, we propose a generic and simple model that broadly illustrates opinion formation and its vulnerabilities.

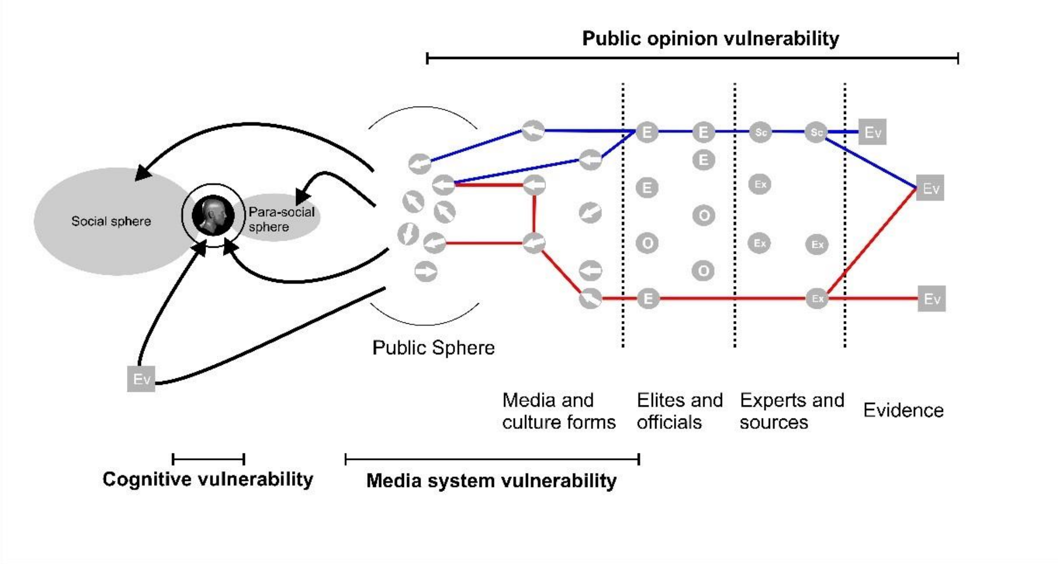


Figure 2.3: An ideal-type model of opinion formation in Western societies.

The epistemic chain

Western society is built on free opinion formation in a public sphere.[17] What this means is a highly complex question that has been debated for centuries.[18] For the purpose of mapping out vulnerabilities, we suggest that opinion formation, for illustrative purposes in the case of one single individual, can be understood as an *epistemic chain* that brings seven different systems into dynamic interplay. Fig. 2.2 (above) illustrates the seven systems that encompass from left to right:

- the *individual* with her unique identity and history and the biases resulting from the interplay of the realistic-, identity-, pragmatic- and modularity-principle[19];

- the *social* and *para-social sphere*, where the individual connects with other individuals either real (social sphere, populated by friends, colleagues, neighbours etc.) or mediated (para-social sphere, para-sociality is the forming of one-sided intimate relationships with figures in the media such as talk show hosts or US presidents[20]);

- the *public sphere,*[21] as experienced by the individual, where the individual not only connects with others, but deliberates with a community;

- *media forms* and forms of *culture* ranging from newspapers and television news to blogs and services like Facebook or Twitter;

- *elites* (in the sense of persons that the focal individual perceives as prestigious) and *officials*;

- *experts* (in the sense of persons that the individual perceives as endowed with expertise regarding a focal issue) and *sources* (whose expertise of a matter rests on 'being there');

- the *scientific system* or similar systems (like independent courts or independent, evidence-based journalism) which are characterized by their potential to operate with opinion- and person-independent *evidence*.

Although Western systems allow everyone to form their own opinion and to make a contribution to others' opinion formation, they rely on the idea that people who wish to be taken seriously do not make unsupported or unsupportable claims – and, moreover, that they are indeed real people with a reputation to lose.[22] What elites or officials say and what is consequently found in the media and in popular culture should be backed up by experts and authentic sources. What experts say should be backed up by evidence, at least where applicable, i.e. in questions that can be investigated scientifically. Sources should be real and genuine. The requirement to substantiate with evidence and argument does not mean, of course, that there can be only version of 'truth': equally substantiated and supported solutions might compete (as the blue and red versions do in Fig. 2.2). But politics and public debate should be a contest of only those ideas that can be reasonably substantiated and supported (given a good deal of tolerance towards ideas that might not fit preconceived notions of reasonableness). If that is guaranteed, and if the reputation of actors based on their contributions can be tracked, the public sphere will be by and large populated with sound and reasonable ideas which in turn will affect the ideas circulated in social media and at the dining table; there will be, ideally, convergence.

The ideal that society should be run in a rational, evidence-based fashion with actors identifiable and accountable, and that the public sphere is in some way a clearing-house where opinions, arguments and evidence are discussed openly, for every citizen to see and to follow, has served liberal democracies well.[23] Few theorists would claim that the ideal has ever been fully attained. But by and large, the acknowledgement of the ideal is traditionally seen as a strength of Western institutions. For the people, its greatest strength lies in the fact that the system makes it unnecessary to exclude certain actors or censor certain content beforehand[24], because there is a general trust that unsupported and unsubstantiated contributions will be revealed, or reveal themselves, as such; that tricksters and shysters will not stay in the game for long.

The emergence of a society-wide ‘fake news’ debate indicates, however, that the general trust in self-correction is to a degree shaken. Increasingly, vulnerabilities of the current system come into focus. For our purposes, we indicate three sources of vulnerabilities, which are depicted in Fig. 2.2.

- *Media system vulnerability*. Western media systems are currently vulnerable for two interrelated reasons. First is the rapid rate of innovations in the media system. The other is the commercial reconfiguration of large parts of the media system. Western media systems face rapid changes in technologies, patterns of media consumption, audience fragmentation, and new economic models.[25] The result is a hybrid media system comprised of ‘old’ (newspapers, television, radio) and ‘new’ (social) media. The state of hybridity, where familiar old and unfamiliar new media forms co-exist in constantly shifting and overlapping ways, makes evaluating news sources and triangulating facts more challenging than ever.[26] What comes on top, is the commercial reconfiguration. The commercial imperative has always been juxtaposed with the need for reliable news sources, of course; public service broadcasting is an important example. However, social media and online news are reliant on commercial models that are not as visible as advertisement breaks, and that are not as widely understood by their users: click-bait news[27] comes to mind. Therefore, information influence activities exploit technological, regulatory and economic vulnerabilities resulting from hybridity and to a degree invisible commercial reconfigurations in Western media systems as opportunities for negatively influencing public opinion and individual perceptions.

- *Public opinion vulnerability*. Western societies have a commitment to free opinion formation, which means that society ultimately relies on the sound judgment of its citizens. Sound judgment by citizens relies to a great degree on judgment of the *persons* acting in the public sphere, however, their reputation over time, their demonstrated expertise, their general habitus. Digital technology does not only empower everyone with a smartphone to take an active part in opinion formation; it also greatly expands the opportunities to do so covertly, anonymously or even impersonating someone else. Algorithms give a new logic to how news agendas are shaped. And while easier access may have broken monopolies, it has also led to the erosion of authoritative landmarks that gave orientation.

- *Cognitive vulnerability*. The human mind has evolved over millennia for conditions other than modern society.[28] Cognitive science and social psychology increasingly make clear that the human mind is not best understood as epistemic or 'truth-seeking' apparatus, but as a system that produces a *viable reality*.[29] We do not seek truth, but viability, i.e. a practical way of living our life.[30] For prehistoric hunter gatherers, conformity with a group was probably far more important than dogged determination to see the world as it is.[31] Confirmation bias,[32] i.e. the human tendency to ignore information that is at odds with what one already believes, acts as a protection mechanism against threats to identity and belonging. In a recent publication, the authors have reduced this complex topic to the interplay of four principles: the reality principle, the identity principle, the pragmatic principle and the module principle.[33] Information influence activities exploit the tensions in the human mind that arise from the conflicting principles to exert influence over individual perceptions, behaviour and decision-making.

2.4 Hybrid threats and grey zones

Information influence activities are conducted by a foreign power (including non-state actors) or their proxies in the context of a hybrid threat- or grey zone-scenario.[34] Here, information influence activities are often only one element in a larger asymmetric strategy of influence that involves targeted use of corruption, investing in political parties, think thanks and academic institutions, cyberattacks, the use of organized crime, coercive economic means, and the exploitation of ethnic, linguistic, regional, religious and social tensions in society.[35] It is important to contextualize, therefore, and understand the nature of hybrid threats and the grey zone of unpeace[36], i.e. a state that is neither peace nor war.[37]

Some hostile external actors (1) possess the ability to conduct influence operations, either directly or through agents; (2) have an appreciation of the vulnerabilities of the epistemic chain (media system vulnerability-public opinion vulnerability-cognitive vulnerability); and (3) intend to exploit these vulnerabilities for the sake of furthering their own agenda. Such actors constitute *threats*. The following four overarching type of threats can be discerned in the literature:

- *Violent extremism*: The general intent is (1) to create a climate of fear; (2) to radicalise; and (3) to incite acts of terrorism. Examples include ISIS/Daesh disinformation and propaganda.

ISIS/Daesh influence activities:[38] ISIS/Daesh has gained a strong reputation with regards to its successful strategic communications. The purpose of ISIS/Daesh strategic communications activities is fourfold: to portray itself as an effective and legitimate organisation, to attract and retain recruits, to explain its ideology, and to instil fear and polarise societies. ISIS/Daesh builds its strategic communications around a meta-narrative which combines several themes including Islamic religious dimensions, conspiracy theories of Western oppression, underdog, and youth culture narratives. The resulting brand is dubbed "Jihadi cool" and has been widely successful in attracting younger audiences. ISIS/Daesh's tactics are creative and make use of modern technology, particularly when it comes to reaching key target audiences in the West. ISIS/Daesh utilizes a network of peers for most of its communications, having moved from a vertical to a horizontal approach towards messaging.

- *Hostile states*: This includes state-sponsored activities that contribute to regional instability. The intent is to pursue political goals through *information warfare* and *hybrid threats*. An example is Russia's annexation of Crimea in 2014.

Russian active measures:[39] Under the umbrella term *active measures*, Russia has been observed using disinformation and propaganda to support hostile actions on the ground, as in the cases of Ukraine and Georgia. Russia's information warfare is based around a meta-narrative constituted by core themes that surprisingly often contradict one another. These core themes are spread by Russia's government-controlled media channels (RT and Sputnik among others). The meta-narrative is strongly anti-interventionist and depicts the West as an aggressive and expansionist entity on the one hand, and as weak and verging on collapse on the other. The aim of this narrative is to undermine the legitimacy of the EU and other Western institutions, and to destabilise the information environment and ultimately political decision-making.

- *Sub-state criminal actors:* In some parts of the world, criminal organisations, such as drug cartels, can be as powerful as the state they use as a cover and operational basis, while simultaneously fighting it for power. Contrary to hostile states, the primary motive of these sub-state actors is commercial. Guevara concludes that "The Mexican drug cartels have blurred the lines between criminality, insurgency, and terrorism further raising the national security importance of this topic."[40]

Mexican drug cartels:[41] In the so-called Mexican Drug War, drug cartels use a range of influence techniques in competition with the state and other competing cartels in order to secure the loyalty and grassroots support of the people. Narco-propaganda is ultimately based on fear induced by violence, but it also employs devices such as graffiti glorifying the narco-lifestyle or narco-themed folk songs (narcocorridos) as well as social media and television shows. Experts have pointed out that there is a ‘public relations battle’ going on, with the drug cartels by and large controlling the way their activities are portrayed in the national media, and the government struggling to reach out to its own citizens.

- *Hackers & profiteers*: Individuals and groups skilful in the logics of digital systems sometimes act disruptively to demonstrate their ability. This could include activities as diverse as hacking the scores of an online multiplayer videogame or damaging vital social infrastructure. These actors are primarily interested in ‘gaming’ or ‘hacking’ systems, sometimes for economic benefit and sometimes simply because they want to show that it can be done. This can contribute to a muddying of the waters in the information environment. Such actors can be knowingly or unknowingly engaged for harmful purposes.

Teenagers working from their bedrooms:[42] During the 2016 US presidential election campaign, investigative journalists found that multiple pro-Trump websites were run out of Macedonia, many of them sensationalist in character and spreading fake news. The young entrepreneurs behind these websites, many not more than 18 years old, made thousands of US dollars from click-based advertisement revenues by flooding the information environment with bogus news and websites, with each click earning just a fraction of a cent.

Regarding the agendas of specific hostile actors, we prefer to avoid speculation. Such assessments are best left to intelligence agencies with access to specialist knowledge and specific information. What can be said on a general level is that actors’ agendas can vary, and could encompass goals as diverse as influencing a country to shift position in sensitive policy issues to better suit the hostile actor, prevent international cooperation to strengthen the hostile actor’s position, weaken governance structures in preparation for a conflict, cause strain on important societal institutions to increase the cost of business, shift political opinions in favour of the hostile actor, or cause polarisation within the population to weaken societal cohesion.[43] At the more amateur level, intentions include ‘gaming’ the system to prove it can be done, or simply exploiting a loophole to gain an advantage until the loophole is closed.

3. Identifying information influence activities

Identifying information influence activities is the first step in counteracting them. For that reason, the following chapter offers an inventory of the strategies, techniques and stratagems typically employed in information influence activities. It begins with a discussion of general influence *strategies*, i.e. the broad and general line, followed by an examination of *techniques* that are used for hostile influence. It concludes with suggestions of how these strategies and techniques are deployed in coordinated fashion, as *stratagems* – that is to say, manoeuvres, ploys or schemes – designed to achieve specific goals, together with case study examples.

3.1 Influence strategies

The following subchapter looks at influence strategies in a broad sense. We do not make claims about the legitimacy or hostility of these strategies in their own right; they are simply approaches to using information to achieve goals.

3.1.1 Positive, negative, oblique

The classic military distinction between offensive and defensive strategy does not transfer easily to information influence. On an abstract level, influence strategies can be categorised by whether they pursue a positive, negative or oblique aim.

- *Positive* or *constructive* strategies try to establish a coherent narrative, either on a general societal level or with selected target audiences. This means that the influence campaign's narrative directly correlates with or complements existing, widely accepted narratives: communism as well as capitalism claimed, for example, that they would build a more prosperous future.

- *Negative* or *disruptive* strategies attempt to prevent the emergence of a coherent narrative or try to weaken or destroy an existing narrative, again either on general societal level or with selected target audiences. Attacks on narratives are normally conducted by selecting contested themes such as e.g. crime or immigration. For disruption to work, these themes must compete with congruent themes in the attacked narrative, but they do not have to be used coherently by the attacker.

- *Oblique* strategies try to draw attention, with the aim of distracting from the key issues. They tend to focus upon the information environment, seeking to dilute, flood or poison it with alternative messages.[44]

The level of operation provides another way of categorizing different strategies by examining whether they operate on a general societal level or on limited targeted audiences. Targeting can occur on any given level in society and be based on any available variable. However, the following three levels provide for a sufficiently differentiated understanding:

- *General societal level (mass audiences):* Influence operations can target society as a whole, by aligning messages with symbols and narratives which are widely shared by a society's population.

- *Sociodemographic targeting (groups):* Sociodemographic targeting is the classic approach used in the advertising profession to identify audiences based on demographic factors such as age, income and education, allowing for more adaptation of messages. This can also be focused on groups or communities committed to specific causes.

- *Psychographic targeting (individuals):* Psychographic targeting refers to techniques where technology is used not only to target audiences as groups of similar people, but as *individuals* with a specific psychographic profile, be they key decision makers or ordinary citizens. Social media offers a range of big data points through which artificial intelligence can search for correlations that are suggestive of e.g. an individual's political views. The ability of social media platforms to enable precision-targeted messaging and advertising based on psychographic targeting is an unprecedented dimension of contemporary information influence activities.[45]

In sum, we can abstract three key dimensions of influence, useful for classifying different strategies. The aim of the strategy constitutes a first dimension (positive, negative, oblique). The level of operation constitutes the second (general, targeted, precision-targeted). A third aspect is the range from *environment-oriented* (general and aimed at altering the information environment) to *message-oriented* (specific and directed towards single individuals/narratives or issues). The table below, which draws inspiration from the work of business strategy theorist Michael Porter,[46] illustrates the generic configurations of influence strategies along these dimensions using generic examples:

Environment-oriented ↔ *Message-oriented*

Influence strategies	**General**	**Targeted** (sociographic)	**Precision-targeted** (psychographic)
Positive, constructive	*Convince target society that influence campaign's new ideology is more attractive than the existing one (fairer, more prosperous etc.)*	*Recruit adherents of the new ideology in a circumscribed target group, e.g. university students*	*Disseminate political propaganda to individuals based on micro data on individual preferences and interests*
Negative, disruptive	*Polarize target society, erode trust in target society's institutions*	*Spread disinformation amongst key policy makers in order to disrupt decision-making processes*	*Utilize harassment, including harassment by bots, to discourage individuals with specific profiles from engaging in public debate*
Oblique, distractive	*Draw attention away from the influence campaign and to other events*	*Keep journalistic debate about a societal issue focused on trivial technical issues*	*Distract specific individuals by precision-targeting them with distractive content*

As a rule, some influence campaigns will pursue different (or multiple) strategies for different target audiences (and individuals, in the case of precision targeting) and may or may not pursue a strategy on general societal level. Conversely, an influence campaign might not expend any effort on influencing target audiences but concentrate solely on poisoning the information environment to a degree that civilized democratic debate breaks down, or that makes isolating individuals more effective. One important aspect of digital platforms is the lower cost of targeting and precision-targeting audiences. According to a statement from Facebook in the U.S. Senate Intelligence Committee's hearing about Russian election interference, Russian advertising on Facebook during the 2016 Presidential election may have reached up to 126 million users at the cost of just $46,000.[47]

3.1.2 Narratives and facts

In the counter influence literature there is debate, and also some confusion, about the 'battlefield' on which to engage information influence activities. While some authors concentrate on the role of facts and fact checking[48], others place greater emphasis on narratives[49], i.e. the ways in which facts – and false information deliberately positioned to appear as facts – are used to support storytelling. Influence strategy can only be adequately understood, however, if facts and narratives are seen as interrelated. Narratives refer to the sequencing, structure, or organisation of signs, codes, and events into a coherent order. They can include both real and imaginary components. Narratives based on facts are more correctly defined as being formed out of *representations*, *interpretations* and *perceptions* of information that is claimed to be factual. They rely, in other words, upon *statements of facts*, which represent a fact more or less well. When added to a sequence of propositions in a narrative, factual statements can be transformed by that narrative. For example, the factual statement "Elvis died in 1977" is modified by the additional line, "and now he lives as a preacher in the Valley".

Factual statements

The concept of 'fact' has increasingly become a source of confusion, particularly in light of terms such as post-truth.[50] For the sake of clarity – and bearing in mind that there are some 2,500 years of philosophical debates into this topic – we offer a simple working definition. A *factual statement* makes the claim to represent something that is or has been the case in the material or social environment. A key element of a factual statement is that it is verifiable within pragmatic boundaries.[51] The statement "Elvis died in 1977" can be verified, but convincing evidence that he faked his death remains elusive.

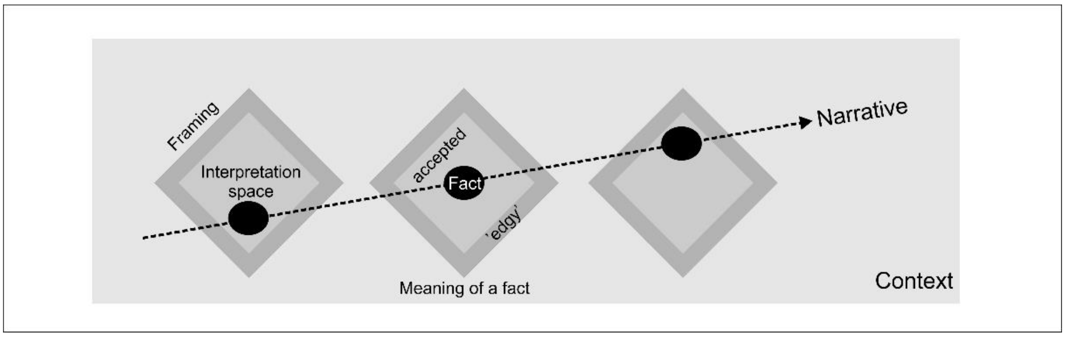


Figure 3.1.2.1: The interpretation space of facts allows for the alignment of facts to form a narrative.

Meaning, context and narrative

In influence campaigns, factual statements, for example in the form of statistics, can be employed to support one's own narratives or to undermine the adversary's. Metaphorically speaking, factual statements are selected and aligned to define the trajectory of a narrative, as in Fig. 3.1.2.1. However, the 'facts' in statements rarely have a bearing on narrative: a burned car might indicate civil unrest or a faulty electrical wiring. When employed to support or undermine narratives, factual statements put the facts in *context*. Often the author will offer an *interpretation*, i.e. implicitly or explicitly attribute a meaning to the factual statement in a process known as 'framing.'[52] Whether the attribution of meaning is convincing does not only depend on the skill of the author, but on the space of accepted and acceptable interpretations. Fig. 3.1.2.1 illustrates that factual statements can be framed to fit within a narrative; disinformation around MH17, for example, drew upon the pre-existing frame that the CIA has previously covered up similar events. In many cases, outrageous claims are not made to be believed, but to widen the corridor of publishable opinions, i.e. to make 'edgy' things thinkable (environmentally oriented strategy). Conspiracy theories are a classic example of this principle. Conversely, debates about a certain fact often obscure another, sometimes far more important question: namely which facts are not presented.

Developing an alternative narrative – often one designed to be damaging to the target community – is an activity fundamentally different from undermining the target's existing mainstream narrative. The positive strategy is constructive and becomes vulnerable because it demands coherence to existing statements and requires factual statements which can be verified. The negative strategy is deconstructive and is therefore easier to wield. The case of the tobacco industry in the second half of the 20th century shows that asymmetric strategies that rely on doubt (i.e. doubt about the fact that cigarettes really cause cancer) are extremely difficult to counter.[53] Russian broadcaster RT, for example, has the tagline "Question More", and positions itself as a counter-cultural voice when compared to Western mainstream media.[54]

Narratives, push and pull

That influence campaigns *push* narratives, i.e. try to convince people, is only one side of the coin. The other side of the coin is that narratives, once firmly embedded in a person's mind, tend to act as the organizing principle by which people make sense out of input, and indeed governs the acquisition of new input. A successful narrative, once ingrained, is to a degree *self-stabilizing*, as people tend to shy away from input that threatens the narrative's integrity: a phenomenon captured and researched under the label *cognitive dissonance*. Put very simply, narratives are pushed onto people, then act as a pull-principle: people are less interested in whether the story is true or false, as long as it fits into their preferred narrative. This effect is well-researched under the label *confirmation bias*.[55] Social media technology, particularly through the personalization of information flows, potentially reinforces this dynamic and runs the risk of contributing to what is commonly referred to as *filter bubbles* and *echo chambers*.[56]

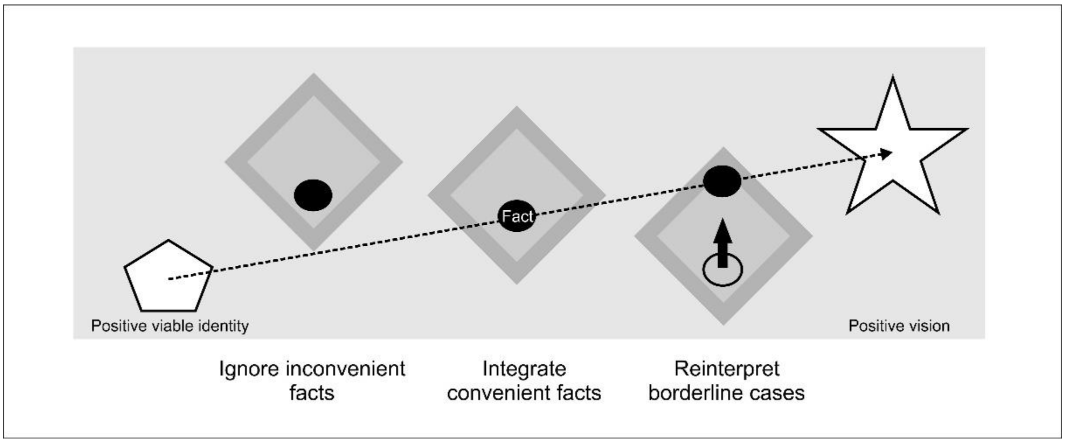

Fact
Positive viable identity
Positive vision
Ignore inconvenient facts
Integrate convenient facts
Reinterpret borderline cases

Figure 3.1.2.2: Meta-narratives provide coherent frames for factual statements to create and shape identities

Meta-narratives

Although the power to integrate observable realities is important, humans do not only accept narratives to the degree they are supported by factual statements and ascertainable facts. [57] On a very general level, the most persistent grand narratives, or meta-narratives, in people's lives are probably the ones they are *socialized* into, i.e. the ones acquired in childhood.[58]

Narratives are not only defined by a trajectory of factual statements, but perhaps more importantly by their endpoints, a positive viable identity for the believer *now*, and an attractive vision *then* (see figure 3.1.2.2). The endpoints themselves might be reasonable, but often they have very little to do with reality. The most successful grand narratives – the great religions which have endured for millennia – are to a large degree *a-factual*, i.e. the truth value of their material content cannot be ascertained. Successful radical ideologies, the secular equivalents of religions – often seem to derive their attractiveness from visions of a *superior future*, which tends to be 'just around the corner', and to which the 'fighter' can contribute here and now (again, the truth statements cannot be verified, but values and injustices can). Thus, those interested in how influence campaigns function are well-advised to study how a narrative gives identity to its adherents.

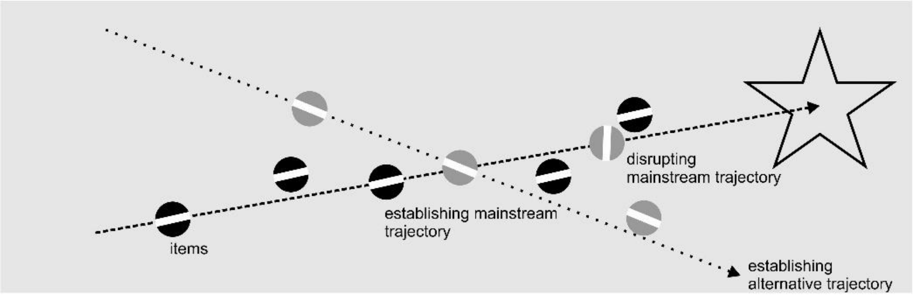


Figure 3.1.2.3: Alternative narratives align facts differently to support their trajectories

When one looks at media coverage, they not only relate facts, but contain a *fractal* of a meta-narrative, i.e. a miniature version of a bigger idea. The same is true for e.g. newspapers or publishing houses: *Professional periodical media products reproduce* their commercially viable meta-narrative over and over again. Fig. 3.1.2.3 illustrates the principle. The black dots represent items that contain stories (or publishers) in alignment with mainstream narratives. The grey dots either represent items that are in alignment with *alternative narratives* or items that disrupt mainstream narratives by presenting a contradicting view. With the increasing fragmentation of the media landscape, this happens in multiple fora. That is why the chain of event-perspective is important for counter influence operators. A seemingly trivial and isolated misrepresentation in one forum might serve as another piece in the mosaic of a destructive narrative in another.

3.1.3 Classic vs. cognitive strategies

Influence is different from force because a message needs to be *cognitively processed* by the target to have an effect. In the classical view, 'processed' meant noticed, understood, believed and in the best of worlds acted upon. In this hierarchy of effects, the credibility of the message, and in extension the credibility of the message's source, were the key determinants for *believing* – and without belief, no action. Awareness, attitude and behaviour change were at the heart of the traditional persuasion industries for most of the 20th century. The credibility of messages and sources still matters, of course, especially when pursuing a positive and constructive strategy for developing a coherent narrative leading toward behaviour change. However, the general rationale that messages need to be believed (or even understood or noticed) is increasingly being called into question, both in its applicability today and in principle.[59] When it comes to applicability, it must be remembered that classic communication models date to a time when a propaganda target had limited access to alternative media sources. Today, in contrast, networked individuals are exposed to a daily barrage of fragmented messages: many, like memes, without ascertainable factual content (they cannot be true or false), many without an identifiable source, many only in the flow because they were shared by a friend or promoted by algorithms. Often, the only action necessary is a simple click. Cognitive strategies may therefore be considered those that reduce the response to a simple impulse – like, don't like, happy, sad, share – and shortcut the traditional hierarchy of effects that lead from awareness to behaviour change.

3.2 Influence techniques

The study of influence techniques is not new, and inventories of influence techniques are not a new idea either:[60] Walter Lippmann's *Public Opinion* (1922)[61] and Harold Lasswell's *Propaganda Technique in the World War* (1927)[62] have been seminal works for the study of propaganda techniques. During the inter-war period, pre-emptive counter-propaganda was important to US foreign policy and the Institute for Propaganda Analysis (IPA) formulated seven basic propaganda devices, colloquially referred to as 'the seven sins', including techniques such as name-calling, testimonial, card stacking and band wagon.[63]

The problem with lists such as the seven sins is that they mix rhetorical moves such as name-calling or populist rhetoric with broader approaches such as recruiting shills for testimonials or the engineering of a fake 'band wagon' or popular movement (astroturfing). Our approach is somewhat different. By studying a wide variety of literature on information influence, as well as a number of cases where such activities can be inferred, we have abstracted a number of common and contemporary *techniques*. The employment of these techniques is indicative of information influence, but it is not conclusive evidence. There are two levels here.

- *Purpose:* On the basic level, techniques themselves are neither good nor bad. As an example, bots can be useful pieces of software for automating tasks for example in customer service. But they can also be used to deceptively amass 'fake social capital' in order to support a disruptive narrative.

- *Acceptable use:* On a more advanced level, tools can be used openly and in accepted ways, or deceptively and with hostile intent. The use of rhetoric is expected from public debaters nowadays, for example, and it probably helps with citizen involvement. Malign rhetoric such as the gish-gallop is not considered acceptable, however.

While many of the techniques here stretch or even break the accepted rules of civilized debate in polite society, they are not proof of hostility per se. An underdog politician employing malign rhetoric is neither automatically an agent of foreign influence, nor should her other, legitimate arguments automatically be dismissed. Organisations should be wary, however, when several techniques are employed simultaneously, or temporally close, by a single source, or by sources with known connections to the benefactor of the techniques. This may indicate that a stratagem is deployed (see 3.3).

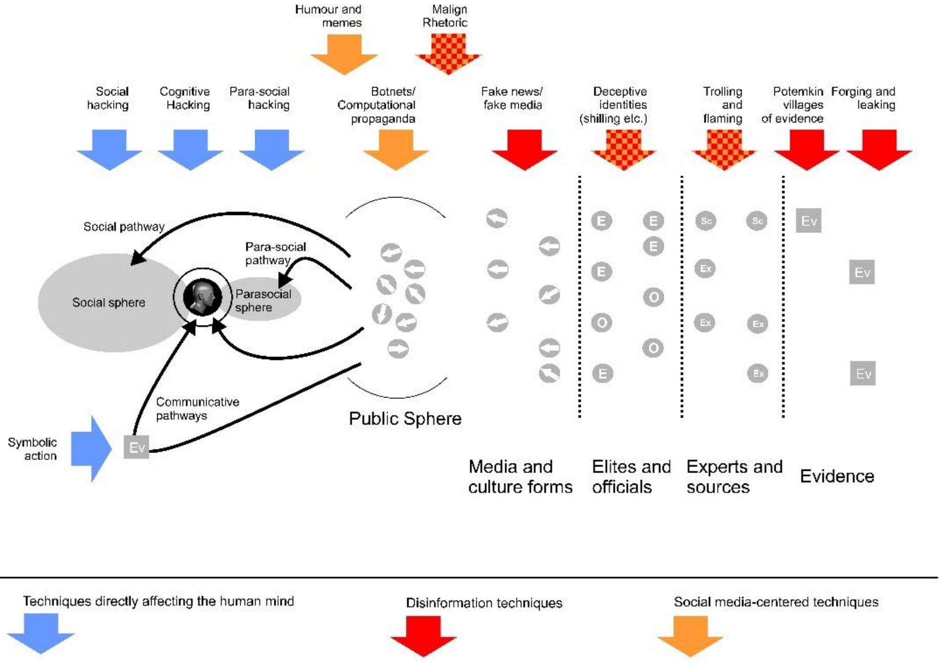


Figure 3.2: Different influence techniques are applied to different vulnerabilities along the model of free opinion formation.

Figure 3.2 builds on the framework established in Chapter 2 in order to map how and where these information influence techniques exploit vulnerabilities in the epistemic chain. It shows, in other words, how each technique seeks to impact upon media system, public opinion and cognitive vulnerabilities. The following techniques are detailed systematically, from left to right according to the image, in the following sections. Taken together, this framework and taxonomy of techniques helps to build the knowledge necessary for the effective identification and countering of information influence activities.

