COUNTERING MALICIOUS USE OF SOCIAL MEDIA

THE ROLE OF COMMUNICATORS IN COUNTERING THE MALICIOUS USE OF SOCIAL MEDIA

James Pamment et al.

ISBN 978-9934-564-31-4

ISBN: 978-9934-564-31-4
Authors: James Pamment, Howard Nothhaft, Henrik Twetman, and Alicia Fjällhed
Project manager: Sebastian Bay
Text editor: Anna Reynolds
Design: Kārlis Ulmanis

Riga, November 2018
NATO STRATCOM COE
11b Kalciema Iela
Riga LV1048, Latvia
www.stratcomcoe.org
Facebook/stratcomcoe
Twitter: @stratcomcoe

James Pamment is Head of the Department of Strategic Communication at Lund University and a senior analyst at the Centre for Asymmetric Threats Studies (CATS) at the Swedish National Defence University.

Howard Nothhaft is associate professor at the Department of Strategic Communication at Lund University.

Henrik Agardh-Twetman is a project manager at the Department of Strategic Communication at Lund Univeristy.

Alicia Fjällhed is a PhD-student at the Department of Strategic Communication at Lund University.

Introduction

This brief discusses the role of communicators in countering the malicious use of social media. It is based on the report 'Countering Information Influence Activities: The State of the Art' (2018) developed by the Department of Strategic Communication at Lund University and published by the Swedish Civil Contingency Agency (MSB).

This brief is divided into three sections: understanding, identifying, and counteracting information influence activities. The *Understanding* section covers defnitions, diagnostics, and vulnerabilities. *Identifying* provides a basis for analysing the narratives, target groups, and techniques used in information influence activities. *Counteracting* covers preparation, action, and learning.

Understanding

Information influence activities should now be counted among the many techniques hostile actors employ to negatively impact democratic societies together with activities such as espionage, cyber threats, and the deployment of irregular forces. These are all part of the toolkit used in what we now call hybrid, asymmetrical, or unconventional warfare. Information influence activities are used to manipulate public opinion, disturb elections, isolate vulnerable social groups, and destabilise entire regions and countries. For the purposes of this brief, we understand information influence activities as the *illegitimate efforts of foreign powers or their proxies to influence the perceptions, behaviour, and decisions of target groups to the benefit of foreign powers*. Our focus is on how information influence activities can be used to exploit the vulnerabilities in media systems, our cognitive biases, and in public opinion formation processes.

While using information to influence others is not a new phenomenon, the speed, reach, and intensity through which information can be disseminated online creates new challenges, especially in relation to social media. Information influence operators leverage their understanding of how the modern information environment functions to achieve the effects they desire. This includes activities such as

> We understand information influence activities as the *illegitimate efforts of foreign powers or their proxies to influence the perceptions, behaviour, and decisions of target groups to the benefit of foreign powers*

manipulation of social media algorithms, data collection for use in sociographic targeting, doxing, impersonation, and other malicious operations.

The concept of *legitimacy* is the cornerstone of our approach to understanding these activities. We use the words 'malicious' and 'illegitimate' to highlight the fact that information influence activities mimic legitimate forms of communication such as advocacy, public relations, lobbying, advertising, and public debate on social media platforms to undermine and pervert the rational formation of opinion. They are illegitimate because they use falsehoods to poison the principles of communication essential to the healthy functioning of democracies, such as free deliberation and debate, factual bases for claims/knowledge, rational public opinion formation, and mutual trust.

Information influence activities also exploit trends in the *digital media system*, such as the inaccessibility of quality journalism behind paywalls, the culture of 'clickbait' headlines, and the prevalence of unvetted citizen journalism. The results include untrustworthy news sources mixing spurious information with real news, news articles designed to provoke outrage, and hostile foreign actors posing as citizens involved in democratic debate. By leveraging the personal data most people expose daily on social media, such falsehoods can be accurately targeted toward specific audiences to create a desired impact.

Information influence further exploits the social character of *public opinion formation*, for example through closed chat groups, friends sharing stories they haven't read, and bots that boost circulation numbers. The results can include secret debates based on false or manipulated sources, the appearance of friends approving of information they share, and the false appearance of a broad public debate. Information influence techniques also exploit *shortcuts in our thinking* (heuristics) either by learning about us through the data we share, or by applying 'nudges' that manipulate our cognitive biases. The results of these techniques can include targeting based on psychological traits and framing information to induce logical shortcuts or to nudge decision-making in some pre-determined preferred direction.

