The Hague Centre
for Strategic Studies

Start with the End

Effect Measurement of Behavioural Influencing in Military Operations

Laura Jasper, Santeri Puttonen, Martijn de Jong, Sophie van der Zee and Michel Rademaker

November 2023

The Hague Centre for Strategic Studies

ERASMUS UNIVERSITEIT ROTTERDAM

Start with the End

Effect Measurement of Behavioural Influencing in Military Operations

Authors:
Laura Jasper, Santeri Puttonen, Martijn de Jong, Sophie van der Zee and Martijn de Jong

Cover image credit:
Dan Asaki, Unsplash

November 2023

The research and collaboration with researchers from Erasmus University School of Economics for this report was completed in July 2023. Events or developments that took place in the period between completion and publication did not affect the findings.

This paper has been written as part of the project Platform Influencing Human Behaviour, commissioned by the Royal Netherlands Army. The aim of this platform is to build and share knowledge on information-based Behavioural Influencing (BI) in the military context, dissecting the ethical, legal and military-strategic issues and boundaries involved. Responsibility for the content rests solely with the authors and does not constitute, nor should it be construed as, an endorsement by the Royal Netherlands Army.

The authors would like to thank the following individuals and organizations for sharing their expertise and experience which have significantly contributed to the overall content and quality of this research paper. Evidently, the content of this paper is the sole responsibility of the authors.

- Lieutenant-Colonel Björn de Heer, seconded at HCSS
- Lieutenant-Colonel Johan Koers, NLD LSO to USAREUR
- Linde Arentze, former HCSS strategic analyst until the end of June 2023

Table of Contents

Glossary

ABM	Agent Based Models
AI	Artificial Intelligence
AJP-01	Allied Joint Doctrine
AMP	Affect Misattribution Procedure
AR	Augmented Reality
BI	Behavioural Influencing
BDA	Battle Damage Assessment
CLD	Causal Loop Diagrams
DID	Difference-in-Difference
IAT	Implicit Association test
INFOMAN	Information Manoeuvre
InfoOps	Information Operations
IoT	Internet of Things
IV	Instrumental Variable
LAWS	Lethal Autonomous Weapons Systems
LLM	Large Language Model
NATO	North Atlantic Treaty Organisation
NLP	Natural Language Processing
PSYOPS	Psychological Operations
RAS	Robotics and Autonomous Systems
RCT	Randomised Controlled Trial
R&D	Research & Development
SDM	System Dynamics Model
StratCom	Strategic Communication
TA	Target Audience
TAA	Target Audience Analysis
VR	Virtual Reality

1. Introduction

Information[1] has long determined the outcome of war, both on the battlefield and by "winning hearts and minds".[2] It dates back as far as 12th century BCE when, according to Greek mythology, the Trojan Horse was used as a deception tactic by the Greeks[3] to sway the Spartan's knowledge and beliefs in order to ultimately influence their behaviour. Later on, during World War II, similar influencing tactics, like deception, were used to achieve strategic objectives. One such example is Operation Fortitude. Operation Fortitude was part of a larger strategy called Operation Bodyguard, where the Allied Forces employed a series of disinformation tactics that targeted the Germans' knowledge, beliefs, and emotions in order to ultimately influence their behaviour. Which in this case was, making sure that they kept their forces and resources focused on Calais rather than Normandy where the real invasion was being planned.[4] The surprise counter offensive launched by the Ukrainian army in Kharkiv, instead of Kherson, can be considered a modern version of Operation Fortitude.[5] These examples show how, throughout history, targeting the cognitive environment in order to influence the physical behaviour has been an important factor in deciding the outcome of war, on and off the battlefield.

These tactics target the cognitive domain of individuals, where the decision-making process takes place, in order to exert changes in their physical behaviour. Ultimately, the aim is to influence the decision-making process of individuals in order to compel them to behave in a certain manner. For example, to boost the morale of your own troops in order to improve performance, or target the morale of hostile troops to make them surrender or defect. Besides carrying out influence operations, assessing the sustained effects by and against hostile forces is equally as important to decide the outcomes of battles, and with that, wars. However, carrying out such measurements to determine effectiveness of such operations remains an empty spot in most influencing operations and constitutes the focus of this paper.

Besides carrying out influence operations, assessing the sustained effects by and against hostile forces is equally as important to decide the outcomes of battles, and with that, wars.

Over time a myriad of definitions and terms have emerged to describe tactics employed in the information environment with the aim to exploit and utilise human cognition for military purposes.[6] Information-based behavioural influencing is the act of meaningfully trying to influence the behaviour of an individual by targeting people's knowledge, beliefs, and emotions. With recent advances in information and communication technology, the boundary between the physical and cognitive battlefield has begun to fade. Ever since the start of the Russian

1 Information is interpreted from a broader strategic viewpoint, where it is used to target knowledge, beliefs and emotions (cognitive) in order to ultimately influence behaviour (physical). It is not interpreted in the narrow way relating to activities in the cyber domain.

2 The phrase winning hearts and minds is associated with British Field Marshal Gerald Templer in the context of counter-insurgency warfare. However, it also refers to the US military campaigns in both the Vietnam and Afghan war, where psychological operations and public relations tactics were used to sway knowledge, beliefs, emotions and ultimately behaviour. Paul Fishstein and Andrew Wilder, 'Winning Hearts and Minds? Examining the Relationship between Aid and Security in Afghanistan' (Feinstein International Center, 2012), https://fic.tufts.edu/assets/WinningHearts-Final.pdf.

3 Homer, trans. Robert Fagles, *The Iliad* (New York: Penguin Classics, 1998).

4 Roger Hesketh, *Fortitude: The D-Day Deception Campaign*, First Edition (Woodstock, N.Y: Abrams Press, 2000).

5 Huw Dylan, David V. Gioe, and Joe Littell, 'The Kherson Ruse: Ukraine and the Art of Military Deception' (Modern War Institute, 10 December 2022), https://mwi.westpoint.edu/the-kherson-ruse-ukraine-and-the-art-of-military-deception/.

6 Alicia Wanless and Michael Berk, 'The Changing Nature of Propaganda: Coming to Terms with Influence in Conflict', in *The World Information War* (Routledge, 2021), 63–80, https://doi.org/10.4324/9781003046905-7.

invasion of Crimea in March 2014, we have witnessed an unprecedented scale of such influencing tactics. This further intensified since Russia's full-scale invasion of Ukraine on February 24th, 2022. The demonstrated spill-over of war into the cognitive domain, is changing not only the outcome but also the character of war itself. It is therefore increasingly important to have an understanding of the participants and audiences of war with the aim to "compel them", in early Prusian general and military theorist Clausewitz's words, "to do our will".[7] This is reflected by the shift in perspective by military organisations across the Atlantic. The most notable example of this is the adaptation of NATO's capstone Allied Joint Doctrine (AJP-01) on Strategic Communication and Information Operations.[8] The AJP-01 serves as the basis for all other NATO doctrine documents as well as those of its member states. NATO's new doctrinal changes highlight the importance of influencing human decision-making, and thus behaviour, in military operations. It also introduces a behaviour-centric approach as their fourth key tenet, alongside the manoeuvrist approach, the comprehensive approach and mission command.[9]

We seek to explore under which conditions, and with which tools the effects of behavioural influencing tactics can be measured

Our current living environment is characterised by rapidly evolving information technology that is changing the way we think, communicate, and ultimately how we live. More so, emerging technologies and the intersection of those such as machine learning and Artificial Intelligence (AI), are adding new dimensions to the information environment in both scale and reach.[10] Unfortunately, it also allows others to use those same channels for more malignant purposes. As we are becoming increasingly dependent on these information and data-driven technologies for the organisation of our society, we are thus also more prone to being manipulated by hostile actors.[11] Applied to the military context these advances in information technology provide us with a rapidly growing possibility to employ information both as a tool and as a weapon. Information-based influencing is no longer only taking place through traditional channels such as newspapers and radio, such as during the Cold War with Operation INFEKTION, where Soviet Union successfully spread a false rumour that the United States had created and was using biological weapons in various parts of the world by planting articles in foreign newspapers.[12] Similar tactics were also more recently used in active war situations such as during the Gulf War. For example, the distribution of "The Road to Freedom" leaflets by aircraft and ground forces, in an attempt by the U.S.-led coalition forces to convince Iraqi soldiers to surrender or defect.[13] These tactics are still being used in conflict situations today. But with the emergence of new online channels, that are often less regulated than traditional platforms, much of this influencing has moved online. Due to this shift, influencing practices increased in intensity.[14] As the ways of communication have gained speed, reach, and above all have become easier.

This paper aims to provide an overview of methods to measure the effect of behavioural influencing operations in a military context. It will not focus on how the effects are reached, but rather what effect is reached in the first place. In doing so, we seek to explore under

7 Carl von Clausewitz, ed. and trans. Michael Howard and Peter Paret, *On War* (Princeton University Press, 1989).

8 Markus Iven, Laura Jasper, and Michel Rademaker, 'Cognitive Effects in Combined Arms: A Case Study of the Division 2025' (The Hague Centre for Strategic Studies, 10 February 2023), https://hcss.nl/report/cognitive-effects-in-combined-arms-a-case-study-of-the-division-2025/.

9 NATO, 'NATO Allied Joint Doctrine (AJP-01)' (UK Government, 2010), https://www.gov.uk/government/publications/ajp-01-d-allied-joint-doctrine.

10 Michael J. Mazarr et al., 'The Emerging Risk of Virtual Societal Warfare: Social Manipulation in a Changing Information Environment' (RAND Corporation, 9 October 2019), https://www.rand.org/pubs/research_reports/RR2714.html.

11 Mazarr et al., 'The Emerging Risk of Virtual Societal Warfare'.

12 Thomas Rid, *Active Measures: The Secret History of Disinformation and Political Warfare*, Illustrated edition (New York: Farrar, Straus and Giroux, 2020).

13 Jeffrey Bryant Jones and Jack N. Summe, *Psychological Operations in Desert Shield, Desert Storm and Urban Freedom* (AUSA Institute of Land Warfare, 1997).

14 Wanless and Berk, 'The Changing Nature of Propaganda'.

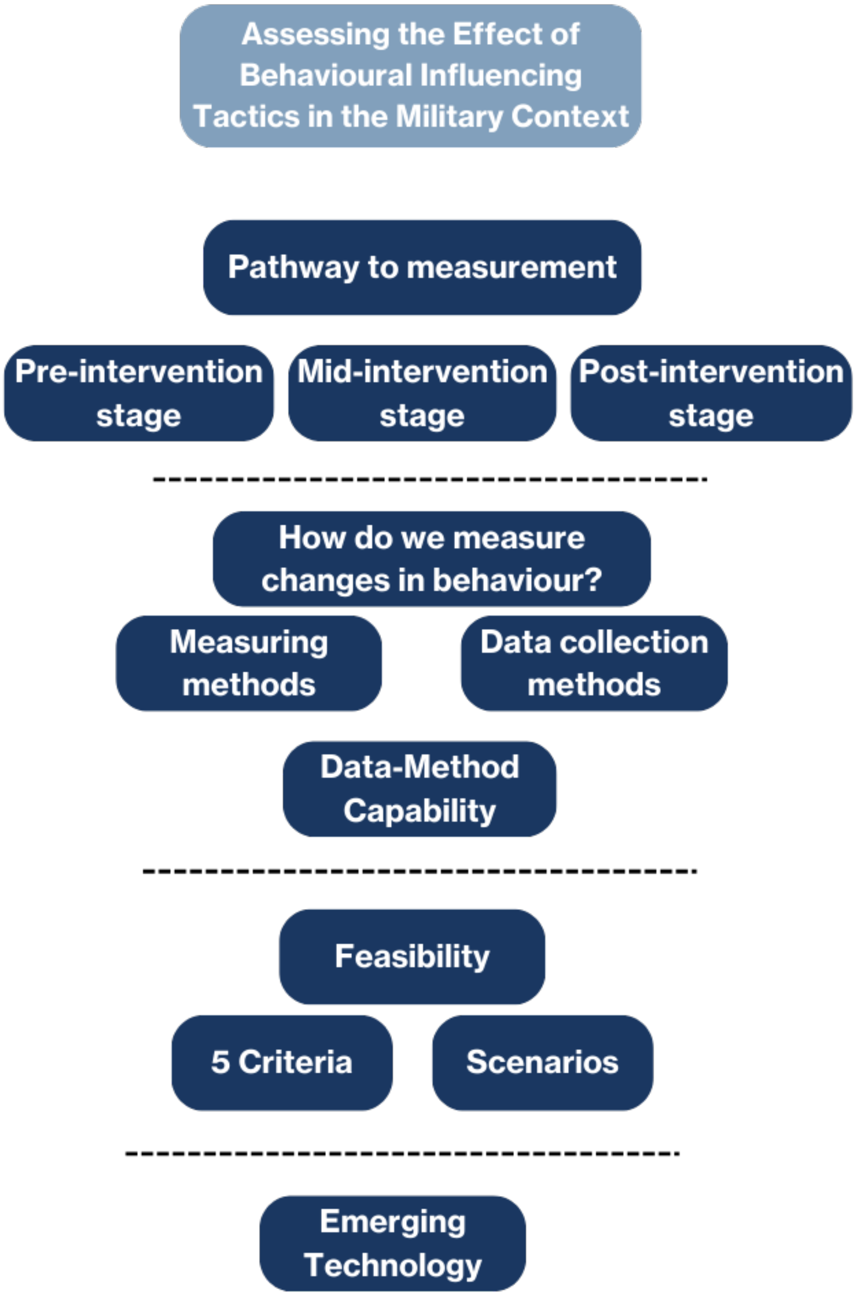
which conditions, and with which tools the effects of behavioural influencing tactics can be measured. This allows the military to determine the effectiveness of their interventions aimed at influencing human behaviour. This aids the effectiveness and predictability of the use of behavioural influencing tactics. As an end result, a 'toolbox' will be constructed of measuring methods that correspond to different circumstances where behavioural influencing tactics are employed. The literature dealing with measuring the effects of behavioural influencing is spread out across different fields, such as behavioural economics, marketing, intelligence studies, psychology, and war studies. This paper will attempt to partially centralise the research dealing with the topic so as to provide an overview of what is currently possible, as well as what can still be possible in the future regarding emerging technologies.