3.2.1 Sociocognitive and psychographic hacking

Advertising, public relations and other forms of professional public communication have always relied on targeted approaches, i.e. on messages that are specifically created to appeal to (more or less) precisely identified groups. Such attempts to "get into the heads of people" have generally been regarded as legitimate, as long as the messages told in one campaign to different target groups are not fundamentally incongruent and, of course, reasonably true.[64]

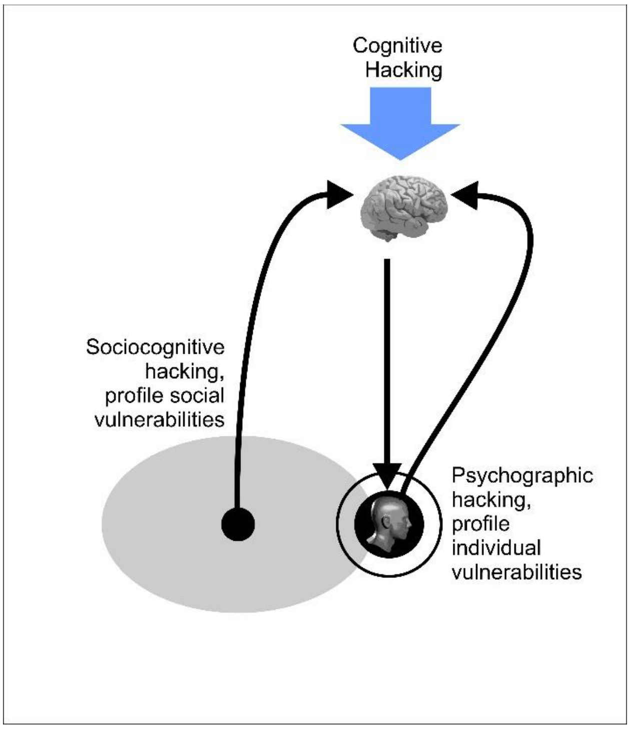

Cognitive Hacking
Sociocognitive hacking, profile social vulnerabilities
Psychographic hacking, profile individual vulnerabilities

Figure 3.2.1: Sociocognitive hacking utilizes messages specifically targeted at individual or group-based vulnerabilities arising for sociocognitive features

Cognitive hacking takes this idea to the extreme. However, contrary to marketing campaigns, cognitive hacking is conducted with *intent* to covertly influence an audience. Moreover, disruptive or distractive information influence campaigners, especially when they remain in the shadows, do not have to offer a coherent narrative or even facts that stand up to scrutiny in the long run: as in Swiftboating, i.e. smear attacks on a political candidate timed so aptly before an election so that the swiftboatee cannot counter the attack anymore, the immediate effect is all that counts.[65] Cognitive hacking comes in two forms. Sociocognitive hacking targets the *cognitive vulnerability* of individuals in communities. Psychographic hacking targets the isolated individual.

Sociocognitive hacking: harnessing outrage

The core of sociocognitive hacking lies in the attempt to activate psychosocial trigger-points, which are the 'cognitive vulnerabilities'[66] of individuals and, in extension, communities. Although the hacking relies to a degree on psychosocial dynamics, it predominantly works by appealing to powerful emotions.

Social unrest initiated by rumours:[67] A case of sociocognitive hacking took place in India in September 2013. It began with a young Hindu girl complaining to her family about being verbally abused by a Muslim boy. Allegedly, the girl's brother and cousin retaliated by killing the boy. While the rumour led to a first round of clashes between Hindu and Muslim communities, an unknown individual posted a video of two men being beaten to death by an angry mob. Fanning the flames, the caption identified the men as Hindu and the mob as Muslim. Rumours that the mob had murdered the brother and cousin spread like wildfire over social media and telephone. In the end, it took 13,000 troops to stop the ensuing violence. In the aftermath, it emerged that the video did not show the brother and the cousin and, although authentic in its content, was not recorded in India.

What makes cognitive hacking highly potent for information influence is not the targeting per se, but the degree to which *emotions* and *fundamental human motives* – fear, anger, hate, anxiety, but also positive emotions such as honour – are targeted. Strong emotions, especially when suppressed, constitute cognitive vulnerabilities, because they suppress reason and restraint: the crowd 'sees red'. Furthermore, cognitive hacking requires a thorough understanding of the target audience and its identity processes, which, in the above example, was centred on religious community. Information is 'weaponized' in order to hack or short-circuit the individuals' and the community's cognitive defences, in the aforementioned case to lower the threshold to violence.[68]

Psychographic hacking: dark ads

While sociocognitive hacking exploits the cognitive vulnerabilities of individuals by manipulating group dynamics, psychographic hacking targets individuals by isolating them. Psychographic hacking relies on social media technology, especially the big data collection and commercial services provided by social media platforms such as Facebook. Facebook 'dark ads' are characterized by two related features:

- They are psychographically precision-targeted, i.e. they are constructed based on an individual user's psychographic profile, aggregated from what the user has been doing on Facebook and elsewhere on the internet.
- They are 'dark' insofar as they are only visible to that user.

Will Moy, director of the independent fact checking website Full Fact, warned against the consequences of dark ads for the British elections: "It's possible to target dark ads at millions of people in this country without the rest of us knowing about it. Inaccurate information could be spreading with no-one to scrutinise it. Democracy needs to be done in public."[69]

Facebook ads in the US election: In the aftermath of the U.S. presidential election of 2016, Facebook disclosed that the company had effectively sold ads to the value of more than USD 100,000 to a Russian company linked to the Kremlin.

The ads ran in the period between June 2015 and May 2017 and were connected to around 500 fake accounts created by the St Petersburg 'Internet Research Agency'. The majority of the 3,000 ads did not refer to the political candidates. Instead, they focused on hot topics such as race, gay rights, gun control and immigration. Following the stratagem of *Polarisation* (see 3.3.7), the ads did not display a coherent pattern. Some expressed support for groups like Black Lives Matters, others agitated against the movement. In the opinion of U.S. senator Mark R. Warner, the aim of the operation was simply "to spread chaos".[70]

Research suggests that psychographic profiling and precision-targeting is a highly potent technology.[71] Chris Sumner, research director and co-founder of the not-for-profit OnlineFoundation conducted experiments that showed that targeting people with ads tailored to their psychographic profile increases the impact of the ads. Sumner also draws attention to the consequences, especially in political campaigning: "The weaponised, artificially intelligent propaganda machine is effective. You don't need to move people's political dials by much to influence an election, just a couple of percentage points to the left or right." The problem with psychographic hacking is that the identities of those targeted, and the messages they are targeted with, remain clandestine.[72]

Summary

- Sociocognitive and psychographic hacking aims to get inside the head of the person who is to be influenced by using psychosocial trigger points and emotions
- Cognitive hacking can be precision-targeted down to the individual level
- Social media data can be used for information influence campaigns to design interventions based on individual sentiments, with information spread 'in the dark'

3.2.2 Social hacking

While sociocognitive hacking aims at short-circuiting an individual or community's rationality by triggering overpowering emotional responses, social hacking exploits people's tribal nature.[73] Having evolved as hunter-gatherers and incapable of surviving on our own, humans tend to be conformers, as Asch's famous experiments have shown.[74] Humans have a tendency to believe and do what others in their in-group believe and do. This is partly due to pragmatism, partly due to the identity-constituting effect of 'belonging', and partly due to the real requirements of society. The consequence for influence operations, especially in social media, lies in the fact that humans are vulnerable to the exploitation of several group dynamics.

Social proof and fake social proof

Social proof is the tendency to believe something not because there are good arguments but because a lot of others seem to believe it. In certain situations, social proof *is* a good argument. If millions of car-owners are happy with their Toyota, their satisfaction stands as proof for the car's quality. The same does not hold for e.g. accusations of criminal activities, however. Even if everyone believes that the accused is guilty, social proof does not constitute proof. To complicate matters, the agreement of experts who have independently weighed the evidence *can* constitute proof.

The problem with social proof is that humans are likely to blur the differences between the three constellations. This tendency can be exploited, especially on social media. Experimental research shows, for example, that posts on social media will get more likes over time if they start with some likes.[75] On platforms such as Facebook and Twitter, social proof in the form of popularity expressed through liking or sharing, is one of the key drivers of algorithmic curation determining what users will be made aware of.[76]

Another aspect of social proof is that it tends to override real evidence: people count numbers instead of arguments. Observers of the climate change-debate have long observed, for example, that the typical setup of debates about global warming, with one expert arguing against one other expert, creates the false impression that the issue of man-made global warming is controversial amongst experts, and that roughly 50 % of the expert community side with each position. A debate setup more reflective of the consensus in the expert community would, in this case, be 99 experts arguing against 1, but the seemingly equal setup with one expert on each side prompts a cognitive shortcut where the social proof overrides evidence.

Bandwagon-effect and spiral of silence

The tendency of humans to conform to a group (while not necessarily thinking of themselves as conformers) gives rise to two dynamic effects, which are often exploited: the *bandwagon-effect* and the *spiral of silence*.

The bandwagon-effect captures the phenomenon that the rate of adoption of ideas increases with the number of people that have adopted the idea. As more and more people buy into a fad or trend, more and more others will want to join. In addition, celebrities and elites will want to 'hop onto the bandwagon' in order to benefit from its popularity. In other words, popular ideas are self-amplifying. With fashion cycles, the rate of adoption itself is the driver, not necessarily the inherent quality of the design. In addition, the bandwagon-effect is an established effect in politics and voting research, where it gives an advantage to the leading party, cancelling the underdog-effect.[77]

In public affairs, the engineering of grass roots-movement, i.e. popular movements that pressure politicians ‘from below’, is a common tactic. *Astroturfing* – i.e. suggesting that there are a lot of folks out there who support a political agenda, while in fact there is no such support[78] – takes the idea not only a step further, but into the realm of deceptive exploitation.

The spiral of silence captures the reverse dynamic: namely that people are extremely sensitive about being isolated or ostracized because of non-conforming beliefs. [79] The theory proposes that humans possess a sensitive “social skin” or “quasi-statistical” sense to determine public opinion.[80] Individuals who feel that their opinions are publicly accepted and in the majority become more outspoken; those with apparently non-conforming minority beliefs increasingly stay silent. This dynamic drives the spiral further: as the silenced belief becomes less and less audible in society, it loses even more ground.

The dynamics created by the bandwagon-effect and spiral of silence are open to exploitation because people’s ‘quasi-statistical sense’ of opinion distribution is most likely only attuned to small groups. When it comes to a complex mediatized society, humans can be easily misled about the actual distribution of opinions, especially if media sources present inaccurate, unbalanced or contradicting representations of public opinion.

Selective exposure: Filter bubbles and echo chambers

Nicolas Negroponte, one of the early theorists of digitalization, predicted already in 1996 that information technologies would become customizable to each and every individual user. Negroponte “envisioned a digital life, where newspapers tailor content to your preferences [...] and media consumption becomes a highly personalised experience”.[81] Experts increasingly draw attention to the dark and dysfunctional side of this emergent personalization and customization.

The term ‘filter bubble’ was popularized by Pariser (2011).[82] In his book, Pariser warned that algorithms which personalize and customize a user’s experience on social media platforms like Facebook might entrap the user in a bubble of his or her own making. Pariser’s warning came at a time when experts began to warn about ideological polarization in social networks.[83] Echo chambers, similarly, refer to organically created internet sub-groups, often along ideological lines, where people only engage with “others with which they are already in agreement”. [84] Together, filter bubbles and echo chambers online can potentially contribute to political division and fragmentation of opinion online. This, however, remains to be conclusively proven or disproven. Contemporary research increasingly challenges the notion that filter bubbles and echo chambers influence opinion formation, as the internet, despite selective algorithms, often presents a broad spectrum of opinions and information to everyone, offsetting the negative effects of personalized user experiences.[85] Further, people rarely rely entirely on the internet and social media as their primary source of information, and for example in the U.S. TV is still the dominant platform for news.[86] More likely is that algorithms contribute to reinforcing pre-existing beliefs by selective exposure rather than creating and shaping new ones.[87]

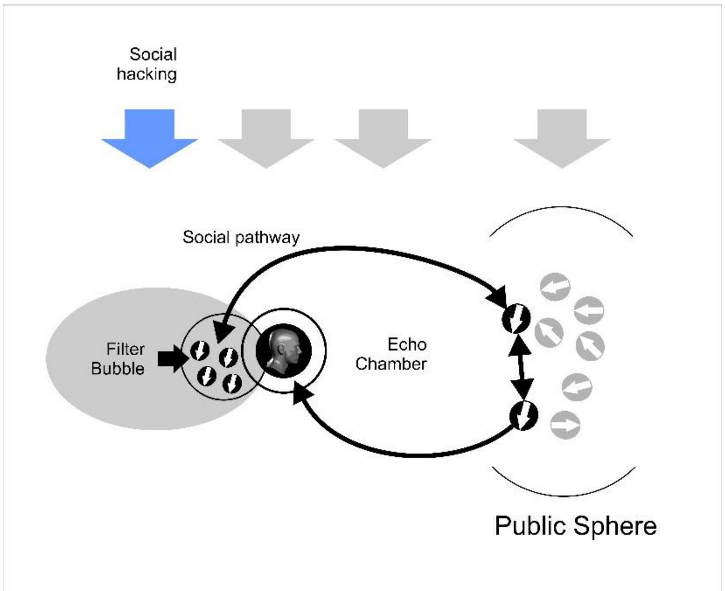


Figure 3.2.2: Filter bubbles and echo-chambers reinforces pre-existing beliefs by limiting exposure to alternative narratives and messages

Still, the possible dysfunctional effect of personalization and customization, and its exploitation potential, derives from the fact that democratic debate requires citizens to be exposed to a variety of viewpoints, with at least some at odds with what they already believe (so-called counter-attitudinal exposure). However, users do not necessarily wish to be exposed to perspectives they do not agree with (lack of user-driven counter-attitudinal exposure); and social networking-sites do not necessarily have an interest to provide it (lack of system-driven counter-attitudinal exposure). Contrary to the ideals of classic journalism, the primary interest of social networking sites does not lie in political education. It lies in keeping the user on the platform, *clicking*. Since exposure to information conflicting with one's beliefs risks the interruption of the flow of clicks,[88] social networking sites have very little interest in exposing users to dissonant information. The aim and function of their algorithms is "to connect people with information they are likely to want to consume, by making some items easier to access than other items [which] curates [...] a personalized stream of content [that fails to offer] users a set of alternatives to choose from."[89]

While polarization constitutes a real danger to democratic debate, research does not support the more alarmist claims at present. Borgesius et al. conclude, for example, that "in spite of the serious concerns voiced – at present, there is no empirical evidence that warrants any strong worries about filter bubbles".[90] The researchers do stress the necessity of debate and further inquiry, however, especially in the likely event that filter personalization develops into one of the main sources of information for society.[91]

Other researchers have found moderate effects from ideological and partisan echo chambers in twitter discussions[92] and in Facebook groups and pages.[93] Dylko et al. offer experimental proof that “system-driven, user-driven, and aggregated customizability technology increased clicks on and time spent reading pro-attitudinal political articles and decreased clicks on and time spent reading counter-attitudinal political articles”.[94] The researchers found empirical evidence that customisable technology increased ideologically driven selective exposure and the likelihood of echo chambers and filter bubbles in the modern media landscape. System-driven selective exposure, “due to its automatic and unobtrusive operation, customizability technology might be particularly effective at reducing cognitive dissonance associated with the avoidance of challenging information”, seemed to have a stronger influence on selective exposure than user-driven customisability.[95] Ideologically moderate target users were shown to be particularly susceptible.[96]

Summary

- Social hacking exploits vulnerabilities arising from social cognitive features of the human mind
- The harnessing of social proof, band wagon effects and selective exposure are examples of social hacking
- Algorithms on social media platforms can enable social hacking by, for example, contributing to *filter bubbles* and *echo chambers*

3.2.3 Para-social hacking

The expression *para-social* captures the idea that humans sometimes begin to experience their objectively one-sided-relationships with personalities in media subjectively as two-sided; that is to say, symmetrical and reciprocal. In other words, viewers begin to believe that the talk show host talks directly to them, that she is a friend. A para-social relationship is the one-sided illusion of a social relation. While para-social relationships became the object of research only with the advent of television – the term was coined in the 1950s[97] –, humans have always formed illusionary relationships, e.g. with far-away kings and mythical figures.

Social media and celebrity cultures have given everybody the tools to build their own spaces of public immediacy, and even intimacy, without having to pass the scrutiny of classical gatekeepers such as journalists. Images on Instagram and Snapchat help to break down the distance between icons such as celebrities and politicians and promotes the illusion of intimacy with their followers and fans. During the 2016 U.S. presidential campaign, Donald Trump's twitter feed seemed to give its followers intimate access to the candidate's unfiltered thoughts on hot topics. By keeping a direct line to his supporters and, arguably, establishing a para-social relationship with many of them, Trump managed to get elected in the face of overwhelming negativity in the established media.

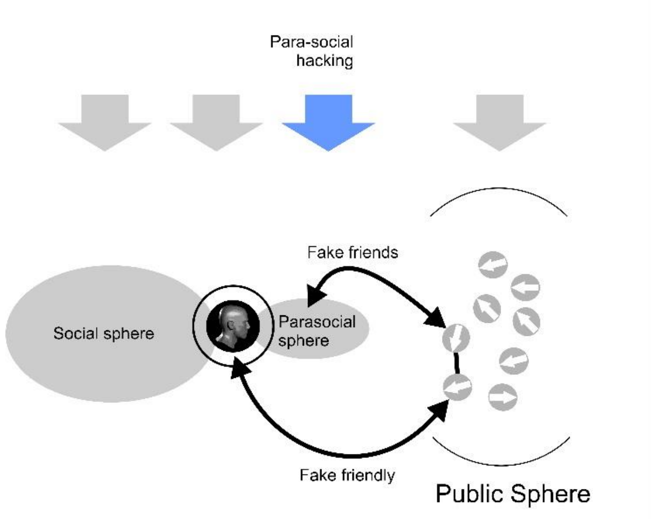


Figure 3.2.3: Para-social hacking exploits our perception of one-sided relationships as reciprocal

Para-social relationships can be exploited in two distinct ways, as well as in hybrid forms. The first, *fake friends*, is the establishment of a para-social relationship between charismatic information influencers and individual members of the target group. The second, *faked friendly*, is the exploitation of networks of friendship, e.g. on Facebook, and the appeal of user-generated content. Users often share content uncritically, sometimes barely reading the headline, thus potentially[98] contributing with their own social capital to the spread of information influence activities. As experts warn, "[c]itizens themselves actively participate in their own disenfranchisement by using social media to generate, consume or distribute false information."[99]

The Digital Caliphate and its 'media mujahidin': According to a Swedish study,[100] ISIS/Daesh propaganda does not follow the classic sender-receiver model but engages peer-to-peer with its target audience. User-generated content is the most important factor for ISIS/Daesh's success on social media. Followers volunteer to produce content, translate articles and spread information, contributing to the overall network. These collaborators are referred to as "media mujahidin". Individuals use multiple channels to increase the speed of the propaganda, while the distributed network structure makes it hard to shut down. Twitter used to be the most common channel for ISIS/Daesh propaganda but during 2016 this changed to Telegram. Facebook and ask.fm are also common platforms. Popular but unrelated hashtags, such as #justinbieber, are often used to make ISIS/Daesh related material appear in popular feeds and catch the attention of new audiences.

A third approach is a hybrid technique that draws on a mixture of *para-social hacking* and *shilling*. Propagandists, be they volunteers or paid, pose as ordinary users on issues in internet fora and comment in seemingly reasonable, decent ways. This helps to build the illusion of a bandwagon. Sometimes, these comments can be less reasonable (see *trolling*). Experts have repeatedly drawn attention to the 'troll factories' maintained by for example Russia and China. Although they do not necessarily try to build friendly relationships, the mere fact that propagandists pose as ordinary people makes their messages less threatening, more authentic and easily shareable. Taken together, para-social propaganda is hard to counter and almost impossible to stop.

Summary

- Para-social hacking refers to the exploitation of para-social relationships where individuals experience one-sided relationships as two-sided
- Social media enables para-social relationships with strangers, celebrities and decision-makers

3.2.4 Symbolic action

Symbolic actions refer to acts that carry symbolic value in the sense that they signal something to an audience to create a response.[101] Through symbolic acts, actions have "the ability to manipulate sense perceptions symbolically [and] permit complex reasoning and planning and consequent efficacious actions."[102] This can be done very crudely (by playing on universally shared symbolic cues such as in terrorist activities) or in a very sophisticated manner (by relating to precise and culturally contingent symbols only relevant to a specific target audience).

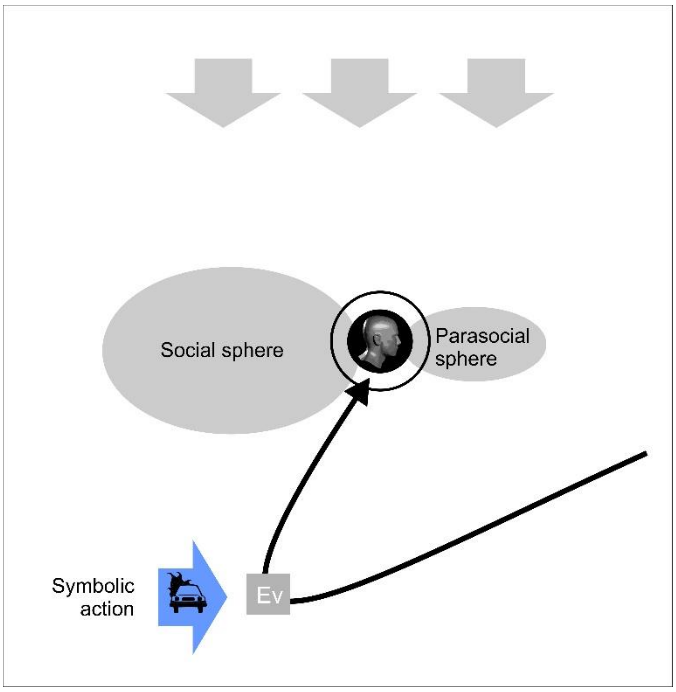


Figure 3.2.4: Symbolic actions insert new evidence into our system of opinion formation

Most actions carry some sort of symbolic value, but not all actions are symbolic actions. In contrast to any ordinary action, symbolic actions are motivated by a communicative logic and a strategic setting. While going to the store may indeed signal that you have run out of groceries, the purpose of your shopping trip is surely not to communicate this to others around you but rather just to stock up your empty cabinets. Conversely, a typical symbolic action used for influence purposes (such as a demonstration, protest, blockade, military exercise, or the public appearance of a politician) is purposefully performed to send a signal to a specific audience. The distinction between legitimate influence through this technique and information influence activities must be identified based on an interpretation of intentions.

Simulated nuclear strike on Sweden:[103] Military exercises provide clear-cut examples of symbolic actions that could fulfil an information influence purpose, conducted on the one hand to train troops and on the other to signal military strength to neighbouring countries. In 2014, Russia held a military exercise in the Baltic Sea, which several experts assessed was simulating a nuclear attack against Sweden at the same time as Sweden hosted a NATO exercise. According to Thomas Reis, security expert at the Swedish Defence University in Stockholm, the intention of this manoeuvre was to signal that "it could get dangerous if you operate military exercises with NATO"[104] in an attempt to influence public opinion against a Swedish NATO membership.

Supporting or disrupting narratives

In contrast to most other influence techniques highlighted in this report, symbolic actions occur in the material environment but have their primary effect in the mediated, or communicated, environment. As 'real' actions and events, symbolic actions can be utilized both to support and disrupt specific narratives. In the case presented above, the symbolic military exercise supported the narrative that Swedish NATO membership is detrimental for Baltic security. In other cases, where an action is less congruent with a narrative, it can instead disrupt by providing evidence that does not fit the story. Symbolic actions can be highly convincing when the symbolic cues of the target audience have been read correctly. As such, they are also hard to dismiss since they are not fabrications or false.

Summary

- Symbolic actions achieve influence by aligning actions in the material environment with symbolic systems, to create powerful signals
- Symbolic actions are effective for supporting or disrupting narratives
- Sophisticated symbolic actions require in depth understanding of target audiences

3.2.5 Disinformation and "fake news"

Disinformation is a technique based on the distribution of false information intended to mislead and deceive.[105] Within disinformation, several subsets of misleading information can be discerned, such as forgeries and leaks (see section 3.2.6) and the recently popularised term "fake news". While contested as an elusive and ubiquitous term with political implications due to its use as a point of attack on politicians and trusted media sources,[106] we refer to disinformation (rather than fake news) in the strict sense of "news articles that are intentionally and verifiably false and could mislead readers"[107] to avoid confusion. Naturally, diagnosing an article as disinformation in this sense requires thorough fact checking and is not something that can be claimed based on ideological disagreement.

While the term fake news is novel, the use of purposely false information disguised as legitimate for deceptive purposes is not – this technique has been used to influence opinions and public perceptions since ancient times. Historically, disinformation exploited public gullibility and played on reader's passions both to promote (or discredit) ideas or individuals, and to attract audience and increase sales.[108] The digitization of news has however fundamentally changed the ways in which disinformation can be used. There are several dimensions of this transformation which highlight how such techniques can successfully exploit various vulnerabilities of the epistemic chain.

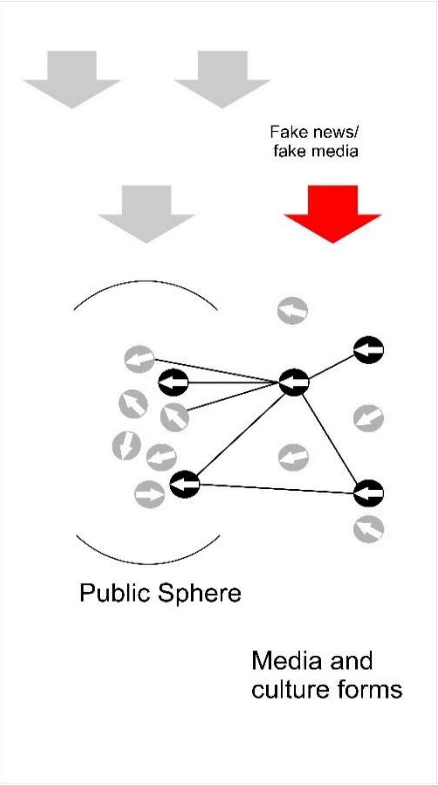


Figure 3.2.5: Disinformation inserts false information to support the trajectories of specific and disruptive narratives

First, the digital environment challenges the connection between news and trained journalists, allowing ordinary people to reach mass audiences through online platforms. To be sure, this is not in itself negative. Rather, inviting more actors to participate in information sharing undoubtedly has a positive democratic effect. On the other hand, it also allows non-journalists to pass off false content as news without information passing through the same mechanisms of scrutiny and review as traditional journalism.[109] In essence, this undermines the institutional media system journalists operate within (*media system vulnerability*).[110] Second, there is an increasingly important financial dimension to disinformation, owing to the financial models of social media and online news platforms, where misleading information is used to attract clicks which generate advertisement revenues.[111] This has historical parallels to the financial models developed by news outlets to cope with the new media landscape brought on by the introduction of the penny press in the early 19th century (*public opinion vulnerability)*.[112] Third, the development of computer technologies and bots increasingly allow fake news to appear legitimate and real by pushing, circulating and engaging with fake stories, attributing them with a false sense of social capital (*cognitive vulnerability*).[113]

Disinformation is easily distributed in the online media environment, especially in social media. The highly polarized and segregated environment on social media (see section 3.2.2), combined with engaging features such as sharing and liking, makes such platforms vulnerable to disinformation by providing ideal conditions for, for example, selective exposure to information, social psychological biases, and confirmation bias.[114] Owing to these conditions, the top-performing disinformation stories generated more engagement on social media than the top real news stories during the lead-up to the U.S. presidential election of 2016.[115] Recent research further suggests that disinformation diffuses "farther, faster, deeper, and more broadly than the truth in all categories of information" due to their perceived novelty and their ability to inspire "fear, disgust, and surprise" in audiences.[116] It may be noted that online news, fake or true, can circulate so widely that they become newsworthy in their own right, thereby reaching traditional or mainstream media as 'going viral' provides a legitimate path to relevancy.

Disinformation appears in many shapes and forms. The following general categories can be discerned:[117]

- *Fabrication:* Fabrication refers to news with no factual basis that is published in a style that misleads the audience to believe it to be legitimate. Fabrications are underlined by explicit intentions of misinforming and deceiving. For fabrications to be believable, they often play on pre-existing narratives and utilize platforms that are either legitimate or have the appearance of legitimacy to the audience.

- *Manipulation:* Disinformation is not only text-based. Visual information, such as photos, video- or audio clips, can be manipulated to deceive an audience and create or support a false narrative.[118] With the development of cheaper and easy to use technology, digital manipulation is becoming more common. It can range from very simple adjustments such as Photoshopping the colour of an item in a picture, to more complex manipulations such as generating convincing audio and video material of digital copies of public figures.[119] In the future, so-called deepfakes, manipulated audio-visual material which is virtually indistinguishable from real material, is expected to increase.[120] Traditional media operate on the lower end of this spectrum, sometimes altering images to remove skin blemishes, for example (and sometimes slightly altering body features, albeit to much critique). The higher end of the spectrum, where techniques such as adding or removing content from photos, merging different photos, or grossly manipulating audio- and video material, is firmly within the confines of what can be understood as illegitimate.

- *Misappropriation:* Misappropriation includes the use of misleading content, false context and false connections.[121] This involves, for example, using unrelated information to frame an issue or an individual in a specific way to fit a narrative, referencing sources that do not contain the alleged information, putting real information into a false context, or using headlines, pictures and other supporting elements that are incongruent with the content. Here, the material may be factual in and of itself, but it is applied deceptively to support or create a false narrative.[122] The fact-checking initiative StopFAKE.org has identified several of the forms of disinformation within the category of misappropriation as the most common form of disinformation in Ukraine.[123]

- *Propaganda:* Propaganda refers to information created with the purpose to influence public perception or public opinions to benefit a public figure, an organisation or a government.[124] In contrast to other categories of fake news and disinformation, propaganda is more often overt in its purpose and focuses on grand strategic narratives.

- *Satire:* Research shows that satire can have a significant impact on public discourse, public opinions and political trust.[125] Satire in the context of disinformation means to ridicule, expose and critique individuals, narratives, or opinions by presenting factual information using humour and exaggeration. Often, this is done without explicit intention to cause harm, but satire has great potential to mislead and deceive its audience nonetheless.[126] Legitimate satirical stories have the primary purpose of entertaining, rather than informing, but the format can easily be exploited to convey specific and controversial ideas to a target audience.[127]

- *Parody:* In contrast to satire, which applies humour to factual information, parody "plays on the ludicrousness of issues and highlights them by making up entirely fictitious [...] stories”[128] with vague plausibility. Parody builds on a shared understanding of the absurdity of its claims between the author and the audience. As such, parodies walk a fine line between the possible and the absurd, sometimes making it hard for audiences to distinguish parody from real information. The well-known parody news site The Onion is still occasionally mistaken for real news, despite the website’s high-profile status.[129]

- *Advertising:* While not traditionally thought of as disinformation, advertising materials are sometimes presented as genuine news, using a misleading format to disguise the commercial interest behind the information. This is referred to as *native advertising*. The contemporary phenomenon of *clickbait* is another type of misleading advertisement, where catchy headlines lure readers onto a commercial site.[130] Such advertisement techniques are generally considered legitimate and accepted as fair means of business, but they can easily be exploited by information influence campaigners, using for example clickbait links similar to commercial links to direct traffic to harmful websites or to spread disruptive or illegitimate content.

Disinformation is generally employed using one of two broad approaches, indicating the different functions these techniques can fulfil in different strategies:[131]

- *Constructive approach:* Disinformation can be used to construct new or alternative narratives or support specific existing narratives by legitimizing false information or evidence, or by replacing certain parts of legitimate stories with false but convincing information. This approach demands a high level of sophistication and can subsequently be hard to spot.

- *Destructive approach:* Disinformation can also be employed in a destructive way, to muddy the waters and to inject noise into an information space to disturb or disrupt legitimate information, or to drown it in a sea of unreliable information to the point where finding legitimate information is like finding a needle in a haystack. Instances of blatantly omitting or removing information in order support a narrative would also fit into the destructive approach.

Major and minor league fakery

In a complex media environment, separating disinformation from real information and legitimate news stories can be difficult, especially as some types of disinformation are based on very minor manipulation. Since factual statements are the backbone of narratives, journalistic craft often involves selection of facts and alignment with a ‘story’. Information influence activities go beyond journalistic standards, however, when influence campaigners try to make the facts suit the narrative:

- by being highly selective with facts, i.e. by leaving out everything that is inconsistent with their narrative,
- by taking facts out of their legitimate context or by exploiting interpretation spaces,
- by making factual statements that are untrue (i.e. lying),
- by creating ‘false facts’ or ‘deepfakes’,
- by denying or limiting attempts to correct, through e.g. trolling or the use of bots (discussed in section 3.2.10 and 3.2.9 respectively),
- by creating fake sites/platforms/media.

The preceding pathways mark a continuum from the minor league of influence to the major league. While the first three approaches involve forms of behaviour that are regularly seen in typical social scenarios, the three ‘higher’ approaches require an increasing investment of resources and are more clearly illegitimate.

Creating false facts: direct action and agitation

Creating ‘false facts’ means to create appearance situation that legitimately allows for a certain interpretation. Burning cars in a city street can be legitimately interpreted as a sign that the police do not control a certain area anymore. However, if the reporter bribed a couple of youths to torch a car to get spectacular video footage, we would speak of a falsehood which has been created with purpose to deceive and mislead.

TV crew bribes youngsters to riot: In 2017, the Independent reported that a Russian television crew attempted to offer Swedish teenagers in the suburb of Rinkeby a bribe of SEK 400 in exchange for ‘some action’, presumably rioting or torching of cars. The event took place shortly after U.S. president Donald Trump had said that Sweden’s “large-scale immigration was not working out”.[132]

Actions in the physical environment can be performed deliberately for the purpose of producing disinformation. Why fabricate a news story around something that has happened when you can just orchestrate an event that supports your narrative and then tell the truthful story of what happened (albeit omitting that the event itself was orchestrated)? The creation of false facts is a particularly powerful aspect of disinformation as it blurs the lines between what's real and what's not by casting doubt not only on the representation of reality offered by media and online, but the actual authenticity of real events. Such strategies have recently been successfully employed by for example ISIS/Daesh and by Russia in the Ukraine.[133] False facts target the epistemic chain by challenging the very notion of evidence based on an observed reality.