Finally, information influence activities contribute to the breakdown of social trust and social cohesion—political debate becomes polarised and political decision-

> Information influence techniques also exploit *shortcuts in our thinking* (heuristics) either by learning about us through the data we share, or by applying 'nudges' that manipulate our cognitive biases

making becomes more difficult. When groups within a society mistrust one another they become more easily agitated, and government institutions lose credibility, which can lead to negative consequences for public health and safety. Information influence activities disrupt social cohesion through many small individual acts ('death by a thousand cuts'), through coordinating a series of activities aimed at achieving a single goal ('information influence operations'), and through combining information influence and hybrid activities over the long-term ('Information influence campaign'). It is important to understand how these strategies work, which vulnerabilities they seek to exploit, and how to counteract them.

We have developed a simple diagnostic tool—the DIDI diagnostic—to help professional communicators recognise information influence activities based on four characteristics that define such activities. Information influence activities are *deceptive*: they involve falsehoods in some way or another; they have the *intention* to exploit vulnerabilities to benefit a foreign power or its proxies; they seek to *disrupt* constructive debate; and they *interfere* in debates or issues in which foreign actors play no legitimate role.

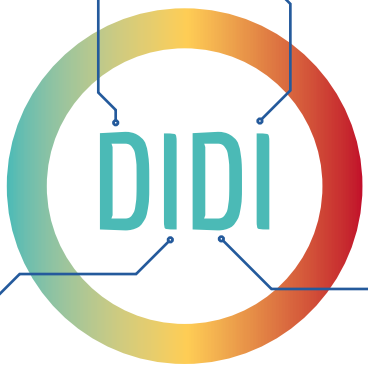
These four defining factors can be used as diagnostic criteria for differentiating legitimate communication from illegitimate information influence. To qualify as information influence, an activity should contain at least two of these factors. You will rarely be able to ascertain the presence of all of them. The appearance of any two factors suggests further investigation is necessary; three or more suggest a reasonable likelihood of information influence activities. The diagnostic checklist is designed to be flexible enough for communication professionals to adapt the criteria to their professional fields and their knowledge and experience of what is 'normal' for the work they do, and to make empowered choices regarding how to proceed.

Deception

Information influence activities aim to deceive audiences by mimicking established forms of media, hiding their true identities, and deliberately introducing false information.

Intention

Information influence activities are exercised with hostile intent and deliberately exploit vulnerabilities that can influence the process of opinion formation; they betray the generosity of liberal democratic systems where free and open public discussion is the nexus for opinion formation and turn one of our greatest strengths into a liability.



Interference

Information influence activities involve actors with little or no legitimate reason to interfere with domestic issues; this encroaches on the sovereignty of the state and the legitimacy of open and free debate.