The first section aims to clarify and define the question of measuring effectiveness of behavioural influencing operations in the information environment. It thus clarifies the used terminology and relevance of this research. Additionally, it introduces effect measurement of behavioural influencing in three stages of the intervention, being pre-intervention, during intervention, and post-intervention.

Hereafter an introduction into different methods that can be used to quantify the effect of behavioural interventions is given. This is followed by an overview of data collection methods to measure knowledge, beliefs, emotions and behaviour. An in-depth discussion of these measuring and data collection methods is provided in the annex at the end of the paper. The combination between a measuring and data collection method is called a Data-Method Capability. These capabilities are assessed on the basis of the military feasibility of implementation for which five different criteria are used. The criteria are applied to the Data-Method Capabilities, on the basis of a context-specific scenario in order to assess their feasibility. The paper ends with a discussion on how and which possible emerging technologies can further impact the measuring of behavioural influencing effects. After which ultimately the conclusion follows.

Finally, two points need to be clarified. Although this paper is focused on military applications, the methodologies are not restricted to military examples. This specifically applies for the case studies and examples corresponding to the different measuring methods, where both military and civilian capabilities are taken into consideration. Second, this paper came to be through a collaboration between HCSS and researchers from the Erasmus School of Economics of the Erasmus University Rotterdam. It is an accumulation of desk research and multiple expert sessions bringing together military and non-military experts and practitioners with backgrounds in behavioural economics, strategic communications and marketing.

Figure 1. Reading Guide



2. Pathway to measurement

Even if an influencing campaign would affect a target audience's knowledge, emotions or beliefs, if the changes do not translate into favourable behaviour either on or off the battlefield, it is of little impact.

Over time, the means actors use to shape the information environment to their advantage have changed rapidly, as have the terms to describe such practices, often lacking a common terminology.[15] Information warfare, deception, disinformation, political warfare, propaganda - these are all examples of terms used to describe the influencing of behaviour through the information environment. This section, in the first instance, aims to better understand the term behavioural influencing and its scope. Subsequently, the section attempts to clarify what it means to conduct Battle Damage Assessments (BDA) when it comes to influencing operations in both high and low intensity combat operations. The literature has demonstrated that certain influence operations targeting people's beliefs and behaviour do have measurable effects[16], in this research we explore what that means for conducting BDA. Since the dawn of warfare, the assessment of sustained effects by and against hostile forces has determined the outcome of battles and hence wars.

This paper takes a behaviour-centric approach, which is in line with NATO's newly adjusted capstone doctrine Allied Joint Publication 01 (AJP-01), where the behaviour-centric approach is one of the four key tenets.[17] For this paper the focus will primarily be on the behaviour of the target audience (TA) as that is where one wants to register effects. This corresponds to the notion that changes in behaviour constitute a change in the decision-making process of the target audience. As every conflict in one way or another is caused by human behaviour, a change in that behaviour is fundamental to shaping the outcome of that conflict. Even if an influencing campaign would affect a target audience's knowledge, emotions or beliefs, if the changes do not translate into favourable behaviour either on or off the battlefield, it is of little impact. Information-based behavioural influencing is thus defined as the act of meaningfully trying to affect the behaviour of an individual by targeting people's knowledge, beliefs, and emotions in the information environment.

The information environment exists next to the physical environment and consists of three dimensions:[18] the physical, cognitive, and virtual dimension. As effects cannot be measured directly in the cognitive dimension, they can be assessed through the physical and virtual dimension to track how behaviour of the target audience changes. The cognitive dimension encompasses the sphere where changes in the decision-making process of the individual takes place through acquired knowledge, beliefs or experience.[19] This is where the information is received, and processed by the individual upon which they respond. The physical

15 Alicia Wanless and James Pamment, 'How Do You Define a Problem Like Influence?', *Journal of Information Warfare* 18, no. 3 (2019): 1–14.

16 Jon Bateman et al., 'Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research' (Carnegie Endowment for International Peace, 28 June 2021), https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-research-pub-84824.

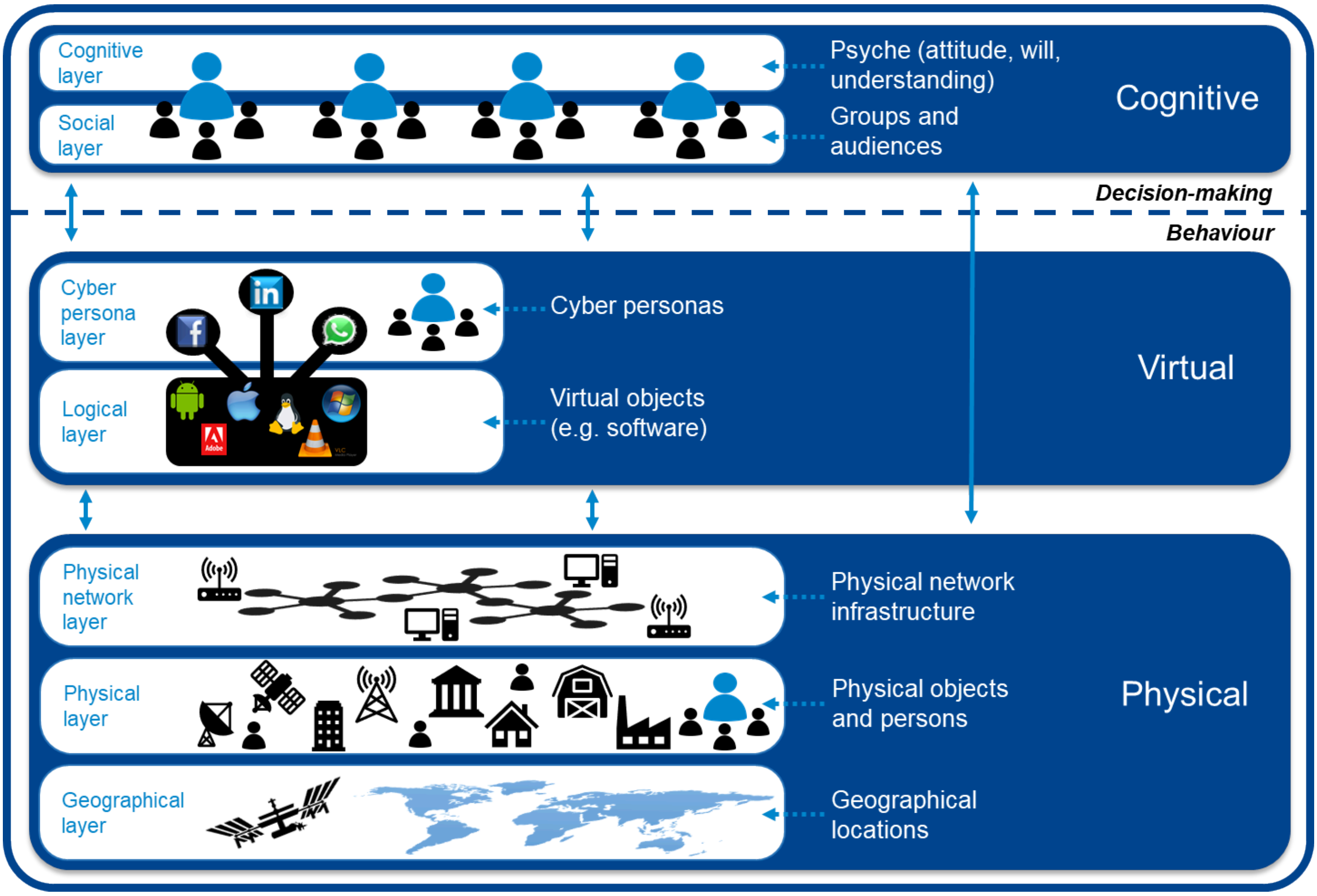
17 NATO, 'NATO Allied Joint Doctrine (AJP-01)'., more information in Iven, Jasper and Rademaker, 'Cognitive Effects in Combined Arms: A Case Study of the Division 2025'

18 P.A.L. Ducheine, Jelle van Haaster, and Richard van Harskamp, 'Manoeuvring and Generating Effects in the Information Environment' (Amsterdam Center for International Law, 2017).

19 Ministry of Defence, 'Doctrine Note 19/04 on Information Manoeuvre' (British Ministry of Defence, 2019).

dimension is that where physical or tangible activity takes place.[20] This includes 'hard' infrastructure, such as command and control centres. Lastly, the virtual dimension is where cyberspace is positioned.[21] This dimension encompasses the 'soft' intangible networks that are needed to process data and information.

Figure 2. The information environment (AJP 10.0, p.38)



Measuring effectiveness should not be limited to the post-implementation actions alone but should rather encompass actions throughout the entire process of implementing the operation. Therefore, it is relevant to introduce different steps in the decision-making process. One concept developed to analyse this military decision-making cycle is the OODA-loop, first introduced by American Air Force Colonel and military strategist John Boyd.[22] The OODA-loop consists of four steps: Observe, Orient, Decide and Act.[23] The aim of information operations is, as an integral part of manoeuvre, to disrupt the OODA-loop of hostile forces leading to

20 Ducheine, van Haaster, and van Harskamp, 'Manoeuvring and Generating Effects in the Information Environment'; Ministry of Defence, 'Doctrine Note 19/04 on Information Manoeuvre'.

21 Peter B.M.J. Pijpers and P.A.L. Ducheine, '"If You Have a Hammer..." Reshaping the Armed Forces' Discourse on Information Maneuver' (Amsterdam Center for International Law, 1 November 2021).

22 Frans Osinga, *Science, Strategy and War: The Strategic Theory of John Boyd*, 1st ed. (Routledge, 2007).

23 Osinga.

an advantage on the battlefield. Therefore, information warfare can be seen as two opposing cogs trying to rotate in different directions. Both adversaries are trying to disrupt the other while going through their own loop. As such, the one whose operational tempo and effectiveness is greater will prevail and gain control of the information space. The last stage of effect measurement will take place between the steps Act and Observe.

Figure 3. Visualisation of two opponents trying to disrupt each other's OODA-loops

Following the clarification of used terminology above, and the introduction of the analytical framework that the OODA-loop provides to think about military decision-making, the remainder of the chapter will focus on the three different phases associated with implementing influencing operations after the decision-making process has taken place. These three stages are: (1) pre-intervention, (2) mid-intervention, and (3) post-intervention, and apply to both low and high intensity combat.

Figure 4. Stages of intervention

2.1. Pre-intervention phase: Start with the End

There are several core principles of conducting Battle Damage Assessment or more generally measuring the effectiveness of any operation. Put simply, in the pre-intervention phase one should already have the post-intervention phase in mind. Along with establishing the goal of the intervention comes identifying and analysing the intended target audience, determining the context of the intervention which all allows for establishing a baseline against which the post-intervention stage is tested. Last in this stage is estimating the likelihood of success.

Establishing the goal of intervention

In the pre-intervention phase one should already have the post-intervention phase in mind.

As behavioural influencing interventions cannot reach just any desirable effect, identifying the objectives of the intervention and what supports their success, by breaking down the main objectives in separate steps, aids the assessment process.[24] In order to know what failure and success will look like, clear, measurable goals ought to be determined in the planning stage of the intervention.[25] In other words, prior to implementing a tactic, its desired outcome or end result needs to be established.