Creating fake media

High on the list of major league disinformation is the creation of fake media in its entirety. The digitization of news has allowed not only for individuals to more easily produce and publish news stories but also contributed to making it easier to establish new, alternative media outlets, which are either partially or fully fake. Alternative media in its traditional sense is nothing new. Many alternative media outlets are fully legitimate and differ from established and dominant types of media by, for example, focusing on niche content, presenting alternative perspectives, or by utilizing different forms of production or distribution.[134] There are however ample possibilities to use such structures for information influence. Two examples are:

- *Imitating or replicating legitimate media outlets:* There are examples of influence campaigns where real media outlet platforms have been fully replicated but changed the content to contain fake news. Such imitations or replications can be very convincing for an audience and aim to harness the legitimacy of a real news outlet. Replicated but false news outlets were, for example, utilized in the Columbian chemicals hoax where duplicate websites similar to CNN spread fake news and disinformation to concerned citizens going online to find information on a local chemical spill (unbeknownst to them, the incident in itself was entirely fake).[135] Another example is provided by the fake news article about an MI6 official which used "the same font, format and banner as an authentic Guardian news story, appears under the by-line of a genuine reporter at the newspaper. Whoever created the hoax also secured a plausible-looking domain name, 'theguardıan.com', with the Turkish character 'ı' masquerading as the 'i' in the paper's web address."[136]

- *Creating or redirecting alternative media:* Information influence campaigns can also utilize legitimate alternative media for influence, by for example establishing new media outlets which initially distribute real news stories to build up their legitimacy and reader base whilst simultaneously spreading disinformation or following a specific narrative. The Russian international news outlet Sputnik News, which is often criticized for biased reporting and sometimes even disinformation, is one example of this.[137] Similarly, pre-existing and legitimate alternative media can be redirected using monetary incentives (simply, they can be bought, either overtly or covertly) to spread fake stories or follow specific narratives, as was the case with the Euronews service, which was identified as shifting narratives depending on funding in certain regions.[138] Finally, traffic can be diverted from legitimate sites to fake sites by using alternative web domains and claiming them to be localisations (e.g. by buying the web domain bbc.nu and tricking users into believing that it is a genuine version of the BBC web services).

Summary

- Disinformation is designed to intentionally deceive and mislead
- Digitization has enabled disinformation to spread at an unprecedented pace and the online environment is especially vulnerable to fabricated stories.
- There are multiple types of disinformation, ranging from the slightly illegitimate activity of using facts selectively, to the highly disruptive activity of creating fake news outlets.

3.2.6 Forging and leaking

Forging or fabricating information is an effective tool for attributing disinformation with false authenticity to credible sources, to negatively influence attitudes and behaviour.[139] Forging can for example employ fake letterheads, official stamps and signatures, and can be combined with the appearance of a secret communique being leaked. Sometimes, leaked information (stolen or otherwise obtained by illegitimate means such as phishing emails) can be used as the basis for a forgery, or presented on their own, without proper context. Most effective is probably a combination of both forgeries and leaks, sometimes referred to "tainted leaks".[140] The seeding of false information into a genuine leak not only delegitimizes and distorts the information environment, but also tests "the limits of how media, citizen journalism, and social media users handle fact checking, and the amplification of enticing, but questionable information."[141]

Macron’s email hack: During the French presidential election of 2017, the then presidential candidate Emmanuel Macron had his emails hacked and leaked. The leak was carefully timed to coincide with the start of the pre-election news blackout and was boosted by multiple bots to spread disinformation. According to the BBC, around 47,000 tweets were posted for three hours following the leak, promoting the hashtag #macronleak which trended in France. However, forged emails had been intermingled with genuine emails by Macron’s team beforehand, which questioned the legitimacy of the leak itself. While the leak was quickly managed by Macron and his staff, it threatened to disrupt the election process in the final hours of public deliberation.[142]

The use of forgeries and leaks, and their combination, is like many other techniques listed here, not new– intelligence services around the world have previously utilized such tools as means of information- or psychological warfare.[143] Under the operational umbrella of *active measures,* forgeries and leaks of varying sophistication were popular instruments of the Soviet Union during the Cold War.[144] Forgeries and leaks usually have a twofold goal: they aim at propagating falsehood and discrediting the parties connected to them, and also at “cultivating distrust among citizens and introducing them to question the integrity, reliability and trustworthiness of the media” and of public institutions and figures.[145] Due to the stickiness of fake information,[146] and citizens’ limited ability to themselves verify leaks and forgeries, such techniques can be impactful even when crudely performed.[147]

Tainting the information environment

In contrast to disinformation, which is essentially fabricated stories, forgeries are falsified *evidence* that taints the information environment by fuelling misleading narratives.[148] Forgeries are adapted to fit into and strengthen pre-existing narratives that challenge the mainstream conception of a phenomenon. This is problematic on a different level than disinformation. News stories can often quite easily be debunked by checking the original sources. But when such sources are fabricated or interpreted without context it becomes much harder to disprove them. Such was allegedly the case of the investigation of the downing of the Malaysia Airlines aircraft MH17, where the Russian Ministry of Defence produced what it claimed was radar evidence to disprove the conclusions of the official report.[149]

Shifting burden of proof

Forgeries and leaks effectively cast doubt on the authority and legitimacy of individuals and institutions. Once forgeries and leaks have caught the attention of the public, "there is a burden on the victim of the disinformation to prove that the leaks are not genuine".[150] Depending on the sophistication of the leak/forgery and the amount of trust enjoyed by the victim in the eyes of the public, this may prove both difficult and time consuming. Regardless of the success of the victim to disprove and debunk incriminating information, audiences may, due to backfire effects,[151] still believe the narrative perpetuated by the leak/forgery.

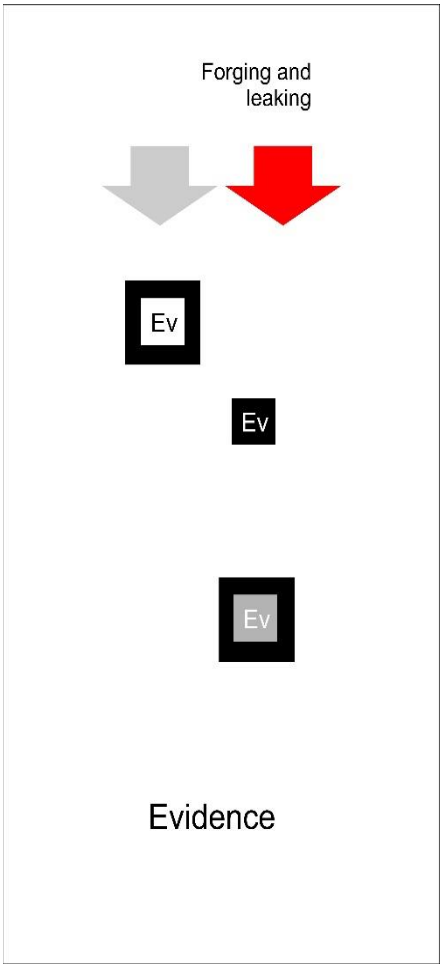


Figure 3.2.6: Sometimes evidence is simply forged. Sometimes, information influence activities leak authentic evidence, such as hacked emails, to journalists. To make matters more complicated, the leak can be tainted, i.e. contains partly authentic, partly forged evidence.

Forged letter from Peter Hultqvist: In the spring of 2016, a forged letter from Swedish Minister of Defence Peter Hultqvist was spread online.[152] The letter used the MoD's official letterhead and a forged signature by the minister. In the letter, the minister offered to sell Swedish manufactured weapons to Ukraine. The origin of the letter remains uncertain but the MoD regard it as part of an information influence activity, according to an interview with SVT.[153] During 2015 and 2016, researchers identified some 26 forgeries in the Swedish information environment.[154]

Policy paralysis

The blurred line between truth and falsehood caused by forgeries and leaks can lead to a situation where decision-making is stifled due to perceived uncertainties in a field. Forgeries and leaks often target specific key communicators and public establishments to undermine their authority and trustworthiness. As such, forgeries and leaks contribute to cynicism about key institutions and "cultivate a fatigue among the population,"[155] leading to what can be referred to as "policy paralysis"[156] – a situation where uncertainty defines the public sphere to the degree that it is difficult for policy makers to perform their functions.

Summary:

- Forgeries and leaks falsely imitate or illegitimately disseminate information for the sake of negatively influencing public perception
- The internet provides a convenient platform to spread and amplify forgeries and leaks
- By blurring the line between truth and falsehood, forgeries and leaks taint the information environment, occupy decision-makers by shifting the burden of proof, and contribute to policy paralysis

3.2.7 Potemkin villages of evidence

In philosophical discussion, the one quality that makes a fact is *truth*. In reality, however, truth often means little more than that a statement is either:

- not disputed (for whatever reason, e.g. because it is dangerous to dispute it or because it is not worthwhile);
- or, in the face of questioning, it is endorsed or guaranteed by a network of interlocking agents who can muster sufficiently convincing *evidence* which, in turn, cannot be challenged (or can only be challenged at a cost disproportionate to the importance of the fact).

Although understanding facts as being true is an important regulatory ideal, for practical purposes it is sometimes helpful to understand facts in the sense of Bruno Latour as being not discovered, but *produced* by "fact-producing apparatuses", i.e. networks of interlocking institutions.[157]

Woozle-effect

In scholarly circles, the so-called Woozle effect, or 'evidence by citation', is well-established. The term goes back to the Winnie the Pooh-stories created by A.A. Milne. In one of the stories, Winnie and Piglet march through the winter forest in search of the mysterious Woozle. Walking around a tree, they hit upon its track. Circling the tree, they find that another Woozle has apparently joined. In the end, Christopher Robin explains that they have of course been following their own tracks.

The key drivers of the Woozle-effect in research, and the vulnerabilities of the scientific process, include:

- The tendency of seeing in data, especially in experiential data such as interviews, what the scholarly community *expects to see*. Since conclusions in the social sciences are often based on plausibility argumentation, it is far easier and safer to make an expected than an unexpected case.
- The assumption that often-cited research equals solid research, that it reflects the consensus in the scholarly community.
- The practice of citing conclusions from studies without disclosing the method by which the conclusion has been reached to an adequate degree (large sample, small sample, experimental design, anecdotal evidence etc.).
- The practice of removing qualifiers in citing research and firming up language, i.e. "this *could have the effect*..." becomes "according to X, this *has* the effect of..."

Thus, within short time and by multiple, self-referential citations, a small-scale exploratory study suggesting a careful conclusion becomes the basis for far-reaching claims such as "It is generally agreed that..." Pseudo-research, or in a benign form 'advocacy research', is easily exploited by propagandists.

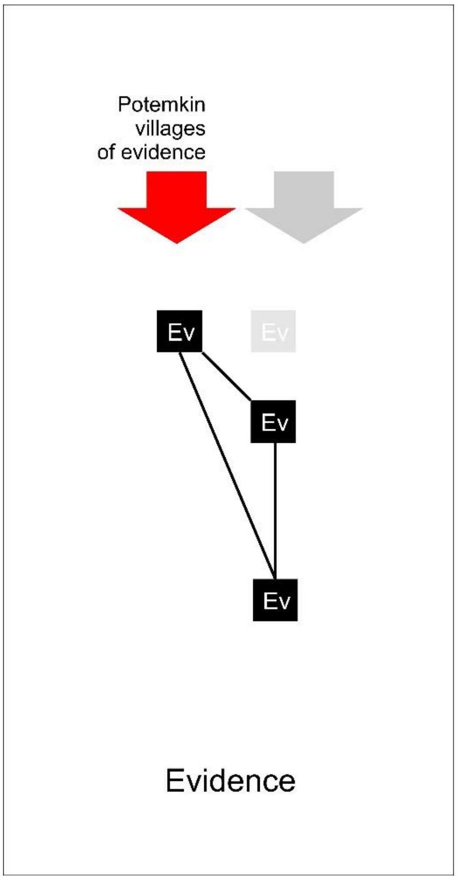


Figure 3.2.7: Potemkin villages are 'fact-producing apparatuses'. By setting up and maintaining bewilderingly complex networks of illegitimate institutions and platforms that reference to each other, influence operators can create (pseudo-)scientific proof for whatever needs to be proven.

Potemkin villages

While the Woozle-effect is due to the way scientific discourses are used within legitimate institutional networks, Potemkin villages are the attempt to set up institutional networks that are controlled by actors conducting information influence. The label of the technique goes back to an expert who described the tobacco industry's approach as setting up "Potemkin villages of science ... a simulacrum of science, but not science itself."[158] Thus, actors with sufficient resources can pursue the strategy of setting up institutions or even complex networks of institutions that serve as endorsers or guarantors. The term ‘village’ is apt, because these arrangements are not only ad-hoc and opportunistic but constitute *structures* and *eco-systems* maintained with a view to long-term efforts.

Tobacco industry’s fake journals: What is humorously described in the Hollywood movie *Thank you for Smoking* is based on real events. The technique of setting up interlocking networks of fake institutions is well-documented for the tobacco industry, which established conferences, workshops, institutes and research centres such as *The Tobacco Institute*, and even created ostensibly peer-reviewed journals such as *Tobacco and Health, Science Fortnightly* and *The Indoor Air Journal*.[159]

In information influence activities, Potemkin villages of evidence will not be limited to the institutional rubberstamping of scientific results but serve a narrative. Thus, a fact-producing apparatus in international affairs may include correspondents and independent journalists, NGOs, refugees, victims, etc., all ‘on the ground’, and all more or less enlisted to serve the narrative. In addition to eyewitness evidence, front organisations, experts, think tanks and research institutions may be utilized to produce studies, working papers, conferences, etc., to solidify the narrative as a product of careful scholarly consideration. Later, articles, op-eds, TV-shows, books, documentary movies etc. solidify the narrative even further: it becomes common knowledge. Furthermore, Potemkin villages also create what is sometimes referred to as *source magnification* – that is, the cognitive response that makes multiple sources, essentially regardless of their quality, “enhance information processing activity and that it is this enhanced processing activity of the message content that mediates persuasion”.[160] Essentially, more sources make the messages more convincing.

Summary

- The Woozle-effect refers to seeing what one is expected to see rather than what is actually there, and assuming that well-cited sources are necessarily true.
- Potemkin villages of evidence refers to the intricate web of deceptive structures that can be utilized as fact-producing apparatuses for specific narratives
- Potemkin villages can consist of complex networks of illegitimate or fake institutions and platforms

3.2.8 Deceptive identities

Persuading experts or celebrities to act as ambassadors for a cause is a common public relations technique. It is accepted, as long as the expert in question really believes in what she is saying or if they genuinely possess expertise. With celebrities, the public should not be misled about the status of the ambassadorship: is it based on a commercial agreement, charity or conviction. In all cases, the key is image transfer and identity work. The expert lends her professional prestige, the celebrity glamour and profile to a cause. In information influence activities, identity can be exploited, however, by three interrelated techniques.

Shilling

A shill is the person in the crowd surrounding the shell-game hustler, the person who makes the game look so easy and wins a lot of money. The shill gives the impression of being independent, but in reality, she is in league with the hustler. The same technique has been applied to other fields, not only casino gambling, but auctioneering online and offline as well as marketing. In marketing, for example, shills write glowing customer reviews or answer user questions (sometimes, in a game of sock-puppetry, they answer their own questions, which they posted under a different identity). In influence operations, shills operate both online and offline. Their key trait is that they *seem* neutral, yet in reality they are dedicated propagandists.

Impersonators and impostors

While shills do not declare their dependencies, impersonators and impostors are deceptive about their true identity. *Impersonators* pretend that they are someone else, i.e. adopt someone else's personal or professional identity. Impersonating a police officer constitutes a serious crime in many countries, but when it comes to other 'official' roles, there is a grey zone, especially in the social media sphere. During a crisis, impersonators might give the impression that they speak officially for a government authority on Twitter, thereby using a false identity to spread disinformation.

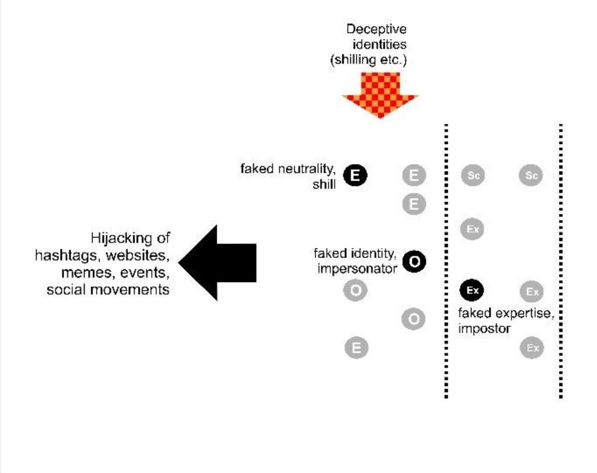

Deceptive identities (shilling etc.)
faked neutrality, shill
E
E
E
Sc
Sc
Ex
Hijacking of hashtags, websites, memes, events, social movements
faked identity, impersonator
O
O
O
E
Ex
Ex
faked expertise, impostor
Ex

Figure 3.2.8: Deceptive identities insert disruptive but convincing elements into the system of opinion formation to create a situation where it becomes difficult to separate reliable elements from untrustworthy ones

Impostors do not pretend that they are someone else, but they pretend to possess expertise or credentials that they do not have. Again, there is a considerable grey zone. Is it acceptable that an expert on alternative medicine is presented as 'Professor so-and-so' on television, although the title is actually an honorary professorship in health journalism, or a PhD in art history?

Hijacking

Hijacking has nowadays become a term that relates less to airplanes and more to hashtags. On a general level, hijacking means that a 'vehicle', be it a website, hashtag, meme, event or social movement, is taken over by an adversary or someone else for a different purpose. The fact that many activities on social media are transparent and participatory makes them particularly vulnerable to organised efforts to hijack.

Capture the flag: Although not an information influence campaign in the full sense, the fate of actor Shia LaBeouf's art project directed against the presidency of Donald Trump provides an illustrative example of hijacking. La Beouf's project began at the Queens Museum in New York. It featured a webcam with live-stream and sympathizers were encouraged to express their rage and desperation by shouting "He will not divide us" into the camera. However, the project was soon hijacked by Trump supporters and others who shouted undesired content, obscenities or simply nonsense – a technique commonly called 'shitposting'. When the museum discontinued the project, La Beouf decided to move the camera to an undisclosed location where it was pointed skyward at a flag with the slogan "He will not divide us". Internet activists then used the patterns of contrails visible in the sky above and other indicators to locate the flag, captured it and replaced it with a baseball cap bearing Donald Trump's slogan "Make America great again". The project was then moved out of the USA.[161]

Summary

- Deceptive identities aim to transfer legitimacy from a legitimate actor or platform to an illegitimate one by shilling, impersonating or hijacking
- Deceptive identities can be first hand (by assuming the role of someone else) or second hand (by being prescribed an identity by someone else – i.e. being cited as an expert in something you are not).

3.2.9 Bots, sockpuppets and botnets

Bots are all around us on the internet – in fact, in 2017 bots were responsible for more than half of all web traffic.[162] A bot (short for robot) refers to a piece of automated computer software that performs highly repetitive tasks along a set of algorithms.[163] The simplest bots are based on a script with pre-determined possibilities, whereas more sophisticated bots can use machine learning and artificial intelligence to process complex requests.[164] Bots with legitimate and beneficial intent are useful tools that can be designed to perform helpful tasks such as collecting data for search engine optimization or market analysis purposes (crawler bots), monitoring system health (monitoring bots), gathering information from different sources to keep users up-to-date on news, events or blogs (aggregator bots) or provide automated customer support (chat bots).[165]

However, the intent of a bot is determined by the individual/organisation behind its creation.[166] While 23 percent of all web traffic can be attributed to good bots, roughly 29 percent can be attributed to their illegitimate relatives.[167] Such bots are used for all sorts of nefarious reasons: spreading disinformation and illegitimate content, price scraping, forum spam, skewing web analytics, distributed denial of service attacks (DDoS), distribution of malware, and other scams. Bots used to support information influence activities can easily mimic organic behaviour to mislead, confuse and influence a public beyond his or her own social network. Some of the more prominent types of bots used for influence purposes are:[168]

- *Hacker bots:* These bots are typically covert bots that users do not engage with. They can be designed as computer-to-computer scripts which attack websites or networks by exploiting security vulnerabilities to inject code, or malware, into the victim's computer or web page. Hacker bots can also be employed to establish botnets by infesting personal computers with malware and contribute to DDoS attacks which can disrupt vital IT-infrastructures. While users rarely interact or come into contact with these bots directly, they affect the online environment in which the individual is active.

- *Spammer bots:* Spammer bots are designed to post content in a forum or commentary section, often including links that may be malicious by for example leading to phishing sites or malware. Such bots are easy to code and are easily scalable, making them ideal for large scale attacks. They are often used as a supplementary tool to spread disinformation and other illegitimate content, or simply to crowd out legitimate content. Spammer bots are however relatively easy to identify and can be banned or removed from platforms. Social media sites sometimes even have sophisticated machine-learning filters to identify spammer bots.

- *Impersonator bots:* Impersonator bots mimic natural user characteristics to give the impression of a real person. Naturally, these are more complex than spammer bots, and require more effort to develop and deploy. Nevertheless, research has shown that impersonator bots can amass significant popularity and credibility among users online by just reiterating simple social activities.[169] They are sometimes referred to as *automated social actors, propaganda bots* or *social bots*, and are commonly used to engage with political content on social media platforms or to scam people. The power of these bots relies on the (often convincing) illusion of a person behind the account, which attributes them with false legitimacy and makes them prime tools for engaging others.

- *Sockpuppets:* While not technically bots, sockpuppets are imposter accounts created by an individual to appear like accounts controlled by someone else (either real people or a made-up characters).[170] In contrast to impersonator bots, sockpuppets are not entirely automated, but partially controlled by a human. Still, using semi-automated sockpuppets, one person can control multiple false accounts to coordinate content across different accounts and platforms to conduct the equivalent of false-flag operations, i.e. covert operations designed to deceive by appearing as though they have been performed by others than those who actually executed them. It is often difficult to distinguish between an impersonator bot and a sockpuppet.

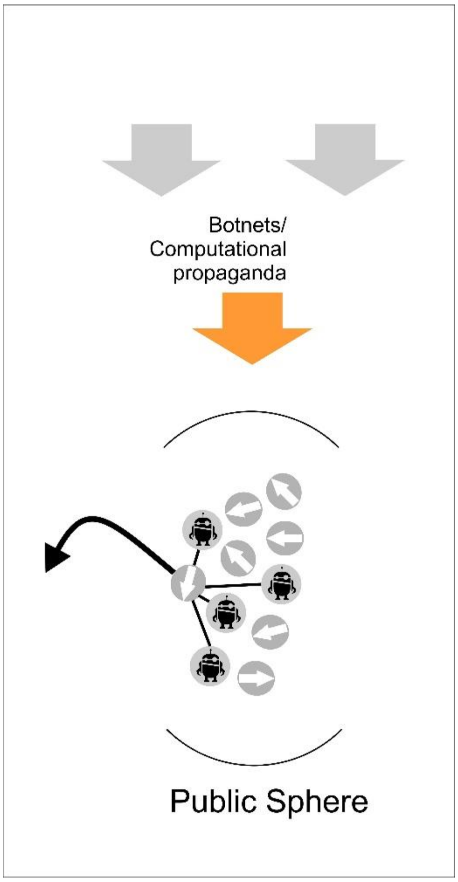


Figure 3.2.9: Bots can be used to generate fake social capital. In the algorithm-driven attention economy, the vote of bots can make the difference between a message being shared or not.

In the context of information influence, bots can be employed in a variety of ways to exploit the vulnerabilities of opinion formation. Some examples of how bots are used include creating social capital, disseminating disinformation, penetrating filter bubbles, and conducting DDoS attacks. Bots are very efficient for amplifying the reach of disinformation online.[171] Spammer bots can effectively and at a low-cost disseminate false information or distribute links directing traffic toward disinformation while also giving a false impression that the content is widely shared among social media users.[172] Similarly, impersonator bots, or sockpuppets, can engage users in pseudo-debates based on disinformation, or get real users to start sharing false content.[173]

Band wagoning

As with social and para-social hacking, social proof is a powerful cognitive mechanism, making it an ideal operational environment for bots.[174] Bots are highly efficient for amassing virtual social capital online to exploit social pressures and cognitive biases[175] by acting as force multipliers, or *false amplifiers*[176] in online discussions. This engineers the appearance of a critical mass of people conforming or sharing a particular view to cause *band wagoning*.[177] Depending on the level of sophistication of a bot, this can be done by, for example, automatically following, re-tweeting, or liking posts from real social media accounts to boost their legitimacy, using spammer bots to reinforce impersonator bots, or using bots to crowd out dissenting opinions to create a false sense of consensus.[178]

Tony Abbot's fake followers: [179] After a suspicious surge in the then Australian opposition leader Tony Abbot's twitter followers, an internal investigation by Abbot's Liberal Party found that someone had been purchasing fake twitter followers for Abbot's account. These followers consisted of spam bots which we later removed by Twitter. An unofficial audit later noted that around 95% of Abbot's 200,000 followers were likely to be fake. According to twitter, there are over 48 million bot accounts on the platform.[180]

Bots can also exploit technical features of social media platforms such as trending algorithms, friend lists, and recommendation features to reach the desired audience. By "riding the wave of algorithmic curation,"[181] bots can repeatedly post and reinforce specific messages via multiple accounts and exploit features such as tags and hashtags to effectively direct content on social media platforms. This can help to push posts to virality, and by that method penetrate individuals' filter bubbles. Impersonator bots that mimic real users are particularly effective in this regard, especially when they manage to get real users to add them as friends or connections on social media.[182]

Botnets

Hacker bots are sometimes used to spread infested code, or malware, to hijack internet-connected devices. A *botnet* is a group of such hijacked devices that can be deployed from a remote location without the knowledge of the device's owner to provide computing resources that can be used for a variety of malicious purposes, such as distributing phishing email and orchestrating distributed denial of service (DDoS) attacks.[183] Bots and botnets can in this way act as force multipliers for other influence techniques, such as performing a DDoS attack on a news website simultaneously as their social media pages are spammed with disinformation, or by distributing phishing emails which are used to obtain information which can be altered and later leaked. During the US presidential election of 2016, for example, multiple "Mirai"-botnets were used to conduct a series of DDoS attacks targeting the candidate's websites to disrupt and disturb the election process.[184]

The tell-signs of a bot

While bots are efficient, they are also vulnerable to exposure. Certain types of bots, such as impersonators, are naturally hard to detect – again, awareness of the technique is the best defence here. Other types have clear tell-signs that help users identify when bots are being used. It is important to remember that the presence of one tell-sign does not necessarily mean that an account is a bot. Still, the presence of multiple signs may be an indication that something suspicious is going on. These are some examples of what to look for when trying to discern if an account is automated or a real user.[185]

- *Account information, user names and handles*: Except for impersonator bots, most bots on social media generate their user names and handles (the account name which, on twitter for example, is used when linking to the user using the @-feature) automatically using random generation. Usernames that seem at odds with other information provided by the user (such as profile pictures) and handles which are made up of seemingly random letters and numbers possibly belong to automatically generated accounts. Further, many bots are created on demand, meaning that accounts can be very young. Thus, having a look at the account's creation date can provide some information about its authenticity. There are however bots that use old accounts that have been purchased or obtained through hacking. Normally, old information is removed from such accounts, so that there may be a wide discrepancy between the date of creation and the first post of the account.

- *Personal information:* Many bot accounts exhibit a high degree of anonymity, with little or no personal information. They are often without a profile picture or using a stolen profile picture. Google image search can be used to validate profile pictures. Impersonator bots are a bit more devious because they try to imitate a real user. Here discrepancies in personal information/pictures can be discerned to tell if the account is a bot.

- *Posting activity:* Spammer bots are often highly active, sometimes with more than 50 posts per day. Look out for accounts with a suspiciously high number of posts per day since creation. Bots that are organised in a botnet can, however, sometimes behave differently, using a critical mass of different accounts with only one or a few posts each to amplify a message. Another approach is to study the time-stamps of individual posts, to see if there are many posts within a very short amount of time, or if the posts are equally distributed throughout a day or an extended period. If activity gaps can be spotted, such as periods of intense posting with long breaks between, the account could be automated.

- *Nature of content:* Bots used to amplify messages by for example retweeting, liking or quoting other users, will share content with other bots. "The timeline of a typical bot will therefore consist of a procession of retweets and word-for-word quotes of news headlines, with few or no original posts."[186] Check user history of posting-patterns and look for the timings of posts to discern if the account seems to be behaving organically or following a script. Also, bots often both like *and* share the same posts, leading to the number of likes and shares being almost identical, which is less usual for posts not amplified by bots.

- *Language:* Most of us are not fluent in multiple and extremely diverse languages. Bots, however, sometimes use automated translation services to disseminate messages in multiple languages (especially commercial bots). Accounts posting similar content in multiple languages are likely to be automated. If you happen to know one of the languages that the account posts in, obvious grammatical errors or incoherent sentences can also indicate automatic translation.

There are also a variety of technical ways to identify bots which are often too complex to benefit everyday users.[187] However, some tools are openly available via online platforms. The website http://botornot.co, for example, allows for automated detection of Twitter bots.

Summary

- Bots are powerful tools of influence due to their ability to exploit social, cognitive and technical vulnerabilities
- Bots are excellent amplifiers for other influence techniques
- Bots are vulnerable to exposure – awareness of tell-signs is key

3.2.10 Trolling and flaming

Anyone who has ever spent time on an online message board, forum or social media platform has probably encountered internet trolls in one form or another – from the obnoxious teen looking for attention, to the scheming cyberbully who knows exactly which buttons to push. A troll is simply a user of an online social platform who deliberately tries to aggravate, annoy, disrupt, attack, offend or cause trouble by posting provocative and unconstructive content.[188] They gravitate to platforms where users interact and thrive on polarized topics and vulnerable groups. The malicious online behaviour of trolls is referred to as 'trolling',[189] or sometimes 'flaming' with the difference that trolling generally targets particularly naïve or vulnerable users while flaming aims to entice any reader in general.[190] Or to put it another way: "[Trolling is] the art of deliberately, cleverly, and secretly pissing people off".[191] In contrast to bots, which sometimes fulfil the same function as trolls, there is a real person behind a troll, who exploits online anonymity to achieve their desired end.[192] Here it is important to remind ourselves that not all people you experience as obnoxious and provocative online are trolls – far from it. The makings of a troll in contrast to someone who just happens to push your buttons is a systematic and intentionally hostility which most ordinary users lack. We distinguish between two types of trolls, where the first is more or less benign but disruptive, while the second is more clearly aligned with a covert purpose:[193]

- *Classic trolls:* In the case of classic trolls, what you see is what you get. Classic trolls are often ordinary people engaged in trolling for the sake of some personal motivation, such as attention-seeking. For this purpose, they employ a variety of techniques, such as 'whataboutism', ad hominem-arguments, straw man arguments (see section 3.2.12), commenting on grammar, derailing the discussion, claiming to be offended, positing offensive and sometime false content, and so on, to provoke others.[194] While often engaging in ideological and political discussions (where invoking emotional responses is easy), classic trolls are not necessarily aligned with any particular ideology or higher purpose, even if they sometimes are. Still, "content is just an instrument in their hands to implement their main purpose [...] to provoke."[195] They can, however, be engaged by actors within the context of an information influence campaign to, unknowingly, contribute to the spread of disinformation.[196]

- *Hybrid trolls:* Hybrid trolls operate under the direction of someone else, most often an organisation, state or a state institution.[197] These trolls fulfil a clear instrumental purpose, often connected to communicating a particular ideology to a particular target audience in a systematic manner. This category includes both the highly organised trolls working in 'troll factories'[198] and individual trolls operating in a less organised manner under the influence of someone else. Hybrid trolls appear the same as classic trolls (making it notoriously hard to tell them apart) but they have different underlying intentions intimately connected to information influence.[199] Hybrid trolls are strategic tools of influence employed consciously to achieve a purpose whereas classic trolls most often are not.

Invoking powerful emotions

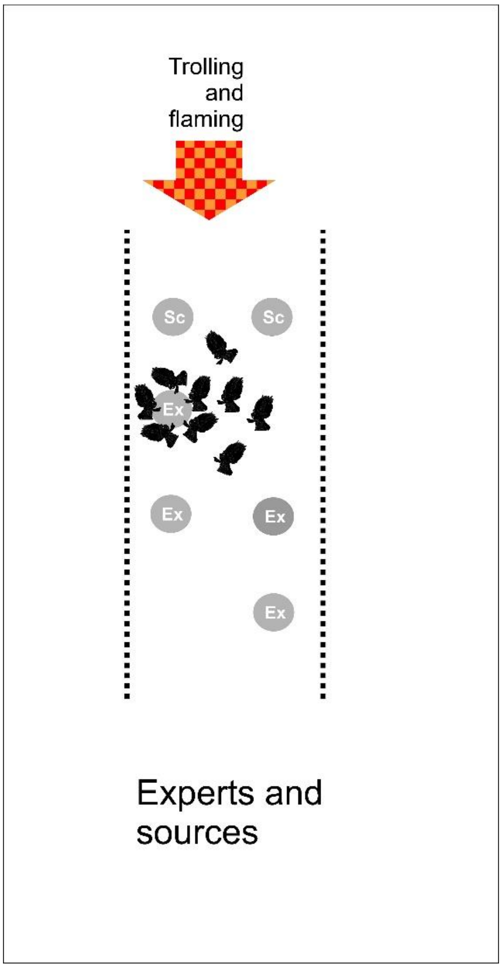


Figure 3.2.10: Although elites and officials can be targeted by trolls, as in principle everybody can be, one of the more dangerous long-term effects of trolls is that they drive reasonable experts out of debates.

Emotional engagement is the bread and butter of trolls. As mentioned earlier, emotions tailor to a cognitive vulnerability – powerful emotions supress reason and restraint in an individual by triggering certain heuristic patterns.[200] Trolls, be they classic or hybrid, exploit this mechanism by provoking as much as possible. This can have one of two effects: either targets or audiences are successfully provoked in which case they are likely to engage in a meaningless discussion controlled by the troll, or they are discouraged and leave the discussion completely. Either way, the troll wins – either their audience is lured into a destructive discussion or they are deterred from discussion. In contrast to the more sophisticated process of cognitive hacking where a thorough understanding of the target audience is required, trolls often use trial-and-error to find the path of least resistance to provoking their targets. If one provocation does not work, just try another one until you hit the sweet spot.