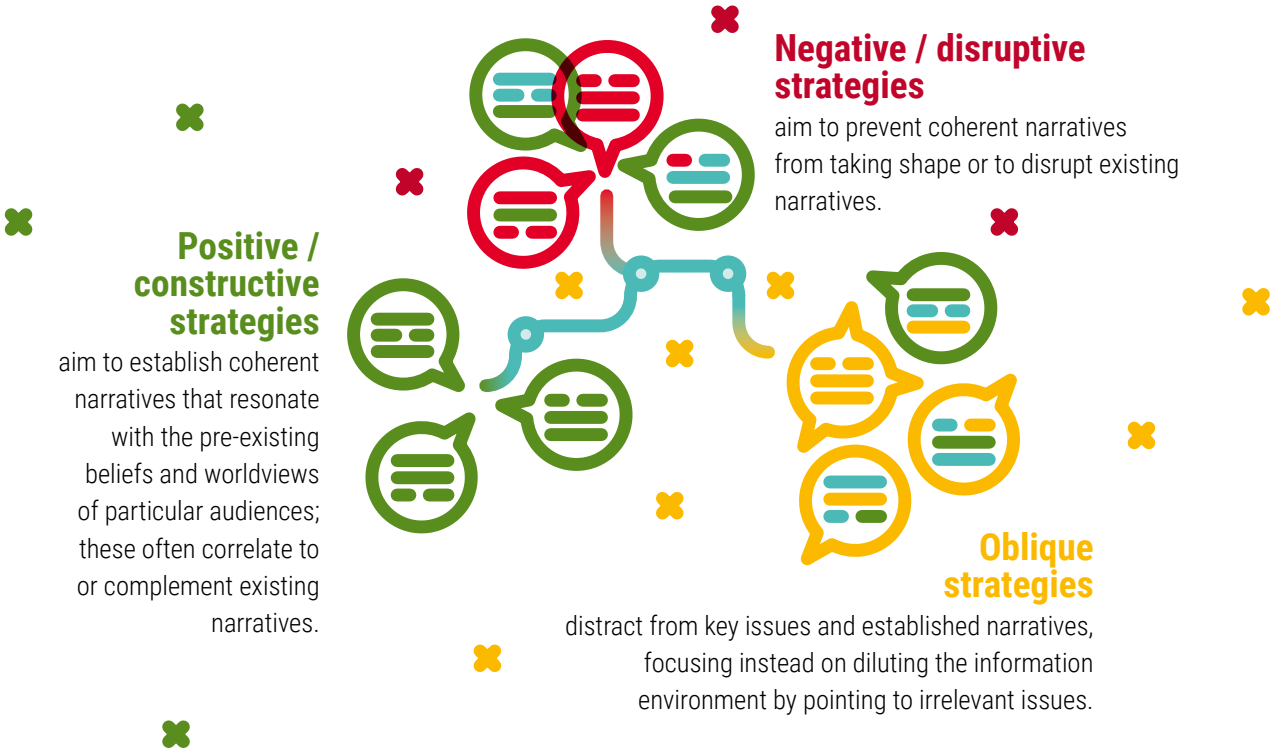
Disruption

Information influence activities disrupt normal democratic dialogue by breaking the normative rules that govern constructive debate by systematically silencing opinions, polarising society, and employing a range of techniques that disrupt discourse.