Seese, Linera, and McQuagge, developed an evaluation method for Psychological Operations programs that focuses on impacting behaviour of foreign target audiences. This method relies on translating the intended goal of the behavioural influencing intervention into quantifiable and measurable effects up-front. Their framework aids in building and understanding how identifying such quantifiable, measurable goals in the pre-intervention phase aids overall effect measurement of such interventions.[26]

For example, the objective of an influencing operation could be to facilitate the offensive operations of a brigade. The desired effect is to cause the main effort of the opposing forces to concentrate in one area which is to the advantage of the brigade. The brigade thus wants the target audience to move from north to south. The quantifiable, measurable goals would be: how many elements of the enemy forces are leaving their current positions? What percentage of the target audience is moving into the desired direction? Failure in this case, or the undesired effect, would be that the target audience either does not move, or moves into a faulty direction. Another example would be, wanting to decrease the number of civilian casualties during active combat operations.[27] This objective can be divided into different goals. For instance, the target audience needs to comply with coalition troops, stay indoors during such operations, and obey a curfew. The desired and measurable effect would be a decrease in the number of civilian casualties, which would require a measurement of the number of civilians injured or killed during active combat operations. The undesired effect would be the number of civilian casualties staying the same or increasing.

24 Christopher Paul, 'Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Worked Example' (RAND Corporation, 22 March 2017), https://www.rand.org/pubs/research_reports/RR809z4.html.

25 Paul.

26 For a detailed overview of their framework please consult: Gregory S Seese, Rafael E. Linera, and Erinn McQuagge, 'Effects-Based Psychological Operations Measures of Effectiveness: Measuring Change and Impact', in *What Do Others Think and How Do We Know What They Are Thinking?*, Strategic Multilayer Assessment (Department of Defense, 2018).

27 *Ibid.*

Identifying and analysing target audience

Identifying the goals of intervention in the previous stage also requires identifying the relevant target audience. The target audience can consist of an individual, a group of individuals, or a mass public.[28] Examples of targeting one individual would be foreign military leadership, detainees, defectors or hostile political leadership. Examples of targeting a group of individuals are hostile troops, foreign civilian population, or insurgents.

After identifying the relevant audience, the next step is selecting, and ultimately analysing that specific target audience. Conducting a target audience analysis (TAA) is essential in creating an understanding of the TA's knowledge, beliefs, behaviour, and how they might respond.[29] TAA creates a profile of how the intended audience behaves, but more importantly how they can be influenced to change that behaviour. It helps build a notion of what is possible to expect in terms of outcomes. A baseline can be created of what is considered normal behaviour or current perceptions of the TA.[30] There are three different tiers of TAA.[31] Tier 1 of TAA is the most useful one of the three. This tier encompasses detailed scientifically-proven research that is done on the basis of multiple sources, in the country of the TA and also in the language spoken by the TA.[32] Tier 2 is one step less detailed, as it, in contrast to tier 1, does not follow a strict methodology. Rather it is primary research recorded from interactions with the TA, done either remotely or on-site. The last tier uses assumed information of the TA, found through secondary research.[33]

TAA creates a profile of how the intended audience behaves, but more importantly how they can be influenced to change that behaviour.

Determine context of intervention

Establishing the intended goal and selecting the relevant target audience are the first steps, which still leave the question of which influencing tactics best support the intended outcome. An overview of different influencing tactics can be found in the annex. Selecting the best tactic per target audience, as well as their overall effectiveness also depends on the context of implementation, more specifically the timeframe and location.[34] In terms of timeframe or horizon, it needs to be established what the duration of the effects should be. This is in line with the first step in the pre-intervention phase, namely establishing the intended outcome which incorporates establishing whether to achieve short-term or long-term effects. These are to be understood as different from the question of short-term and long-term exposure to the influencing tactics which is discussed in the mid-intervention stage. Whether the effect should be long-term or short-term, as well as when the effect should ideally occur, all in turn influence the design of the intervention. For example, creating short-term effects will often require a very specific and perhaps narrow target audience, where the intended effect ideally is visible within days. Creating long-term effects often incorporates larger groups and the time-to-effect in turn rather takes several months or years. This leaves out the medium-term effects, which is in line with the established research-gap of medium-term effects, meaning weeks or a few

28 Eric V. Larson et al., 'Foundations of Effective Influence Operations: A Framework for Enhancing Army Capabilities' (RAND Corporation, 27 May 2009), https://www.rand.org/pubs/monographs/MG654.html.

29 Steve Tatham, 'Target Audience Analysis', *Three Swords Magazine* 28 (2015): 50–53.

30 Nato Standardization Office, 'AJP-3.10.1 Allied Joint Doctrine for Psychological Operations Edition B Version 1' (NATO, 2014), https://www.gov.uk/government/publications/ajp-3101-allied-joint-doctrine-for-psychological-operations.

31 Nato Standardization Office.

32 Tatham, 'Target Audience Analysis'.

33 Tatham.

34 A note is to be made here, that this research focuses on the effectiveness of behavioural influencing as a whole and not between different influencing tactics.

months, is missing in the literature.[35] The timeline of effects is associated with the intensity of the conflict or warfare, rather than with the military level of implementation.

Another factor is location, or where the intended effect should take place. Besides physical environment or location, which is assumed to be related to where the intended TA is situated, there are also the non-physical or tangible environmental features, such as culture, language and history.[36] For example, a campaign designed for a West-African audience could be ineffective when applied in the context of East-Asia. Taking into account the location characteristics, in line with the target audience, is thus vital for the overall success of the intervention.

Estimated likelihood of success

Simulations can provide valuable insights into the potential outcomes and effectiveness of the behavioural influencing operations before they are implemented in real-world scenarios.

In some situations, simulations can provide valuable insights into the potential outcomes and effectiveness of the behavioural influencing operations before they are implemented in real-world scenarios. For instance, suppose one is planning an operation in a conflict region. The objective could be to influence the local population's behaviours and attitudes as quickly as possible by propagating information to that population within a desired time frame. For this purpose, it is important to select a small number of influential people in the population, either opinion leaders or unknown information propagators, who are likely to spread a message on social networks for maximum social influence. An agent-based model (ABM) can then be constructed to simulate the interactions and behaviours of individual agents representing the local population, which is useful to determine how quickly specific messages would spread depending on the individuals in the social network. By putting different options into the simulations, military analysts can assess and quantify the potential impacts and outcomes of each option. This helps in identifying the most promising strategies and optimising resource allocation for maximum effectiveness. Simulations also provide an opportunity to test the robustness and adaptability of behavioural influencing operations under different scenarios and conditions. This allows military planners to develop contingency plans accordingly. Finally, simulations can facilitate the exploration of potential unintended consequences and side effects, enabling refinements of interventions before they are implemented.

2.2. **Mid-intervention phase: Monitor and Adjust**

The mid-intervention stage mainly focuses on monitoring whether the operation is going as planned. When this is not the case the operation should be stopped or adjusted accordingly. Which implies a certain degree of flexibility while carrying out the intervention. This flexibility is supported by real-time monitoring or multiple intermediate moments of measurement depending on long or short-term exposure to the intervention, which makes the assessment an iterative process in effect measurement.

35 Bateman et al., 'Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research'.

36 Paul, 'Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade'.

2.3. Post-intervention phase: Battle Damage Assessment

For both kinetic and non-kinetic warfare, the absence of a discernible effect does not necessarily mean that no effect has occurred, even if measuring such effects can be difficult.

Battle Damage Assessment (BDA) is used to evaluate the sustained damage of an attack by one's own forces, as well as by the struck hostile target.[37] BDA has traditionally been used to assess the effectiveness of kinetic forces, by estimating the number of lives and equipment lost and damage done to infrastructure.[38] Historically, conducting BDA was fairly straightforward as battles were fought close quarters, restricted in time and space, where commanders had an unobstructed view of the entire battlefield which made it easy to calculate losses.[39] However, as weapons developed, armies gained the ability to strike targets far away from the frontlines. This complicated BDA. The impact of a strike could no longer be directly observed. The problem is further magnified with the increasing scale and pace of combined operations and the development of precision munitions that can incapacitate targets without leaving much visible evidence. It is an important tool for any military mission as its ultimate goal is to assess whether the intended objective has been reached with the force employed, or whether further actions are necessary.

For both kinetic and non-kinetic warfare, the absence of a discernible effect does not necessarily mean that no effect has occurred, even if measuring such effects can be difficult. In an ideal case, kinetic and non-kinetic warfare BDA should not be limited to a mere tally, e.g. counting destroyed tanks or surrendering soldiers, but it should be measured at the level of effect: How did an operation affect the enemy's ability to wage war? However, due to limited resources and information available, this is not feasible in most cases. Still, even a rough, probabilistic BDA will provide to be a useful insight for future operations. Just as with kinetic warfare, the absence of observable effects does not necessarily indicate ineffectiveness. Even if the only observable evidence that a precision munition strike leaves behind is a barely visible entry hole into a building, the damage inside the building could be tremendous. A behavioural influencing operation might not directly result in a measurable change in behaviour, but it may influence knowledge, beliefs and emotions. Not having observable "damage" is especially challenging when it comes to behavioural influencing. As the main impact of an intervention can happen on all these different levels: knowledge, emotions, beliefs and behaviour. Some of these impacts cannot be directly observed, although they do unwrite changes in behaviour. However, the effect of employing any behavioural influencing tactic is ultimately determined by attaining the intended change in the target audience's behaviour. Therefore, any effort to measure the effectiveness of behavioural influencing tactics should in the end come down to changes in the outcome measures that either underpin or directly relate to behaviour. As information-based behavioural influencing operations in the military context, are increasingly moving online, so is the way their effects are measured and assessed. This will be addressed in the chapter regarding different types of data collection methods.

What does it mean to be effective?

As mentioned already in the pre-intervention stage, it needs to be well defined what the intended outcome of the behavioural influencing operation is. In other words, what are the goals, and measures of effectiveness? This raises the question of what it means to be effective, and which different assessment measures are possible. **Figure 5** outlines four different

37 Alan Paul Ostenberg and Fedora Maria Harmon Baquer, 'A Model for Battle Damage Assessment in Command and Control Warfare' (Master's Thesis, Monterey, Naval Postgraduate School, 1994).

38 Ibid.

39 James G Diehl and Charles E Sloan, 'Battle Damage Assessment: The Ground Truth', *Joint Force Quarterly*, no. 37 (2005): 59–64.

types of assessment measures that can be used for behavioural influencing operations: measure of activity, performance, effect, and success.

Measure of Activity is the broadest assessment measure as it tracks all activity that has taken place since the implementation of the influencing operation. This, however, does not imply causality. The activity is nearly an indicator for the events that have taken place. One step more detailed in terms of assessment is the Measure of Performance. The performance measure evaluates the accuracy and progress of activity. The third assessment measure, examined for this study, is the Measurement of Effect. The Measure of Effect, or MoE, tracks the changes in the outcomes measures emotions, beliefs, knowledge, and behaviour. The MoE should be taken into comparison with the intended goal of the operation, as outlined in the pre-intervention stage. The last stage of assessment is the Measure of Success. This assessment is subjectively carried out to determine whether the influencing operation is successful against the objectives and the desired end state.

Establishing causal relationships is important for future planning, increased efficiency, and accountability.

These assessment measures do not take into account spill-over effects, or otherwise put collateral damage assessment. Although it should be noted that the term spill-over is not entirely reflective of reality, since it implies that these effects are unexpected and uncalculated for within the set-up of the operation, which is not always the case. Second hand influencing can be both intended and unintended as it is possible that a certain audience cannot be reached directly and therefore one group is influenced in order to reach another. Additionally, one of the key difficulties in establishing the intended and observed effect is taking into account the potential impact of external factors. As conflict and war situations are volatile and uncontrolled settings, establishing a clear link between the used influencing tactic and the intended outcome and thus attributing causality can be especially difficult. The main priority, especially in the military context, is to achieve the objective, how to get there is often a secondary priority. Still, establishing causal relationships is important for future planning, increased efficiency, and accountability. Therefore, in the next chapter different methods are introduced to measure effects of behavioural influencing operation.

Figure 5. Different assessment measures

3. How do we measure changes in behaviour?

Understanding the created effect of a behavioural influencing operation helps to plan future operations, allocate resources more efficiently, and plan counter-offenses.

Empirical evidence of how behavioural influencing operations affect their target audiences is limited. This is due to limited accessibility of the target audience, the unstable environment in which military operations take place, or factors such as political and cultural context that are often not taken into account.[40] The current difficulty of measuring changes in behaviour in the military context complicates the effective deployment of these tactics. Yet, measurement is needed if one wants to attain maximum effect of targeting people's cognitive level in order to effectively change physical behaviour. Understanding the created effect of a behavioural influencing operation helps to plan future operations, allocate resources more efficiently, and plan counter-offenses. This section sets out to provide a range of measurement and data collection methods that can be utilised to measure the impact of a BI operation.

3.1. Overview of measuring methods

When measuring the effect of any type of behavioural influencing operation (hereinafter referred to as "intervention"), one might be inclined to just compare the target audience before and after the intervention. This simple pre and post analysis will give an indication on how the situation has changed after implementing the intervention. While this analysis will give very simplified results, this method could be used to gain a rough estimate of the impact of the program if the resources are scarce. A rough measurement is better than no measurement at all.