Reinforcing polarisation

To provoke, trolls often adopt extreme positions in political issues and play on pre-existing polarisations in society to dichotomize comment sections, chat rooms and forums. This can be achieved by, for example, criticising political figures, inserting value loaded trigger words into posts, and constructing a context based on misinformation or disinformation. Such strategies can easily polarize what was originally an uncontroversial debate. This phenomenon is observable in the Swedish online environment where trolls often enter discussions on any given political issue and shift the discussion toward immigration issues (a case of ‘whataboutism’).[201] Regardless of the relationship (or lack of thereof) between the discussed issue and immigration, this triggers an emotional response in some users with pro-migration views who feel the need to respond to the false accusation made by the trolls: this is known as “feeding the troll”. At the same time, users with anti-immigration sentiments side with the trolls, owning to a sense of confirmation of their world views by the troll’s argument (confirmation bias). Suddenly the discussion has derailed into a heavily polarized conversation about immigration between highly emotionally engaged users. Swedish Radio P1 recently reported on this type of behaviour on the Italian news outlet La Stampa’s social media platforms.[202] Whenever La Stampa publishes news about Sweden, their social media commentary sections are derailed by trolls who criticise Sweden’s immigration policies, repeating the same reoccurring talking points over and over again. While the messages are in Italian, La Stampa have traced the troll comments to servers in Ukraine and Lithuania. La Stampa considers the motivation of these hybrid trolls to be to target the image of Sweden as a successful country, wishing to reframe Sweden as a country in crisis due to immigration.[203] Once users have been emotionally engaged in the topic they are easily primed into adopting extreme positions by trolls who masterfully exploit their cognitive and social vulnerabilities.

‘50 cent party’ in China: China operates perhaps the most comprehensive troll army there is – the so-called ‘50 cent party’. The ‘party’ was rumoured to consist of as many as 2 million individuals commissioned by the government to post on social media. While research is inconclusive as to the size of the ‘party’, researchers have shown that China engages government employees in social media posting, publishing around 448 million posts per year.[204] In contrast to many other trolls, the ‘50c party’ does not engage in argument, but rather pursues a strategy of *cheerleading*, where positive messages about China are used to crowd out dissenting opinions. Simply put, it is a form of ‘trolling by distraction’.[205]

While some trolls only aim to provoke in order to initiate an emotional response in their audience, examples from China show that trolls sometimes opt to distract than to provoke in order to quell emotions. Trolls can, especially when organised, easily derail a discussion by going “off-topic” and fill-up a forum or commentary section with unrelated content. This exploits cognitive heuristics such as availability bias to distract the audience from the issue at hand.

Summary

- Trolls are online users or bots who deliberately try to invoke emotional reactions in their audiences by posting unproductive or provocative content
- Many trolls are attention seekers, but some trolls have political agendas and are used as tools of information influence campaigns
- Trolls can contribute to the polarization of debates, the silencing of opinions, and distraction from important topics, in order to disrupt public opinion formation.

3.2.11 Humour and memes

In *Winning the Information War*, Lucas and Pomeranzev point out the feature that distinguishes contemporary propaganda from its often 'dull and stiff' predecessor during the Cold War: "Modern [...] propaganda is cleverly targeted, technically adept and cynically fact-free. It is also enjoyable."[206] In a world of instant gratification, capturing the attention of your audience is increasingly important, be it for commercial, political or hostile purposes.[207] One way of doing this is by using humour and 'memes' as tools of communication. Content or symbols that previous experience indicates as either good or bad instinctively capture our attention, due to cognitive cues which direct our attention to potentially rewarding or harmful features in our surrounding environment.[208] Humour, by drawing on culturally shared positive references, does exactly this. As a good laugh is both physically and mentally rewarding, humour is also apt for keeping our attention.

Contemporary public spaces are filled with entertainment. Videos, cartoons, comics, TV shows, movies, funny images, apps – humour is an integral part of culture and a "communication tool that entertains, attracts attention [and] serves as light relief".[209] But, at the same time, humour can serve to covertly manipulate and influence 'hearts and minds' to advance goals and agendas not recognized by the audience. Humour is particularly powerful in this regard as it

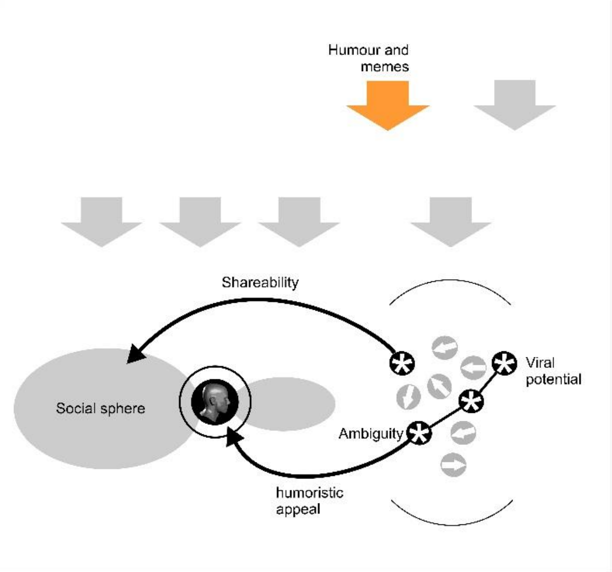


Figure 3.2.11: Humour and memes are useful tools of influence for their potential to spread virally and legitimize edgy narratives

leaves people with their guard down. Humour often plays on people's experiences and perceptions to provide alternative versions of reality, or to *shift frames*.[210] Simply put, humour influences ideas, ideas form beliefs, and beliefs generate and influence political positions and opinions (eventually translated into behaviour).[211]

On the internet, the humorous pictures with catchy taglines known as 'memes' have recently been recognized as powerful instruments of influence.[212] VICE Magazine even described them as "the stuff by which reality is made and manipulated".[213] Memes, however, precede the rise of the internet – the word itself, coined by Richard Dawkins, refers to "units of cultural transmission"[214] which can be spread from person to person within a culture with the aim of conveying meaning. As such, memes are more than just funny pictures, they are carriers and disseminators of cultural ideas and practices.[215] The potential of memes for information influence activities primarily rests on three factors:

- *Their immediate humoristic appeal:* We are hard-wired to direct our attention towards expressions of humour, especially humour that fits within our cultural environment.[216] Memes that utilize visual images combined with catchy taglines thereby have an immediate appeal to us. Even if we do not agree with the message perpetuated by the meme, we may still recognize its humoristic value and thereby perpetuate it. This makes memes hard to avoid.

- *Their viral potential and natural process of variation:* Memes are designed to be shared, "taking advantage of the fact that propaganda spread through interpersonal ties is more successful than that generated by a top-down apparatus".[217] This implies that memes have the potential to be accepted as accurate representations of reality due to the fact that they are popular (availability bias) and come from within one's social network (social confirmation). Proponents of *memetics* (the study of meme theory) further theorize that a meme's viral potential is contained in the evolutionary nature of memes.[218] In this sense, memes are understood as cultural equivalents of genes that evolve. Memes go through a process of variation (there are multiple popular memes), mutation (memes are changed and adapted over time to fit current context), competition (only popular memes can go viral), and inheritance (the cultural ideas of popular memes are passed on).[219] Memes are thereby resilient, adaptable and infectious. Once a meme has started to gain traction, the ideas contained in it are hard to stop.

- *Their ambiguity:* Memes do not have a fixed meaning. Rather, they are *living structures* residing in our brains to which physical images or symbols are connected.[220] This attribute memes with high degrees of ambiguity – they can mean different things to different people, and in different cultural contexts, or even at different points in time. In this sense, memes are difficult to control, and their effects are tough to predict.

Legitimizing controversial ideas – the subversive buffer

From a communication perspective, humour generates new meanings which are less offensive than the sum of their content. This is made possible by what is referred to the *subversive buffer*.[221] As humour is perceived as a communicative act with the purpose of entertaining rather than conveying serious ideas, "a humorous message is understood to be less offensive than a non-humorous one".[222] This makes humour and memes ideal for legitimizing 'edgy' or controversial ideas and narratives. The subversive buffer of humour has limited scope and will not withstand any type of content (which was painfully demonstrated by the public outcry that followed American comedian Gilbert Gottfried's joke about the terrorist attacks on 9/11 merely weeks after the incident),[223] but as long as a meme or any other instance of humour is perceived to be within general limits of what's acceptable to an audience, there are no limits to the content.

'Swedistan'-meme on 4chan: The meme-driven information operation referred to as "Operation Swedistan" was initiated by anonymous users of the controversial imageboard 4chan in 2017 in response to news that some Swedish public schools had banned the Swedish flag claiming it is potentially offensive to ethnic minorities. The operation attempted to, via the use of humorous memes, build opinion for replacing the Christian cross on the Swedish flag with an Islamic crescent to make Sweden a more inclusive county for Muslim immigrants. According to the controversial conservative online media webpage Squawker "[the operation] follows the usual 4chan strategy of attempting to trick extremist liberals into siding with a progressive cause that would seem ridiculous or even outright appalling to the average person".[224]

Infiltrating social spheres

We tend to accept information and views from within our social sphere more readily that equivalents coming from the outside of our social sphere.[225] In a so-called non-personal setting new "perspectives [have] to be established in a typically longer communication process" than in a personal setting.[226] Memes are particularly effective tools of influence in this regard as they are spread and shared widely on social media, which provides an entry point for externally generated ideas into an individual's social- and para-social spheres. The message of the meme reaches its target from within his or her sphere while the sender is displaced.[227] The evolutionary spread of memes contributes not only to their survivability but also to their believability and legitimacy with their target audience.

Internet insurgency

Memes can be considered tools of internet insurgents since they by ridiculing, humouring and joking “weaken monopolies of narratives and empower challenges to centralized authority”.[228] This implies that memes essentially are cultural artefacts without authority in and of themselves but with the potential to challenge the establishment by ridiculing and making fun of it. The successful use of memes by Trump supporters in the 2016 U.S. presidential election testify to the power of memes to subtly challenge the establishment by systematically ridiculing ‘crooked’ Hillary Clinton (combining humour with malign rhetoric in the form of name calling, see section 3.2.12), “[chipping] away at the wall built around institutional authority”.[229] Unawareness of the discursive power of memes and humour makes them especially useful for information influence.

Summary

- Humour is a powerful tool to attract attention and raise sensitive issues
- Memes are more than just funny pictures; they are cultural ideas that are hard to contain once they have started to spread
- Humour and memes are useful for legitimizing edgy or controversial ideas and opinions

3.2.12 Malign rhetoric

The distinction between rhetoric and dialectics as well as the differentiation between philosophy and sophistry dates back to classical antiquity. For at least 2,000 years, humans have been aware of ruses that actors in the public sphere employ in order to persuade an audience, as opposed to logically convincing it. Today, a certain amount of rhetoric is accepted in public debate, as long as it makes an issue more comprehensible and engages people. Malign rhetoric, in contrast, exploits the often-fragmented nature of conversations in the contemporary public sphere, especially on social media, in order to muddy the waters and frighten away the reasonable.

A complete account of moves and ruses of malign rhetoric would require a separate report, but some examples drawn from classical propaganda techniques as well as a contemporary overview of techniques illustrate the way malign rhetoric undermines rational, problem-oriented debate:[230]

- *Name-calling:* As one of the classical propaganda devices, name-calling discredits an adversary with words “calculated to lower their prestige of credibility”.[231] Such words naturally vary – in the context of propaganda in the early 20th century, derogatory words such as 'fascist' and 'warmonger' were common. Today, other words may have stronger negative connotations online.

- *"Whataboutism":* Whataboutism is a double rhetorical move. It consists, firstly, of deflecting an argument by drawing attention to a similar phenomenon which ostensibly does not get as much attention. For example, if a journalist talks about the dangers of right wing-activism, the whatabouter raises the question "… and what about left wing-activism?" In a second move, the person in question can be implicitly or explicitly accused of hypocrisy. It should be noted, however, that the what about-question can be legitimate; it just should not be used to deflect a debate.

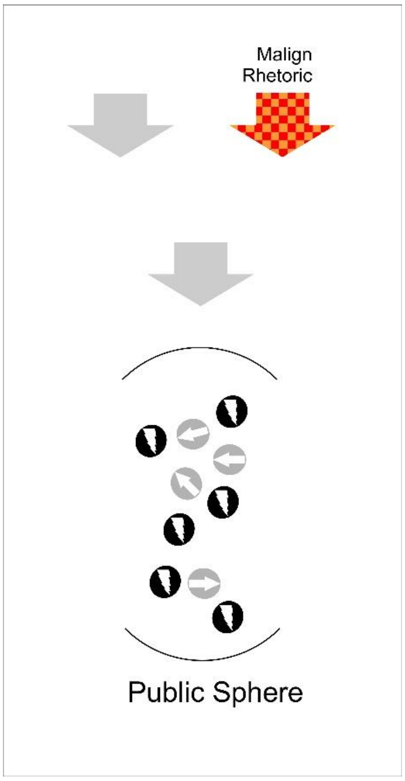


Figure 3.2.12: Malign rhetoric is not constructive in finding solutions, but it can be entertaining as it caters to the 'Schadenfreude' of others being humiliated. Another effect is that it poisons the opinion climate so that reasonable debaters shy away from the issue.

- *Ad hominem:* Ad hominem means 'against the person' and is the terminus technicus for attacks directed against the character or personality of an adversary in a debate. An ad hominem-attack shifts the attention away from the argument to the speaker (by using for example name-calling, see above) or by calling into question the adversary's personality (serial liar, mentally ill). It should be noted, however, that doubting a person's competence to have an opinion on an expert matter is not necessarily an ad hominem-attack (e.g. if a geneticist accuses a layperson of not correctly understanding epigenetics that is not necessarily ad hominem).

- *Gish gallop:* The gish gallop, named after the creationist debater Duane Gish. It consists of overwhelming an opponent with a flood of arguments, facts and sources, many of which are spurious. The force of the Gish gallop does not derive from the strength of the argument, but from the number of 'facts' cited; it is proof by verbosity.

- *Transfer:* Transfer is a rhetorical ruse which can be used in both positive and negative ways. Positively, it is aimed at "unjustifiably associating an argument with an admired category of thought, such as religion or patriotism".[232] Negatively, it conversely associates arguments with controversial and negatively connoted categories of thought.

- *Strawman:* The 'strawman' is a form of ad hominem-attack and consists of attributing extreme and untenable positions to the adversary which the adversary does not hold. It is a technique of arguing against a false adversity, or a real adversity temporarily ascribed false characteristics that are useful for building an argument.

Summary

- Malign rhetoric captures lingual ruses aimed at undermining reasonable and legitimate debate and silencing opinions
- Different rhetorical ruses are used for different purposes, but the general intention is to undermine, delegitimize and distract adversaries

3.3 Influence stratagems

The inventory of techniques presented so far maps out influence techniques in accordance with how they exploit the vulnerabilities of opinion formation in Western society. But specific organisations will rarely face one technique in isolation. As influence campaigns have become more complex and indirect, organisations in reality face *arrangements* of influence techniques in time and space. And although the possible combinations of influence techniques are theoretically infinite, it is becoming increasingly clear that information influence campaigns tend to apply similar combinations, dramaturgical patterns, or 'playbooks'. The arrangement of techniques in dramaturgical patterns is the reason why the detection of and engagement with information influence should take place on the level of the *chain-of-event*.

In our terminology, these dramaturgical arrangements are referred to as 'stratagems'. The word stratagem is obviously related to the term strategy; *strategema* means "act of generalship". In contrast to strategy as a more neutral term, stratagem invokes the meaning of trickery, of outwitting an adversary, with the Trojan Horse perhaps the first documented ruse of war. A common translation of stratagem is *ploy* or *ruse*. The term 'ruse de guerre' is legally defined in the Hague Convention, which expressly prohibits *perfidious ruses* such as e.g. booby-trapping religious objects but allows deceptive manoeuvres like setting up dummy forces or moving landmarks to confuse the enemy. Psychological warfare is also considered legitimate.

In order to illustrate how techniques are arranged in dramaturgical patterns, we discuss several examples of typical stratagems that are used in information influence campaigns.

Please note that this section explores some examples of stratagems but is not exhaustive. Nor are they presented in any particular order. This section may be developed further in future revisions of this report as new stratagems are defined.

3.3.1 Black propaganda

Influence operations came of age in the two World Wars, where they were used by all sides. The distinction between white, grey and black propaganda is widely recognised by scholars and practitioners. *White propaganda* clearly states its source, openly pursues its objective and is by and large factually accurate, like e.g. an airdropped pamphlet that warns the inhabitants of a city that the area will be bombed. *Grey propaganda* partly or fully obscures its origin and/or objective and is not always totally accurate. *Black propaganda* actively aims to deceive the target audience about the origins of the information.

During the Second World War, the Japanese distributed leaflets on the Philippine Islands that looked like they were official information material handed out to US soldiers. The leaflets informed US military personnel that Philippine women would readily sell their bodies for little money or scraps of food but warned that many females were sick with venereal diseases. US soldiers were thus advised to limit themselves to respectable wives or virgins as sexual partners. The Japanese-produced leaflets were not aimed at US soldiers at all, of course, but were designed to enrage the civilian population against the occupiers.[233]

This example illustrates how the techniques of *sociocognitive hacking*, *forging and leaking* as well as *direct action* can be combined into a single

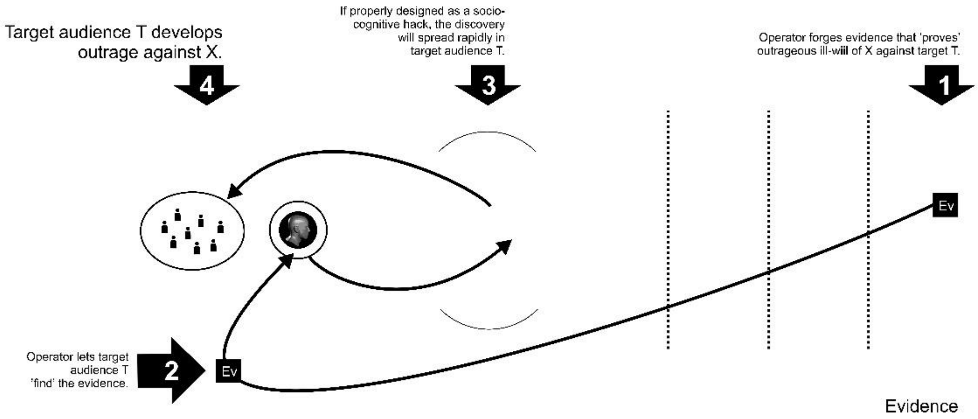

Target audience T develops outrage against X.
4
If properly designed as a socio-cognitive hack, the discovery will spread rapidly in target audience T.
3
Operator forges evidence that 'proves' outrageous ill-will of X against target T.
1
Ev
Operator lets target audience T 'find' the evidence.
2
Ev
Evidence

Figure 3.3.1: Black propaganda spreads clandestine disinformation in order to deceive a target audience.

stratagem. The sharp end of the operation lay with the sociocognitive hack appealing to strong emotions (outrage, hate) and fundamental human motives, but the operation would not have been possible without forging what looked like a directive to U.S. soldiers.

3.3.2 Point and shriek

A stratagem that organisations frequently encounter in dealing with outspoken activists is *point & shriek* or its variant, *bait, point & shriek.*[234] In its contemporary form, the stratagem takes advantage of the extreme sensitivity to perceived 'injustice' in certain groups in contemporary society, groups which are often also highly active on social media, and well aware of the viral dynamics of the hybrid media space.

In Iraq in 2006, U.S. Special Forces – or, to be more precise, critical observers of the war in the Arab World and the West – were subjected to an attack by *point & shriek*.[235] On March 26, 2006, during an operation named 'Valhalla', a battalion of the 10th US Special Forces Group engaged a death squad of the Jaish al-Mahdi (the 'Mahdi Army' or JAM). In the ensuing firefight, U.S. and Iraqi troops killed a number of enemy fighters, captured 17, destroyed a weapons cache and rescued a badly wounded hostage without taking any serious casualties. The real engagement, however, took place afterwards. By the time the U.S. and Iraqi forces had returned to their compound, "someone had moved the bodies and removed the guns of the JAM fighters back at their compound so that it no longer looked as if they had fallen while firing weapons. They now looked as if they had fallen while at prayer. Someone had photographed the bodies in these new poses and the images had been uploaded to the web, along with a press release explaining that American Soldiers had entered a mosque and killed men peacefully at prayer."[236] A U.S. military stated later: "Literally they had their

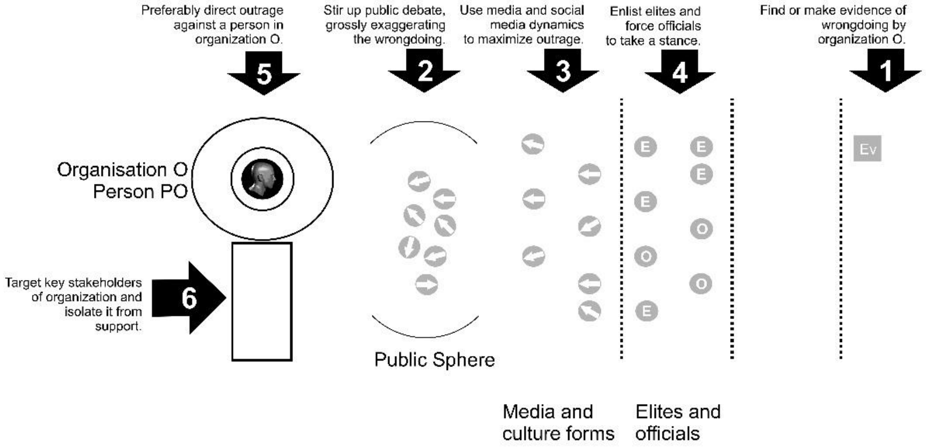

Preferably direct outrage against a person in organization O.
5
Stir up public debate, grossly exaggerating the wrongdoing.
2
Use media and social media dynamics to maximize outrage.
3
Enlist elites and force officials to take a stance.
4
Find or make evidence of wrongdoing by organization O.
1
Organisation O
Person PO
E
E
E
E
O
O
O
E
Ev
Target key stakeholders of organization and isolate it from support.
6
Public Sphere
Media and culture forms
Elites and officials

Figure 3.3.2: Point and shriek aims to both divert and create outrage by stirring up public debate through exaggeration and provocation.

story, their propaganda, out on the wires before the assault force was back at the compound."[237] In contrast, it took the US. Military three days to confront the adversary's narrative. Fortunately for the U.S. forces, the unit was accompanied by a combat camera unit that had filmed the operation and could prove the U.S. version. Nevertheless, the ensuing investigation lasted about 30 days, during which the entire battalion was 'benched': a major success for a humble information influence campaign.

The stratagem of point and shriek normally constitutes arrangements with *sociocognitive hacking* and *fake news* as their starting point, followed by the attempt to boost the more or less fake story using every means available, with virality and worldwide trending hashtags the Holy Grail. The Jaish al-Mahdi communicators knew that populations in the Arab World and in the West would be outraged about a cowardly attack on unarmed men in prayer, and that the news media would take the story up eagerly, especially when backed up by visual evidence. Information influence activities may be expected to distort and amplify innocent or ambiguous statements in ways that are impossible to predict.

3.3.3 Laundering

Information laundering refers to a collection of techniques aimed at de-contextualising information so that it can be used in disinformation campaigns. By laundering is meant something similar to money laundering – the process of legitimizing dirty money by obscuring its illegal origins – adapted to the information sphere. In this case, the process can involve taking genuine information and laundering it through intermediaries to become false information or taking false information and laundering it through apparently credible news sources to make it appear legitimate. Intermediaries cite these sources with minor changes to the text each time, gradually peeling away the original context and meaning. A hostile information source can then refer to these intermediaries as its sources for the falsified quote. In addition to making information appear more legitimate, information laundering contributes to *source magnification*.[238] It is common for this technique to mix *fake news, misappropriation, manipulation* and*fabrication* with *Potemkin villages* and *Woozles*, via mis-translation and de-contextualisation of the original quotes.

For example, in July 2015, Sputnik published the article "Sweden Getting Ready to Fire Missiles at Russian Troops from Gotland Island".[239] An article in the Russian online news outlet Regnum is cited as the source, though the original quotes can be traced back to an earlier Swedish Radio article. The article cites the governor of Gotland as saying, "From Gotland we could, for example, fire missiles and cover our ships sailing towards St. Petersburg". The ambiguity in this sentence, when removed from the context of the original interview, is utilised to give the impression that the statement is about Swedish aggression toward Russia rather than vice versa, as was the intention. A French version of the Sputnik article went with the headline, “Swedish Official: The Island of Gotland is Well-Placed to Bomb Russia”. Information laundering in this example supports disinformation by claiming that Swedish military based on Gotland are not there for defensive purposes, but to launch an attack on Russia.

A second example mixes these techniques with *hacking*. When hackers accessed servers at the Climatic Research Unit of the University of East Anglia shortly before the Copenhagen Climate Summit in 2009, they leaked thousands of emails and documents (technique of *leaking*, although it is not entirely clear whether the hackers acted in conjunction with climate change-deniers). However, it was mainly two sentences that climate change-deniers focused on. In one email, climate scientist Kevin Trenberth wrote: “The fact is that we can’t account for the lack of warming at the moment and it is a travesty that we can’t.”[240] In another email, climate scientist Phil Jones explained that he had used a ‘trick’ employed earlier by climate scientist Michael Mann ‘to hide the decline’. Various investigating bodies found the statements to be innocent in their rightful contexts. However, taken out of context, supported by the machinery of climate change-denial, employing *malign rhetoric* and following the *trolling*-technique of focusing on *persons* (as opposed to institutions), they were sufficient to fuel a worldwide controversy, suggesting a conspiracy driven by climate scientists.

3.3.4 Flooding

A stratagem counter influence experts increasingly draw attention to was highlighted in the 2016 *Firehose of Falsehood* report on Russia during events in the Ukraine and Crimea.[241] Fig. 3.3.4 illustrates the idea of creating confusion and consequently inaction by overloading principal actors and decision-makers with information. The firehose of falsehood relies at its core on *fake news*, underpinned by *Potemkin villages of evidence*, *laundering* and *woozles*, or not genuinely sourced at all. *Symbolic action and agitation* can be used to create ‘fake evidence’ on the ground to complicate the situation even further. Experts identify four distinct features of this information: high-volume and multichannel; rapid, continuous, and repetitive; lacks commitment to objective reality; lacks commitment to consistency. [242] The overall impact of this collection of techniques is of *flooding* the information space.

The key principle of flooding is that it is not credibility, but rather complexity-based. Quantity over quality, speed over credibility and repetition instead of argumentation are the maxims. The flooding stratagem overflows the target media system and public sphere with *high-volume, multi-channel disinformation* in order to overload fact-checking capacities and to crowd out emerging plausible narratives. Botnets are utilized to amplify the flood; armies of trolls organised in so-called troll factories play their part, too. Although the stratagem exploits cognitive biases that give a credibility-advantage, such as being first with a story, those behind the campaigns do not seem too concerned about being 'found out' in the end,

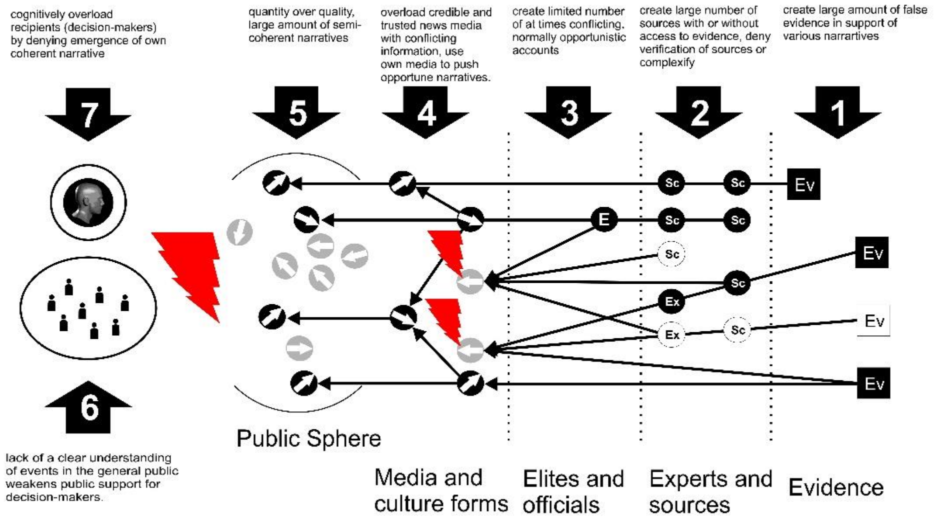

cognitively overload recipients (decision-makers) by denying emergence of own coherent narrative
quantity over quality, large amount of semi-coherent narratives
overload credible and trusted news media with conflicting information, use own media to push opportune narratives.
create limited number of at times conflicting, normally opportunistic accounts
create large number of sources with or without access to evidence, deny verification of sources or complexify
create large amount of false evidence in support of various narratives
7
5
4
3
2
1
Sc
Sc
Ev
E
Sc
Sc
Sc
Ev
Sc
Ex
Ex
Sc
Ev
Ev
6
lack of a clear understanding of events in the general public weakens public support for decision-makers.
Public Sphere
Media and culture forms
Elites and officials
Experts and sources
Evidence

Figure 3.3.4: Flooding creates an information overload so that actors cannot reasonably assess which information is credible or not. A variety of techniques are inserted at different points of the epistemic chain to create maximum confusion.

nor do they seem to care about consistency: even media outlets directly controlled by hostile actors often offer accounts at odds with each other or opportunistically change their narrative in accordance with what runs best.

3.3.5 Cheerleading

While flooding is a stratagem primarily aimed at another societal system, *Cheerleading*[243] is a stratagem normally employed towards one's own society. Flooding operates with a large but limited number of more or less spuriously substantiated *narratives*, pushed in multiple channels and amplified by *botnets*, in order to overload the target system's capacity to differentiate credible from incredible. Cheerleading, in contrast, is not concerned with credibility per se. It utilizes *social dynamics*, especially the *spiral of silence*. In a way, Cheerleading means *echo-chambering* an entire society.

Experts have estimated that the Chinese government utilizes government employees as a social media army, posting millions of social media posts per year.[244] Posing as average citizens, the members of the so-called 50c army (so-called because they are rumoured to be paid 50 cents per comment) secretly insert comments into the stream of social media, as if they were genuine opinions. According to expert estimations, the activities of the 50c-army amount to one in every 178 posts made on social media in China.[245]

When evidence about the existence of the 50c-army emerged, observers assumed that the 'internet commentators' were utilized to engage in arguments with regime critics online, to defend the regime, its leaders and policies. Newer research making use of leaked emails shows, however, that the 50c-army consists mainly of identifiable government officials, and that the Chinese government's approach mirrors the *oblique* influence strategic.[246] King et al. argue: "In contrast to prior claims, we show that the Chinese regime's strategy is to avoid arguing with sceptics of the party and the government, and to not even discuss controversial issues. We show that the goal of this massive secretive operation is instead to distract the public and change the subject, as most of these posts involve cheerleading for China."[247]

The subtlety of the Chinese approach – as it is reconstructed by Western researchers, but not denied by the Chinese government either – lies in the fact that criticism on social media is permitted, as long as it does not lead to mobilization. As soon as collective action is in the air, the government 'jumps in' and 'derails the conversation' by flooding it with other news and entertainment activity: an approach termed 'trolling by distraction.'[248]

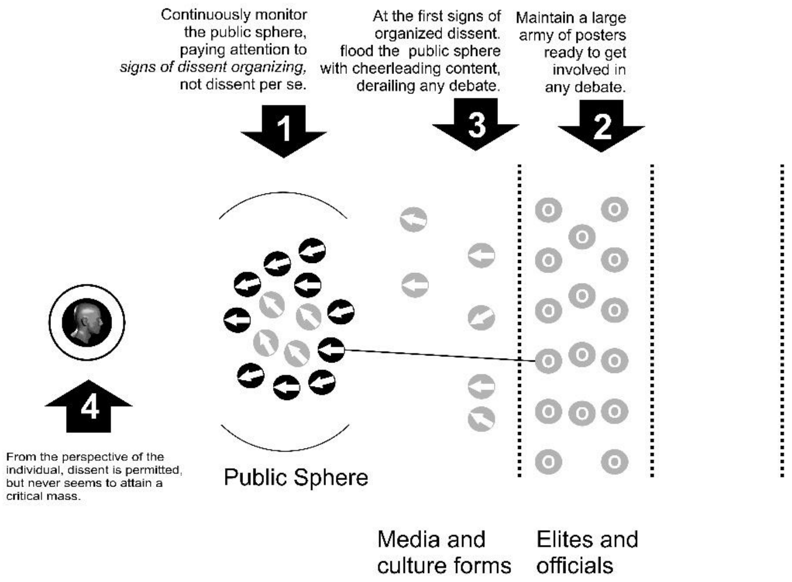

Continuously monitor the public sphere, paying attention to *signs of dissent organizing,* not dissent per se.
At the first signs of organized dissent, flood the public sphere with cheerleading content, derailing any debate.
Maintain a large army of posters ready to get involved in any debate.
1
3
2
4
From the perspective of the individual, dissent is permitted, but never seems to attain a critical mass.
Public Sphere
Media and culture forms
Elites and officials

Figure 3.3.5: Cheerleading crowds out legitimate opinions by overflowing the information space.

3.3.6 Raiding

In military terms, a raid refers to a sudden attack without the intention of holding ground but rather with the purpose of attacking and retreating in order to surprise, confuse and exhaust the enemy. This description largely captures the influence stratagem of *raiding* as well. Raiding is an influence manoeuvre which rallies and musters forces in order to coordinate an attack on an information arena, be it online or in the real world, to crowd out and silence opinions and exhaust others via disruption. This can be done with a variety of tools, such as *spammer bots*, *trolls*, *peer-to-peer engagement*, or *symbolic actions*. Raiding can also be complemented with, for example, *DDoS attacks* to shut down online platforms, forcing users to the channels which are about to be raided.

Swiftboating is an example of a raiding strategy. During John Kerry's 2004 presidential campaign, his political enemies spread doubt about Kerry's military record. A book, television advertisements, a special interest lobby group and personal testimonies flooded the information space. The allegations were proven false, but nonetheless affected the election process. The stratagem was used primarily as an offensive tool to spread doubt, and by the time the claims were debunked, it was too late. A characteristic of raiding is that it does not seek to hold its ground, merely create an information *surge* for a short space of time.