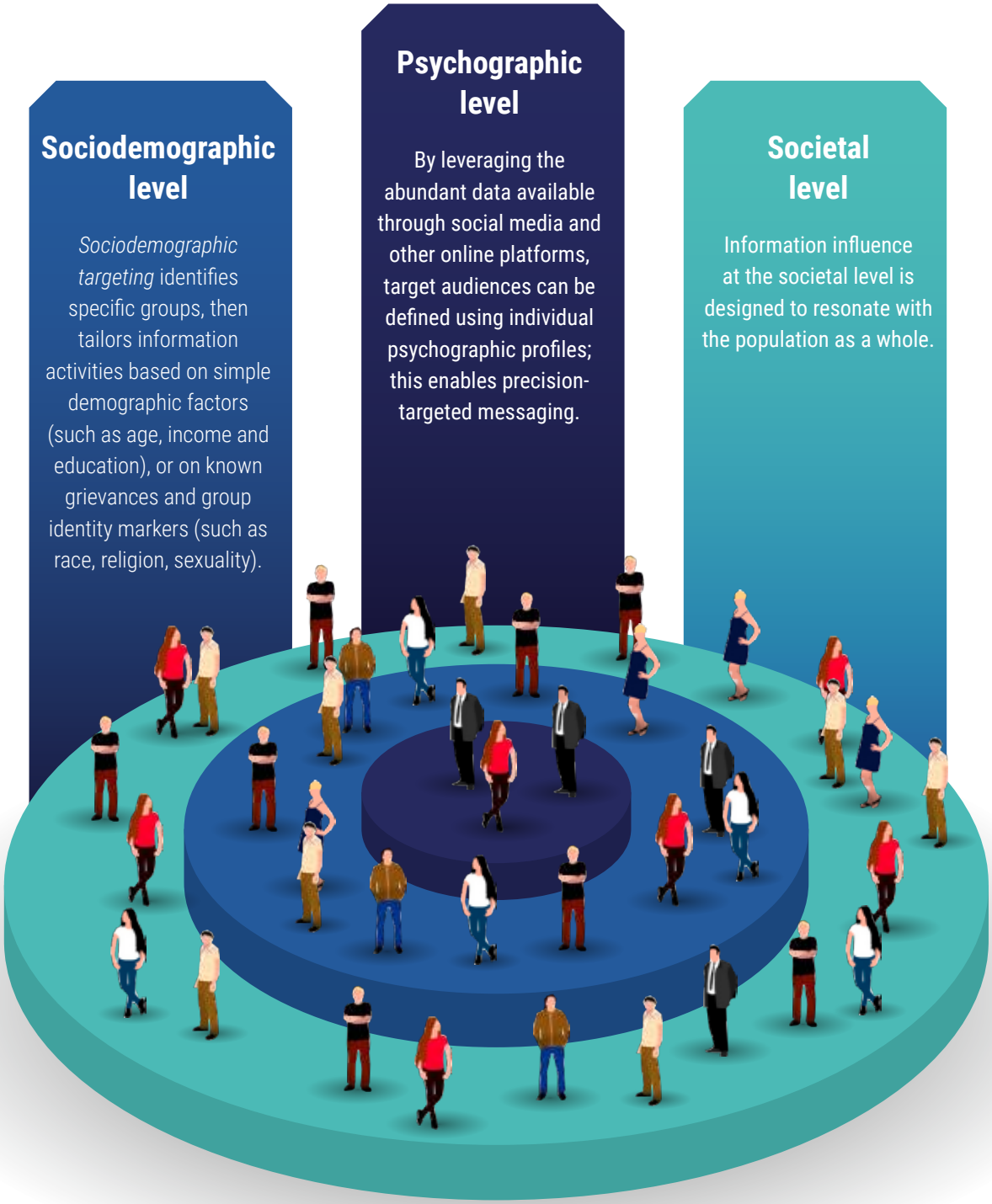
Identifying

Once communicators are equipped with a baseline understanding of what information influence activities are, it becomes easier to recognise and identify such activities. We now turn to an analysis of suspected information influence. Three parameters are particularly useful for identifying information influence: *narrative strategies*, *target audiences*, and *influence techniques*. As with the above DIDI diagnostic tool, we do not seek conclusive proof of information influence. Rather, we aim to provide communicators with pragmatic steps to simplify their work and allow them to focus on sound, evidence-based judgements as a basis for action.

The first step is to assess the overarching narrative strategy of a communication activity—this will help you gain insight into the possible purpose and context of the activity. The classic distinction between offensive and defensive strategy does not apply to narrative. Rather, influence activities can be categorised as positive, negative, or oblique, depending on the goal of the operation. Identifying narrative strategies makes it easier to discern the logic behind hostile communications, which in turn can provide important insights for preparing counter messages.



Equally important is to consider how these narratives resonate and with whom. Examining possible target audiences provides insight into the level of operation of an influence activity. The following levels allow for sufficiently differentiated categorisation:



Identifying the level at which audiences are being targeted can help reveal the reach and intention of information influence activities. This together with an understanding of the narrative strategy can provide a reasonable assessment of the goal and potential impact of a suspected influence activity.

Finally, knowledge of commonly used techniques is crucial for identifying information influence activities. Many of the techniques used today depend on the exploitation of vulnerabilities in social media and/or the malicious use of personal data, and are deployed over digital media platforms. These techniques are not necessarily good or bad in and of themselves. For example, bots can play a legitimate role in communications if used transparently and constructively, and micro-targeted ads based on psychographic data can provide value for both consumers and advertisers. However, if they disseminate illegitimate messages to specific audiences or skew social media flows to polarise or influence public opinion these techniques can be disruptive and dangerous. When analysed in relation to the DIDI diagnostic, narrative strategies, and target audiences, identification of the techniques used can help you form a rational assessment. Two or more of these techniques employe relation to each other may indicate the presence of a broader information influence operation or campaign and is cause for concern.

Sociographic & Psychographic Hacking	Commonly used in advertising and public relations, the terms sociocognitive and psychographic hacking refer to the covert influencing of audiences using messages created to appeal to precisely identified groups or individuals. This effect is achieved by activating psychosocial trigger-points and exploiting cognitive vulnerabilities, for example by engaging with known grievances for specific groups to provoke an emotional response through precision-targeted advertisements (dark ads).
Social Hacking	Social hacking techniques, such as social proof, bandwagon effects, and selective exposure, exploit group-dynamics and cognitive biases embedded in our tribal nature.
Para-social Hacking	Similar to sociographic hacking, para-social hacking exploits the biases arising from our para-social relationships, and is mainly experienced online. During the US presidential campaign of 2016 para-social hacking was successfully used to construct fake Facebook groups where legitimate users unwittingly contributed to the spread of disinformation.
Symbolic Action	Some actions can be used for communicative effect in addition to the original objective of the action itself. Symbolic actions, such as acts of terrorism, are often motivated by a communicative logic.