However, the results of a simple pre and post intervention analysis can be biased and misleading. Because there is no other group to compare the target audience to, it is not enough by itself to establish a causal effect between the intervention and the observed change. There might be a lot of external factors affecting the target audience at the same time, meaning that a part, or all, of the changes in behaviour would have happened even if the intervention was never implemented. Under specific circumstances it could be argued that the intervention is the only factor meaningfully impacting a target audience and therefore all the changes in behaviour can be contributed to the intervention. However, in these cases there should be an extensive amount of evidence that there were no outside factors impacting the results.

The establishment and degree of a causal relationship may appear inconsequential or irrelevant. Especially in a high intensity military conflict, where reaching objectives is the top priority, figuring out exactly how these were established might seem of secondary importance.

40 Bateman et al., 'Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research'.

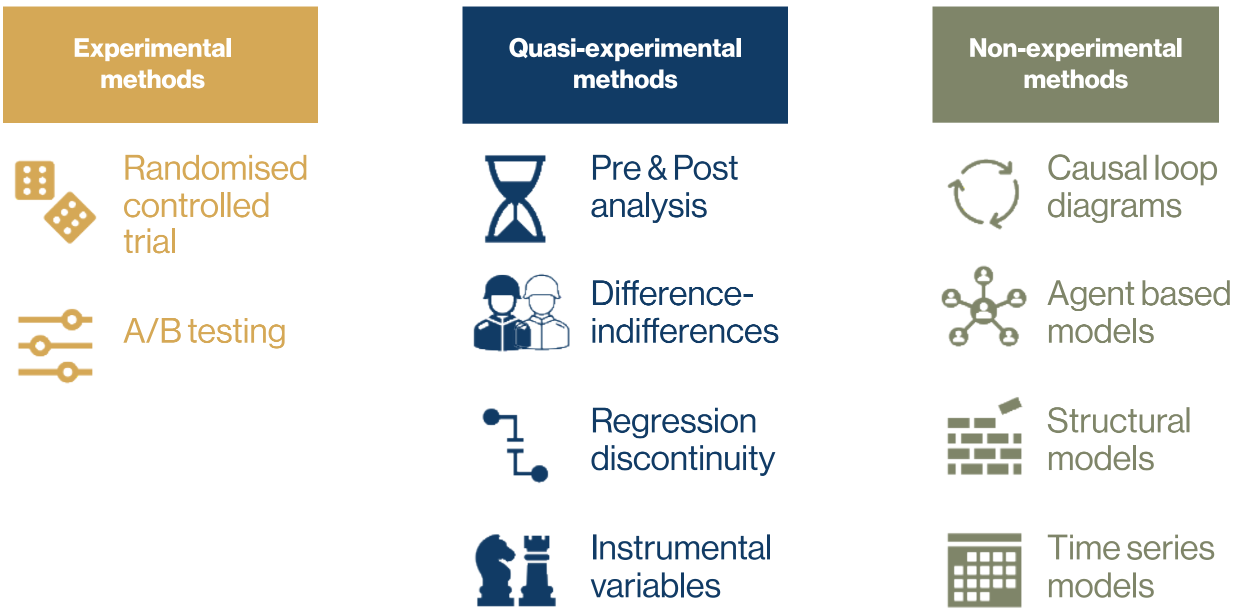
In order to efficiently measure the effects of behavioural influencing operations, it is crucial to be able to correctly identify the causal effect between the intervention and the subsequent change in behaviour.

However, establishing causality is important as it allows for measuring the actual effect that was reached with the intervention. Understanding the impact of a specific intervention, will help plan future operations, make more efficient decisions and increase accountability.

Therefore, in order to efficiently measure the effects of behavioural influencing operations, it is crucial to be able to correctly identify the causal effect between the intervention and the subsequent change in behaviour. To correctly measure the effect of an intervention, one would not only have to measure how the behaviour of the target audience changes after being exposed to the intervention, but also how their behaviour would have changed if they had not been exposed to the intervention. This is called the counterfactual. However, as target audiences cannot simultaneously be both exposed to an intervention and not exposed to an intervention, one can never truly measure the actual counterfactual. Nevertheless, different methods can be utilised when planning an intervention to get as close to an estimate of the impact of a behavioural influencing operation.

In the following section three clusters with ten types of measuring methods are provided that aim at measuring changes in the outcome and measure knowledge, beliefs, emotions and behaviour of the target audience. This overview of the different measuring methods is a brief one, the annex provides a more in-depth explanation and discussion of the possible methods, with corresponding reading material.

Figure 6. Measurement methods



Cluster 1: Experimental methods

The ultimate goal of the BI influencing operations should be considered when making a trade-off between a precise measurement and the effectiveness of an intervention.

The first cluster refers to methods that include some kind of intervention where the TA is randomly divided into groups.[41] Experimental methods can be used to measure the effect of an intervention by exposing only some of the groups to the intervention, or to determine the most effective intervention by exposing different groups to different interventions. The different groups are compared post intervention, to measure the effect of the intervention(s). These methods, and specifically Randomised Controlled Trials (RCTs), are known to be the 'gold standard' for causal identification of the effect of the intervention.[42] While experimental methods offer a reliable way to establish causality, they can be challenging to implement in the field as they require the ability to randomly apply the intervention to only some parts of the TA while avoiding any spillover effects.

Using experimental methods does come with some challenges. Not exposing the entire TA to an effective intervention, or only exposing them to inferior interventions, will lead to less reach and inefficiency. In kinetic warfare this is comparable to only engaging a part of the enemy forces while letting the rest operate freely. Thus, the ultimate goal of the BI influencing operations should be considered when making a trade-off between a precise measurement and the effectiveness of an intervention. In an intense war for survival that for example Ukraine is currently engaged in, it might make more sense to focus on maximising the reach of an intervention, sacrificing some measurability. However, in a prolonged, less intense conflict with long term goals of stability, like the NATO allies' involvement in Afghanistan[43], it can make more sense to sacrifice the effectiveness of individual BI operations in order to achieve long term efficiency.

Table 1. Experimental measuring methods[44]

Cluster	Method	Explanation	Main limitations	When to use
Experimental methods	Randomised controlled trial	Compare randomly assigned groups to each other post intervention to measure the effect of the intervention.	Requires control over TA and environment. Resources and ethical concerns. Challenging to design complex interventions. Usually not feasible in the field.	When high control over target audience.
	A/B testing	Compare randomly assigned groups to each other to measure the effects of different interventions.	Requires adequate control over audience and environment for proper randomization.	When high control over audience and there is multiple interventions.

Case study: Social media has been criticised for creating "echo chambers" that only expose individuals to content they agree with, leading to further polarisation of their political beliefs. A study from Duke University examined how exposure to opposing political ideology affects an individual's attitudes and beliefs. To measure the effect, part of the participants were randomly assigned to follow a twitter bot exposing them to opposing ideological views for a month in exchange for money. It was found that exposure to opposite political ideologies increased instead of decreased political polarization.

41 Mildred L Patten and Michelle Newhart, *Understanding Research Methods: An Overview of the Essentials*, 10th ed. (Routledge, 2018).

42 Eduardo Hariton and Joseph J. Locascio, 'Randomised Controlled Trials—the Gold Standard for Effectiveness Research', *BJOG : An International Journal of Obstetrics and Gynaecology* 125, no. 13 (December 2018): 1716, https://doi.org/10.1111/1471-0528.15199.

43 NATO, 'NATO and Afghanistan', NATO, 2022, https://www.nato.int/cps/en/natohq/topics_8189.htm.

44 Christopher A. Bail et al., 'Exposure to Opposing Views on Social Media Can Increase Political Polarization', *Proceedings of the National Academy of Sciences* 115, no. 37 (11 September 2018): 9216–21, https://doi.org/10.1073/pnas.1804840115.

Cluster 2: Quasi-experimental methods

Quasi-experimental methods relax the requirement of random group assignment that is necessary for experimental methods.[45] These methods rely on several other strategies to causally identify the effect of an intervention. Often, additional assumptions are required.[46] These include the assumptions that the chosen control group is similar to the group exposed to the intervention and that there are no external factors affecting the results. Because of these assumptions, the causal evidence can be weaker, leading to less precise BDA. However, compared to experimental methods, relaxing the randomization requirement makes some of the quasi-experimental methods more suitable to be conducted in the field. The assumptions made while using the method should always be well understood and ideally be tested for validity.

Table 2. Quasi-experimental measuring methods[47]

Cluster	Method	Explanation	Main limitations	When to use
Quasi-Experimental methods	Pre-Post analysis	Compare behaviour before and after intervention.	Exogenous factors can affect the change within a group over time leading to unreliable measurements.	When resources and environment allow for only minimal measurement.
	Difference-in-difference	Compare similar groups to each other pre and post intervention to try and isolate the effect of the intervention.	While it's possible to get close, causality is hard to establish.	When there is no control over randomization but a close counterfactual available.
	Regression discontinuity	When the intervention is dependent on some benchmark, compare subjects just above and below the benchmark to try and isolate the effect.	Requires a specific set up where intervention is based on some measurable benchmark and causality can never truly be proven.	When the intervention is dependent on some measurable benchmark.
	Instrumental variables	Use a proxy for the independent variable to isolate and measure the effect of the independent variable on the variable of interest.	A strong proxy for the independent variable is not always available.	When the dependent variable can't be used because of exogeneous correlation leading to measurement bias.

Case study: Following the Euromaidan protests, Ukraine, and its 2014 elections became a frequent topic in Russian news coverage. At the same time, concern over Russian influence in the informational space grew in Ukraine. As a result, the Ukrainian authorities banned the broadcasts of Russian television in an attempt to decrease the impact of Russian influence on the Ukrainian information space. Despite this, in some areas near the Russian border it is possible to gain reception of Russian analogue television. Researchers from New York University utilised this variation in television reception to measure the effect of Russian television on Ukrainian electoral behaviour. They found consistent evidence that availability and consumption of Russian television increases electoral support for pro-Russian parties. The television coverage manages to strengthen the views of individuals with strong pro-Russian priors, while it is less or even counter-effective on those with pro-Western priors. This implies that introducing media with a noticeable bias into a highly politicised environment could lead to further polarisation.

45 Peter H. Rossi, Mark W. Lipsey, and Howard E. Freeman, *Evaluation: A Systematic Approach*, 7th ed. (SAGE Publications, 2003).

46 Thomas D. Cook and Donald T. Campbell, 'The Causal Assumptions of Quasi-Experimental Practice: The Origins of Quasi-Experimental Practice', *Synthese* 68, no. 1 (1986): 141–80.

47 Leonid Peisakhin and Arturas Rozenas, 'Electoral Effects of Biased Media: Russian Television in Ukraine', *American Journal of Political Science* 62, no. 3 (2018): 535–50.

Cluster 3: Non-experimental methods

Non-experimental methods refer to methods that do not include a specific intervention. Instead, TAs, resources and environments are measured and observed in their natural setting.[48] This approach can be useful when the aim is not to distinguish specific causal relationships, but to observe how changes to one variable will affect the outstanding situation as a whole. This applies when the characteristic one is interested in cannot be manipulated (e.g., gender), or when the general topic of interest is broad and exploratory.[49]

Table 3. Non-experimental measuring methods[50]

Cluster	Method	Explanation	Main limitations	When to use
Non-Experimental methods	Causal loop diagrams & System dynamics models	Model the interactions between different variables in a system. Can be used to simulate the effects of different interventions.	Building a well specified model will require substantial previous knowledge about the system.	For simulation when the causal relationship within a system is well known.
	Agent based models	Model the interaction between agents and simulate how a possible intervention would affect and spread through a group of agents.	Trade-off between the simplicity and realism of the model.	For simulating interventions when the way the TA interact with each other is known.
	Structural models	Models incorporating economic theory and assumptions to estimate causal relationships between variables and simulate how the TA responds to changes.	Requires theoretical assumptions that might not hold in practice.	When there is enough theory available to make reliable assumptions needed to build structural models.
	Time series models	Models that utilise chronological data collected over time to make future predictions, analyse trends or identify anomalies. Can be used to disentangle short- and longer-term effects.	Establishing causality and obtaining data consistently through time might be challenging.	When the variable of interest is highly correlated with historical values and enough data is available.

Case study: The Anglo-Irish war, also known as the Irish War of Independence, is often considered as the first modern urban insurgency. The conflict escalated in 1916 with the Easter Rising insurrection. Although the insurrection failed, the aftermath of the conflict increased dissatisfaction with the British rule gradually leading to a guerilla war. The war ultimately came to an end in 1921 with the Anglo-Irish treaty and the creation of the Irish Free State. Using the Anglo-Irish war as a starting point, Edward Anderson built a system dynamics model that can be used as a basis to model the dynamics of insurgencies. The model consists of three main factors: incident suppression, insurgent creation, and war-weariness. The model was able to replicate the dynamic behaviour of the Anglo-Irish War and was further used to simulate how different policy choices would have affected the insurgency.

48 Carrie Cuttler, 'Overview of Non-Experimental Research', in *Research Methods in Psychology* (KPU, 2019), https://kpu.pressbooks.pub/psychmethods4e/chapter/overview-of-non-experimental-research/.