Albeit not a case of information influence *per se*, a popular internet meme labelled *Pool's closed* arose in the aftermath of a massive online raid in 2006 called The Great Habbo Raid which further exemplifies the stratagem[249] The raid was organized by the hacktivist *Anonymous* via the image board *4chan* and consisted of flooding the then popular social networking site Habbo Hotel (which was structured as a game environment) with dark-skinned avatars in business suits who obstructed entry points to popular hangouts on the platform, such as the pool, denying other users access. The raid was launched after rumours had spread that the social media site had moderators prone to racial profiling against dark-skinned avatars. Detailed instructions of how to participate were distributed across a variety of internet platforms to rally support for the raid, which caused the platform to temporarily shut down. A typical method for a raid is for individuals to meet on one platform in order to organise themselves, and then migrate suddenly to another platform, thereby overwhelming it.

3.3.7 Polarisation

Polarising a debate involves influence activities that support two opposing extremes of a specific issue to force mainstream opinions into one of the two extremes. This is achieved by supporting pre-existing extreme perspectives, by for example using *social* and *parasocial hacking*, *spamming, trolling, leaking,* spreading *fake news* and using *memes* in support of both sides. It can also be done by *creating* extremes, by for example having an actor adopt multiple identities to 'perform' a debate that appears genuine (*sock puppetry*). This can be mixed with techniques such as *impersonation, trolling*, *disinformation* and *social hacking* to undermine public opinion formation.

According to the recent indictment pertaining to Russian interference in the U.S. 2016 Presidential election (see 3.4 below), the Russian Internet Research Agency used fake accounts to set up a series of "thematic group pages on social media sites" addressing a range of sensitive political issues in order to spread disinformation and influence the voter behaviour of specific audiences.[250] Two such groups on Facebook were "Heart of Texas" and "United Muslims of America". These groups together attracted hundreds of thousands of followers online. Both groups were utilized to spread disinformation and political propaganda, albeit with opposing positions – while Heart of Texas followed a pro-Trump and anti-immigration narrative, United Muslims of America was inherently pro-Hillary and supportive of the advancement of Islamic culture in the U.S. This contributed to polarizing the political debate. The Facebook-groups were also utilized as platforms for organising political rallies for opposing groups at the same place and time, fuelling opposition in the physical environment as well as online. Essentially, deceptive identities online were used to mask that the same actor was fuelling both sides of an unproductive debate, to provoke and polarise.

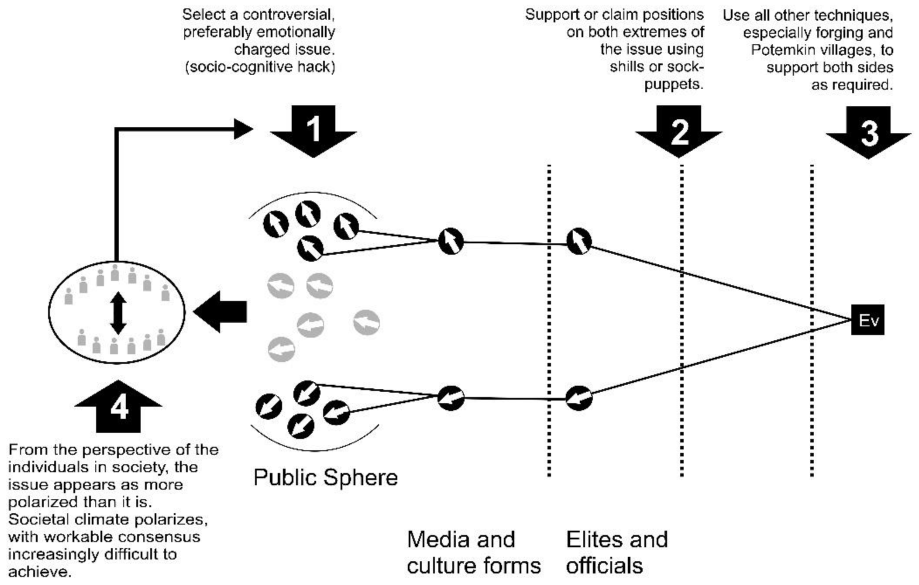

Select a controversial, preferably emotionally charged issue. (socio-cognitive hack)
Support or claim positions on both extremes of the issue using shills or sock-puppets.
Use all other techniques, especially forging and Potemkin villages, to support both sides as required.
1
2
3
4
Ev
From the perspective of the individuals in society, the issue appears as more polarized than it is. Societal climate polarizes, with workable consensus increasingly difficult to achieve.
Public Sphere
Media and culture forms
Elites and officials

Figure 3.3.7: Polarisation targets controversial and emotionally charged issues to support both extremes in the debate, widening the gap between different positions.

3.3.8 Hack, mix, release

The stratagem *hack, mix, release* captures complex influence operations combining the hacking of IT-systems with the tainting of that information to undermine or falsely incriminate individuals or institutions. Typically, the stratagem is initiated by *hacking*, or using *botnets* for *spear phishing* or injecting malware, to obtain information, such as internal documents, emails or classified information. This information can then be diluted with *forgeries*, i.e. tainted, to later be *leaked* to the public. A combination of *fake* *news*, *bot-supported* social media distribution, *memes* and *trolls* can also be used to amplify the effect of the stratagem.

Hack, mix, release is effective due to its utilization and manipulation of credible information. It is highly deceptive since released information can be cherry-picked from a much larger set of information to fit into a specific narrative, and since it places the burden of proof on the victim who is forced to dedicate substantial resources to disprove or contextualize the tainted leak. As indicated by earlier examples, the hack, mix, release stratagem has recently been frequently employed in relation to national elections. Both in the US and in France, spear phishing was employed to access and leak information to influence public opinion negatively for Hillary Clinton and Emmanuel Macron respectively.[251] In the Macron case, the taint was part of a counter-influence strategy aimed at discrediting the leak once the hack had been discovered. Forgeries were quickly and efficiently identified by journalists who had been made aware of the ruse, effectively minimizing the effects of the stratagem.

3.4 Information influence activities in practice: a case study

February 2018's FBI indictments against 13 Russian citizens and 3 Russian agencies provide one of the clearest examples of how information influence activities can look. It should be made clear that this case study does not pass any judgement on the veracity of the indictments. It rather uses this case to explore information influence techniques on the grounds that it is one of the few examples where intelligence communities have made detailed evidence available in the public sphere. It is furthermore connected to a crucial societal institution – the democratic election process – in this case the 2016 US Presidential Election.

The "Mueller indictment"[252] provides a striking example of how the application of multiple information influence techniques in a stratagem may look in practice. The comprehensive indictment furthers the conclusions of the earlier Intelligence Community Assessment (ICA) report "Assessing Russian Activities in Intentions in Recent U.S. Elections"[253] and details a coordinated and multifaceted influence campaign which combined covert intelligence operations with overt propaganda efforts by state-funded media as well as paid social media users.

The indictment centres upon the Kremlin-linked Internet Research Agency (IRA), also known as the "troll factory".[254] With its "strategic goal to sow discord in the U.S. political system"[255], the IRA initiated its operation in 2014 by conducting a comprehensive target audience analysis and psychographic mapping of US social media sites dedicated to politics and other sensitive issues. This was done to discern metrics such as reach, audience engagement, frequency of posts, nature of content etc. This baseline social media intelligence provided a point of departure for the operational design which largely revolved around information influence activities on social media platforms, with the apparent objective of "impairing, obstructing, and defeating the lawful governmental functions of the United States by dishonest means in order to enable the Defendants to interfere with U.S. political and electoral processes, including the 2016 U.S. presidential election".[256]

Social media accounts designed to attract U.S. audiences and appearing to be operated by Americans were created by employees at the IRA. Deceptive identities were used to create the illusion of para-social groups in the form of grassroots movements and political organisations. These accounts were used to set up groups and pages dedicated to "divisive U.S. political and social issues". According to the indictment, stolen identities of real U.S. persons were also used to gain legitimacy on social media. Some accounts even mimicked official accounts, such as the fake twitter account "Tennessee GOP" which claimed to be controlled by a political party. To hide the origin of the account, an elaborate IT infrastructure with VPN tunnels was established. The resulting mixture of techniques drew upon psychographic hacking to create para-social spheres which were reliant upon deceptive and forged identities.

Operators of these accounts were instructed to create "political intensity through supporting radical groups, users dissatisfied with [the] social and economic situation and oppositional social movements""[257] by for example posting divisive content, spreading false news stories and disinformation, trolling, using memes and malign rhetoric and so on. Social media ads were also purchased, using false U.S. identities, to promote and spread such content. The IRA has been shown to advocate for opposing views on pages related to a diverse variety of sensitive political issues, such as immigration, equality and religion. This is typical of the sockpuppet technique. The indictment estimates the IRA-controlled groups, at the time of the U.S. election, attracted hundreds of thousands of genuine followers online. This reach was supported by just 90 staff at the IRA, who worked in 12 hour shifts 24/7 with a $2 million budget. Supposedly 40 people worked on US political influence at any one time, with the target of producing at least 80 comments and 20 shares a day.[258] This suggests a constant flow of interventions in US domestic political debates numbering several thousand each day, from just one unit of one organisation.

Once legitimacy had been established for the deceptive social media accounts, groups and pages, the IRA used them to organise and coordinate political rallies, "while pretending to be U.S. grassroots activists who were located in the United States but unable to meet or participate in person."[259] Such rallies included as diverse rallies as the "March for Trump"-rally on June 25 2016, and the "Trump is NOT my President"-rally on November 12th.[260] Polarisation was a key stratagem, on the grounds that sowing confusion and uncertainty was more effective than supporting a single candidate or a specific ideology. The false social media accounts were used to coordinate such events with both legitimate social media users running political groups and pages and local campaigners. The use of rallies and meetings points to a form of agitation that turns an advantage on social media into a real-world advantage.

Keeping in mind that these allegations have yet to be proven, the suspected information influence campaign described in the Mueller Indictment aptly shows how multiple techniques such as social-, psychographic-, and para-social hacking, symbolic action, disinformation, deceptive identities, bots, sockpuppets, trolls, memes, agitation, and malign rhetoric were combined into a polarisation-based stratagem. While the impact and effect of the alleged information influence campaign remain to be determined, the indictment provides some of the clearest evidence available of how a complex information influence campaign targeted at disrupting democratic institutions could play out over an extended period of time, systematically exploiting cognitive, media system and public opinion vulnerabilities.

4. Counteracting information influence activities

The preceding chapter focused on identifying information influence activities. It presented strategies, techniques and stratagems commonly employed in information influence campaigns to enable communicators in public sector organisations to consider some of the ways in which they may be targeted. The concluding case study based on the 2018 Mueller indictments demonstrates how a vulnerabilities-based approach to identifying information influence activities can help to interpret complex arrangements of techniques. The purpose of this chapter is to form an overview of the counter measures that have been suggested in the literature, and to present our suggestions for preparatory and counter activities at the communicator's level. The perspective therefore shifts from an understanding of the techniques used to exploit societal vulnerabilities, to the perspective of communicators and how they may respond to this threat.

4.1 Approaches to countering information influence activities

As may be expected from a broad and multidisciplinary literature, there is not one answer to the question of how information influence activities should be countered. Rather, academics, organisations, institutions and authorities offer a plethora of suggestions, instructions and advice to consider based on different aspects of available research. Some of these are politically motivated insofar as they seek to persuade policymakers to take a specific course of action. Others are speculative and dramatic, helping to draw attention to the potential severity of the threat. Others still concentrate purely on one narrow area or technique of information influence activities, such as civil society and journalists joining forces to create fact-checking resources. Subsequently, there is no conclusive best-practice or systematic repository of successful counteractions.

To position our approach, we first identify several ideal-type approaches that may be found in the literature.[261] Without making any specific claims regarding their appropriateness for a governmental policy, these approaches provide an overview of the many different lines of action available in response to information influence. They have jointly informed our choices in this chapter, and the reader may wish to consider which approaches are most relevant for their context when assessing our suggestions.

- *Civil society approach*: The civil society approach suggests that individuals and civil society should be empowered to resist information influence activities. In recognizing that state-driven responses could suffer from a range of compromising biases,[262] this approach argues for a bottom-up method focused primarily on mobilising civil society to reject hostile influence activities. Civil society is therefore expected to share the burden of raising awareness of citizens, educating for improved source criticism, identifying disinformation, and supporting a resilient, robust and reliable media system.[263]

- *Facts first approach*: This approach suggests that fact-checking, debunking and deconstructing disinformation should constitute the core of countering information influence activities.[264] The prevalence of this approach can be seen from the recent surge in fact-checking initiatives by both states, media institutions and civil society actors. In Sweden, for example, some of the largest media outlets have recently joint up to collaborate in a fact checking initiative in preparation for the national election in September 2018.[265] The simple premise of this policy is that disinformation should be countered by ensuring that citizens have access to facts.

- *Collaborative approach*: As information influence techniques often demonstrate, there is strength in numbers.[266] The collaborative approach advocates for the establishment of more national and international networks to jointly increase our capacity to counter information influence activities by, for example, supporting information and experience sharing, establishing financial structures to scale up capacity development, and to improve coordination between like-minded actors and institutions.[267] The collaborative approach has in part been adopted by, for example, the EU which has recently set up the *High-level group on Fake News and Online Disinformation*[268] as well as the *East Stratcom Task Force* which acts as a hub for initiatives against disinformation.

- *Counter narrative approach:* Counter narratives are designed to provide alternative and believable frames of reference in order to prevent hostile narratives from gaining traction within a population. The counter narrative approach suggests a focus on defining, formulating and perpetuating strategic narratives.[269] With reference to information influence activities from violent extremist groups such as ISIS/DAESH, this could include, for example, formulating a narrative which communicates your values to a target audience subjected to the hostile narrative, tailored to expose the vulnerabilities of the hostile narrative.[270] As noted by the EU, this is a less aggressive approach than counter-propaganda (see below), which needs to be constructed from research and analysis to discern how narratives are to be designed.[271]

- *Counter-propaganda approach*: Reminiscent of the Cold War, the counter-propaganda approach advocates for tactical and strategic messaging conducted by state institutions to "push back against [specific messages]" and "forcibly preventing the adversary's ideas from circulating within one's own society"[272] by messaging of some kind. While authors such as Cull highlight that counter-propaganda approaches sometimes also include "broader responses which seek to alter the environment in which the messages circulate"[273], we have instead included such extreme activities under the hard-liner approach (see below). Rather, we see counter-propaganda as the attempt to directly counter information influence activities using targeted tactical and strategic messaging on a state-level.

- *Raising the threshold approach*: Raising the threshold means dis-incentivising information influence activities by, for example, establishing resilient government structures with high legitimacy in society, actively pursuing and punishing the perpetrators, as well as strengthening the population's vigilance and will to resist. The aim is to raise the costs (economic, political, labour) associated with information influence activities, so that actors think twice about the value of such activities. This broadly mirrors Sweden's approach to total defence.[274]

- *Ignoring approach*: In contrast to the counter narrative approach, the ignoring approach suggest an "inward-looking and protective" [275] strategy, which simply seeks to minimize the reach of information influence activities by denying them attention and not engaging with them. This approach places faith in the democratic institutions of society, choosing to disregard information influence activities altogether. This does not, however, exclude activities in the civil society approach[276] (see above) but such activities would be based on a rationale related to public education, and not directly related to countering information influence activities.

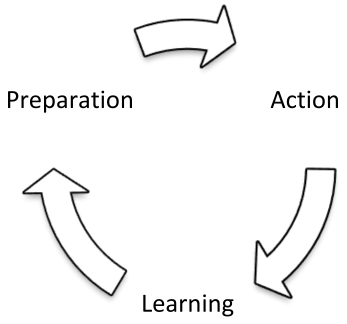
- *Regulatory approach*: Many issues related to information influence stem from the legal grey zone within which it operates. The regulatory approach advocates for minimizing this grey zone by establishing clearer and stricter regulations.[277] This could include, for example, imposing regulation on social media companies, as has been hotly debated in the U.S.[278] or by changing legal structures to more accurately account for acts of information influence, as is being suggested by president Macron in France, where a recently presented proposal suggests to "grant judges emergency power to remove or block certain content deemed to be "fake".[279]

- *Hard-liner approach:* Finally, the hard-liner approach suggests fighting fire with fire. This controversial approach includes measures such as, for example, imposing strict regulations to social media companies, internet providers and media actors; hitting back with proactive information influence activities; and the possibility of "jamming, corrupting, degrading, destroying, usurping, or otherwise interfering with the ability of the [hostile actor] to broadcast and disseminate their messages"[280] by means such as aggressive lawfare, kinetic operations, electronic warfare and cyber operations.

The listed approaches differ broadly with regards to their understanding of the problem, their preferred counter activities as well as the actors leading the response, be it individual, society, organisation or state. While some, such as the ignoring approach, suggest a relatively passive and unobjectionable line of action, others, most notably the hard-liner approach, are basically a call to arms which prescribes dramatic and controversial action. All have merit in their own right at the broader policy level, but none offer a complete and actionable approach for communicators which is suitable in relation both to the communicator's role, the legitimacy of communicative activities, as well as the democratic values we wish to preserve. Rather, a synthesis of approaches which situates legitimate counter activities in the context of public sector communicators is the approach we have used to design the countermeasures outlined here.

4.2 The communicator's mandate

Against the backdrop of these many understandings of how to counter information influence activities, this chapter outlines our suggested approach. This approach borrows from the ideal-types above, from some more than others, and situates counter measures at the level of the communicator as well as within a Swedish context to make recommendations that are practical and actionable. Our approach departs from an understanding of the communicator's mandate (introduced below) and is divided into three interconnected stages:



- *Preparation*: What can be done pre-emptively to minimize the effects of information influence activities and boost immunisation against interference from a long-term perspective?

- *Action*: What are a communicator's options when it comes to responding to information influence activities in the short- and medium-term?

- *Learning:* How do you utilize lessons learned and experiences to improve future activities?

The aim is to provide a full spectrum approach to countering information influence activities, without being overly prescriptive or dogmatic. To be clear, this chapter does not provide *the* approach to countering information influence activities, but rather outlines a set of suggestions drawn from the literature to help support each individual communicator's own decisions. It should be noted that while there is little specific and tested advice available, especially at the communicator's level, the suggestions made below are inevitably normative insofar as they represent lines of actions and activities we believe to be appropriate rather than scientifically proven best practices. There is no definitive playbook. It will be up to each organisation and individual communicator to reflect on our recommendations and assess how they can be integrated and applied within their operational context.

Before elaborating on counter activities, it is important to clarify that different techniques will only be appropriate in certain contexts. Communicators have different mandates depending on where they work. It may be appropriate to mount a political defence in some organisations, where as in others it is not. It falls to the responsibility of the communicator to determine which activities are most appropriate depending upon their mandate. For example, in section 4.4, we outline four stages of response, some of which are appropriate for all organisations (Assess, Inform), and some of which may only be appropriate in certain instances (Advocate, Defend). In these later examples, we urge caution.

Communicators play an important role linking government to the people. Effective communication is about strengthening that link and should never be used to limit open democratic debate. One might argue that the exploitation of vulnerabilities outlined in chapters 2 and 3 is best countered by renewed confidence in public institutions. In this respect, a key goal of counter influence is to restore trust in organisations that are being undermined through illegitimate means. Much of the public debate on information influence activities has emphasised the importance of source criticism among the public as a remedy to disinformation. However, we also believe that the responsibility of public sector communicators is to communicate in a legitimate manner. This strengthens the bond between society and citizens, making citizens more resilient to information influence activities.

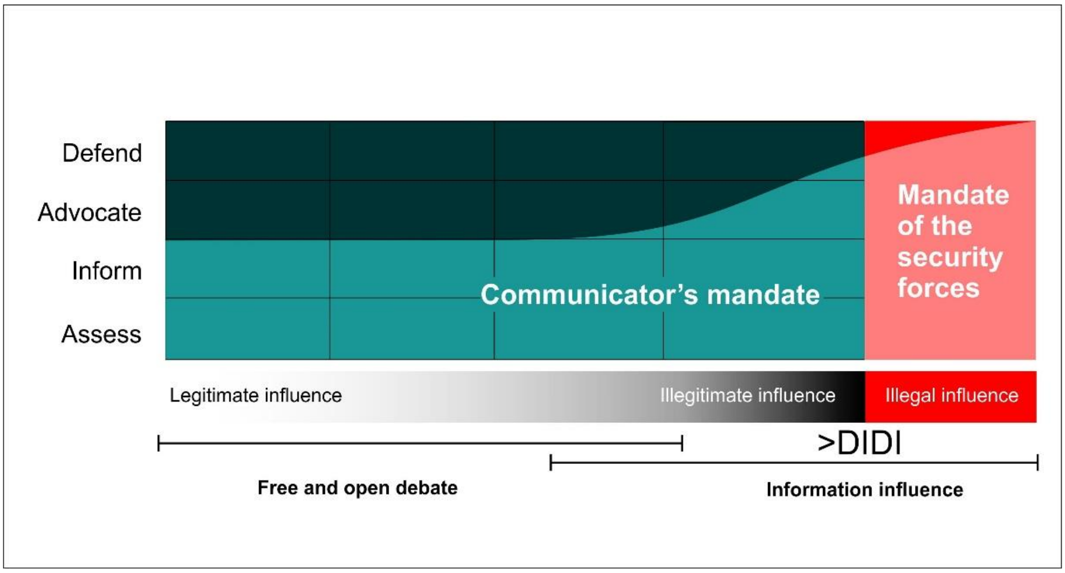


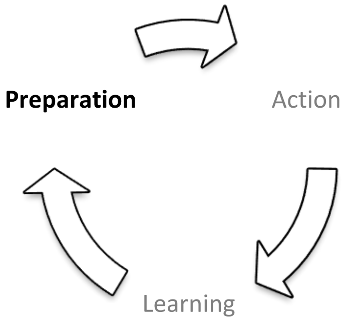
Figure 2.2: A model of the communicator’s mandate in relation to the different levels of counter action.

Some important and challenging considerations should be made in relation to the communicator’s mandate. First, information influence activities are difficult to prove. It may not ever be possible to determine whether foreign actors are behind an influence activity. In many cases, even after such influence has been identified, security services prefer not to release the information into the public domain. This presents a challenge to communicators who believe that they may be facing information influence activities. The Mueller investigation into Russian meddling in the 2016 U.S. presidential election is a rare example of where intelligence has been used to publicly confront a hostile actor rather than to give an edge in future counter-intelligence activities; even then, it has taken the vast resources of the U.S. government a great deal of time and political energy to build a case. In general, public sector communicators must use their judgement about the techniques being used against them on the basis of their mandate to retain the trust of the general public. The less savoury the techniques used, the more reasonable it is to apply a counter influence technique in response, even if information influence activities cannot be conclusively proven.

Second, the problem of proving information influence activities is complicated by the fact that the exploited vulnerability is usually located in domestic debates. Foreign actors seeking to exploit societal vulnerabilities typically identify domestic proxies to work through, perhaps by joining a group while obscuring their identities and interests, supplying that actor with funding, social proof, technological support, or narratives. In most cases, these proxies have the legitimate right to communicate on these issues. Denying these actors the right to speak on these issues is therefore not an option. However, use of information influence techniques by any actor can merit a measured response of the kind outlined in this chapter. The legitimacy of the techniques used by all sides therefore becomes an important consideration. A legitimate domestic actor who uses forged documents to falsely discredit somebody can be legitimately met with a counter-influence technique such as debunking. These choices must ultimately fall to the assessment of the communicator, based upon their mandate.

DIDI (as introduced in chapter 2) is one simple diagnostic tool for determining whether communication techniques fall outside of what might be considered legitimate communication. It is far from perfect. Besides thinking about identifying information influence activities, it is worth considering how one's own response might fit within such a tool. Many communicative techniques are legal but are not considered legitimate for respectable actors to undertake. It's worth noting, for example, that in a recent New York Times article about the use of bots to boost celebrity followers on twitter, most of the celebrities asked about the techniques denied knowledge or preferred not to comment on a technique that could be conceived as 'shady'.[281] Some techniques are simply more legitimate than others. Although communicators in different organisations have mandates that grant them use of different levels of counter-influence techniques, the approach we advocate is of only using techniques that one would be willing to discuss transparently at a later stage. Techniques that could potentially embarrass an organisation, or that you are not comfortable discussing openly, are not recommended.

4.3 Preparation



Preparation is perhaps the most essential part of any type of contingency or crisis management plan. Preparation allows for quick and efficient management of issues, even unforeseen or unexpected ones, by establishing structures, processes, functions and mindsets at individual, organisational and national levels.[282] Just like immunisation can be more effective than treatment of some diseases, preventing problem is more effective than having to react to it unprepared.[283] Preparation is, however, a time consuming process that demands resources, especially when it concerns complex and unpredictable phenomena such as information influence activities. The following sections introduce a selection of tools for communicators to use to prepare themselves, their organisation and their colleagues in relation to information influence activities.

4.3.1 Societal and organisational preparedness

The main component of preparedness will always be knowledge of the threats and vulnerabilities that one faces. Such activities can be difficult to identify, since they "straddle the preconceived divisions between ... cyber-attacks, political communication, election interference and disinformation".[284] It can be challenging to distinguish between regular political debate and hostile efforts to manipulate those debates, and that is arguably the point of threats located in the 'grey zone'. Consequently, it is difficult to relate local issues to these broader geopolitical controversies, and to engage people who already have their working days filled with challenging tasks. In many respects, preparedness is as simple as spreading a warning about the kinds of threats that are out there, how they function, and how they might impact upon the work of individuals and organisations.

Organisational preparedness begins with an awareness of the threat. It is also closely associated with an awareness of risks and vulnerabilities, which should already be part of the organisation's strategic planning. Crisis managers and communication officers may have already developed contingency plans that overlap with, or could be readily adapted to, countering information influence activities. Knowledge of information influence techniques should be cascaded through organisations to ensure that everybody is aware of their modus operandi, in the same way that there is a general awareness to look out for unattended baggage at airports in relation to terrorism. In other words, part of the aim of preparedness should be to create a mindset of vigilance, which can in turn support the creation of an environment of resilience. This should not lead to undue anxiety or add to the workload, but rather adds an additional dimension to the understanding of warning signs for an organisation.

There is a great deal of support available for organisations who wish to prepare themselves for coping with information influence activities. At the level of international institutions, some major initiatives have been designed to study how information influence functions. For example, the European Centre of Excellence for Countering Hybrid Threats is a membership association co-sponsored by the European Union and NATO, and is based in Helsinki.[285] The NATO Strategic Communications Centre of Excellence, based in Riga, also produces analysis of information influence activities.[286] In Sweden, the Swedish Civil Contingencies Agency (MSB) has responsibility for educating and coordinating responsible actors to identify and counter information influence activities.[287] These institutions, and others like them, support organisational preparedness and capacity building through sharing expertise, developing tools, training, cooperation and raising awareness. It should be noted that the question of coordination and coherence between institutions based in different countries is often raised as a potential area for improvement.[288] For example, Edward Lucas notes that the Nordic and Baltic states plus Poland (the so-called front line states of Europe) have a GDP roughly one third greater than Russia, but "strategic incoherence" when it comes to a common defence strategy.[289]

Closely associated with institutional responses is legislative responses. What role should governments play in narrowing the range of exploits that hostile actors can use? Censorship has historically been a key method for blocking undesired messaging; however, it runs contrary to the values of Western societies to censor legitimate opinions, even if they are propagated in illegitimate ways. It may be possible in some instances to disrupt a propaganda machine by hindering the circulation of information at key points in the information distribution process, through regulation or economic (dis-)incentives rather than censorship.[290] Social media platforms in particular are typically seen as areas for intensified regulatory action, where for example advertising revenues play a key role in driving traffic including fake news.[291] Other regulatory methods can include re-labelling news agencies as propagandist, as occurred in late 2017 with the U.S. forcing RT and Sputnik to register under the Foreign Agents Registration Act (FARA). Twitter recently revealed that it will no longer accept advertising from these sources due to their illicit influence in 2016's U.S. presidential election.[292]

Many of these approaches increasingly relate to long-term questions of societal resilience. The reality is that the best defence against information influence activities is to build societal capacity over the long term. This includes themes such as education, and the ways in which children are taught source criticism and media literacy. [293] This is particularly relevant to communicators, and a short resource is included below under the heading "debunking" (4.3.3). Research into information influence activities, and the interaction between researchers, think tanks, media and government, are also important areas for cooperation over the medium to long term. [294] Overall developments within the media landscape, such as funding models that promote lower quality news ("clickbait"), access to quality (often paid) journalism, and impartial news sources are also important factors. Furthermore, the conduct of leaders and public figures, and particularly politicians and individuals with influence, plays a part in shaping confidence in public institutions. Many of these questions are about the future of democracies in light of globalisation and new media technologies and cannot be changed overnight because of possible external threats.

Finally, the big picture question of governmental policy must be addressed. As Matt Armstrong has observed,[295] there is a tendency in the information influence debate for the propagandists to set the agenda, and for governmental responses to be reactive rather than proactive. Part of the problem for Western governments is that, while they may have policy goals relating to their relationships with hostile actors (e.g. with a given nation-state), this is not always sufficiently developed into a vision that can define policy for the relationship with that nation-state, and actors protected by it, in the informational space. What is the policy goal related to countering information influence activities? Like previous more general examples, policy for the information sphere could include:

- “Bring it on”: Increase the robustness of institutions and public resilience to raise the threshold/cost of information operations in a given country
- “Fight fire with fire”: Retaliate in kind by conducting information operations toward the hostile actor or its key stakeholders.
- “Name and shame”: Release information on those behind information operations and their characteristic methods at regular intervals.
- “Information arms war”: Raise the technological hurdles to information operations until the information sphere reaches an equivalent state to MAD.

As these different perspectives suggest, individuals and organisations make up just one small part of the overall question of societal preparedness. However, the principle of resilience should be built on a cornerstone of “total defence”; that is to say, of the full participation of civilians in strengthening society. [296] While many of these larger issues are contingent on political will and international collaboration over the long term, much can be achieved in the short and medium terms simply by empowering organisations to assume responsibility for their spheres of influence. This will not, however, replace the long term need for a clear statement of policy direction for the information sphere.

Summary

Preparedness is the most important step in countering information influence campaigns. It involves:

- Institutional, organisational and individual preparedness;
- Vigilance and awareness;
- A sense of shared responsibility for a society’s “total defence”.
- A governmental policy toward information influence actors

4.3.2 Raising awareness

It may sound like a cliché from a self-help brochure, but the first step toward dealing with a problem is admitting that the problem exists. A common theme throughout the literature on information influence activities is the sudden realization that influence operations have become a regular part of the public sphere of Western countries. If Crimea and the Trump election represent turning points in public awareness of these issues, it should be clear that the issues themselves are not new.[297] Raising awareness through tangible examples is therefore an important counter measure for information influence activities simply because it seeks to establish a common understanding of the problem.

Awareness can involve several elements. First, it involves establishing the source of information influence activities, its goals, and the narratives and techniques that are used. The key component for raising awareness is to understand the target groups, whether intentional or unintentional, and the contexts in which influence is deployed.[298] A decision can then be made about how to inform those groups that they are the subject of information influence activities. Particular focus should be placed on informing decision-makers, journalists, public officials, and other key communicators in society.[299] The ultimate aim of such work is not to malign or intimidate hostile actors, but rather to strengthen societal resilience through knowledge. In other words, it builds on the principle that the best defence against hostile manipulation of open societies is awareness and education. Some recurrent themes in research suggest that:

- Leadership, whether from people in positions of formal authority or informal "thought-leaders", is important to raising awareness of the problem. Governments should officially acknowledge examples of information influence activities in their countries. [300]

- A "joined-up" response between governments, business, civil society and academia is desirable, although different demographics have different levels of confidence in each of these institutions. Problems of finding common ground between these different actors should not be underplayed.[301]

- Training and education at all levels of society is essential. The frequent, systematic discussion of issues such as fake news, source criticism, and the techniques of manipulation as social issues is an important means of strengthening awareness. This approach was particularly successful for the allies during the Second World War.

- Tools for checking facts, tracking sources and revealing influence are helpful for creating consensus around the nature and scope of the problem.

- Public exposure of cases, illicit funding and networks can demonstrate information influence activities through, for example, revealing patterns of funding that link the intent of hostile foreign actors to legitimate domestic actors who unwittingly support that agenda. Notably, this generally falls within the mandate of journalists rather than communicators, but depending on the severity of a situation, exposure may in some cases also be a useful tool for a communicator.

- Media platforms have a responsibility to their users to raise awareness of disinformation conducted on their platform. As private companies, they also have the right to ban or remove content that would normally be protected by freedom of expression laws. It may become increasingly important to maintain a dialogue with media platforms regarding how hostile influence activities are flagged.

- Raising awareness may involve bursting filter bubbles or engaging with those who are unlikely to believe that they are subject to information influence activities.

The societal resilience approach

One of the key arguments for the raising awareness approach is that information influence activities can be mitigated by establishing trust and legitimacy in government, public institutions, and the media.[302] If citizens trust their major public institutions, they are less likely to believe false or manipulated stories; or at least more likely to believe corrected information by those public institutions. Managing the reputation and legitimacy of public institutions is therefore a central component in a resilient society. However, it is important to consider who the best messenger is for different target groups, since confidence in government varies greatly. When faced with questionable information, citizens should know which government agencies, civil society actors and news sources they can turn to for reliable information. Factors such as education and source criticism are closely associated with resilience.

The information warfare approach

A second argument for raising awareness centres on the notion that all societies are now in a state of “hybrid” or “information” warfare. From this approach, competing narratives are fighting it out for legitimacy on an information battlefield, and narratives are “weaponized”. Awareness from this perspective is seen as both tactical and strategic. Citizens are at the front line of a confrontation that takes place on social media, transnational broadcasting networks, and on the pages of newspapers.[303] Individual and collective responsibilities therefore include the choice of media one consumes, the decision to share or recirculate information, and thought-leadership within social circles. Raising awareness becomes a social responsibility shared by all members of the community.

Israeli public diplomacy: As part of Israel's public diplomacy and diaspora outreach during times of war, civilians with pro-Israeli sentiments are encouraged to post stories from Israeli sources on their social media accounts, and to be vocal in criticising members of their networks who post negative stories. Citizens are empowered to become thought-leaders within their online communities, actively representing pro-Israeli narratives in debates and curating their newsfeeds as part of the war effort. This has included the use of algorithms "to identify negative voices and contain the spread of violent content" and "to identify the voices of reason & interlink them to spread the positive message of Israel," and to provide these tools to supporters as a form of information warfare.[304]

Summary

Raising awareness is an essential step in countering information influence campaigns. It involves:

- Identifying key audiences and stakeholders and making them aware that they are subject to information influence campaigns;
- Supporting societal resilience by strengthening and supporting the legitimacy of public institutions and the media;
- Encouraging individual responsibility through education and media literacy.