Disinformation	Disinformation is the deliberate creation and/or sharing of false information with the intention to deceive and mislead the audience. Disinformation ranges from slightly illegitimate activities (such as selective use of facts) to highly disruptive activities (such as content manipulation).
Forging and Leaking	Forgeries and leaks falsely imitate or illegitimately disseminate information for the sake of negatively influencing public perception. These techniques are often used in conjunction, blurring the line between truth and falsehood, for example a 'tainted leak' is when illegally obtained information is selectively released in tandem with false content.
Potemkin Villages	The term 'Potemkin village' refers to an intricate web of deceptive structures that can be used as fact-producing apparatuses for specific narratives. Online Potemkin villages commonly consist of a network of websites passing false information among themselves to the point where it is impossible for the reader to discern the origin and legitimacy of the information.
Deceptive Identities	The credibility of information is often evaluated on the basis of its origin. Imitating or impersonating legitimate sources of information is therefore an effective method of giving credence to false or deceptive information. This is particularly powerful on social media where fake accounts are common.
Technical Exploitation	Leveraging modern technologies, such as sophisticated algorithms, automated accounts (bots), machine learning, and artificial intelligence, enables the manipulation of online information flows. Technical exploitation is often used as a force multiplier for other influence techniques such as disinformation. As the NATO StratCom CoE report on Robotrolling shows, a substantial percentage of accounts posting about security-related issues in multiple European states use such techniques to spread disruptive narratives.
Trolling	Trolling is deliberate aggravation, disruption, and provocation by users of online social platforms. When used for influence purposes, trolls are employed to polarise discussions, silence legitimate opinions, and distract from important topics.
Humour & Memes	Humour is a powerful tool that attracts attention and can legitimise edgy or controversial ideas and opinions. The use of memes, or humorous pictures that spread cultural ideas, is a highly accessible, shareable, and 'infectious' method of spreading disinformation. A case in point is the meme-driven influence operation referred to as 'Operation Swedistan', which was coordinated over the controversial online forum 4chan to reinforce the narrative of 'Sweden as a haven for Islamic extremism'.
Malign Rhetoric	Malign rhetoric captures linguistic ruses aimed at undermining legitimate debate, silencing opinions, and delegitimising or distracting adversaries. Personal (ad hominem) attacks, 'whataboutism', and the 'gish-gallops' are classic examples.

Analysis of narrative strategies, target audiences, and influence techniques help professional communicators map how influence activities exploit societal vulnerabilities to achieve an effect in their field. These techniques are rarely used in isolation to target only one audience using only one narrative. Rather, influence operations and campaigns most often combine a multitude of techniques into a complex chain-of-events, or *stratagem*. While such combinations are theoretically infinite, some stratagems are frequently encountered in contemporary influence operations.

Common stratagems include:

Laundering

Information laundering refers to the process of legitimising false information or altering genuine information by obscuring its origin. Often this involves passing genuine information through a series of intermediaries (such as fake news or foreign language websites), gradually distorting it and feeding it back to legitimate channels through Potemkin villages.

Point & Shriek

The point & shriek stratagem builds on tactics used by social activists, taking advantage of perceived injustices within certain social groups and heightening emotion around these issues to disrupt rational discourse.

Flooding

Flooding creates confusion by overloading actors with spurious and often contradictory information.