49 Cuttler.

50 Edward G Anderson, 'A Proof-of-Concept Model for Evaluating Insurgency Management Policies Using the System Dynamics Methodology', *Strategic Insights* 6, no. 5 (2007).

3.2. **Overview of data collection methods**

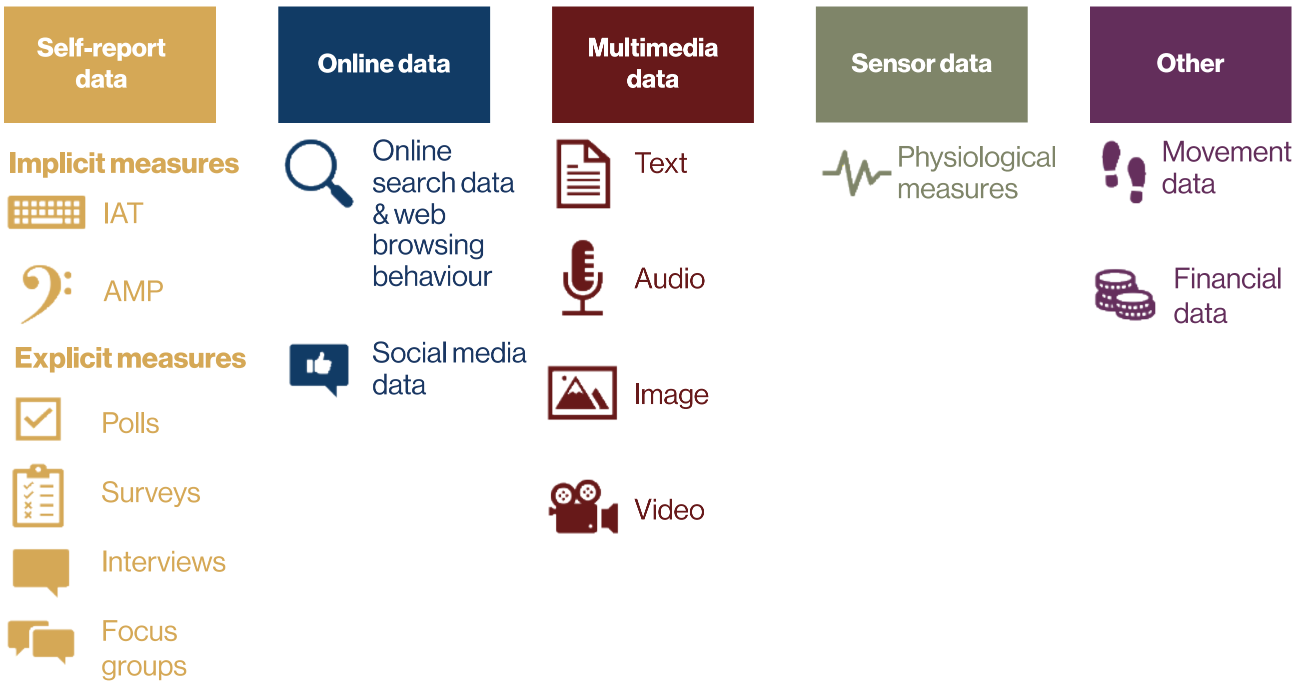
It is essential that there is a clear metric(s) that can be measured in order to track the operation's success both over time and in an instant depending on what is needed.

This section provides fifteen relevant data collection methods that can be used in conjunction with the measuring methods discussed in the previous section. The data collection methods have been grouped into five clusters as illustrated in **Figure 7**. A more extensive explanation of individual data collection methods can be found in the annex. Each description of a cluster starts with a brief explanation of the specific cluster and its respective data collection methods and is followed by a table with each cell containing an explanation of how each data collection method could be utilised to measure a specific outcome measure. Additionally, each table contains a brief case study that is used as an example to illustrate how these methods have been previously used in similar contexts.

The tables are built around the four relevant outcome measures: emotions, beliefs, knowledge and behaviour. Knowledge is what people know, beliefs refer to people's opinion on what they know, emotions refer to how people feel about what they believe and know, and behaviour is how people act based on what they know, feel, and believe. These four measures were chosen as they represent the main functions targeted by BI operations. The aim of a BI operation is to change a TAs behaviour by disrupting and influencing their decision-making processes. While changes in behaviour is the main measure of a BI operation, emotions, beliefs and knowledge can be measured to determine the impact on decision making processes.

Please note that this list is non-exhaustive. Other methods can be used, and new methods can be developed over time. The method that should be used is dependent on the context and the intent of the BI influencing operation in question. The key questions when deciding which factors to measure are twofold: (1) What needs to be measured: knowledge, beliefs, emotions, or behaviour? (2) What would success look like? It is essential that there is a clear metric(s) that can be measured in order to track the operation's success both over time and in an instant depending on what is needed.

Figure 7. Data collection methods



Cluster 1: Self-report data

The first cluster of data collection methods refers to methods where the target audience being observed (voluntarily) reports their emotions, beliefs, knowledge and behaviour. In simple terms, individuals possess the most intimate knowledge of their own experiences and perspectives, so simply asking can be an effective approach to gain valuable insights into their thoughts and behaviours. The application of self-report measures is broad as they can be used to measure not only observable factors, but also internal decision-making processes. Self-report measures are efficient and practical, as the complexity of the collected data can be adjusted to match the context and the implementation is simple and requires only a TA willing to answer questions.[51]

While an individual does have the greatest insight about themselves, it does not mean that the answer they provide will be fully accurate

Despite its advantages, there are some concerns regarding self-reporting data that should be considered. The biggest concern with self-report data is the credibility of the answers. While an individual does have the greatest insight about themselves, it does not mean that the answer they provide will be fully accurate. There are a number of reasons why an individual might, knowingly or unknowingly, not provide a fully credible answer. The credibility of an answer might be limited based on the respondent's memory, or respondents might deliberately be imprecise because of socially desirable responding or self-presentation concerns.[52] There are also privacy concerns that might limit the use of self-reporting, due to legal limitations regarding certain questions, or respondents' unwillingness to discuss sensitive topics. Different questioning techniques have been developed to mitigate the latter issue.[53] Self-reports require the cooperation of the respondent, which further limits its usage amongst hostile audiences. Using diaspora respondents could be a method to try and circumvent this, although this requires sophisticated methods to link answers from the diaspora to answers from these hostile audiences.

Self-reporting methods can be divided into implicit and explicit measures. Implicit self-report measures, like implicit association tests and affect misattribution, are methods where the respondent is not directly asked to state their emotions or beliefs about a given topic, but instead they are prime d indirectly. This can be used to measure underlying beliefs about sensitive topics that an individual might not be willing to openly admit to, or not consciously aware of.

Explicit self-report measures are methods that directly ask the target audience to self-report their emotions, beliefs, knowledge or behaviour. The way this can be implemented varies from simple polls to multi-person focus groups. The approach chosen should be based on the goal of the measurement and the resources available. A simple poll might be sufficient for a straightforward measurement where the options are limited, while a focus group could be used to obtain more creative and exploratory answers. The extent of information obtained can also be controlled by the use of closed and open-ended questions.

51 Richard W. Robins, R. Chris Fraley, and Robert F. Krueger, *Handbook of Research Methods in Personality Psychology*, Handbook of Research Methods in Personality Psychology (New York, NY, US: The Guilford Press, 2007).

52 Robins, Fraley, and Krueger.

53 These include randomised response, see Stanley L. Warner, 'Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias', *Journal of the American Statistical Association* 60, no. 309 (2012): 63–69, https://doi.org/10.1080/01621459.1965.10480775. , list experiments, see Judith Droitcour Miller, 'A New Survey Technique for Studying Deviant Behavior' (Dissertation, George Washington University, 1984)., Crosswise models, see Jun-Wu Yu, Guo-Liang Tian, and Man-Lai Tang, 'Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis', *Metrika* 67, no. 3 (1 April 2008): 251–63, https://doi.org/10.1007/s00184-007-0131-x., and Bayesian truth serum, see Dražen Prelec, 'A Bayesian Truth Serum for Subjective Data', *Science* 306, no. 5695 (15 October 2004): 462–66, https://doi.org/10.1126/science.1102081.

While closed ended questions can be easier to analyse and compare, open ended questions can offer richer content that can be analysed through Natural Language Processing (NLP) models.[54]

Table 4. Self-report data collection measures[55]

Data category	Data type	Emotions	Beliefs	Knowledge	Behaviour
Self-report data	Implicit	Emotions can be measured implicitly by examining which emotions individuals associate with a specific topic of interest. Underlying emotions can be obtained from open ended questions through NLP models.	Implicit association tests can be used to observe the type of beliefs an individual associates with a specific topic of interest. Affect misattribution measures how an individual projects their beliefs about a topic onto an ambiguous figure.		
	Explicit	Explicit self-report measures can be used to gather information about a TAs emotions. Emotions can be measured in a general level or related to specific events, situations, people or ideas.	Asking a TA to state their beliefs can be a good way to gain insight into their values. Different techniques have been developed to facilitate reporting culturally sensitive beliefs.	By asking direct questions about a specific topic, it is possible to measure the level of knowledge a TA holds.	Self-report measures can be used to gather information about a TAs past and intended behaviour, as well as hypothetical behaviour through vignettes. Different techniques have been developed, to facilitate reporting culturally sensitive behaviours.

Case study: The war in Afghanistan is the longest war in US history. After the quick toppling of the Taliban government, the Taliban remained in the rural parts of the country, starting a widespread insurgency against Afghan and coalition troops. US forces remained in country to help establish a democratic government and keep the Taliban from re-seizing power. U.S. leaders recognised the importance of fostering cooperation with the local population. To win their support, multiple influencing operations were carried out. In the beginning the operations achieved some success, but by 2005 there was an increasing amount of disappointment towards the administration and coalition forces. In 2012 the U.S. Marine Corps asked the RAND National Defence Research Institute to estimate the effectiveness of the U.S. military information and psychological operations. To measure effectiveness, RAND utilised a wide range of polls, surveys and interviews conducted with the local population, former Taliban members and U.S. personnel working on the operations. Estimating the causal effects of individual operations was tricky and rarely feasible as polling happened on a general scale, not connected to individual operations. Despite this, the researchers were able to recognise successful and unsuccessful operations, providing general recommendations on how to improve informational operations in the future.

54 These models include for example simple dictionary methods like LIWC and unsupervised learning models like Latent Dirichlet Allocation, see David M Blei, Andrew Ng, and Michael Jordan, 'Latent Dirichlet Allocation', *Journal of Machine Learning Research* 3 (2003): 993–1022.

55 Arturo Munoz, *U.S. Military Information Operations in Afghanistan: Effectiveness of Psychological Operations 2001-2010* (RAND Corporation, 2012), https://www.jstor.org/stable/10.7249/mg1060mcia.

Cluster 2: Online data

Online data refers to data that can be collected by observing and measuring a target audience's online activity. In the context of this paper, this revolves around information seeking and online networks. Analysing online content will be discussed under multimedia data. The data obtained can be quantitative or qualitative. For instance, number of website visits, likes or Google searches are examples of quantitative data. They can be used to measure individuals' knowledge, beliefs and behaviour by measuring the type of content that a target audience is looking for, consuming and engaging with. Qualitative data, like analysing search terms or social media networks can be used to further explore the target audience's online behaviour and connect it to a greater context. While data on a larger scale is usually available to the public, there might be some legal limitations to collecting more detailed, individual level data. However, there are some ways to alleviate these privacy and data protection concerns. These include using an intermediary or synthetic data.[56]

Table 5. Online Data Collection Measures[57]

Data category	Data type	Emotions	Beliefs	Knowledge	Behaviour
Online data	Online search data and web browsing behaviour		Analysing online searches and website visits can be indicative of the beliefs the TA holds.	Online searches and web browsing can be analysed to measure the level of awareness and knowledge around a topic. It also helps identify what type of information the TA is consuming.	
	Social media data	Social media engagement metrics and content can explicitly express and subsequently measure emotions.	Beliefs are widely discussed on social media. By analysing interactions, measurement metrics and networks it's possible to understand a TAs ideology and beliefs.	Analysing which content and users TAs interact with can be used to observe knowledge the TA holds and pursues.	

Case study: Since 2013 there has been significant ISIS, or ISIS inspired, terrorist attacks worldwide. Many believe that radicalisation through the internet and social media plays a role in the increase of attacks. A group of researchers set out to investigate whether interest in Islamic terrorism is preceded or followed by a terrorist attack. The amount of interest towards Islamic terrorism was measured based on the amount of Google searches of related key words. The researchers found a significant uptick in searches preceding terrorist attacks. This implies that there is a link between online information seeking and terrorist attacks, which further supports the notion that online data can be used as a measure to predict behaviour.

56 An overview of these emerging methods can be found in chapter 5

57 Carl E. Enomoto and Kiana Douglas, 'Do Internet Searches for Islamist Propaganda Precede or Follow Islamist Terrorist Attacks?', *Economics and Sociology* 12, no. 1 (2018): 233–47, https://doi.org/10.14254/2071-789X.2019/12-1/13.