4.3.3 Debunking

Debunking is the act of correcting false information with accurate information. If allowed to circulate without correction, false information can gradually shape worldviews based on lies, which undermines the functioning of democracies. In the US, statements by politicians are ranked on the Politifact *Truth-o-meter* from "true" to "pants on fire".[305] Such efforts increase transparency and allow voters to inform themselves as to which politicians are most trustworthy. Almost all recent reports on disinformation suggest that fake news is best met with the truth. However, it may nonetheless be argued that such approaches primarily reach those who are pre-disposed to finding out the truth (see 3.1.2). As debates have raged following the election of Trump, concepts such as "alternative facts" suggest that debunking individual fake news stories remains vulnerable to the sheer volume of disseminated stories (*flooding*). As one report suggests, "Don't expect to counter the firehose of falsehood with the squirt gun of truth".[306]

For any kind of organisation, it is clearly a priority to quickly and accurately correct any misconceptions that circulate; something that has always been part of the everyday work of communicators. This means that disinformation is met by organisations based upon their areas of formal responsibility. Businesses wish to protect the reputations of their products and brands, while government agencies protect their areas of civic responsibility. More systematic international efforts have appeared since 2014. Dozens of initiatives now collect false news stories, analyse their sources, and correct false statements. Some approaches draw on large networks of volunteers tracking disinformation across multiple languages. Others use automated methods, network analysis, and metadata analysis to track sources. Debunking is therefore a counter influence technique that can vary greatly in its forms and objectives, despite sharing a common concern with responding to lies with accurate information.

Some important lessons have been derived from recent experiences of debunking. These include:

- Repeating false information can create familiarity, leading to a situation in which fake news becomes more memorable than the truth. It is therefore important to concentrate on the facts and narratives that you wish to communicate.

- Not every piece of false information needs to be corrected. Organisations easily fall prey to the stratagem of bait, point & shriek, i.e. it is their response to a minor piece of disinformation that is then branded as 'repressive' or 'anti-democratic'. Sometimes an over-eager response can give credibility to disinformation, giving the impression of something to hide or "no smoke without fire".

- Too much corrective information can overwhelm the target audience. Respond proportionately, with clear and simple messages.[307]

- Disinformation works best when it fits neatly within pre-existing worldviews or expectations. Accurate information may not be accepted because it clashes with these views. Corrective information should therefore be presented in ways that consider how and why the false story seemed credible. What are the audience's dispositions? Who do/don't they trust? What aspects of the truth are they least/most likely to resist?[308] Question the frame, not just the content.

- Fact checking is time-consuming and expensive. Such efforts may be unsustainable over the long-term, particularly since there is limited evidence that (1) meeting fake news with facts always changes peoples' minds, and (2) it is possible to reach all target groups with this information.[309]

- Rather than simply countering false information with the truth, it may be more important in some cases to create conditions that facilitate debate, scrutiny and critical reflection.[310]

In recent years, a number of truth trackers have been developed by civil society and government funded groups. These provide an excellent resource as a starting point for organisations seeking to counter disinformation. Generally, we recommend that communicators make independent checks of facts and sources as a matter of routine. In some cases, particularly if looking for broader patterns of behaviour or narratives, it may be useful to use repositories of fact-checking information, such as:

- Politifact
- FactCheck.org
- EU vs Disinformation
- Stopfake.org
- Snopes
- Viralgranskaren
- The Sunlight Foundation
- Flack Check
- Truth Or Fiction
- Hoax Slayer
- Fact Checker by Washington Post
- Faktiskt.se

Transparency tools

One technique used by truth trackers is that of making concealed information accessible to the public eye. ‘Open Secrets’ is one example where the service tracks the flow of money in US politics and its influence on elections and public policy. Another example of a similar approach is ‘The Sunlight Foundation’, which uses open data as a tool to make the US government and politics more transparent and accountable. Similar tools could become relevant to tracking covert financing from hostile foreign sources, for example.

Truth tracking tools

A second category of truth trackers are those identifying questionable stories and proving their false provenance. How they identify appropriate stories can differ greatly:

- *Who said it?* Some select stories on the basis that they come from influential voices in public debates. By reviewing statements made by politicians, experts, columnists, bloggers, political analysts and the hosts and guests of talk shows, to mention a few examples, they pinpoint statements that are deemed appropriate for further review (e.g. 'FactCheck', 'Politifact', 'Pundifact' and 'Fact Checkers of Washington Post').

- *Viral*. Some make their selection on the basis of their viral nature. Popular stories circulating though e-mails and on social media are identified by the use of a combination of technical tools, observations made by the staff and tips from the public. These stories often include an element of danger or threat, or issues that become to be seen as digital urban legends (e.g. 'Snopes', 'Truth or Fiction', 'Hoax slayer' and the Swedish truth tracker 'Viralgranskaren').

- *Thematic perspectives*. A final approach is by thematic perspectives. For example, some trackers focus on examining content about scientific claims (e.g. 'SciCheck'), climate change (e.g. 'Climate Feedback'), pro-Kremlin propaganda ('EU vs Disinformation'), or disinformation spread about a specific nation (e.g. Ukraine, 'StopFake.org').

After identifying relevant stories, the editorial board starts a multiple-step process in order to conclude whether the claim is false or true, and if false displaying non-partisan information to back that claim. The process involves one or a combination of the following approaches: requesting more information from the person or organisation making the claim; collecting non-partisan expert opinions and/or data from relevant databases; searching for the original sources of data; peer review of findings.

Summary

- Presenting facts is a crucial technique for countering false information. Some approaches to fact checking are organisation-centric and rely on different organisations taking responsibility for their areas of operation, and others work in the general social interest.

- Countering lies with facts can be problematic. It can be expensive, time-consuming, and may not reach the most vulnerable audiences. Furthermore, engaging with falsehoods can reinforce those stories through repetition.

- It is important to consider how both lies and facts fit within existing worldviews. Debunking should seek to create opportunities for reflection and debate and should not solely rely on polarising corrections.

- There are many existing fact-checking sources that can be used as resources by communicators.

4.3.4 Risk and vulnerability analysis

Information influence activities constitute a threat against society as foreign actors seek to disrupt and undermine democratic institutions and open public debate. In Sweden, public organisations are required to conduct a risk and vulnerability analysis.[311] The purpose of the risk and vulnerability analysis is "increasing awareness and knowledge among policymakers and operation managers about threats, risk and vulnerabilities within their own area of responsibility, and to create the basis for the creation of an action plan."[312] We propose that information influence activities should be included as a natural part of that risk and vulnerability analysis: to help organisations prepare for, prevent and manage information influence, to identify organisational vulnerabilities and to create recommendations for appropriate counter-measures. To help guide Swedish public institutions in their risk and vulnerability analysis on a general level (including threats as diverse as natural disasters, pandemics and water scarcity), MSB has created a guide that divides the process in four steps: (1) determine the analysis' point of departure, (2) risk assessment, (3) vulnerability assessment, and (4) risk management.

- *Point of departure:* This includes (1) defining the organisation's role and responsibilities (for example what geographical and sectoral areas the organisation is bound to protect and from what types of threats), (2) determining the applied methods used to identify and assess threats, and (3) presenting the analysis delimitations and chosen perspectives. For the specific purpose of identifying and assessing information influence activities, we propose that public organisations take a point of departure from this report to identify threats and make the assessment of the threats' severity based on the extent to which it disrupts and undermines an institution's work.

- *Risk assessment:* The next step consists of the organisation making an inventory of possible threats and an assessment of the threats' probability and severity. Organisations may use different tools for this analysis, such as a structural or functional model to identify threats and a quantitative or qualitative approach as basis for the assessment. Based on the list of threats, the organisation then decides which scenarios to include in an overall evaluation of the organisation's crisis management abilities and whether or not to take preventative actions to mitigate the risk of the hypothetical threats becoming a reality. Red scenarios (high probability and severity) are usually considered to require immediate actions while green scenarios (low probability and severity) could be perceived as accepted risks by the organisation.

- *Vulnerability assessment:* The next step is to assess vulnerabilities. Unlike the risk assessment stage that focuses on hypothetical risk scenarios that threaten the organisation, the vulnerability assessment takes its point of departure in the organisation itself.

Based on an analysis of how different scenarios may affect the organisation, organisational vulnerabilities are identified. As MSB describes: "The emphasis in a vulnerability analysis should be to analyse what consequences a certain event brings and how the organisation manages, resist and recovers from it."[313]

- *Risk management:* This final step refers to the actual activities conducted during a crisis. The following sections of this report outline many examples of possible counter-influence activities.

Summary

- Risk & vulnerability assessments are necessary for all organisations, and should include information influence campaigns
- MSB has prepared a handbook with further information about how to prepare such an analysis. It may be found at https://www.msb.se/RibData/Filer/pdf/25893.pdf

4.3.5 Target audience analysis

Knowledge of target audiences is essential to countering disinformation. From this perspective, communicators play an elevated role for their organisations in providing the expert knowledge required to identify and counter hostile threats.[314] The key principle behind this approach is that in many cases, the response should not be aimed at the source of the threat, but at the audience that is targeted by the hostile actor. In other words, it is about providing support to critical stakeholders and audiences who are exposed to manipulation, and not necessarily directly engaging with the aggressor.[315] This means knowing who those audiences are, as well as understanding how to reach them, the narratives that resonate with them, and their patterns of behaviour, motivations, fears and expectations. Some important considerations include:

- Audiences and public groups do not simply exist; they are "produced" by shared behavioural traits (common views, beliefs and interests), relationships to an organisation or issue (e.g. stakeholder, consumer or observer), and different forms of interconnection (networks, platforms and technologies). Knowledge of which of these factors structures key audiences can help to determine an appropriate response.
- All organisations should create stakeholder maps in order to understand their key audiences and their interests. This includes potential critics and adversaries. Data from media monitoring services should be used to keep these stakeholder maps up to date.

- All organisations should create maps of the narratives and counter-narratives that relate to their work, according to audiences. They should also create lists of credible intermediaries for specific subject matters, their points of contact with audiences that are most vulnerable to manipulation, and an internal guide to handling hostile threats that both communicators and leadership are familiar with. [316]

- Preparation – in the form of understanding audiences, their communication channels, behaviour, motivations, and narratives – can improve responses to sudden crises. This work should be grounded in the values that the organisation stands for. A strong self-identity is one of the key components of effective rebuttal or argumentation, and values act as the building blocks for positive (counter-)narratives. [317]

- In some cases, refuting falsehoods merely reinforces them, and fact checking can be too slow. Alternative messaging is therefore appropriate: change the story. Positive messaging can be appropriate to certain audiences, particularly if it can be used to drown out disinformation.[318] Note that this is similar to the "point and shriek" stratagem and hence may not be an appropriate technique for a government agency to use.

Summary

- Analysis of key audiences should be part of both the preparatory and ongoing work of organisations
- The target of counter influence activities will rarely be the source of the campaign. Key audiences are far more important

4.3.6 Strategic narratives and messaging

Telling your own story is an essential part of any response to information influence activities. This goes beyond debunking false information with facts and seeks to emphasise the ways in which identity and values shape an organisation's behaviour. Such approaches develop a broader framework and context for making sense of claims about an organisation, and whether they are true or false. If audiences have a positive attitude toward an organisation's identity and values, they may be more likely to question, or seek clarification, for information that runs contrary to those expectations. Narratives form a crucial role in shaping worldviews, yet they often consist of simple messaging that builds over time into stories unique to the perspectives of different audiences. It is therefore important to consider the interaction between tactical messaging and strategic narratives from a holistic perspective.

Tactical messaging

Messaging refers to discrete statements and responses crafted for a defined scenario. Organisations typically create messages for specific circumstances, such as to proactively explain a new initiative or product, or to react to a crisis. Press releases, for example, consist of a number of simple messages that the organisation wishes to project. Interviews are often used as opportunities to advocate key messages. Some messaging campaigns have been incredibly effective: "Don't drink and drive" is one of the most successful and famous examples. Such approaches differ from strategic narratives insofar as messaging usually has a single, defined utility. Furthermore, messages are usually created in small batches as part of a limited agenda for promoting a specific issue. This is straightforward for small organisations that only pursue a single activity but can lead to vulnerabilities for organisations that have complex identities and roles.

Raising awareness of fake stories, and debunking them with facts, are a good example of the way in which messaging can be used to counter disinformation. Slogans, mission statements, and taglines are also forms of messaging, and these can be used to explain why an organisation exists and why certain rumours or claims are false. An organisation's engagement with members of the public is often derived from messaging that has been cleared by communication officers; dialogue is typically derived from agreed messaging. Clearly then, messaging is important both as a proactive and reactive technique. However, it tends to be tactical insofar as it has limited objectives and can have a tenuous relationship to the broader stories that circulate about an organisation. It is therefore important to consider how individual examples of messaging fit within, and contribute to, the identity, values and narratives that an organisation wishes to project.

Strategic narratives

Narratives refer to the sequencing, structure, or organisation of signs, codes, and events into a coherent order. This process cannot be wholly controlled, and depends upon the perceptions, experiences and exposure of audiences to messages and stories from a variety of sources. Organisations often seek to manage these perceptions strategically: hence the concept of strategic narratives.[319] It is common for organisations to outline their vision, purpose, values and goals in order to explain who they are and what they want both to their own staff and to those outside of the organisation. In issues of reputational threats, strong strategic narratives play a crucial role in shaping resilience to falsehood.

Strategic narratives build on the notion that storytelling can help to make sense of reality in such a way as to structure reality. These stories become the intellectual tools that enable us to make sense of new information. For example, during the Cold War it made sense to speak of East and West, of the First, Second and Third Worlds, and of the Iron Curtain. Such concepts form the building-blocks through which people understand the world and make decisions; they become "common sense". Narratives conceptualize and define complex events by taking these common-sense building blocks and sequencing them into a coherent order. This involves both rational elements and elements that draw upon emotions, ideas, beliefs and prejudices. A strong strategic narrative builds on multiple components simultaneously.[320]

Some recurrent themes in research suggest that:

- Messages perform an important tactical role, and narratives are strategic. Specific messages should be aligned with the overarching strategic narrative in order to strengthen – and not contradict – the coherence of that narrative.
- Coherence can be boosted through analysis. It is important for organisations to analyse and understand which factors contribute to the overall coherency of their preferred narratives, as well as the harmful messages and narratives that circulate about their work.
- A strong strategic narrative is derived from a clear sense of organisational identity, values and goals.
- Attacks on narratives should be countered by upholding those values that the organisation stands for (and that underpin the narrative) and demonstrating that they have the resilience to cope with threats.

Summary

- Short-term messaging and long-term narratives are crucial elements of a response to information influence campaigns
- A strong sense of organisational identity, values and goals should inform all communication work, especially in cases of information influence campaigns that threaten that identity and those values.

4.3.7 Social media

Chapter 3 dealt with examples of how to identify techniques that use social media to conduct information influence activities. Bots, sockpuppets and trolls are three examples of problems that must be identified before they can be countered and all three are employed on social media platforms. Social media can be challenging in this sense because the rules of engagement are different to real life: it is difficult to be sure who is behind a social media account, from where their information is sourced, and whether their network represents a contingent of genuine public opinion or whether it is in fact fabricated. Speed of response is also a factor. Though crafting a clear and persuasive message is a key part of outreach on social media, messaging is usually juxtaposed with other elements of a post. Due to these challenging circumstances, social media is worthy of counter influence activities of its own.

The bread and butter elements of a social media post are *messaging* (the core message), *tagging* (creating a search term for an item), *name calls* (tagging a person or organisation's account), *linking* (providing a hyperlink to a different part of the internet), and *attaching* multimedia files such as an image or video. A typical social media post will contain one or more of these elements, which together contribute to positioning the post within a network of accounts and ideas. These elements can be complementary or contrasting and help to narrow or broaden the networks activated. Likewise, a link attached to such a post could lead to a genuine news story or to a questionable or manipulated source, and this is impossible to predict if the link uses a shortened URL. Blogs are commonly used as sources, and social media can be used to drive traffic to blog sites often without clear indications as to who is behind them or what their interests are. Complex combinations of these elements can further narrow down the circulation of a post to specialised communities of interest while simultaneously hiding the intent of the posters or original sources.

- *Which hashtags are being used against you?* This should be identified in order to understand how disinformation is circulating. The question of whether to engage through the same hashtag, or one that is preferred by your organisation, is one that can only be determined based on context. It is analogous to the question of whether to debunk disinformation or tell a different story.

- *Which name calls/tags are involved in disinformation?* It is essential to determine whether they are your antagonist or your audience. An organisation should protect its reputation with its audience, but the question of whether to engage with the source or agent of a hostile threat is more sensitive.

- *What links are being used in information influence posts?* An analysis of the sources that are being used can support strategic decision making regarding how best to respond. Some blog sources may be seeking pathways to become legitimate or newsworthy via Twitter, Facebook and other social media platforms.

- *Are multimedia attachments being used?* Do they have the potential to become negative memes? Is it possible to track their source, or to find evidence of manipulation, through metadata or a Google image search?

Organisations harness these elements and techniques as part of their everyday image management work. They promote their brands, products and services with hashtags and name calls, answer general or specific queries from customers, respond to crises, and use them to manage their reputations. Proactive social media work includes building networks and establishing hashtags that enable an organisation to get messages out to the right people. Generic posts for handling crises can be prepared and cleared beforehand, and in that way ensure a prompt response when an unforeseen event occurs.

Municipal guidelines for social media: One way of navigating the complex terrain of information influence online, and for establishing structures for conducting proactive communication, is to include these aspects in the organisation's social media strategy or social media guidelines. In Sweden, most municipalities already have an established social media strategy which provides an excellent starting point for further developments.[321]

Social media also enable an organisation to listen for potential threats or vulnerabilities to their reputations in real time. It is therefore both an advocacy tool for dialogue and messaging, and an open source intelligence tool for understanding important trends. Standard techniques for countering information influence on social media are analogous to techniques used when dealing with negative reporting, angry customers or an emerging crisis. Messaging, tags, name calls, links and attachments are the principle tools of engagement.

- Use of tags and name calls when countering information influence should be decided based on careful consideration of the target audience. This determines not only which audiences are reached, but which accounts are likely to share or comment upon your messages.

- Messages should be crafted based on an organisation's identity and values, with consideration of how it fits within different possible strategic narratives.

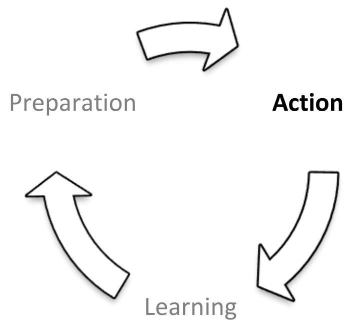
- Develop an organisational strategy for guiding activity around questions such as: when to speak and when to be silent; which audiences should be prioritised or ignored; when general messages should be used, and when they should be tailored; the use of humour, informal language or irony; to what extent the techniques and sources of information influence activities should be revealed.

- Quick responses on social media are made possible by prepared, pre-cleared messaging. For example, the Metropolitan Police sent its first tweet just seven minutes after the Westminster terrorist attack of March 2017. It gave accurate information about the unfolding situation but was based upon a message prepared for similar scenarios.[322]

Summary

- Social media is an important forum where information influence activities take place.
- It has its own logic and components, which should be harnessed in counter influence activities
- Many of the traditional rules of good communication are equally valid for countering information influence on social media

4.4 Action



Section 4.3 outlined a number of steps to improve preparedness for an information influence attack. Some of these steps are very long term, such as improving source criticism at societal level. Others are more directly related to the medium-term planning of organisations regarding their analysis of stakeholders and the narratives they wish to project. This section considers how organisations might respond to an information influence attack once it has happened. It is not a one-size-fits-all list of steps for communicators to follow. Rather, it highlights some crucial techniques that can be used depending on the nature of the crisis. In other words, it is down to the communicator and their leadership to determine which techniques are most appropriate for a given context. We explore four levels of approach, parts of which will apply to many information influence situations. The four steps are:

- Assess: communication activities that reflect a need for outreach despite a lack of knowledge about the situation
- Inform: communication activities that offer basic information about the situation

- Advocate: communication activities that argue a case or perspective about the situation

- Defend: activities that seek to protect the organisation by means other than arguing a case, such as blocking users

It should be noted that the entire array of communication techniques listed in chapter 3 can also be used by an organisation wishing to proactively respond to information influence activities. As we noted in the introduction, it is the *intention*, rather than the *technique*, that determines whether something is hostile. However, such techniques are not recommended on the grounds that an organisation's legitimacy depends in large part upon how it represents itself to the public. Section 4.2 further discusses the communicator's mandate from this perspective. We recommend that public sector communicators restrict their choices to the four raised here, in careful consideration of their mandate.

4.4.1 Assess

The first level of response includes techniques used by the organisation to respond to possible, but not yet confirmed, information influence attacks. The assessment is either conducted by the organisation itself (see *Fact check*) or by engaging external actors and facilitating their independent investigation (see *Transparent investigation*). The organisation may choose to release an initial statement, pending further investigation (see *Holding statement*).

- *Fact check*: A first, entirely uncontroversial step in handling a possible information influence campaign is to ascertain, to the best of one's ability, the facts. This is an obvious step during any kind of crisis, and information influence activities are no exception.

- *Transparent investigation*: Credibility may be derived by allowing an unbiased examination of the facts. This could include an independent inquiry, for example, allowing reputable external actors such as journalists the opportunity to visit a site and interview staff. It could also involve open archives for independent research. The aim here is to use transparency as a means of establishing the facts entirely independently. It should be noted that many public-sector organisations are bound to public transparency in such instances.

- *Holding statement*: Holding statements are an initial statement that attempts to buy some time for an organisation to ascertain certain facts. This can include internal communications.

4.4.2 Inform

The second level of response includes techniques used by the organisation to inform the public and key stakeholders. The facts may be outlined in public statements (see *Correct*) or in reference to independent actors as sources (see *Refer*). This level also includes briefings, which are a discreet means of circulating information.

- *Correct*: Once the facts have been ascertained, a formal statement outlining those facts should be prepared. This may include a Q&A format that directly responds to false allegations.

- *Refer*: In cases where independent actors can corroborate facts, it may be useful to refer to them as sources. Likewise, references to unbiased websites may provide useful sources through which stakeholders can corroborate information for themselves.

- *Brief*: Slightly less savoury, but often essential, is to provide briefings to journalists and key stakeholders. Briefings tend to be off-the-record and are for the purpose of providing facts or crucial contextual information that cannot yet be released publicly for some reason.

4.4.3 Advocate

The third level of response includes techniques used by the organisation to advocate their position. At this point, the organisation may release an official statement (see *Statement*), a persuasive dossier (see *Démarche*), or relate the event to a broader narrative (see *Storytelling*). The organisation may also engage in conversations with key stakeholders (see *Dialogue*), identify key actors to gain access to important stakeholder groups (see *Multipliers*), facilitate meetings between stakeholders (see *Facilitation*) or use existing events (see *Piggybacking*) to advocate a certain position.

- *Statement*: Depending upon the communicator's mandate within an organisation, it may be appropriate to advocate a certain position or narrative. A statement is the least controversial advocacy approach. A statement lays out the facts in a way that is aligned with the goals of the organisation.

- *Démarche*: This refers to a prepared dossier on a specific issue. It is designed to make a case, and hence is often persuasive in nature. Démarches are used to explain a position using multiple sources of evidence, and sometimes include suggestions for arguments and counter-arguments to support that position.

- *Dialogue*: An important option for a communicator is to engage in dialogue with key stakeholders and/or members of the public. This might, for example, involve responding to comments on a website or social media.

- *Storytelling*: The communicator may wish to relate any specific information influence event to a broader narrative about the organisation or the hostile context. Storytelling should be developed in reference to organisational values and strategic narratives established at the preparatory stage.

- *Facilitation*: An organisation can play an important role as a facilitator. This refers to organizing events or meetings that bring different stakeholders together to discuss a specific problem. An organisation facing false accusations related to constitutional matters could, for example, facilitate a dialogue between politicians, lawyers and scholars to clarify the legal and ethical grounds of the case.

- *Multipliers*: Multipliers are key actors who act as gatekeepers for important stakeholder groups. They are essential for amplifying information across networks. For example, a credible journalist for an industry newspaper can multiply information about an issue to a small but crucial specialist readership. On social media, hashtags and account names have become essential for amplifying messages.

- *Piggybacking*: Existing events, initiatives or debates can provide an opportunity to add an organisation's perspective. Piggybacking is a widely used technique similar to facilitation that could see a debate about constitutional matters, for example, added to an existing conference about human rights.

4.4.5 Defend

The fourth level of response includes overtly defensive techniques. The organisation may choose to ignore the situation (see *Ignore*) or take actions against the information influence attacker by reporting (see *Report*), blocking (see *Block*) or exposing (see *Expose*) the person or organisation in question. Such techniques are more problematic than those raised in the previous three sections and should be used sparingly.

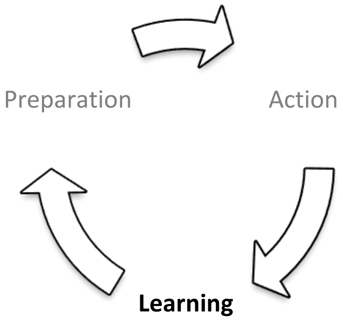
- *Ignore*: Defensive opportunities are limited for public sector organisations and have the risk of giving the impression of something to hide. In some cases, however, they will be necessary. Sometimes, ignoring a social media troll may be the best option.

- *Report*: It may be necessary to report an information influence attacker to police or to the owner of a specific platform.

- *Block*: Blocking an actor who threatens others could be a reasonable response. However, communicators should be acutely aware of the need to respect freedom of speech and refer to the appropriate governing code of conduct before blocking a user.

- *Expose*: Although not recommended, a strategic response to information influence activities could include exposing the individual behind an account, for example. Less controversially, de-personalised data could be revealed to provide examples of the kinds of attacks that an organisation regularly faces.

4.5 Learning



It is not always possible to determine whether or to what extent an event may be defined as hostile foreign influence. Therefore, collecting and documenting these events is essential. Examples of threats and vulnerabilities, and of best practice in counteracting them, should be shared so that society as a whole can learn.

Some form of evaluation process is therefore desirable. This should be shared (1) with other communicators in similar roles, (2) with your organisation's leadership, (3) with authorities tasked with identifying information influence activities (e.g. MSB), and (4) in some cases, with the general public.

It may be an appropriate step in the future for stakeholders to agree upon a common method for collecting and disseminating cases. As a basic guide, we recommend that the following information should be recorded for learning purposes:

- Describe the context and background to the case.

- Describe the actors and networks involved. Do not speculate about who is behind the influence operation.

- To what extent did the case meet the DIDI definition?

- What was the nature of the vulnerability being exploited?

- Describe the influence techniques used, including activity chains and narratives. Does the case fit within a broader pattern of campaigns?

- What do you think were the intended effects? What evidence do you have to support this?

- What do you think would happen if the influence operation was not counteracted?
- What countermeasures did you take? What were their effects?
- What lessons do you take from this example?
- Be sure to save evidence or data related to the case.

Consider how this case can be used to train your own staff to meet future challenges. Use these cases when updating preparatory work and strategies.

4.6 The limits of counter-strategy

Countermeasures are limited by the fact that they respond to somebody else's agenda. In this regard, the entire principle of countering information influence activities has a premise that is problematic, since the aggressor may appear to be setting the conditions under which a nation's democracy can or cannot properly function. It makes more sense, generally speaking, to focus on upholding democratic values which depend upon debate and free speech. In lieu of legal or normative frameworks determining the validity of, for example, a bot's freedom of expression, we recommend a robust but measured response. We recommend an approach based upon minimising systemic vulnerabilities; raising the threshold for information operations through preparedness; developing proportionate and sensible communicative responses that place the audience (rather than adversary) in focus and uphold shared societal values; and of learning from successful and unsuccessful cases. When assuming this perspective, some additional considerations are related to i) how people think (cognition); ii) legal and regulatory questions; and iii) the likelihood that techniques will remain at least one step ahead of responses.

4.6.1 Cognitive limits

Information influence activities often showcase a profound understanding of how the human mind works by effectively exploiting heuristic patterns, cognitive features and social cues to distort both individual and collective decision-making processes. Counter influence activities often (and regrettably) lack a similarly sophisticated understanding. A typical counter approach is the debunking approach, which sometimes assumes that the human mind works like a computer hard drive, where one piece of information can simply be overwritten by another. This mistaken understanding furthers the notion that exposure of hostile activities, debunking of myths and the provision of facts will automatically disarm information influence activities.[323] However, research warns us of the risk that counter influence activities can strengthen the effects of information influence in people's minds, if done the wrong way.[324] To bridge the discrepancy between what we intuitively think works and what really works to disarm information influence activities, we first need to understand the factors that impact upon the effectiveness of counter measures.

As has been discussed in Chapter 2, influence activities exploit vulnerabilities in media system and public opinion formation in order to influence individual cognitive processes. Therefore, it is "not just *what* people think that matters, but *how* they think".[325] The human brain has throughout evolution developed heuristic patterns in order to rapidly transform cognitive input into mental output through different decision rules. Such heuristics may not be accurate, but they are often effective for making split second decisions under constraints of limited time and knowledge.[326] However, heuristics can lead to faulty conclusions in more complex situations where focus, reasoning, and weighing up alternatives is important.[327] Three cognitive heuristic effects are particularly important for understanding the limits of counter influence strategies:[328]

- The *familiarity backfire effect* states that repeated exposure to a piece of information increases the chances of an individual accepting the information as true. This implies that activities aimed at countering influence and disinformation should be wary of repeating the initial piece of misleading information.

- The *overkill backfire effect* stipulates that information that is easy to process is more likely to be accepted as true, which is why providing long and sophisticated arguments against a piece of disinformation can have the reverse effect.

- The *worldview backfire effect* highlights the cognitive processes related to identity that cause people to unconsciously process information in a biased way, including through confirmation bias. This implies that we selectively seek information that is congruent with our worldview and that we consequently dismiss information that threatens our worldview, no matter its correspondence to reality.[329]

The impact of such cognitive shortcuts is amplified by the limitations of the human memory.[330] Human memory is not a stored artefact like a photograph but is rather continually written and re-written in the act of remembering. It is engaged in a constant "process of (re)construction that is vulnerable to both internal and external influences".[331] Information overload, for example, forces the mind to revert to heuristic shortcuts in order to deal with the abundance of information.[332] As not every piece of information can be stored, primacy is given to information that is congruent with pre-existing memories, which strengthens the worldview backfire effect. It is easier to remember a piece of information that fits than one that doesn't.

Social psychological factors also influence the effects of counter influence activities, as parts of our decision-making process stem from social cues.[333] As social beings, we are inherently inclined to believe and behave in conformity with others around us, relying not only on information stored in our individual minds but also on knowledge established socially by the 'collective mind'.[334] This creates a series of problematic biases such as group think, group polarization, and halo effects.[335] Individuals are likely to reject information coming from outside of their social sphere.[336] The credibility of the messenger has been shown to be of great importance. [337] Coupled with cognitive heuristics, such social factors have a profound impact on the effectiveness of counter influence strategies, as it implies that false beliefs, disinformation or propaganda are sticky phenomena.[338] Once they have penetrated society, they are hard to remove.

As Chapters 3 and 4 discussed, counter influence strategies are more complex than simply "meeting" or "debunking" false claims. It consists of reactive and proactive initiatives that strengthen the societal capacity to identify, prepare for, and counteract information influence activities. It is a mind-set of vigilance, as well as a collection of functions and processes aimed at limiting exploitation of vulnerabilities. Any attempt at countering information influence should recognise that social psychological factors impact the effect of counter activities. It is important to consider the relationship between threats (their techniques and intentions) and vulnerabilities (*media system, public opinion* and *cognitive*). Counter-strategy is therefore not a panacea.

4.6.2 Legal and ethical limits

A comprehensive account of the legal domain within which activities to counter information influence occur is provided by Winther (2016) in his report on the law of freedom of expression in relation to counter influence.[339] Here, it is sufficient to remind ourselves that any statement or account intended for a public audience is protected. The fundamental law on freedom of expression builds on the principle that every citizen has the right to freely express their opinions publicly, and as such the law does not allow authorities of executive agencies to take measures to review, impede or prohibit the disclosure of publications in forms covered by the law (print-, broadcast- or digital media). Effectively, this implies that influence activities that fall within the scope of the fundamental law on freedom of expression should always be addressed on the arena of open and free debate, without impediments to freedom of expression.[340]

Despite the intuitive difference between information influence and legitimate democratic debate, influence activities are often wholly within the remit of the law. Influencing others, even if it is disruptive and has the potential of causing harm to societal institutions, is generally permitted in democratic societies under institutions such as freedom of speech and expression. Legality does not, however, equate to legitimacy, and while influence activities may be legal they can simultaneously be considered illegitimate. It may even be the case that legitimate domestic actors at times use some of the techniques we have highlighted in chapter 3. An important question is therefore why legitimate domestic actors would choose to use similar techniques to those used in information influence. An associated question is to what extent such techniques should be considered acceptable in a democratic society.

The reality is that most information influence activities and counter influence techniques will be fully legal. However, the rules and norms differ greatly depending on who acts, and why. When it comes to governmental actors, we advise communication that is ethically beyond reproach. It is deeply problematic for a democratic state to restrict freedom of speech in any way, and public-sector communicators should be aware that even simple tools such as so-called truth trackers can be controversial if they in any way seem to curtail the right to alternative perspectives. Efforts for countering information influence activities should never have the effect of silencing public debate or creating fear in people of being labelled propagandists for the sake of having opinions which conform to hostile narratives. Open and democratic debate must always be protected and encouraged. Again, care must be taken to consider issues related to the grey-zone between what is legitimate and what is not.

4.6.3 Playing catch-up

Finally, it is important to observe that the field of information influence activities is characterised by rapid developments in techniques. This report takes its point of departure in vulnerabilities precisely because they are more stable than the actual techniques. Even so, many techniques show historical continuities. Impersonation, for example, was considered a dangerous tool during the Second World War. Swedish citizens were advised to keep a sceptical mind when listening to the radio as talented actors would pose as their official representatives and spread disinformation. The same deceptive basis for information influence activities stands today, but in new forms such as fake social media accounts and deepfake/faceswap technologies. There should be no doubt that the near future will bring new tools better suited to accomplish the same ends, with hostile actors exploiting the latest technological advances to carve out a temporary advantage in the information space.