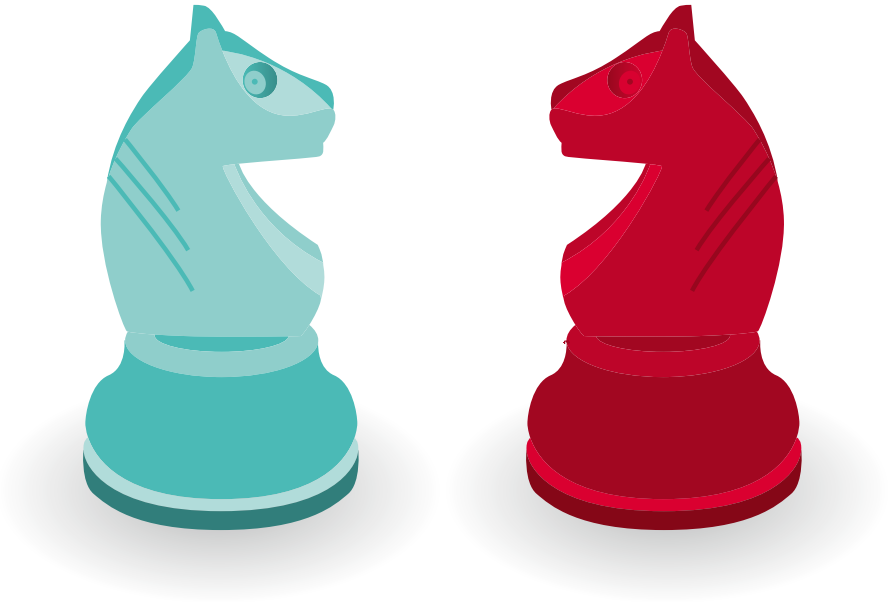
Polarisation

By using a series of deceptive identities, it is possible to support opposing sides of a specific issue to create or reinforce grievances, heighten emotional response, and force mainstream opinion toward greater extremes.

Counteracting

As may be expected from a broad and multidisciplinary literature in a relatively new field, there is no conclusive answer regarding the best way to counter information influence activities, especially not from the perspective of an individual communicator. This is all too apparent at the policy level—the literature regarding online safety and fake news offers a plethora of suggestions and advice. Social media platforms themselves also offer their users guidance on how to prevent the exploitation of personal information. We have synthesised insights gleaned from various sources into a congruent approach that situates counter-measures at the level of the communication professional and provides actionable advice.

Our approach is divided into three steps: *preparation*, *action,* and *learning*. While preparatory activities should be conducted pre-emptively to immunize your organisation from malicious influence attempts, the activities listed under action and learning should be undertaken pro-actively in response to suspicious events. As such, they should be considered in relation to previous sections of this brief, especially the DIDI-diagnostic, when determining whether or not an event merits response and at what level. The less legitimate 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. It should be remembered that the overriding goal of counter influence measures is to restore trust in organisations that are being undermined through illegitimate means.



1

Preparation

Pre-emptive measures with a long-term perspective that can minimise the effects of information influence

Societal and organisational preparedness

Raising awareness

Fact-checking/ debunking

Risk/vulnerability analysis

Target audience analysis

Strategic narratives and tactical messaging

Social media

2

Action

Immediate communication tools to respond to information influence in the short- and medium-term

Assess

Inform

Advocate

Defend

3

Learning

Structures for utilising experiences and lessons-learned to improve future countermeasures

Describe

Reflect

Share

Include information influence in your organisation's strategic planning and set up organisational structures with clear responsibilities and mandates.

Raise awareness among your colleagues and peers about information influence activities to boost resilience.

Establish structures and processes for fact checking and correcting false information regarding your organisation.

Map your organisational vulnerabilities in relation to information influence and identify potential risks. Use the results to guide risk management.

Improve knowledge of your prioritised target audiences to understand their vulnerabilities and how you can reach them effectively.

Develop your own strategic narratives and tactical messaging to build long-term trust for your organisation.

Prepare social media guidelines and standard operating procedures for online behaviour, both in terms of communication and monitoring.

Assess the situation. Investigate what falsehoods are being spread and clarify what is false about the story. Use techniques such as fact checking and internal discussions, and release a holding statement if necessary. Note: not all disinformation motivates a response—some can just be ignored.

Provide basic factual information about the situation through techniques such as a Q&A, public statements, or referrals to official actors who can independently confirm facts. For most common situations it is enough to *Assess* and *Inform*. The aim is to restore credibility through stating facts.

This is a third level response, appropriate for more severe situations, depending on your mandate within your organisation. Relate the situation to a broader narrative (storytelling) and clearly assert your organisational values. Look for opportunities to *advocate* your position or perspective and initiate a dialogue with influencers.

A fourth level response only appropriate in extreme cases. Protect your organisation by using overtly defensive techniques such as ignoring, reporting, blocking, or exposing. Always be transparent about your decisions and confirm your mandate with your bosses before using these techniques. *Note that we do not go beyond this level in our guidance; however, it is also possible to consider proactive counter measures such as Deterrence.*

Collect and document information related to an information influence situation or event in as much detail as possible.

Analyse the situation in relation to your vulnerabilities by identifying influence tools used, their effects, the audiences targeted etc.

Disseminate your insights among relevant actors within your organisation and beyond to promote collective learning.