Cluster 3: Multimedia data

Multimedia data refers to all media including text, audio, imagery and video. As the collection of multimedia data includes but is not limited to media found on the internet and social media, there is some overlap between multimedia and online data. To make a clear distinction between the two, following the purposes of this paper, online data is limited to the analysis of online activity and networks, i.e. how a TA searches for, engages with and distributes content. Meanwhile, multimedia data refers to the analysis of the content itself. If a text post is shared on a social media platform, multimedia data refers to the analysis of the text itself while online data refers to analysing who the post was shared with and who engages with it. Additionally, multimedia data does not only refer to media obtained from online sources like the internet, but any type of multimedia data imaginable: security camera content, newspaper articles, satellite imagery and phone conversations, among others.

A number of AI tools have been developed to analyse multimedia data. These tools can be used to identify and categorise individuals, places, objects, emotions and other content. With the large amounts of data available, analysis can be highly resource and time consuming. However, the development of AI and machine learning are expected to alleviate this issue, as they can be used as tools to analyse large amounts of data more efficiently.[58]

Table 6 - Multimedia Data Collection Measures[59]

Data category	Data type	Emotions	Beliefs	Knowledge	Behaviour
Multimedia data	Multimedia data	Emotions can be identified from text, images, video and audio by analysing visual, audio and linguistic cues.	Beliefs can be identified by analysing the content of text, images, video and audio with the help of AI and machine learning.	Knowledge of the TA can be identified by analysing the content of text, images, video and audio with the help of AI and machine learning.	Individuals and behaviour can be identified from text, images, video and audio with the help of AI and machine learning.

Case study: A lot of radical extremists and hate groups use the internet to spread their message. ISIS propaganda took on many forms on social media in the mid 2010s. A group of researchers set out to investigate which type of propaganda has the biggest impact on its audience and is most likely to radicalize. As the propaganda included a lot of audiovisual content, the researchers utilised deep learning models to identify violent imagery. For audio data they used supervised text classification to categorise the material based on the content of the message. Finally, the group used machine learning to identify pro-ISIS tweets from the users exposed and not exposed to the propaganda. The study found that exposure to non-violent propaganda increased the support for ISIS among the general audience, while violent propaganda had a negative effect. However, violent propaganda was still effective among extremists.

58 Charlie Kawasaki, '6 Ways AI Can Make Sense of Sensor Data in 2020', C4ISRNet, 14 February 2020, https://www.c4isrnet.com/thought-leadership/2020/02/14/6-ways-ai-can-make-sense-of-sensor-data-in-2020/.

59 Tamar Mitts, Gregoire Phillips, and Barbara F. Walter, 'Studying the Impact of ISIS Propaganda Campaigns', *The Journal of Politics* 84, no. 2 (April 2022): 1220–25, https://doi.org/10.1086/716281.

Cluster 4: Sensor data

Generally, sensor data can be used to refer to any type of data obtained through sensors. However, to avoid overlap with other data types, this paper only discusses sensors that can be used to obtain physiological measures. Physiological measures refer to bodily functions that can be either voluntary or involuntary. While they cannot be used to access an individual's thoughts, like knowledge or beliefs, they can be used to gain insight into some internal processes of an individual. Involuntary physical responses can indicate an individual's emotional state, while brain activity and eye tracking can be used to measure engagement and decision making. Physical data can also be a useful tool to detect emotions and behaviours that an individual is trying to hide, for example lying. However, most sensors that measure bodily functions need to be attached to the individual being measured. As this can be invasive and uncomfortable, strong cooperation with the TA is required which limits the use of these measures among certain target audiences.

Table 7. Sensor data collection methods[60]

Data category	Data type	Emotions	Beliefs	Knowledge	Behaviour
Sensor data	Physiological data	Heart rate, pupil dilation, sweat gland activity, blood pressure and brain activity can be used to identify the emotional state and its intensity of an individual. Some data can be collected through wearable devices like smart watches.			Heart rate, pupil dilation, sweat gland activity and blood pressure can be used to identify behaviour associated with lying. Detectable brain wave activity can be linked to changes in behaviour.

Case study: Effective propaganda has the ability to affect people on an unconscious level, bypassing conscious thoughts or decision-making processes. To test unconscious impact of propaganda, researchers measured the physical reactivity of women to propaganda produced by far-right, far-left and jihadist groups. The subjects were shown a variety of propaganda material compelling women to join extremist organisations. The gaze and pupil dilation of the women were tracked in order to identify what their attention was focused on and the level of impact it had on them. The researchers found consistent pupil dilation when the subjects were presented with images of gruesome violence. The researchers assess similar images can be used as a gateway function to additional ideological persuasion.

Cluster 5: Other data

For an indicator to serve as a good measure of success for any operation, it needs to be credibly linked to the intervention itself as well as the desired outcome. As behavioural influencing operations can be used to achieve a number of different goals, the different indicators that can be used to measure success are nearly unlimited. Because of this, it would be impossible to address every single data collection method possible. Instead as an example, two data types are discussed: movement and financial data, that represent explicit observable behaviour. After all, the ultimate goal of behaviour influencing is to influence behaviour.

Movement data refers to the analysis of the location and movements of a TA. One way to gain an advantage on the battlefield is to attack and defend favourable positions. Therefore, an objective of a BI operation could be to influence the location and movement of an opponent's troops. In that case, a measurement of success could be as simple as observing whether the opponents' troops have relocated to the desired location or not.

60 Mojtaba Heidarysafa et al., 'Exploring the Experiential Impact of Online Propaganda Using Eye-Gaze and Pupil Dilation : A Comparison across Three Ideological Groups' (U.S. Department of Justice Office of Justice Programs, 2019).

Financial data refers to the analysis of financial flows into, out of, and within a TA. Financial flows are a good way to understand the internal and external workings of an organisation. A BI operation can aim to disrupt a hostile organisation. Changes in its financial structure and flows can be one way of measuring success. One of the main objectives of Ukraine's information warfare has been to win over public and international support. One of the ways this could be measured is the number of international sanctions posed on Russia, and their effect on Russia's economy.

Table 8. Behaviour data collection methods[61]

Data category	Data type	Emotions	Beliefs	Knowledge	Behaviour
Other	Movement data				TAs travel patterns and locations can be identified through GPS tracking, satellite imagery, UAVs and geolocating.
	Financial data				Tracking financial transactions can be used to identify the spending habits of individuals and gain information about the networks and organizational structures of TAs.

Case study: In the fall of 2022, Ukraine launched a counteroffensive in Kharkiv, a northeast part of Ukraine that caught the Russian military by surprise. The attack was preceded by an extensive information operation aimed at diverting Russian troops. Kyiv repeatedly suggested that the objective of the counteroffensive was recapturing the city of Kherson in the south of the country. Ukrainian intelligence was able to monitor the locations and movements of Russian units to measure the success of the information operation. The deception was successful. The Russians reinforced their troops in Kherson, with some units being diverted from the Kharkiv region. The Ukrainians exploited this by attacking the undermanned Russian positions in the northeast and capturing a significant amount of ground.

Table 9. Overview data collection methods per outcome measure

Data category	Data type	Emotions	Beliefs	Knowledge	Behaviour
Self-report data	**Implicit**	**X**	**X**		
	IAT				
	AMP				
	Explicit	**X**	**X**	**X**	**X**
	Polls				
	Surveys				
	Interviews				
	Focus groups				
Online data	Online search data and web browsing behaviour		X	X	
	Social media data	X	X	X	
Multimedia data	Multimedia data	X	X	X	X
Sensor data	Physiological data	X			X
Other	Movement data				X
	Financial data				X

61 Santelises.

3.3. **Data-Method Capability (DMC)**

While the measurement method is used to establish whether or not the intended objective of the behavioural influencing method is achieved, data collection methods make it possible to translate the achieved effect into measurable data that is needed to determine the reached effect. Therefore, in order to operationalise effect measurement of behavioural influencing operations, one needs both the measuring and data collection method combined. Together, these two factors give the ability to successfully deploy effect measurement of the impact of an operation. In this paper, the term "Data-Method Capability" is used when discussing the combined use of a measurement method with a data collection method in order to establish which effect the behavioural influencing operation has reached.

While the measurement method is used to establish whether or not the intended objective of the behavioural influencing method is achieved, data collection methods make it possible to translate the achieved effect into measurable data that is needed to determine the reached effect.

Choosing the right Data-Method Capability is crucial to successful effect measurement. The relationship between measurement and data collection method is incredibly flexible, allowing for numerous ways to combine and utilise them together. Most data collection options can be used in combination with specific measuring methods, as long as it meaningfully measures the effect of the operation and there are enough resources to conduct data collection. Multiple data collection methods can be utilised simultaneously to measure the sustained effect of an operation. For example, to measure the effectiveness of a campaign encouraging political mobilisation, in addition to measuring actual voting behaviour it's possible to gather data on self-reported voting and information seeking.[62]

Because of the flexible way data collection and measuring methods can be combined, there are multiple possibilities to create Data-Method Capabilities (DMC). Therefore the choice in combination between the measuring and data collection method rests upon the context specific situation of behavioural influencing intervention. Every operation is unique, and the capability used should be chosen based on the operational context they are applied to. Operational context refers to the goals of the operation and the resources available. These largely dictate which capabilities can and should be used.

However, some DMCs are more likely to be used than others. For instance self-report measures can often be used together with RCTs to measure emotions, beliefs, knowledge or behaviour. This was the case in the experimental methods case study, as researchers set up a randomised control trial where some of the participants were assigned to follow a Twitter bot exposing them to opposite political views.[63] The effect of the intervention was determined by comparing how the political views of those exposed to the intervention changed compared to non-exposed individuals. The participants self-reported their political views by filling out a survey.

Additionally, multimedia data can be used well in conjunction with a difference-in-differences set up. Such a DMC was utilised in the multimedia case study, where researchers utilised AI and machine learning to identify pro-ISIS propaganda and the users that were exposed to it.[64] The researchers measured the effect of the propaganda by comparing the exposed individuals to a valid comparison group before and after the exposure. The level of support for ISIS was determined by analysing the content of their tweets.

62 Robert M. Bond et al., 'A 61-Million-Person Experiment in Social Influence and Political Mobilization', *Nature* 489, no. 7415 (September 2012): 295–98, https://doi.org/10.1038/nature11421.

63 Bail et al., 'Exposure to Opposing Views on Social Media Can Increase Political Polarization'.

64 Mitts, Phillips, and Walter, 'Studying the Impact of ISIS Propaganda Campaigns'.

Alternatively, some measuring and data collection methods do not work as well together. As an example, a DMC involving physical measures and System Dynamics Models[65] are rarely useful. SDMs are mainly used to model the dynamics of large-scale systems over long periods of time, while physical measures mainly measure an individual's momentary emotional reactions. Momentary emotional reactions are rarely representative of longer-term emotional impact, which SDMs are modelling.

Table 10. Examples Data-Method Capabilities

	RCT	A/B Testing	Pre-Post analysis	Difference-indifference	Regression discontinuity	Instrumental variables	Causal loop diagrams	Agent based models	Structural models	Time series models
Self-report data	✓									
Online search data and web browsing behaviour										
Multimedia data			✓							
Physiological data							X			
Movement data										
Financial data										

The next section discusses in greater detail different feasibility criteria that can limit the usage of specific capabilities.

65 For detailed information on SDM's please consult the annex.

4. Feasibility

In this section, the feasibility of the different measurement methods and data collection strategies is determined on the basis of a set of criteria and context-specific scenarios. The combined Data-Method Capabilities for establishing effects are thus now introduced into the military context. The military environment is prone to rapid changes, with scarce resources and information diluted by the fog of war.[66] Therefore this section aims at providing a thought-framework to help determine which effect measuring capabilities can be used, and on which occasions. It will thus assess whether DMCs are appropriate for different situations.

The five feasibility criteria are intended to provide a thought framework to help assess the combined choice of measurement and data collection method based on the specific context of the scenario.

The feasibility of each DMC will be evaluated based on five criteria: time, flexibility, data availability, cost effectiveness, and expertise. As suggested by **Figure 8** below, which provides an overview of the criteria, the criteria do not carry equal weigh when it comes to assessing the feasibility of a DCM. In other words, some criteria are more important in deciding whether or not a specific combination of measuring and data collection can be used. The weight that was given to the criterion is based on the impact they have on choosing the DMC. For example, the two criteria that are deemed to be the most important in assessing feasibility are the time and data availability constraints. Limited time and limited data availability impact the possible DCM combinations gravely. If these criteria are not sufficiently satisfied there is little to do to change that. For example, the time horizon of an information operation and the distance or accessibility to the TA collecting data are most often fixed and hard to change. Simultaneously, a criterion such as expertise is more easily adjustable. When the needed expertise is not yet present, this can be adjusted easier than when the needed data is not available. The five feasibility criteria are intended to provide a thought framework to help assess the combined choice of measurement and data collection method based on the specific context of the scenario. These criteria and their implementation have been designed outside of the intelligence process.