Current discussions about possible future developments relate foremost to digital manipulation. Techniques on the horizon suggest that sophisticated bots may one day simulate authentic human online conversations, and that emerging tools for audio and video manipulation will allow hostile actors to create multimedia content that can convincingly mislead audiences. A major risk is that videos of public figures and political leaders making statements that have been manipulated could create short term crises, where the speed and accuracy of a counter response is crucial. Similarly, it is clear that changing online business models have played a significant role into the recent proliferation of fake news. It is likely that future alterations to online advertising processes, for example, could produce unexpected responses from those adept at 'hacking' or exploiting systems.

As we have repeated in this report, tools are neither good nor bad by nature. Just as the future may hold a number of new information influence techniques, so may new counter techniques help organisations to both identify and counter information influence activities. For example, new AI techniques may provide tools to detect, label and perhaps even remove manipulated content from media platforms. Big Data analysis may assist in attempts to reveal "filter bubbles" on social media and openly display not merely politically polarised views but the nuances and range of arguments between them. New digital tools may help to assist organisations and individuals in assessing whether an audio or video recording is the real deal or manipulated. Counter influence strategies may be destined to be one step behind information influence activities, but it is important to ensure that they remain no further than one step behind.

5. Conclusion

The structure and disposition of the report reflects a simple reality: much more is known about the techniques and conduct of information influence activities than is known about how to counter it effectively. This is not least because we are always playing ‘catch-up’ with last year’s information influence techniques. Keeping up with case studies of information influence activities, making sense of their underlying techniques, and positioning those techniques within taxonomies is worthwhile, since it can prepare us for the work of countering what may be coming today and tomorrow. But it is work that is iterative, that is best developed through collaborative stakeholder networks, and that needs to be shared in formats that are responsive and up to date.

Perhaps most importantly, it is clear that counteracting information influence activities cannot be reduced to a simple checklist of activities. Rather, it should be the enlightened response of educated and informed communicators skilled at their jobs to determine the best course of action in each instance. Successful examples must be recorded, analysed and shared. The subtitle of this report, “The State of the Art”, reflects the idea that counter influence is an art rather than a science. Ultimately, it is the art of counter influence that will shape the resilience of society to these threats, and that will determine whether our cognitive, public opinion and media system vulnerabilities are vulnerabilities or in fact strengths.

Recommended readings

This list provides an entry point for the interested reader to broaden and deepen their understanding of particularly relevant aspects of the report. The readings on this list does not capture all areas covered by the report, but broadly cover the topic of information influence.

- Luca Maria Aiello et al., “People Are Strange When You’re a Stranger: Impact and Influence of Bots on Social Networks,” *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media* 697, no. 483,151 (2012): 10–17.

- John Cook and Stephan Lewandowsky, The Debunking Handbook, 2012,http://www.skepticalscience.com/docs/Debunking_Handbook.pdf.

- Thomas Elkjer Nissen, “Social Media’s Role in ‘Hybrid Strategies’” (Riga: NATO Stratetic Communications Centre of Excellence, 2016).

- Ulrik Franke, “Information Operations on the Internet: A Catalog of Modi Operandi” (FOI Totalförsvarets forskningsinstitut, March 2013).

- Lisa Kaati, “Det Digitala Kalifatet En Studie Av Islamiska Statens Propaganda” (FOI Totalförsvarets forskningsinstitut, May 2017).

- David Lazer et al., “Combating Fake News: An Agenda for Research and Action” (Combating Fake News: An agenda for Research and Action, Cambridge: Harvard University, 2017), https://shorensteincenter.org/wp-content/uploads/2017/05/Combating-Fake-News-Agenda-for-Research-1.pdf.

- Stephan Lewandowsky et al., “Misinformation and Its Correction: Continued Influence and Successful Debiasing,” *Psychological Science in the Public Interest* 13, no. 3 (December 2012): 106–31, https://doi.org/10.1177/1529100612451018.

- Edward Lucas and Peter Pomeranzev, “Winning the Information War” (Center for European Policy Analysis, 2016).

- Daniel Milo and Katarína Klingová, “Countering Information War Lessons Learned from NATO and Partner Countries: Recommendations and Conclusions” (Bratislava: Globsec, 2016), https://www.globsec.org/wp-content/uploads/2017/09/countering_information_war.pdf.

- Björn Palmertz, “Europeiska Perspektiv På Förmågan Att Möta Påverkanskampanjer Från Främmande Makt - Delrapport 1” (Stockholm: Center for Asymmetric Threat Studies (CATS), Swedish National Defence College, 2016).

- Björn Palmertz, “Theoretical Foundations of Influence Operations: A Review of Relevant Psychological Research” (Stockholm: Center for Asymmetric Threat Studies (CATS), Swedish National Defence College, n.d).

- Christopher Paul and Miriam Matthews, “The Russian ‘Firehose of Falsehood’ Propaganda Model - Why It Might Work and Options to Counter It,” Expert insights on a timely policy issue (RAND Corporation, 2016), http://www.rand.org/content/dam/rand/pubs/perspectives/PE100/PE198/RAND_PE198.pdf.

- Sergey Sanovich, “Computational Propaganda in Russia - The Origins of Digital Misinformation,” Working Paper, Computational Propaganda Research Project (Oxford: Oxford Internet Institute, 2017).

- Steve Tatham, “The Solution to Russian Propaganda Is Not EU or NATO Propaganda but Advanced Social Science to Understand and Mitigate Its Effects in Targeted Populations,” Policy paper (Riga: National Defence Academy of Latvia, Center for Security and Strategic Research, July 2015).

- Rand Waltzman, “The Weaponization of Information - The Need for Cognitive Security” (Santa Monica, CA: RAND Corporation, 2017).

- Pontus Winther, “Yttrandefrihetsgrundlagen Och Möjligheterna Att Möta Påverkanskampanjer Från Främmande Makt” (Myndigheten för samhällsskydd och beredskap, 2016).

Myndigheten för samhällsskydd och beredskap

651 81 Karlstad Tel 0771-240 240 www.msb.se

Publ.nr MSBXXX - Månad År ISBN 978-91-7383-XXX-X

Endnotes

Please note that endnotes have in part been generated by third-party referencing software using the Chicago Manual of Style (17th ed.). Therefore, some errors may have escaped the authors' scrutiny. Please inform the authors if you discover errors in the endnotes.

[1] "Strategisk Utblick 7: Närområdet Och Nationell Säkerhet," *Strategisk Utblick 7: Närområdet Och Nationell Säkerhet*, 2017; Regeringen och Regeringskansliet, "Nationell säkerhetsstrategi," Text, Regeringskansliet, January 8, 2017, http://www.regeringen.se/informationsmaterial/2017/01/nationell-sakerhetsstrategi/.

[2] See for example: "Strategisk Utblick 7: Närområdet Och Nationell Säkerhet"; Militära underrättelse- och säkerhetstjänsten (MUST), "Årsöversikt 2016" (Stockholm: Försvarsmakten, 2016), https://www.forsvarsmakten.se/siteassets/3-organisation-forband/hogkvarteret/must/must-arsoversikt-2016.pdf; Regeringskansliet, "Nationell säkerhetsstrategi"; Regeringen och Regeringskansliet, "Försvarspolitisk inriktning – Sveriges försvar 2016-2020," Text, Regeringskansliet, April 23, 2015, http://www.regeringen.se/rattsdokument/proposition/2015/04/prop.-201415109/; Dagens Nyheter, "Stefan Löfven Om Utländsk Påverkan Inför Valet 2018 - DN.SE," accessed July 28, 2017, http://www.dn.se/nyheter/politik/stefan-lofven-om-utlandsk-paverkan/; Regeringen och Regeringskansliet, "Sveriges säkerhet i en ny värld," Text, Regeringskansliet, January 14, 2018, http://www.regeringen.se/tal/2018/01/sveriges-sakerhet-i-en-ny-varld/; Regeringen och Regeringskansliet, "Stärkt psykologiskt försvar och åtgärder mot påverkansoperationer," Text, Regeringskansliet, January 14, 2018, http://www.regeringen.se/artiklar/2018/01/starkt-psykologiskt-forsvar-och-atgarder-mot-paverkansoperationer/.

[3] "Msb.Se - About MSB," accessed December 5, 2017, https://www.msb.se/en/About-MSB/.

[4] The term psychological defence (*psykologiskt försvar*) was coined in 1953 in a public report on hostile psychological warfare. The term is still employed to capture activities related to psychological aspects with in the total defence structure inclunding for example information operations, influence operations and psychological warfare. For more, see Niklas H Rossbach, "Psykologiskt Försvar - Avgörande För Svensk Försvarsförmåga," Strategisk Utblick 7: Närområdet Och Nationell Säkerhet (FOI Totalförsvarets forskningsinstitut, 2017).

[5] "Msb.Se - MSB Och Psykologiskt Försvar," accessed December 5, 2017, https://www.msb.se/sv/Insats--beredskap/Psykologiskt-forsvar/MSB-och-psykologiskt-forsvar/.

[6] Committee on Foreign Relations, "Putin's Asymmetric Assault on Democracy in Russia and Europe: Implications for U.S. National Security" (Washington DC: United States Senate, January 10, 2018).

[7] Regeringskansliet, "Nationell säkerhetsstrategi."

[8] Regeringskansliet, "Försvarspolitisk inriktning – Sveriges försvar 2016-2020."

[9] Regeringskansliet.

[10] Justitiedepartementet, "Regleringsbrev För Budgetåret 2017 Avseende Myndigheten För Samhällsskydd Och Beredskap" (Regeringen, 2016).

[11] The database departed from a collection about 200 pieces of relevant literature provided by MSB. Using a snowballing technique the database has been extended to broaden its scope and include relevant materials from other fields as well as more journalistic and practical accounts. In its current state, the database include over 1,000 sources which have been sampled for the report, with primacy given to sources which include or combine multiple other sources (such as other literature reviews) and sources published by impactful institutions within the field (such as NATO, EU, RAND, CEPA etc). In the database, the materials are roughly sorted based on origin of the text (academic, media, think tank etc), but significant work remains to be done in this regard.

[12] This is MSBs internal working defininition of "informationspåverkan" (information influence campaigns) as presented in MSB DNR2017-11794 "*Förberedande lägesbild valet 2018*" (translated by authors).

[13] Försvarsdepartementet, *En strategi för Sveriges säkerhet: försvarsberedningens förslag till reformer* (Stockholm: Fritzes, 2006).

[14] For a working definition of 'deception' see Jiri Valenta, "Soviet Use of Surprise and Deception," *Survival* 24, no. 2 (March 1982): 50–61, https://doi.org/10.1080/00396338208442019.

[15] NATO, "Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia" (Riga: NATO Stratetic Communications Centre of Excellence, 2016).

[16] Please note that while we are rarely able to know an actors true intent and objective, it remains that intent can nonetheless pragmatically be inferred by actors subject of information influence activities. For example, we cannot know for sure why antagonists establish false media outlets online, but we can reasonably infer by their effect that the intent is decieve a target audience and that the objective is to change the same audience's opinions. To be sure, our inferals may sometimes be wrong, which is why intent is not the sole criteria for diagnosing information influence activities, but rather one of four, and why the benefit of the doubt should always be afforded.

[17] OSCE, "Propaganda and Freedom of the Media: Non-Paper of the OSCE Office of the Representative on Freedom of the Media" (Vienna: OSCE, 2015).

[18] For an overview of key texts see Jostein Gripsrud et al., eds., *The Idea of the Public Sphere: A Reader* (Plymouth: Lexington Books, 2010).

[19] Howard Nothhaft, James Pamment, and Henrik Agardh-Twetman, "Hostile Influence in Western Democracies: A model of systemic vulnerabilities", in Cornelieu Bjola and James Pamment (eds.), *Countering Online Propaganda and Extremism: The Dark Side of Digital Diplomacy,* (Routledge, forthcoming).

[20] "Horton, D and Wohl, R. (1956) 'Mass Communication and Para-Social Interaction: Observation on Intimacy at a Distance', Psychiatry, 19, Pp. 215-229.," n.d., http://images.lib.monash.edu.au/ats1280/04117091.pdf.

[21] Gripsrud et al., *The Idea of the Public Sphere: A Reader*.

[22] Classic conceptions of democracy often are ultimately based on the Athenian polis or the Roman Senate, where it was taken for granted that debaters would contribute in person and as persons (not as representatives of a party) as well as being held personally responsible for the policies advocated by them (for a classic acount see Hannah Arendt, The Human Condition).

[23] A classic account is found, of course, in Jürgen Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*, Sixth Printing edition (Cambridge, Mass: The MIT Press, 1991).

[24] Hence Habermas's famous dictum that a public sphere that excludes certain actors beforehand is not only deficient, but not a public sphere at all. Habermas.

[25] For a fascinating insiders' view see: Ryan Holiday, *Trust Me, I'm Lying: Confessions of a Media Manipulator*, 1 edition (New York: Portfolio, 2013).

[26] Andrew Chadwick, *The Hybrid Media System: Politics and Power* (OUP USA, 2013).

[27] Marju Himma-Kadakas, "Alternative Facts and Fake News Entering Journalistic Content Production Cycle," *Cosmopolitan Civil Societies: An Interdisciplinary Journal* 9, no. 2 (July 2017): 25–41, https://doi.org/10.5130/ccs.v9i2.5469.

[28] John Tooby and Leda Cosmides, "The Psychological Foundations of Culture," in *The Adapted Mind*, by Jerome H. Barkow, John Tooby, and Leda Cosmides (New York: Oxford University Press), accessed May 17, 2018, https://www.cep.ucsb.edu/papers/pfc92.pdf.

[29] Gerhard Roth, *Fühlen, Denken, Handeln: Wie das Gehirn unser Verhalten steuert*, 6th ed. (Frankfurt am Main: Suhrkamp Verlag, 2001).

[30] Roth.

[31] Jerome H. Barkow, Leda Cosmides, and John Tooby, eds., *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Reprint edition (New York: Oxford University Press, 1995).

[32] Raymond S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.," *Review of General Psychology* 2, no. 2 (1998): 175.

[33] Howard Nothhaft, James Pamment, and Henrik Agardh-Twetman, "Hostile Influence in Western Democracies: A model of systemic vulnerabilities", in Cornelieu Bjola and James Pamment (eds.), *Countering Online Propaganda and Extremism: The Dark Side of Digital Diplomacy,* (Routledge, forthcoming).

[34] Representative for many others: Thomas Elkjer Nissen, "Social Media's Role in 'Hybrid Strategies'" (Riga: NATO Stratetic Communications Centre of Excellence, 2016); John Chambers, "Countering Gray Zone Hybrid Threats: An Analysis of Russia's 'New Generation Warfare' and Implications for the US Army" (New York: Modern War Institute, West Point, October 18, 2016); Maciej Bartkowski, "Nonviolent Civilian Defense to Counter Russian Hybrid Warfare" (Washington DC: John Hopkins University, 2015).

[35] Edward Lucas and Peter Pomeranzev, "Winning the Information War" (Center for European Policy Analysis, 2016), https://www.semperfidelis.ro/e107_files/public/1470461530_2186_FT4490_peter_pomerantsev_edward_lucas_-_aug._2016_-_winning_the_information_war_-_the_full_report.pdf.

[36] Chambers, "Countering Gray Zone Hybrid Threats: An Analysis of Russia's 'New Generation Warfare' and Implications for the US Army."

[37] For an excellent discussion of the grey-zone, see: Daniel Jonsson, "Typfall 5: Utdragen Och Eskalerande Gråzonsproblematik" (Stockholm: FOI Totalförsvarets forskningsinstitut, 2018).

[38] Antonio Missiroli et al., "Strategic Communications - Countering Russia and ISIS/Daesh," ISSUE (Brussels: European Union Institute for Security Studies, July 2016).

[39] Missiroli et al.

[40] America Y Guevara, "Propaganda in Mexico's Drug War," *Journal of Strategic Security* 6, no. 5 (2013): 22.

[41] Guevara.

[42] Samanth Subramanian, "Meet the Macedonian Teens Who Mastered Fake News and Corrupted the US Election," WIRED, accessed December 11, 2017, https://www.wired.com/2017/02/veles-macedonia-fake-news/.

[43] Jonsson, "Typfall 5: Utdragen Och Eskalerande Gråzonsproblematik."

[44] Nicholas J. Cull, "Counter Propaganda - Cases from US Public Diplomacy and Beyond," Legatum Institute Transitions Forum (London: Legatum Institute, July 2015).

[45] Timothy Revell, "Dark Ads Pick You out: Political Beliefs Are an Easy Target on Facebook, Finds Timothy Revell," *New Scientist* 235, no. 3137 (August 5, 2017): 8–8; Alex Hern, "Facebook 'dark Ads' Can Swing Political Opinions, Research Shows," *The Guardian*, July 31, 2017, sec. Technology, http://www.theguardian.com/technology/2017/jul/31/facebook-dark-ads-can-swing-opinions-politics-research-shows; "The 'Dark Ads' Election: How Are Political Parties Targeting You on Facebook?," The Bureau of Investigative Journalism, accessed November 23, 2017, https://www.thebureauinvestigates.com/stories/2017-05-15/the-dark-ads-election-how-are-political-parties-targeting-you-on-facebook.

[46] Michael E. Porter, *Competitive Strategy: Techniques for Analyzing Industries and Competitors* (Simon and Schuster, 2008).

[47] U.S. Senate, "Hearings | Intelligence Committee," November 1, 2017, https://www.intelligence.senate.gov/hearings/open-hearing-social-media-influence-2016-us-elections.

[48] John Cook and Stephan Lewandowsky, *The Debunking Handbook*, 2012, http://www.skepticalscience.com/docs/Debunking_Handbook.pdf; Cook and Lewandowsky.

[49] Tanya Silverman et al., “The Impact of Counter-Narratives” (Institute for Strategic Dialogue, 2016), https://www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives_ONLINE_1.pdf.

[50] For a recent and comprehensive overview of philosophical and academic discussions on how to understand and define a fact, plase see chapter two of Åsa Wikforss, *Alternativ Fakta - Om Kunskapen Och Dess Fiender* (Falun: Fri tanke, 2017).

[51] Hagen Schölzel and Howard Nothhaft, “The Establishment of Facts in Public Discourse: Actor-Network-Theory as a Methodological Approach in PR-Research,” *Public Relations Inquiry* 5, no. 1 (January 1, 2016): 53–69, https://doi.org/10.1177/2046147X15625711.

[52] Da Scheufele, “Framing as a Theory of Media Effects,” *Journal of Communication* 49, no. 1 (March 1, 1999): 103–22, https://doi.org/10.1111/j.1460-2466.1999.tb02784.x.

[53]Naomi Oreskes and Erik M. Conway, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming* (A&C Black, 2011).

[54] “RT,” RT International, accessed March 15, 2018, https://www.rt.com.

[55] Nickerson, “Confirmation Bias.”

[56] Eli Pariser, *The Filter Bubble: What The Internet Is Hiding From You* (Penguin UK, 2011).

[57] John R. Hibbing, Kevin B. Smith, and John R. Alford, *Predisposed: Liberals, Conservatives, and the Biology of Political Differences* (Routledge, 2013).

[58] Contrary to conventional wisdom, there is evidence that the grand narratives underlying people’s political orientation are not only products of socialization and enculturation, but to a degree biologically predetermined (not in specific content, but in the values they actualize, e.g. whether change is generally seen as a threat or an opportunity): Some studies of genetically identical twins show surprising similarities in political orientation even when siblings were brought up separately and in very different households. See Hibbing, Smith, and Alford.

[59] Daniel L. Schacter, *The Seven Sins of Memory: How the Mind Forgets and Remembers*, 1st edition (Boston, Mass.: Mariner Books, 2002).

[60] Cull, “Counter Propaganda - Cases from US Public Diplomacy and Beyond.” for the following see Cull

[61] Walter Lippmann, *Public Opinion* (Harcourt, Brace, 1922).

[62] Harold Dwight Lasswell, *Propaganda Technique in the World War* (Peter Smith, 1927).

[63] Cull, “Counter Propaganda - Cases from US Public Diplomacy and Beyond.”

[64] Ronald D. Smith, *Strategic Planning for Public Relations* (Routledge, 2013).

[65] Political Dictionary, “Swiftboating,” *Political Dictionary* (blog), October 28, 2009, http://politicaldictionary.com/words/swiftboating/.

[66] Rand Waltzman, “The Weaponization of Information - The Need for Cognitive Security” (Santa Monica, CA: RAND Corporation, 2017).

[67] Mark Magnier, “Hindu Girl’s Complaint Mushrooms into Deadly Indian Riots,” *Los Angeles Times*, September 9, 2013, http://articles.latimes.com/2013/sep/09/world/la-fg-india-communal-20130910.

[68] Waltzman, “The Weaponization of Information - The Need for Cognitive Security.”

[69] Maeve McClenaghan, “The ‘Dark Ads’ Election: How Are Political Parties Targeting You on Facebook?,” The Bureau of Investigative Journalism, 2017, https://www.thebureauinvestigates.com/stories/2017-05-15/the-dark-ads-election-how-are-political-parties-targeting-you-on-facebook.

[70] Scott Shane and Vindu Goel, “Fake Russian Facebook Accounts Bought $100,000 in Political Ads,” *The New York Times*, September 6, 2017, sec. Technology,

https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html; Thomas Nilsson, "Ryska Facebook-Annonser Skulle Spä På Sociala Och Politiska Motsättningar," 2017, https://www.resume.se/nyheter/artiklar/2017/09/27/ryska-annonser-skulle-spa-pa-sociala-och-politiska-motsattningar/.

[71] Jay McGregor, "Why Facebook Dark Ads Aren't Going Away," Forbes, accessed November 23, 2017, https://www.forbes.com/sites/jaymcgregor/2017/07/31/why-facebook-dark-ads-arent-going-away/; Hern, "Facebook 'dark Ads' Can Swing Political Opinions, Research Shows."

[72] On some platforms, such as Facebook, it is possible for the individual user to obtain some information on why they have received a specific ad by clicking on the (very small) drop-down menu at the top right of the ad and selecting "Why am I seeing this ad". The information under this tab will tell you which parameters the webpage is using to determine the ad's relevancy to you specifically.

[73] Other authors sometimes refer to this as *social engineering*, as is the case in for example: Elkjer Nissen, "Social Media's Role in 'Hybrid Strategies.'"

[74] Solomon E. Asch, "Opinions and Social Pressure," *Scientific American* 193, no. 5 (November 1, 1955): 31–35, https://doi.org/10.1038/scientificamerican1155-31.

[75] Vincent F. Hendricks, "All Those Likes and Upvotes Are Bad News for Democracy," The Conversation, 2013, http://theconversation.com/all-those-likes-and-upvotes-are-bad-news-for-democracy-21547; Lev Muchnik, Sinan Aral, and Sean J. Taylor, "Social Influence Bias: A Randomized Experiment," *Science* 341, no. 6146 (2013): 647–51, https://doi.org/10.1126/science.1240466.

[76] Simon Collister, "Analysing Algorithms in Public Relations Research: Contexts, Challenges and Innovative Methodologies," accessed December 13, 2017, http://www.academia.edu/18268550/Analysing_Algorithms_in_Public_Relations_Research_Contexts_Challenges_and_Innovative_Methodologies.

[77] Vicki G. Morwitz and Carol Pluzinski, "Do Polls Reflect Opinions or Do Opinions Reflect Polls? The Impact of Political Polling on Voters' Expectations, Preferences, and Behavior," *Journal of Consumer Research* 23, no. 1 (1996): 53–67; Ian McAllister and Donley T. Studlar, "Bandwagon, Underdog, or Projection? Opinion Polls and Electoral Choice in Britain, 1979-1987," *The Journal of Politics* 53, no. 3 (1991): 720–41, https://doi.org/10.2307/2131577.

[78] "Astroturfing," Public Relations Wiki, accessed December 14, 2017, http://pr.wikia.com/wiki/Astroturfing.

[79] Katherine Miller, *Communication Theories: Perspectives, Processes, and Contexts*, 1 edition (Boston, Mass: McGraw-Hill Humanities/Social Sciences/Languages, 2001).

[80] Elisabeth Noelle-Neumann, *The Spiral of Silence: Public Opinion - Our Social Skin, 2nd Edition*, 2nd edition (Chicago: The University of Chicago Press, 1993).

[81] Homero Gil de Zúñiga and Trevor Diehl, "Citizenship, Social Media, and Big Data: Current and Future Research in the Social Sciences," *Social Science Computer Review* 35, no. 1 (February 2017): 3–9, https://doi.org/10.1177/0894439315619589.

[82] Pariser, *The Filter Bubble*.

[83] Ivan Dylko et al., "Impact of Customizability Technology on Political Polarization," *Journal of Information Technology & Politics*, August 4, 2017, 1–15, https://doi.org/10.1080/19331681.2017.1354243; Natalie Jomini Stroud, "Polarization and Partisan Selective Exposure," *Journal of Communication* 60, no. 3 (August 19, 2010): 556–76, https://doi.org/10.1111/j.1460-2466.2010.01497.x.

[84] Jonathan Bright, "Explaining the Emergence of Echo Chambers on Social Media: The Role of Ideology and Extremism," *ArXiv:1609.05003 [Physics]*, September 16, 2016, http://arxiv.org/abs/1609.05003.

[85] The SOM-institue annual survey in Sweden did not, for example, find evidence of filter bubbles or echo chambers, see Annika Bergström, "Sanningar Och Myter Om Användningen Av Sociala Medier" (SOM-seminariet, Göteborg, April 25, 2017),

https://som.gu.se/digitalAssets/1624/1624264_sanningar-och-myter-om-sociala-medier.pdf; For a popular summary of the current debate in Sweden, please see: “Forskare sticker hål på myten om filterbubblor,” Forskning & Framsteg, accessed February 26, 2018, https://fof.se/artikel/forskare-filterbubblan-ar-en-myt.

[86] Jesper Strömbäck, “Demokratin Och Det Förändrade Medielandskapet: Mot Ökade Kunskapsklyftor Och Deltagandeklyftor?,” 2015, http://diva-portal.org/smash/get/diva2:803440/FULLTEXT01.

[87] In his essay on the relationship between democracy and the changing media landscaape, Strömbäck highlight selective exposure is becoming increasingly problematic in today’s media system generally, as a more open and diverse media system allows individuals to look for information to confirm their world view. See Strömbäck (above).

[88] Natalie Jomini Stroud, “Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure,” *Political Behavior* 30, no. 3 (September 1, 2008): 341–66, https://doi.org/10.1007/s11109-007-9050-9.

[89] Emilee Rader and Rebecca Gray, “Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed” (ACM Press, 2015), 173–82, https://doi.org/10.1145/2702123.2702174.

[90] Frederik J. Zuiderveen Borgesius et al., “Should We Worry about Filter Bubbles?,” *Internet Policy Review*, March 31, 2016, https://policyreview.info/articles/analysis/should-we-worry-about-filter-bubbles.

[91] Borgesius et al.

[92] Pablo Barberá et al., “Tweeting from Left to Right: Is Online Political Communication More than an Echo Chamber?,” *Psychological Science* 26, no. 10 (2015): 1531–1542; Andrei Boutyline and Robb Willer, “The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks,” *Political Psychology* 38, no. 3 (June 1, 2017): 551–69, https://doi.org/10.1111/pops.12337.

[93] Susan Jacobson, Eunyoung Myung, and Steven L. Johnson, “Open Media or Echo Chamber: The Use of Links in Audience Discussions on the Facebook Pages of Partisan News Organisations,” *Information, Communication & Society* 19, no. 7 (July 2, 2016): 875–91, https://doi.org/10.1080/1369118X.2015.1064461.

[94] Dylko et al., “Impact of Customizability Technology on Political Polarization.”

[95] Dylko et al.

[96] Dylko et al.; Stroud, “Polarization and Partisan Selective Exposure.”

[97] “Horton, D and Wohl, R. (1956) ‘Mass Communication and Para-Social Interaction: Observation on Intimacy at a Distance’, Psychiatry, 19, Pp. 215-229.”

[98] Ulises A. Mejias and Nikolai E. Vokuev, “Disinformation and the Media: The Case of Russia and Ukraine,” *Media, Culture & Society*, 2017, 0163443716686672.

[99] Mejias and Vokuev.

[100] Lisa Kaati, “Det Digitala Kalifatet En Studie Av Islamiska Statens Propaganda” (FOI Totalförsvarets forskningsinstitut, May 2017).

[101] David A. Lake and Robert Powell, *Strategic Choice and International Relations* (Princeton University Press, 1999).

[102] Murray Edelman, *Politics as Symbolic Action: Mass Arousal and Quiescence* (Elsevier, 2013).

[103] Hugo Anderholm, “Why Is Russia Simulating Nuclear Strikes on Sweden?,” Vice, October 17, 2014, https://www.vice.com/sv/article/dpwk4q/why-is-russian-military-hanging-out-on-swedish-territory.

[104] Anderholm.

[105] L. John Martin, “Disinformation: An Instrumentality in the Propaganda Arsenal,” *Political Communication* 2, no. 1 (January 1982): 47–64, https://doi.org/10.1080/10584609.1982.9962747.

[106] This point has been expressed clearly by Oxford Reuters Institute's Rasmus Nielsen in an interview with Oxford Today, where Nielsen problematise the concept. Swedish journalist Jack Werner also provide a discussion of the issue on his blog. We chose to utilise the term in its strict sense despite such criticisms because we believe the term to be so commonly acknowledged that avoiding it would cause more confusion than clarity, although we agree with the aforementioned authors' (as well as others') points of criticism. Please see: Oxford Today, "Tackling Fake News - Interview with Rasmus Nielsen," Oxford Today, January 26, 2018, http://www.oxfordtoday.ox.ac.uk/features/tackling-fake-news#; and Jack Werner, "Därför ska du inte använda begreppet "fake news"," *Jack Werner* (blog), January 27, 2018, http://kwasbeb.se/2018/01/darfor-ska-du-inte-anvanda-begreppet-fake-news/.

[107] Hunt Allcott and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *The Journal of Economic Perspectives* 31, no. 2 (2017): 211–35, https://doi.org/10.2307/44235006.

[108] David Uberti, "The Real History of Fake News," *Columbia Journalism Review*, 2015, https://www.cjr.org/special_report/fake_news_history.php.

[109] "'Anyone Can Know': Citizen Journalism and the Interpretive Community of the Mainstream PressJournalism - Sue Robinson, Cathy DeShano, 2011," accessed December 6, 2017, http://journals.sagepub.com/doi/pdf/10.1177/1464884911415973.

[110] David Lazer et al., "Combating Fake News: An Agenda for Research and Action" (Combating Fake News: An agenda for Research and Action, Cambridge: Harvard University, 2017), https://shorensteincenter.org/wp-content/uploads/2017/05/Combating-Fake-News-Agenda-for-Research-1.pdf.

[111] Edson C. Tandoc, Zheng Wei Lim, and Richard Ling, "Defining 'Fake News': A Typology of Scholarly Definitions," *Digital Journalism*, August 30, 2017, 1–17, https://doi.org/10.1080/21670811.2017.1360143.

[112] Uberti, "The Real History of Fake News."

[113] Tandoc, Lim, and Ling, "Defining 'Fake News.'"

[114] Lazer et al., "Combating Fake News"; Filippo Menczer, "The Spread of Misinformation in Social Media" (ACM Press, 2016), 717–717, https://doi.org/10.1145/2872518.2890092; Jacob Ratkiewicz et al., "Detecting and Tracking Political Abuse in Social Media.," *ICWSM* 11 (2011): 297–304.

[115] Allcott and Gentzkow, "Social Media and Fake News in the 2016 Election."

[116] Soroush Vosoughi, Deb Roy, and Sinan Aral, "The Spread of True and False News Online," *Science* 359, no. 6380 (March 9, 2018): 1146–51, https://doi.org/10.1126/science.aap9559. It should be noted that the article only looked at fact checked articles (which only comprise a very narrow subset of legitimate articles), which limits the generalisability of the conclusions.

[117] Drawing on Tandoc, Lim, and Ling, "Defining 'Fake News'"; Claire Wardle, "Fake News. It's Complicated.," *First Draft News*, 2017, https://firstdraftnews.com:443/fake-news-complicated/; Lazer et al., "Combating Fake News"; "How to Fight Fake News and Misinformation? Research Helps Point the Way," MediaShift, December 28, 2016, http://mediashift.org/2016/12/fight-fake-news-misinformation-research-helps-point-way/; Uberti, "The Real History of Fake News."

[118] Wardle, "Fake News. It's Complicated."

[119] "Fake News: You Ain't Seen Nothing yet," *The Economist*, July 1, 2017, https://www.economist.com/news/science-and-technology/21724370-generating-convincing-audio-and-video-fake-events-fake-news-you-aint-seen.

[120] Nick Bilton, "Fake News Is About to Get Even Scarier than You Ever Dreamed," The Hive, 2017, https://www.vanityfair.com/news/2017/01/fake-news-technology.

[121] Tandoc, Lim, and Ling, "Defining 'Fake News'"; "Fake News."

[122] Tandoc, Lim, and Ling, "Defining 'Fake News.'"

[123] Maria Haigh, Thomas Haigh, and Nadine I. Kozak, "Stopping Fake News: The Work Practices of Peer-to-Peer Counter Propaganda," *Journalism Studies*, April 25, 2017, 1–26, https://doi.org/10.1080/1461670X.2017.1316681.

[124] Tandoc, Lim, and Ling, "Defining 'Fake News.'"

[125] Paul R. Brewer, Dannagal Goldthwaite Young, and Michelle Morreale, "The Impact of Real News about 'Fake News': Intertextual Processes and Political Satire," *International Journal of Public Opinion Research* 25, no. 3 (September 1, 2013): 323–43, https://doi.org/10.1093/ijpor/edt015.

[126] Wardle, "Fake News. It's Complicated."

[127] Tandoc, Lim, and Ling, "Defining 'Fake News.'"

[128] Tandoc, Lim, and Ling.

[129] Tandoc, Lim, and Ling.

[130] Tandoc, Lim, and Ling.