Conclusion

The exploitation and abuse of social media to illegitimately influence decision-making and opinion-formation clearly fits within the broader framework of information influence activities and can be fruitfully approached from a strategic communications perspective. Our approach does not address the larger challenge of designing adequate oversight and regulatory frameworks for social media and user data, nor does it address the issue of responsibility of social media companies or the underlying conflicts between states, However, our model for counteracting information influence activities provides a framework that professional communicators in the public sector can use to identify, assess, and counter many of the effects of such malicious behaviour. This is a small but vital aspect of protecting open societies from the harmful effects of hostile information influence activities.

Even so, counteracting information influence activities cannot easily be reduced to a checklist. Ideally, our counter-measures would be the enlightened response of educated, informed, and skilled communicators, who seek to determine the best course of action in each instance. Successful examples must be recorded, analysed, and shared. The subtitle of our research 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 social resilience to these threats, and will determine whether the vulnerabilities in our cognitive, public opinion, and media systems are in fact strengths to be nurtured.

About this brief

This brief outlines the findings of the Department of Strategic Communication at Lund University and Swedish Civil Contingencies Agency (MSB) research report 'Countering Information Influence Activities: The State of the Art' (2018) and relates them to the NATO StratCom COE projects on countering abuse of social media and malicious use of data. The original report was commissioned by MSB as part of Sweden's preparation for the 2018 elections and sought to deepen and widen our understanding of how information influence works and how it might be countered from a communicator's perspective.

Further Reading

For further reading we recommend “Countering Information Influence Activies: A Handbook for Communicators” published by the Swedish Civil Contingency Agency (MSB).

Swedish Civil Contingencies Agency

Countering information influence activities

A handbook for communicators

ATT MÖTA INFORMATIONSPÅVERKAN – HANDBOK FÖR KOMMUNIKATÖRER 7

Introduction

The annexation of Crimea and the conflict in Ukraine has shown how security threats today can assume a radically different character than what we usually associate with international conflict. There, means other than traditional military means were employed to achieve specific strategic goals. Influence campaigns is the term used to describe this new type of security threat. In influence campaigns, foreign powers exploit societal vulnerabilities to achieve their goals without the need for military force. Influence campaigns are a phenomena that we need to defend ourselves against to safeguard the goals of Sweden's security; the life and health of our population, the functionality of society, and our ability to preserve fundamental values such as democracy, rule of law and human rights and freedoms.

MSB defines influence campaigns as coordinated activities by foreign powers, including the use of misleading or inaccurate information, to influence political and public decision-making, public opinion, or opinions in another country, which may affect Sweden's sovereignty, the goals for Sweden's security or other Swedish interests negatively. An influence campaign consists of several influence activities, of which information influence is one. This handbook helps you as a communicator to become aware of what information influence is, how it works, and what you can do to counter this type of threat.

The use of information to influence others is not new. Industries such as public relations and advertising use information to influence the personal decisions of people around the world every day, such as the choice to buy a certain brand or to support a political candidate. We, as citizens, expect such communication to take place in the open, to be based on accurate and truthful information, and to be presented in a way that allows us to make informed choices.

Not all influence activities play by these rules. Information can be used covertly and deceptively by foreign powers to undermine processes critical to the fabric of democratic societies. This is what we refer to as information influence activities. There are many contemporary examples from around the world where such activities have been identified, as in the recent presidential elections in the U.S. and France. From a big picture perspective, these activities are part of how countries vie for influence in international affairs. They are hostile activities but not acts of war. Indeed, they are still sometimes referred to as information warfare, or as operations taking place in a grey zone between war and peace. They are still considered hostile as they often intend to undermine public confidence in critical societal institutions, isolate vulnerable communities, and contribute to social and political polarisation.

Prepared and published by the
NATO STRATEGIC COMMUNICATIONS
CENTRE OF EXCELLENCE

The NATO Strategic Communications Centre of Excellence (NATO StratCom COE) is a NATO accredited multi-national organisation that conducts research, publishes studies, and provides strategic communications training for government and military personnel. Our mission is to make a positive contribution to Alliance's understanding of strategic communications and to facilitate accurate, appropriate, and timely communication among its members as objectives and roles emerge and evolve in the rapidly changing information environment.

Operating since 2014, we have carried out significant research enhancing NATO nations' situational awareness of the information environment and have contributed to exercises and trainings with subject matter expertise.

www.stratcomcoe.org | @stratcomcoe | info@stratcomcoe.org