Due to the high context dependence when it comes to effect measurement of behavioural influencing operations in the military context, making generalizable statements about methods based on these criteria, averaged across scenarios, is not feasible. Therefore, the introduced five feasibility criteria will be addressed in a specific scenario on the brigade level. When going through the scenarios, the intended outcomes and the relevant target audience will first be addressed. Then, the possible data collection strategies will be determined, based on the specific context of the intervention. As the last step for this exercise, the five feasibility criteria are assessed. As the feasibility of the measuring capabilities is context specific, these scenarios are developed to illustrate what is feasible depending on the context of operation.

66 The fog of war refers to the uncertainty that characterises war situations and was first introduced by Prussian war theorist Carl von Clausewitz. See von Clausewitz, *On War*.

Figure 8. Feasibility Criteria

4.1. **Criteria**

Time

The first criterium of feasibility is time. It is split into three sub-parts: lead-in time, execution, and evaluation.

Lead in-time refers to the time before the implementation of the intervention. This includes the time needed to identify the intended goal of the intervention in terms of the needed outcome measures. The greater the lead in-time, the more preparation is needed before the method can be implemented. For some methods, like the pre- and post-analysis, the lead in time will consist of only assessing the current situation and making baseline measurements. Other methods like an RCT and DID require a longer lead-in-time as the target audience needs to be randomised into groups or a valid control group needs to be found. Finally, some methods require a substantial lead-in-time. It may take time to correctly build and define ABMs and SDMs, and it can take a long time to gather enough data for a time series model. For example, more lead-in time would mean more preparation is needed beforehand. This is not ideal in case of a very limited time frame to prepare intervention.

Execution refers to the implementation of the intervention and is associated with the fourth step in the OODA-loop: Act. When ABMs and SDMs are used for simulations, the execution time is as fast as a push of a button. Execution of interventions outside of simulations are more volatile. The perceived time horizon is dependent on the intensity of the conflict or warfare of engagement. As well as the level in the military organisation that is tasked with carrying out the operation. For example, when engaged in high intensity warfare the time horizon of implementation to evaluation is short, as the window between carrying out the operation and needing to witness the sustained effect is closely tied together. While engaging in a longer-term operation, where the time-to-effect is calculated to take longer, the importance of time is less stringent than when dealing with high intensity situations. This again highlights the importance of context dependency.

The third and final step is evaluation. Evaluation refers to the time it will take from the implementation to the point where it is possible to measure its impact. Evaluation falls between two OODA-loops where the effect of the previous loop is evaluated, and the feedback is used to launch the next loop. The evaluation period depends on the intervention, the data collection method, and the impact we are interested in. A simulation can take a couple of seconds, short term impact can be observed within minutes or days, and long-term effects can take up to months or years.

A good influencing operation is flexible and can be adjusted to match the changes in the informational landscape, ideally based on real-time monitoring.

Flexibility

Flexibility refers to how easily a method can be adjusted, which is inherently also tied to the previous criteria. As a short time horizon, for example in high intensity conflict, will leave less room for adjustments. The information environment is rapidly changing, especially in a military context. A study conducted by RAND points out that when dealing with information operations, one should be prepared to fail fast.[67] This means that if the operation is not working, it should be stopped or adjusted. A good influencing operation is flexible and can be adjusted to match the changes in the informational landscape, ideally based on real-time monitoring. For simulation-based methods, adjustments to the models are easy to make. For more structured methods like an RCT or DID, adjustments could negatively affect the set-up.

Data availability

Unlike kinetic warfare, information warfare is not by definition tied to a specific location. Information warfare can take place on a local or theatre level. When an operation occurs in a local theatre it means that it is conducted in a specific geographical area including sea, air, and land. Local theatre operations usually revolve around the tactical and operational level. However, operations can also take place at theatre level. These tend to have a more strategic component. If an operation is happening on a local theatre level, it means that there is a physical presence within the area of operations. However, information operations can also be executed on theatre level without any physical presence, for example online.

Data availability incorporates distance to the target audience. This refers to the physical presence of the target audience versus the physical distance of the corps, division, or brigade carrying out the influencing operation to the target audience. This is related to whether or not the military actors are considered to be on a local or theatre level. Depending on the Data-Method Capability, it might be preferable to be physically close to the TA. This applies for example when using self-report data such as interviews or focus groups. However, for determining effects, the distance to the TA can be easily bridged by moving online. The only data collection criterion that requires a close proximity presence would be sensor data.

The distance criterium is thus inherently linked to data availability. For data availability, the question is whether there is direct access to data, indirect access to data, or no access

67 Paul, 'Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade'.

to data. No access to data to carry out the measurement only leaves the possibility of conducting a simulation exercise to estimate likelihood of success. Indirect access to data can have different implications depending on the scenario. Instead of measuring the outcome measures, emotions, knowledge, beliefs or behaviour, of the TA in the first order, you can for example measure proxy variables that come close to the actual indicators when direct access is not possible. For example, focusing on engagement metrics on social media instead of directly observed behaviour. Or if there is no direct access to the TA, a potential proxy group can be diaspora in other geographical areas. But indirect access can also mean that instead of observing the outcome measures first hand, they are observed through an intermediate medium such as camera images, or audio recordings.

The data availability criterion also incorporates the architecture that is needed to carry out data collection. Having access to data is only step one. The data needs to be harvested, which, depending on the data collection method, can require different types of equipment. After this the data still needs to be processed, stored, and made ready for analysis. Before the collected data can be used as input for the respective measurement method.

Cost-benefit analysis

Trade-offs have to be made between the accuracy of the measurement method and the costs of implementation.

As resources are usually limited in a military context, it is important that they are used effectively. In high intensity warfare, the first priority is to complete the mission and achieve the desired effect. Measuring how this is exactly done is at best a secondary priority, and a rough estimate is usually sufficient enough. Therefore, while some measurement methods might produce more accurate results, they might not be feasible if the resources required for implementation are too great. Trade-offs have to be made between the accuracy of the measurement method and the costs of implementation.

While the measurements obtained from an RCT are highly accurate, because of randomisation the effort needed to implement this setup can be considerable. The same applies to A/B testing but especially for some online interventions, reliable results can be obtained fast and cost effectively. Compared to an RCT, DID requires less resources to implement and under the right circumstances produces measurements comparable to RCT. Pre & post analysis requires little resources to implement, but the quality of results is highly dependent on the context. SDMs and ABMs can be tailored to fit the needs of the operations. The more accurate the results need to be, the more effort has to be put into building the models. Finally, for time series models gathering large time series datasets can require a lot of time and resources, and depending on what is measured, the measurements can be accurate or less accurate.

Expertise

One limiting factor for the usage of different methods is the amount of expert knowledge required to use them. There are different kinds of expertise that can be required, such as expertise to use the models, expertise to carry out the data collection, expertise to process and analyse the collected data, expertise in behavioural analysis for interpreting results, and expertise in military operations to assess whether the intended goals have been achieved. If a measuring method requires certain expertise, it might not be feasible for levels of the military that lack it. While pre-and post-analysis does not necessarily require any expertise to use, time series models might, depending on the model used. RCTs and A/B testing require some expert knowledge to make sure the randomisation is done correctly and there are no spillover effects. Correspondingly, it requires some expertise to identify a valid control group for Difference-in-Difference. ABM & SDM require a great deal of expertise to create, as a lot of knowledge about the target audience and overall system is needed in order to make accurate assumptions. However, expertise for implementation is not universal across the different stages of implementation. Expertise of, for instance, the measuring methods could be needed in the higher up level when designing the intervention, but not at the grassroot level where the intervention is carried out.

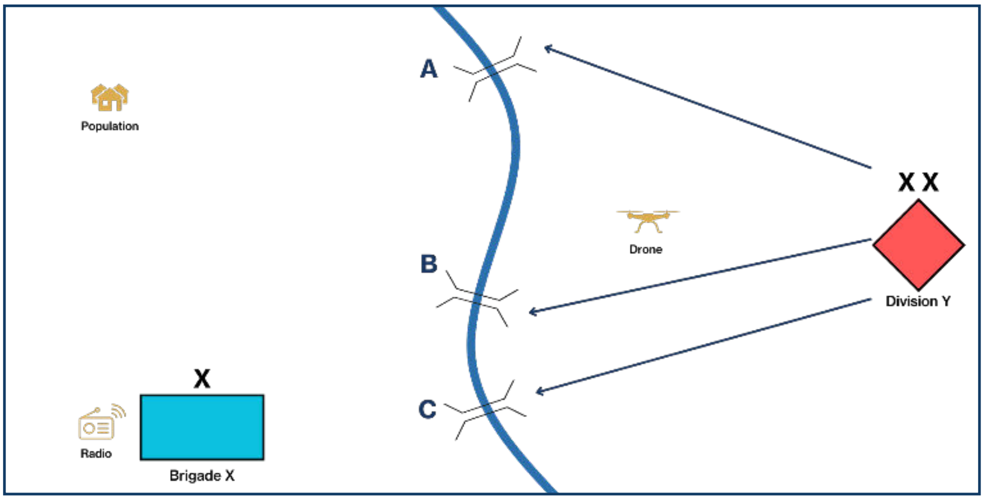
4.2. **Scenario**

Making generalizable statements about methods, averaged across scenarios is not feasible, due to the high context dependency.

Making generalizable statements about methods, averaged across scenarios is not feasible, due to the high context dependency. Therefore, the introduced five feasibility criteria will now be addressed in a specific scenario on the brigade level. The aim of this exercise is to go through the different steps leading up to measurement and to illustrate how the five feasibility criteria help determine the best Data-Method Capability. This specific scenario takes place on the brigade level. For reference of military units, a brigade consists of a few battalions and usually comprises anywhere between approximately 2,000 to 5,000 troops. Two or more brigades make up a division. The intention of providing this scenario is to illustrate how the previously introduced measurement, data collection methods and feasibility criteria can be introduced in a relevant military context.

The scenario goes as follows. There is a brigade X that needs to defend against an enemy division Y. Brigade X receives the order to defend, after which it has 36 hours until the enemy division Y will most likely attack. Those 36 hours thus constitute the time horizon for conducting the influencing operation. The target audience of this operation is enemy division Y.

Figure 9. Brigade level scenario



The terrain where the operation takes place is the following. Approximately 10 km behind the frontline, three bridges cross over a river. From north to south, these are bridge A, B, and C. As **Figure 9** depicts, brigade X is situated west of the river. Division Y finds themselves east of the river. Part brigade X's concept of operation is to lure division Y north towards bridge A. This should set all conditions in place for a concentration of the enemy's main effort east of bridge A. All three bridges are being prepared for destruction, with the explosives already in place.

In order to lure the enemy's main effort to bridge A, brigade X conducts three main information activities. All these activities revolve around influencing division Y to concentrate main effort around bridge A. The information operation is thus deemed successful when division Y has moved their main effort to bridge A within the 36 hours before attack.

The three information activities are the following. First, they inform the local population in the area of brigade X, thus west of the river, that bridge A is not prepared for destruction. As the time horizon is 36 hours, in the pre-intervention phase, brigade X is assumed to either already have established or to find out the most effective communication strategies to convey this message to the local population in the least amount of time necessary. Second, brigade X sends an unencrypted message on its combat net radio that bridges B and C are prepared for destruction. Third, X uses a drone to drop a piece of paper with a brigade defence plan on it in which the enemy attack is expected to be on bridge B and C. Besides the information activities, brigade X positioned their own troops with main effort south of the area, towards bridge C.

Now that the goal of the information operation, target audience, context of intervention, and the means to carry it out are known. We look at how the information operation can be assessed for its effectiveness, all leading up to the main goal of the intervention being influencing division Y towards bridge A.

Starting with the data collection, there are three possible ways to collect data in this scenario. The first being online data, specifically open-source social media data. This allows brigade X to follow how the message of the first information activity is spread amongst the population, with the aim to also reach division Y. It possibly includes imagery or video footage of division Y that is spread through social media. The use of these images is considered to be multimedia data. The second possible form of data collection is also multimedia data, and in particular audio data from the local radio. Following the local radio allows X to determine whether or not they were successful in convincing the local population that bridge B and C will be up for destruction while bridge A supposedly is left intact. Lastly, brigade X can conduct reconnaissance to collect movement data of the TA. This allows them to track whether or not division Y is in the intended area of bridge A, which would be in line with the leaflets that were dropped.

In terms of measuring methods, due to a low control over the target audience, combined with the short-term horizon, the quasi-experimental methods are most fitting in this scenario. The most viable option here is pre-post analysis. Pre-post analysis is the simplest way to measure the impact of the information operation on behavioural change. In this context, it consists of observing the TA before and after the intervention, and comparing the two to each other.[68] In this particular case, this would be observing the location of division Y before and after implementing the three information activities. The simplicity of the analysis makes it a useful tool when the resources and the control over the target audience is limited.