[131] J. Alexander and J. Smith, "Disinformation: A Taxonomy," *IEEE Security Privacy* 9, no. 1 (January 2011): 58–63, https://doi.org/10.1109/MSP.2010.141.

[132] The Independent, "Swedish Teenagers Claim Russian TV Crew Offered to Bribe Them to Cause Trouble after Trump Comments" accessed November 23, 2017, http://www.independent.co.uk/news/world/europe/wedish-teenagers-russian-tv-name-of-channel-crew-money-action-camera-donald-trump-refugee-rape-a7615406.html; Foreign Policy, "Russian TV Crew Tries to Bribe Swedish Youngsters to Riot on Camera," *Foreign Policy* (blog), accessed November 16, 2017, https://foreignpolicy.com/2017/03/07/russian-tv-crew-tries-to-bribe-swedish-youngsters-to-riot-on-camera-stockholm-rinkeby-russia-disinformation-media-immigration-migration-sweden/.

[133] Missiroli et al., "Strategic Communications - Countering Russia and ISIL/Daesh."

[134] John D. H. Downing, *Radical Media: Rebellious Communication and Social Movements* (SAGE, 2000).

[135] Adrian Chen, "The Agency," *The New York Times*, June 2, 2015, sec. Magazine, https://www.nytimes.com/2015/06/07/magazine/the-agency.html.

[136] Matthew Moore Media Correspondent, "Pro-Kremlin Hoaxers 'Posted Fake Guardian Article Online,'" *The Times*, August 15, 2017, sec. News, https://www.thetimes.co.uk/article/pro-kremlin-hoaxers-posted-fake-guardian-article-online-xxm8tbjc9.

[137] European Union Institute for Security Studies (EUISS), "EU Strategic Communications with a View to Counteracting Propaganda - EU Law and Publications," Paper (Brussels: DIRECTORATE-GENERAL FOR EXTERNAL POLICIES, June 20, 2016), https://publications.europa.eu/en/publication-detail/-/publication/26fb6e07-3772-11e6-a825-01aa75ed71a1/language-en.

[138] European Union Institute for Security Studies (EUISS).

[139] ENISA, "Disinformation Operations in Cyber-Space," Cyber security info note, 2017, https://www.enisa.europa.eu/publications/info-notes/disinformation-operations-in-cyber-space.

[140] Adam Hulcoop et al., "Tainted Leaks: Disinformation and Phishing With a Russian Nexus," The Citizen Lab, May 25, 2017, https://citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/.

[141] Hulcoop et al.

[142] Megha Mohan, "Macron Leaks: The Anatomy of a Hack," *BBC News*, May 9, 2017, sec. BBC Trending, http://www.bbc.com/news/blogs-trending-39845105.

[143] ENISA, "Disinformation Operations in Cyber-Space."

[144] United States Information Agency, "Soviet Active Measures in the 'Post-Cold War' Era 1988-1991" (Washington DC: United States Information Agency, 1992), http://intellit.muskingum.edu/russia_folder/pcw_era/index.htm.

[145] ENISA, "Disinformation Operations in Cyber-Space."

[146] Cook and Lewandowsky, *The Debunking Handbook*.

[147] Hulcoop et al., "Tainted Leaks."

[148] Neil MacFarquhar, "A Powerful Russian Weapon: The Spread of False Stories," *The New York Times*, August 28, 2016, sec. Europe, https://www.nytimes.com/2016/08/29/world/europe/russia-sweden-disinformation.html.

[149] "KREMLIN INFOWAR LAID BARE: MH17 Findings Expose Depths of Russian Disinformation," accessed December 11, 2017, http://bunews.com.ua/politics/item/kremlin-infowar-laid-bare-mh17-findings-expose-depths-of-russian-disinformation.

[150] Hulcoop et al., "Tainted Leaks."

[151] Man-pui Sally Chan et al., "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation," *Psychological Science*, September 12, 2017, 0956797617714579, https://doi.org/10.1177/0956797617714579.

[152] "Hultqvists Underskrift På Förfalskat Brev," SVT Nyheter, accessed December 4, 2017, https://www.svt.se/nyheter/svtforum/frammande-makt-forfalskade-brev.

[153] "Hultqvists Underskrift På Förfalskat Brev."

[154] Martin Kragh and Sebastian Åsberg, "Russia's Strategy for Influence through Public Diplomacy and Active Measures: The Swedish Case," *Journal of Strategic Studies* 40, no. 6 (September 19, 2017): 773–816, https://doi.org/10.1080/01402390.2016.1273830.

[155] Hulcoop et al., "Tainted Leaks."

[156] MacFarquhar, "A Powerful Russian Weapon."

[157] Hagen Schölzel and Howard Nothhaft, "The Establishment of Facts in Public Discourse: Actor-Network-Theory as a Methodological Approach in PR-Research," *Public Relations Inquiry* 5, no. 1 (January 1, 2016): 53–69, https://doi.org/10.1177/2046147X15625711; Howard Nothhaft, "Dealing in facts", in Dejan Vercic and Elizabeth Bridgen (eds.) *Experiencing Public Relations – Internal voices,* (Routledge, London, 2018).

[158] Oreskes and Conway, *Merchants of Doubt*.

[159] Oreskes and Conway.

[160] David J. Moore and Richard Reardon, "Source Magnification: The Role of Multiple Sources in the Processing of Advertising Appeals," *Journal of Marketing Research* 24, no. 4 (1987): 412–17, https://doi.org/10.2307/3151389.

[161] Avi Selk, "Internet Trolls Have Declared War on Shia LaBeouf's Anti-Trump Project," chicagotribune.com, April 3, 2017, http://www.chicagotribune.com/bluesky/technology/ct-shia-labeouf-4chan-wp-bsi-20170403-story.html.

[162] Igal Zeifman, "Bot Traffic Report 2016," Incapsula Blog, January 24, 2017, https://www.incapsula.com/blog/bot-traffic-report-2016.html.

[163] K. Michael, "Bots Trending Now: Disinformation and Calculated Manipulation of the Masses [Editorial]," *IEEE Technology and Society Magazine* 36, no. 2 (June 2017): 6–11, https://doi.org/10.1109/MTS.2017.2697067.

[164] "Types of Bots: An Overview of Chatbot Diversity | Botnerds.Com," *Botnerds* (blog), accessed November 30, 2017, http://botnerds.com/types-of-bots/.

[165] "Different Types of Bots | What Are Bad Bots | Bot Defenition," ShieldSquare, accessed November 23, 2017, https://www.shieldsquare.com/what-are-the-different-types-of-bots/; "Types of Bots."

[166] Michael, "Bots Trending Now."

[167] Zeifman, "Bot Traffic Report 2016"; Adrienne LaFrance, "The Internet Is Mostly Bots," *The Atlantic*, January 31, 2017, https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043/.

[168] "Different Types of Bots | What Are Bad Bots | Bot Defenition"; Norah Abokhodair, Daisy Yoo, and David W. McDonald, "Dissecting a Social Botnet: Growth, Content and Influence in Twitter," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (ACM, 2015), 839–851; Sergey Sanovich, "Computational Propaganda in Russia - The Origins of Digital Misinformation," Working Paper, Computationa Propaganda Research Project (Oxford: Oxford Internet Institute, 2017).

[169] Luca Maria Aiello et al., "People Are Strange When You're a Stranger: Impact and Influence of Bots on Social Networks," *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media* 697, no. 483,151 (2012): 10–17.

[170] Kris Shaffer, "Spot a Bot: Identifying Automation and Disinformation on Social Media," *Medium* (blog), June 5, 2017, https://medium.com/data-for-democracy/spot-a-bot-identifying-automation-and-disinformation-on-social-media-2966ad93a203.

[171] Sanovich, "Computational Propaganda in Russia - The Origins of Digital Misinformation."

[172] Chengcheng Shao et al., "Hoaxy: A Platform for Tracking Online Misinformation," *ArXiv:1603.01511 [Physics]*, 2016, 745–50, https://doi.org/10.1145/2872518.2890098.

[173] Ratkiewicz et al., "Detecting and Tracking Political Abuse in Social Media."

[174] Johnnatan Messias et al., "You Followed My Bot! Transforming Robots into Influential Users in Twitter," *First Monday* 18, no. 7 (June 19, 2013), http://firstmonday.org/ojs/index.php/fm/article/view/4217.

[175] Lazer et al., "Combating Fake News."

[176] For a more in depth discussion on how false amplifiers are used on social media, see Facebook's own publication: Jen Weedon, William Nuland, and Alex Stamos, "Information Operations and Facebook" (Facebook, April 27, 2017), https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf.

[177] Cull, "Counter Propaganda - Cases from US Public Diplomacy and Beyond."

[178] Abokhodair, Yoo, and McDonald, "Dissecting a Social Botnet."

[179] Michael, "Bots Trending Now."

[180] Onur Varol et al., "Online Human-Bot Interactions: Detection, Estimation, and Characterization," *ArXiv Preprint ArXiv:1703.03107*, 2017; and Michael Newberg, "Nearly 48 Million Twitter Accounts Could Be Bots, Says Study," March 10, 2017, https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html.

[181] Alex Hern, "How Social Media Filter Bubbles and Algorithms Influence the Election," *The Guardian*, May 22, 2017, sec. Technology, http://www.theguardian.com/technology/2017/may/22/social-media-election-facebook-filter-bubbles.

[182] Yazan Boshmaf et al., "Design and Analysis of a Social Botnet," *Computer Networks* 57, no. 2 (2013): 556–578.

[183] "What Is a DDoS Botnet | Common Botnets and Botnet Tools | Incapsula," accessed November 30, 2017, https://www.incapsula.com/ddos/botnet-ddos.html.

[184] Alison DeNisco Rayome | November 8, 2016, and 7:48 Am Pst, “Hackers Attempt DDoS Attacks on Clinton and Trump Campaign Websites Using Mirai Botnet,” TechRepublic, accessed November 30, 2017, https://www.techrepublic.com/article/hackers-attempt-ddos-attacks-on-clinton-and-trump-campaign-websites-using-mirai-botnet/.

[185] Drawing on: Shaffer, “Spot a Bot”; @DFRLab, “#BotSpot: Twelve Ways to Spot a Bot,” *DFRLab* (blog), August 28, 2017, https://medium.com/dfrlab/botspot-twelve-ways-to-spot-a-bot-aedc7d9c110c; Lutz Finger, “How To Spot Social Media Bots - They Are Often Lonely,” Forbes, accessed November 14, 2017, https://www.forbes.com/sites/lutzfinger/2015/02/24/how-to-spot-social-media-bots-they-are-often-lonely/.

[186] @DFRLab, “#BotSpot.”

[187] A. Paradise, R. Puzis, and A. Shabtai, “Anti-Reconnaissance Tools: Detecting Targeted Socialbots,” *IEEE Internet Computing* 18, no. 5 (September 2014): 11–19, https://doi.org/10.1109/MIC.2014.81; Abokhodair, Yoo, and McDonald, “Dissecting a Social Botnet”; Varol et al., “Online Human-Bot Interactions”; Aiello et al., “People Are Strange When You’re a Stranger”; Clayton Allen Davis et al., “BotOrNot: A System to Evaluate Social Bots” (ACM Press, 2016), 273–74, https://doi.org/10.1145/2872518.2889302.

[188] Elise Moreau, “Internet Trolls and the Many Ways They Try to Ruin Your Day,” Lifewire, accessed December 1, 2017, https://www.lifewire.com/types-of-internet-trolls-3485894.

[189] Bryn Alexander Coles and Melanie West, “Trolling the Trolls: Online Forum Users Constructions of the Nature and Properties of Trolling,” *Computers in Human Behavior* 60 (July 2016): 233–44, https://doi.org/10.1016/j.chb.2016.02.070.

[190] Susan Herring et al., “Searching for Safety Online: Managing ‘Trolling’ in a Feminist Forum,” *The Information Society* 18, no. 5 (October 2002): 371–84, https://doi.org/10.1080/01972240290108186.

[191] “The Art of Trolling: A Philosophical History of Rhetoric | The Artifice,” accessed November 27, 2017, https://the-artifice.com/art-of-trolling/.

[192] “How to Deal with Trolls,” accessed November 27, 2017, https://www.webroot.com/au/en/home/resources/tips/pc-security/you-cant-win-an-argument-with-a-troll.

[193] NATO, “Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia.”

[194] Moreau, “Internet Trolls and the Many Ways They Try to Ruin Your Day”; NATO, “Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia.”

[195] NATO, “Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia.”

[196] Haigh, Haigh, and Kozak, “Stopping Fake News.”

[197] NATO, “Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia.”

[198] Shaun Walker, “Salutin’ Putin: Inside a Russian Troll House,” *The Guardian*, April 2, 2015, sec. World news, http://www.theguardian.com/world/2015/apr/02/putin-kremlin-inside-russian-troll-house.

[199] NATO, “Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia.”

[200] Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011).

[201] For a basic overview of different trolling techniques see: “Trollsnack – En Liten Guide till Näthatets Retorik,” SVT Nyheter, accessed November 27, 2017, https://www.svt.se/kultur/medier/trollsnack.

[202] Sveriges Radio, “Trollattacker om Sverige i Italien - Nyheter (Ekot),” 2018, http://sverigesradio.se/sida/artikel.aspx?programid=83&artikel=6893460.

[203] Sveriges Radio.

[204] Gary King, Jennifer Pan, and Margaret E. Roberts, "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument," *American Political Science Review* 111, no. 03 (August 2017): 484–501, https://doi.org/10.1017/S0003055417000144.

[205] King, Pan, and Roberts; Kaveh Waddell, "'Look, a Bird!' Trolling by Distraction," *The Atlantic*, January 27, 2017, https://www.theatlantic.com/technology/archive/2017/01/trolling-by-distraction/514589/.

[206] Lucas and Pomeranzev, "Winning the Information War."

[207] Rob Weatherhead, "Say It Quick, Say It Well – the Attention Span of a Modern Internet Consumer," *The Guardian*, February 28, 2014, sec. Media Network, http://www.theguardian.com/media-network/media-network-blog/2012/mar/19/attention-span-internet-consumer.

[208] For a discussion on how this works for TV commercials, see: Nigel Hollis, "Why Funny TV Commercials Work," *The Atlantic*, October 27, 2011, https://www.theatlantic.com/business/archive/2011/10/why-funny-tv-commercials-work/247117/.

[209] NATO, "StratCom Laughs. In Search of an Analytical Framework" (Riga: NATO Stratetic Communications Centre of Excellence, 2017).

[210] NATO.

[211] Michael B. Prosser, "Memetics–A Growth Industry in US Military Operations" (MARINE CORPS UNIV QUANTICO VA SCHOOL OF ADVANCED WARFIGHTING (SAW), 2006).

[212] Joel Harding, "Can NATO Weaponize Memes?," *To Inform Is to Influence* (blog), April 14, 2017, https://toinformistoinfluence.com/2017/04/14/can-nato-weaponize-memes/; jensganman, "MEME MAGIC & SHITPOSTING – Så Kommer Valet 2018 Att Vinnas," *Säg Att Du Skojar* (blog), February 13, 2017, https://jensganman.wordpress.com/2017/02/13/meme-magic-shitposting-sa-kommer-valet-2018-att-vinnas/; Jacob Siegel, "Is America Prepared for Meme Warfare?," Motherboard, January 31, 2017, https://motherboard.vice.com/en_us/article/xyvwdk/meme-warfare; "Applications of the Memetic Perspective in Inform and Influence Operations | Small Wars Journal," accessed November 30, 2017, http://smallwarsjournal.com/jrnl/art/applications-of-the-memetic-perspective-in-inform-and-influence-operations.

[213] Siegel, "Is America Prepared for Meme Warfare?"

[214] Richard Dawkins, *The Selfish Gene* (Oxford University Press, 1989).

[215] Prosser, "Memetics–A Growth Industry in US Military Operations."

[216] NATO, "StratCom Laughs. In Search of an Analytical Framework."

[217] Alice Marwick and Rebecca Lewis, "Media Manipulation and Disinformation Online," *New York: Data & Society Research Institute*, 2017.

[218] "Applications of the Memetic Perspective in Inform and Influence Operations | Small Wars Journal."

[219] Prosser, "Memetics–A Growth Industry in US Military Operations"; Dawkins, *The Selfish Gene*.

[220] Dawkins, *The Selfish Gene*.

[221] NATO, "StratCom Laughs. In Search of an Analytical Framework."

[222] NATO.

[223] "Gilbert Gottfried on His Infamous 9/11 Joke and 'Too Soon,'" Vulture, accessed December 1, 2017, http://www.vulture.com/2016/02/gilbert-gottfried-on-his-911-joke-too-soon.html.

[224] Jack Kenrick, “4chan Launches ‘Operation Swedistan’ In A Last Ditch Attempt To Save Sweden,” *Squawker* (blog), November 14, 2017, https://squawker.org/culture-wars/swedistan/.

[225] Björn Palmertz, “Theoretical Foundations of Influence Operations: A Review of Relevant Psychological Research” (Stockholm: Center for Asymmetric Threat Studies (CATS), Swedish National Defence College, n.d).

[226] NATO, “StratCom Laughs. In Search of an Analytical Framework.”

[227] NATO.

[228] Siegel, “Is America Prepared for Meme Warfare?”

[229] Siegel.

[230] “Trollsnack – En Liten Guide till Näthatets Retorik”; Cull, “Counter Propaganda - Cases from US Public Diplomacy and Beyond.”

[231] Cull, “Counter Propaganda - Cases from US Public Diplomacy and Beyond.”

[232] Cull.

[233] Paul M. A. Linebarger, *Psychological Warfare* (Pickle Partners Publishing, 2015).

[234] An analysis of these techniques can be found in Vox Day, *SJWs Always Lie: Taking Down the Thought Police* (Kouvola: Castalia House, 2015); Vox Day, *SJWs Always Double Down: Anticipating the Thought Police (The Laws of Social Justice Book 2)* (Kouvola: Castalia House, 2017), https://www.amazon.com/SJWs-Always-Double-Down-Anticipating-ebook/dp/B075BGGKLG/ref=sr_1_1?s=digital-text&ie=UTF8&qid=1513269994&sr=1-1&keywords=always+double+down.

[235] See Waltzman 2017, Dauber, Cori E. 2009 for the following.

[236] Cori E. Dauber, “The Truth Is out There: Responding to Insurgent Disinformation and Deception Operations,” *Military Review* 89, no. 1 (January 1, 2009): 13.

[237] Dauber.

[238] Moore and Reardon, “Source Magnification.”

[239] Sputnik, “Sweden Getting Ready to Fire Missiles at Russian Troops From Gotland Island,” 2015, https://sputniknews.com/military/201507161024701419/.

[240] Kevin Trenberth, “Statement: Kevin Trenberth on Hacking of Climate Files,” CGD’s Climate Analysis Section, 2010, https://web.archive.org/web/20100611224809/http://www.cgd.ucar.edu/cas/Trenberth/statement.html.

[241] Christopher Paul and Miriam Matthews, “The Russian ‘Firehose of Falsehood’ Propaganda Model - Why It Might Work and Options to Counter It,” Expert insights on a timely policy issue (RAND Corporation, 2016), http://www.rand.org/content/dam/rand/pubs/perspectives/PE100/PE198/RAND_PE198.pdf.

[242] Paul and Matthews.

[243] King, Pan, and Roberts, “How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument.”

[244] King, Pan, and Roberts.

[245] Waddell, “‘Look, a Bird!’ Trolling by Distraction.”

[246] King, Pan, and Roberts, “How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument.”

[247] King, Pan, and Roberts.

[248] Waddell, “‘Look, a Bird!’ Trolling by Distraction.”

[249] “Pool’s Closed,” Know Your Meme, accessed December 12, 2017, http://knowyourmeme.com/memes/pools-closed.

[250] Robert S. Mueller, “Indictment Case 1:18-Cr-00032-DLF United States of America v. Internet Research Agency” (United States District Court for the District of Columbia, 2018), https://www.justice.gov/file/1035477/download.

[251] Mohan, “Macron Leaks”; Hulcoop et al., “Tainted Leaks.”

[252] Mueller, “Indictment Case 1:18-Cr-00032-DLF United States of America v. Internet Research Agency.”

[253] Office of the Director of National Intelligence, “Assessing Russian Activities and Intentions in Recent US Elections,” January 6, 2017, https://www.dni.gov/files/documents/ICA_2017_01.pdf.

[254] Neil MacFarquhar, “Inside the Russian Troll Factory: Zombies and a Breakneck Pace,” *The New York Times*, February 18, 2018, sec. Europe, https://www.nytimes.com/2018/02/18/world/europe/russia-troll-factory.html.

[255] Mueller, “Indictment Case 1:18-Cr-00032-DLF United States of America v. Internet Research Agency.”

[256] Mueller.

[257] Mueller.

[258] Adam Taylor, “Analysis | The Russian Journalist Who Helped Uncover Election Interference Is Confounded by the Mueller Indictments,” *Washington Post*, February 18, 2018, https://www.washingtonpost.com/news/worldviews/wp/2018/02/18/the-russian-journalist-who-helped-uncover-election-meddling-is-confounded-by-the-mueller-indictments/; MacFarquhar, “Inside the Russian Troll Factory.”

[259] Mueller, “Indictment Case 1:18-Cr-00032-DLF United States of America v. Internet Research Agency.”

[260] Mueller.

[261] This taxonomy synthesises views from a variety of different contexts to provide ideal types of different approaches and does not perfectly correspond to any one actor’s approach or modus operandi. Rather, one actor can at the same time operate within multiple approaches, and utilize tools from different approaches in a coordinated manner.

[262] Jean-Baptiste Jeangène Vilmer, “La lutte contre la désinformation russe : contrer la propagande sans faire de contre-propagande ?” (l’Institut de recherche stratégique de l’École militaire (IRSEM), 2017), http://www.jbjv.com/IMG/pdf/JBJV_2017_-_La_Lutte_contre_la_desinformation_russe.pdf.

[263] Lucas and Pomeranzev, “Winning the Information War.”

[264] Lucas and Pomeranzev; Alistair Shawcross, “Facts We Can Believe In: How to Make Fact-Checking Better” (London: The Legatum Institute, 2016).

[265] “Medier startar samarbete mot falska nyheter,” SVT Nyheter, January 25, 2018, https://www.svt.se/kultur/medier/svenska-nyhetsmedier-startar-samarbete-for-faktagranskning-under-valet.

[266] Jonathan Stray, “Defense Against the Dark Arts: Networked Propaganda and Counter-Propaganda", 2017, http://jonathanstray.com/networked-propaganda-and-counter-propaganda.

[267] Atlantic Council, “Democratic Self-Defence against Disinformation Operations (Draft)” (Atlantic Council, 2017); Lucas and Pomeranzev, “Winning the Information War.”

[268] “Experts Appointed to the High-Level Group on Fake News and Online Disinformation,” Digital Single Market, accessed February 20, 2018, https://ec.europa.eu/digital-single-market/en/news/experts-appointed-high-level-group-fake-news-and-online-disinformation.

[269] Edward Lucas and Ben Nimmo, “Information Warfare: What Is It and How to Win It?” (Center for European Policy Analysis, 2015).

[270] Alicia Kearns, “The Democratisation of Hybrid Warfare and Its Implications for Defeating Violent Extremism and Tackling Propaganda,” in *Countering Online Propaganda and Violent Extremism: The Dark Side of Digital Diplomacy*, ed. James Pamment and Corneliu Bjola (Oxon: Routledge, Forthcoming).

[271] European Union Institute for Security Studies (EUISS), “EU Strategic Communications with a View to Counteracting Propaganda - EU Law and Publications.”

[272] Cull, “Counter Propaganda - Cases from US Public Diplomacy and Beyond.”

[273] Cull.

[274] For an overview of the Swedish approach , see Niklas H. Rossbach “Psychological Defence: Vital for Sweden’s Defence Capability“ in “Strategisk Utblick 7: Närområdet Och Nationell Säkerhet.”(FOI, 2017)

[275] Maria Hellman and Charlotte Wagnsson, “How Can European States Respond to Russian Information Warfare? An Analytical Framework,” *European Security* 26, no. 2 (April 3, 2017): 153–70, https://doi.org/10.1080/09662839.2017.1294162.

[276] Hellman and Wagnsson.

[277] Daniel Fried and Alina Polyakova, “Democratic Defense Against Disinformation” (Washington, DC: The Atlantic Council of the United States, 2018), http://www.atlanticcouncil.org/images/publications/Democratic_Defense_Against_Disinformation_FINAL.pdf.

[278] Shannon Bond, “Half of Americans Want to Regulate News on Social Media,” Financial Times, January 16, 2018, https://www.ft.com/content/d3d63680-fa2e-11e7-9b32-d7d59aace167.

[279] James McAuley, “France Weighs a Law to Rein in ‘Fake News,’ Raising Fears for Freedom of Speech,” *Washington Post*, January 10, 2018, sec. Europe, https://www.washingtonpost.com/world/europe/france-weighs-a-law-to-rein-in-fake-news-raising-fears-for-freedom-of-speech/2018/01/10/78256962-f558-11e7-9af7-a50bc3300042_story.html.

[280] Paul and Matthews, “The Russian ‘Firehose of Falsehood’ Propaganda Model - Why It Might Work and Options to Counter It.”

[281] Nicholas Confessore, “The Follower Factory,” *The New York Times*, January 27, 2018, sec. Technology, https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html, https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html.

[282] The importance of considering strategic communication as these elements (function, process, mindset) is elaborated in: NATO, “NATO Strategic Communication Handbook v1.0” (NATO, 2017).

[283] With reference to information influence campaigns this point has been argued strongly by the Oxford scholar Rasmus Nielsen, see for example: Oxford Today, “Tackling Fake News - Interview with Rasmus Nielsen.”

[284] Steve Tatham, “The Solution to Russian Propaganda Is Not EU or NATO Propaganda but Advanced Social Science to Understand and Mitigate Its Effect in Targeted Populations” (Riga: National Defence Academy of Latvia, 2015), http://www.stratcomcoe.org/steve-tatham-solution-russian-propaganda-not-eu-or-nato-propaganda-advanced-social-science.

[285] The European Centre of Excellence for Countering Hybrid Threats, https://www.hybridcoe.fi/

[286] NATO Strategic Communication Centre of Excellence, https://www.stratcomcoe.org

[287] Myndigheten för samhällsskydd och beredskap, https://www.msb.se/

[288] Missiroli et al., “Strategic Communications - Countering Russia and ISIS/Daesh.”

[289] Edward Lucas, “Baltic Sea Security Report - The Coming Storm” (Center for European Policy Analysis, June 2015).

[290] Cull, “Counter Propaganda - Cases from US Public Diplomacy and Beyond.” Christopher Paul and Miriam Matthews, “The Russian ‘Firehose of Falsehood’ Propaganda Model - Why It Might Work and Options to Counter It,” Expert insights on a timely policy issue (RAND Corporation, 2016); Gunnar Sjöstedt and Paula Stenström, “Vilseledning På Internet” (Stockholm: Styrelsen för Psykologiskt Försvar, 2002).

[291]Fried and Polyakova, “Democratic Defense Against Disinformation.”

[292] BBC, “Twitter Bans Ads from RT and Sputnik,” *BBC News*, October 26, 2017, sec. US & Canada, http://www.bbc.com/news/world-us-canada-41766991.

[293] Björn Palmertz, “Europeiska Perspektiv På Förmågan Att Möta Påverkanskampanjer Från Främmande Makt - Delrapport 1” (Stockholm: Center for Asymmetric Threat Studies (CATS), Swedish National Defence College, 2016).

[294] Tatham, “The Solution to Russian Propaganda Is Not EU or NATO Propaganda but Advanced Social Science to Understand and Mitigate Its Effect in Targeted Populations.”

[295] NATO StratCom COE, *Quo Vadis “Question More”? | The Riga StratCom Dialogue 2017*, 2017, https://www.youtube.com/watch?v=XOUGsYnT5zM.

[296] Niklas H Rossbach, “Psykologiskt Försvar - Avgörande För Svensk Försvarsförmåga,” Strategisk Utblick 7: Närområdet Och Nationell Säkerhet (FOI Totalförsvarets forskningsinstitut, 2017).

[297] Edward Lucas, *The New Cold War: Putin’s Russia and the Threat to the West*, 3 edition (New York, NY: St. Martin’s Griffin, 2014).

[298] Garth S. Jowett and Victoria O’Donnell, *Propaganda & Persuasion* (SAGE, 2011).

[299] Palmertz, “Europeiska Perspektiv På Förmågan Att Möta Påverkanskampanjer Från Främmande Makt - Delrapport 1.”

[300] Daniel Milo and Katarína Klingová, “Countering Information War Lessons Learned from NATO and Partner Countries: Recommendations and Conclusions” (Bratislava: Globsec, 2016), https://www.globsec.org/wp-content/uploads/2017/09/countering_information_war.pdf.

[301] High level Group on fake news and online disinformation, “A Multi-Dimensional Approach to Disinformation” (Belgium: European Union, 2018), http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271; Fried and Polyakova, “Democratic Defense Against Disinformation.”

[302] Björn Palmertz, “Att Identifiera, Förstå Och Möta Påverkanskampanjer Från Främmande Makt En Översikt Av Verktyg Och Metoder” (Stockholm: Center for Asymmetric Threat Studies (CATS), Försvarshögskolan, 2015).

[303] Lucas and Pomeranzev, “Winning the Information War.”

[304] Twiplomacy, “Coder or Diplomat? @EladRatson, the King of Algorithmic Diplomacy. @IsraelMFA Developed Software: - To Identify Negative Voices and Contain the Spread of Violent Content - to Identify the Voices of Reason & Interlink Them to Spread the Positive Message of @Israel. #DDconf2017pic.Twitter.Com/7PQqPuwHqY,” Tweet, *@Twiplomacy* (blog), December 3, 2017, https://twitter.com/Twiplomacy/status/938368576544788480.

[305] Politifact, http://www.politifact.com/

[306] Paul and Matthews, “The Russian ‘Firehose of Falsehood’ Propaganda Model - Why It Might Work and Options to Counter It.”

[307] Chan et al., “Debunking.”

[308] Cook and Lewandowsky, *The Debunking Handbook*.

[309] Shawcross, “Facts We Can Believe In: How to Make Fact-Checking Better.”

[310] Chan et al., “Debunking.”

[311] Jonas Eriksson et al., *Vägledning för risk- och sårbarhetsanalyser* (Karlstad: Myndigheten för samhällsskydd och beredskap, 2011).

[312] Eriksson et al.

[313] Eriksson et al.

[314] Gerry Osborne, “Strategic Communications: Insights from the Commercial Sector,” NATO StratCom Centre of Excellence, (Riga: NATO StratCom Centre of Excellence, 2017).

[315] Paul and Matthews, “The Russian ‘Firehose of Falsehood’ Propaganda Model - Why It Might Work and Options to Counter It.”

[316] Silverman et al., “The Impact of Counter-Narratives.” Missiroli et al., “Strategic Communications - Countering Russia and ISIS/Daesh.” NATO, “New Trends in Social Media” (Riga: NATO Stratetic Communications Centre of Excellence, 2016).

[317] Palmertz, “Europeiska Perspektiv På Förmågan Att Möta Påverkanskampanjer Från Främmande Makt - Delrapport 1.”

[318] Stray, “Defense Against the Dark Arts.”

[319] Miskimmon, Alister, Ben O'Loughlin, and Laura Roselle. *Strategic narratives: Communication power and the new world order*. Vol. 3. Routledge, 2014.

[320] Pamment, James. "Strategic narratives in US public diplomacy: A critical geopolitics." *Popular Communication* 12.1 (2014): 48-64.

[321] See for example Örebro municipality’s social media guidelines for an example of how this could look: Örebro kommun, “Riktlinjer För Sociala Medier,” 2016, https://www.orebro.se/download/18.2bea29ad1590bf258c52a22/1484207071633/Riktlinjer+f%C3%B6r+sociala+medier+i+%C3%96rebro+kommun.pdf.

[322] Ian Griggs, “Inside the Met Police Comms Response to the Westminster Attacks,” 2017, https://www.prweek.com/article/1432250.

[323] Cook and Lewandowsky, *The Debunking Handbook*.Cook and Lewandowsky, *The Debunking Handbook*.Cook and Lewandowsky, *The Debunking Handbook*.Cook and Lewandowsky, *The Debunking Handbook*.

[324] Cook and Lewandowsky.

[325] Cook and Lewandowsky.

[326] Gerd Gigerenzer and Peter M. Todd, *Simple Heuristics That Make Us Smart* (Oxford University Press, 1999).

[327] Kahneman, *Thinking, Fast and Slow*.

[328] Cook and Lewandowsky, *The Debunking Handbook*.

[329] Cook and Lewandowsky.Cook and Lewandowsky.Cook and Lewandowsky.Cook and Lewandowsky.

[330] Schacter, *The Seven Sins of Memory*.

[331] Chan et al., “Debunking.”Chan et al., “Debunking.”Chan et al., “Debunking.”Chan et al., “Debunking.”

[332] Waltzman, “The Weaponization of Information - The Need for Cognitive Security.”

[333] Palmertz provides a comprehensive overview of the social psychological features relevant to information influence campaigns. See: Palmertz, “Theoretical Foundations of Influence Operations: A Review of Relevant Psychological Research.”Palmertz, “Theoretical Foundations of Influence Operations: A Review of Relevant Psychological Research.”

[334] Steven Sloman and Philip Fernbach, *The Knowledge Illusion: Why We Never Think Alone* (Riverhead Books, 2017).

[335] Kahneman, *Thinking, Fast and Slow*.

[336] Sharon S. Brehm and Jack Williams Brehm, *Psychological Reactance: A Theory of Freedom and Control* (Academic Press, 1981); Stephan Lewandowsky et al.,

"Misinformation and Its Correction: Continued Influence and Successful Debiasing," *Psychological Science in the Public Interest* 13, no. 3 (December 2012): 106–31, https://doi.org/10.1177/1529100612451018.

[337] Cull, "Counter Propaganda - Cases from US Public Diplomacy and Beyond."

[338] Chan et al., "Debunking."

[339] Pontus Winther, "Yttrandefrihetsgrundlagen Och Möjligheterna Att Möta Påverkanskampanjer Från Främmande Makt" (Myndigheten för samhällsskydd och beredskap, 2016).

[340] Winther.