68 Stratton, "Quasi-Experimental Design (Pre-Test and Post-Test Studies) in Prehospital and Disaster Research."

Looking at the five feasibility criteria helps explain why pre-post analysis is the most viable option for brigade X. The first feasibility criterium is time. The time horizon for this operation was 36 hours. The time criteria is split into three sub-parts: lead-in time, execution, and evaluation. Lead in-time refers to the time before the implementation of the intervention, which is considered the most important factor for this scenario given the limited time horizon before the attack by division Y. This includes the time needed to identify the intended goal of the intervention in terms of the needed outcome measures. In this case, this is the movement of division Y towards bridge A. Given the 36 hours, the lead-in time gives a limited window for preparation. For pre- and post-analysis, the lead in time consists of only assessing the current situation of where division Y is currently situated, as well as making baseline measurements of their location and how they are expected to move. This leaves the final two steps of execution and evaluation. Both are considered to be short in time horizon due to the intensity of the conflict. Which is why pre-post analysis, given that the needed data availability and expertise is present, is a fitting option.

For pre- and post-analysis, the lead in time consists of only assessing the current situation of where division Y is currently situated, as well as making baseline measurements of their location and how they are expected to move.

The next criteria is flexibility. Flexibility refers to how easily a method can be adjusted. The short time horizon of this scenario, a notion that generally applies in high intensity conflict, leaves less room for adjustments of the method. The third criterium, data availability, for this scenario is considered to be satisfied. Physical distance to the TA is considered to be close, especially given the expected attack in 36 hours. Therefore, access to the TA for collecting data is considered as direct, either on a first order through reconnaissance by brigade X themselves, or via the observed social media data and radio transmission. The latter two are considered to provide data access more indirectly, as the social media data and radio transmissions are also used to observe whether the information activity targeting the local population has been successful. More difficult to satisfy will be the architecture that is needed to collect and analyse the harvested data. Especially when it comes to the resources available. Which is linked to the fourth criterium of cost-benefit. Pre-post analysis requires little resources to implement, but the quality of results is highly dependent on the context. Additionally, with the predicted attack of division Y, the resources allocated to carrying out the data collection and pre-post analysis might be scarce. So, the accuracy of measurement, and with that the quality of results, might not be the highest compared to other DCMs. However, satisfactory for this particular scenario.

The last criterium revolves around the needed expertise for carrying out the measurement. Pre-post measurement does not necessarily require extended modelling expertise to carry out. However, the most difficult areas regarding expertise to satisfy in this scenario will be expertise to carry out the data collection and expertise to process and analyse the collected data. As mentioned under data collection, the harvesting, storing, and analysing of the data within 36 hours in an active combat situation is less easy to satisfy and comes with additional requirements that are outside the scope of this example.

5. Discussion on emerging technologies

Important to keep in mind here is that prediction, especially about the importance and impact of emerging technologies, remains very difficult. As technological trends very rarely evolve in a linear fashion, and in, this particular case, the synergies between emerging technologies are as important as the emergence of the technologies themselves.

Our current living environment is characterised by rapidly evolving information technology that is changing the way we live, think, communicate, and ultimately how we fight. This section aims to touch on the importance of emerging technologies in the military context specifically when it comes to effect measurement of influencing operations. What is possible today, tomorrow, and the day after tomorrow in terms of influence and its measurement, is changing rapidly due to these emerging technologies. The methods and practices previously proposed for measurement might thus be enhanced or exchanged sooner rather than later due to these developments. Important to keep in mind here is that prediction, especially about the importance and impact of emerging technologies, remains very difficult. As technological trends very rarely evolve in a linear fashion, and in, this particular case, the synergies between emerging technologies are as important as the emergence of the technologies themselves. For instance, this applies to the synergy between AI and Autonomous Systems. Therefore, any expected impact or progress due to the implementation of these technologies are preliminary, and further changes should be taken into account as their impact at this point in time remains speculative and is still to be determined.

This section first maps out the links between military and civilian industries when it comes to technological innovation. The spill-over between these industries is used as a theoretical framework to think about innovations taking place on both sides and what this link means for effect measurement of influencing operations. The intertwining between these industries are evident throughout all levels of defence, both kinetic and non-kinetic. It has led to a non-stop search for innovation in the digital arena, specifically when it comes to the use of digital technologies and AI in military systems.[69] After scoping out the historical and current intertwining of innovation technologies between the civilian and defence industries, the section looks at what this link means for effect measurement of influence operations. Here the focus is in particular on information technologies, AI, and Large Language Models. While much is being written on the usage and impact of these emerging technologies on carrying out influencing operations, the focus of this discussion is how they can aid the question of effect measurement.

69 William Merrin and Andrew Hoskins, 'Tweet Fast and Kill Things: Digital War', *Digital War* 1, no. 1–3 (2020): 184–93.

5.1. Synergies between military and civilian innovation

There has been a long lasting and evolving relationship between military and civilian industries when it comes to innovation. The link between military and civilian innovation is, and has been, a two-way street with each sector impacting the other. This two-way street can generally be divided into two categories: spin-offs and spin-ins.[70] Spin-offs describe technology that is originally designed for military use but is later also transferred to civilian applications.[71] An example of this is the Global Positioning System (GPS). Spin-ins describe the opposite transfer; civilian technology that is used for military purposes.[72] As an example, AI technology is largely being utilised in various military applications, also in cooperation with private companies.[73] An example of this is the Pentagon's Project Maven, where the US Department of Defence worked together with Google LLC to develop an AI system to interpret video imaginary from drones, before Google pulled back due to concerns raised by their employees.[74] Project Maven was later taken over by another private company named Palantir.[75] Palantir's AI technology is used close to the front line, aiding the Ukrainian Army to increase accuracy and speed of artillery strikes.[76]

The link between military and civilian innovation is, and has been, a two-way street with each sector impacting the other.

This interlinkage between civilian and military industries, however, is not new. During the Cold War, with geopolitical tensions high and the arms race in full swing, innovation was largely driven by the defence sector. During this time, the relationship between the civilian and defence industries have been described as "the Military Industrial Complex", a definition coined by Dwight Eisenhower.[77] With the financial backing of the Pentagon, technologies were developed for military use that slowly trickled down to civilian applications.[78] Until the end of the Cold War, the spin-off transfer was dominant. This dynamic changed at the end of the Cold War as military expenditure declined. The defence industry was largely concentrated to a few specialised companies, shifting away from continued cooperation with large commercial enterprises. As a result, civilian R&D emerged as the new driving force behind innovation, with a spin-in into military applications.[79] Examples of this is the application of AI for defence purposes, more specifically the development of Robotics and Autonomous Systems (RAS). With decreasing investments into defence R&D, the defence sector has little to no chance to keep up with the technological development of commercial companies[80],

70 Maaike Verbruggen, 'The Role of Civilian Innovation in the Development of Lethal Autonomous Weapon Systems', *Global Policy* 10, no. 3 (2019): 338–42, https://doi.org/10.1111/1758-5899.12663.

71 Lewis M. Branscomb et al., *Beyond Spinoff: Military and Commercial Technologies in a Changing World* (Harvard Business Publishing, 1922).

72 Verbruggen, 'The Role of Civilian Innovation in the Development of Lethal Autonomous Weapon Systems'.

73 Emily Gilbert, 'Military Geoeconomics: Money, Finance and War', in *A Research Agenda for Military Geographies* (Edward Elgar Publishing, 2019), 100–114, https://www.elgaronline.com/display/edcoll/9781786438867/9781786438867.00014.xml.

74 Daisuke Wakabayashi and Scott Shane, 'Google Will Not Renew Pentagon Contract That Upset Employees', *The New York Times*, 1 June 2018, sec. Technology, https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html.

75 Becky Peterson, 'Palantir Grabbed Project Maven Defense Contract after Google Left the Program: Sources', Business Insider, 10 December 2019, https://www.businessinsider.com/palantir-took-over-from-google-on-project-maven-2019-12.

76 George Grylls, 'Ukraine Is Outflanking Russia with Ammunition from Big Tech', 12 September 2023, sec. news, https://www.thetimes.co.uk/article/ukraine-is-outflanking-russia-with-ammunition-from-big-tech-lxp6sv3qz.

77 Charles J. Dunlap, 'The Military-Industrial Complex', *Daedalus* 140, no. 3 (2011): 135–47.

78 William J. Lynn, 'The End of the Military-Industrial Complex: How the Pentagon Is Adapting to Globalization', *Foreign Affairs* 93, no. 6 (2014): 104–10.

79 Branscomb et al., *Beyond Spinoff: Military and Commercial Technologies in a Changing World*.

80 Lynn, 'The End of the Military-Industrial Complex'.

or its adversaries.[81] Therefore, having access to cutting edge technology requires a good relationship with the private sector.[82] An example of such effort is the Defence Innovation Accelerator for the North Atlantic (DIANA) initiative by NATO, announced in 2022.[83] DIANA provides a platform for cooperation on Emerging and Disruptive Technologies between military and industry innovators and investors, on topics such as big data, AI, autonomy, and quantum technologies.[84] The goal of the Accelerator is to provide innovators of deep tech and dual-use technologies in NATO countries funding, and ultimately fast track the adaptation of their technologies to defence and security purposes.[85]

For example, the US, a leader in defence R&D, has faced issues with companies finding cooperation with the defence sector both legally challenging and morally questionable.[86] The military still continues to invest in new technologies as an angel investor. It is suggested that a large part of new innovations have its roots in government funding.[87] For example the development of touchscreens and virtual assistance have been at least partly funded by the military.[88] Therefore, the innovative cooperation between the defence and private sector may be deeper than what might appear on the surface with the current dynamic being described as a "Military-Commercial Complex".[89]

The same technologies that will impact the influencing interventions themselves will also aid how their effects are measured.

5.2. Building the plane while flying it

Advances in information technology provide rapidly expanding possibilities to employ information both as a tool and as a weapon. Through these advances, information-based behavioural influencing is no longer only taking place through traditional channels such as newspapers and radio, but is increasingly moving online making use of the newest emerging technologies. As described above, dynamics and synergies between the military and civilian innovation space play an essential role in these developments. Besides the rapidly evolving capabilities used for information-based behavioural influencing, the ways their effects can be measured are changing too. Which means that the same technologies that will impact the influencing interventions themselves will also aid how their effects are measured. The focus of this paragraph remains on the latter. Three broad strands of emerging technologies are highlighted, those are Artificial Intelligence, Natural Language Processing (NLP), and Large Language Models (LLMs). It is important to note here that these technologies are interlinked.

AI is considered the overarching field of computer science that deals with a broad range of subfields, techniques, and methods that focus on building systems that operate independently. NLP in turn is a specific application of AI and linguistics.[90] NLP, also known as

81 Tabby Kinder, 'Silicon Valley Chiefs Urge Pentagon Procurement Overhaul', *Financial Times*, 26 June 2023, sec. Tech start-ups, https://www.ft.com/content/45da39f2-4e05-46f1-96f4-813fbba79b16.

82 Gilbert, 'Military Geoeconomics'.

83 NATO, 'NATO Approves 2023 Strategic Direction for New Innovation Accelerator', NATO, 10 December 2022, https://www.nato.int/cps/en/natohq/news_210393.htm.

84 NATO, 'About DIANA', About Diana, accessed 12 September 2023, https://www.diana.nato.int/about-diana.html.

85 NATO, 'NATO Approves 2023 Strategic Direction for New Innovation Accelerator'.

86 Lynn, 'The End of the Military-Industrial Complex'; Gilbert, 'Military Geoeconomics'.

87 Richard Hawkins, 'Marianna Mazzucato The Entrepreneurial State: Debunking Public vs Private Sector Myths', *Science and Public Policy* 42, no. 1 (1 February 2015): 143–45, https://doi.org/10.1093/scipol/scu071.

88 Gilbert, 'Military Geoeconomics'.

89 Gilbert.

90 Prakash M. Nadkarni, Lucila Ohno-Machado, and Wendy W. Chapman, 'Natural Language Processing: An Introduction', *Journal of the American Medical Informatics Association: JAMIA* 18, no. 5 (2011): 544–51, https://doi.org/10.1136/amiajnl-2011-000464.

computational linguistics, uses computational techniques in order to learn, understand, and produce content, based on human languages.[91] LLMs are a recently developed application of AI and NLP, that is trained on large amounts of text-based datasets to understand, synthesise, and use that content in order to create new content.[92] Such as, answering questions, or generating text. Well-known recent examples of LLMs are ChatGPT-3 and 4.

Lastly, some technologies are introduced which can aid data protection and privacy while collecting the data that is needed to carry out effect measurement. Data protection is one of the aspects that highlights the importance of meaningful human oversight when it comes to employing AI-based technologies. As measuring effects requires data collection, and the specific uses of AI, NLP, and LLMs require access to large datasets, the question of privacy and data protection becomes increasingly more important. This closing section of the paper functions as a discussion space to highlight potential uses of emerging technologies regarding measuring effects of behavioural influencing, it thus not touches upon constraints, nor risks of implementation.

Data protection is one of the aspects that highlights the importance of meaningful human oversight when it comes to employing AI-based technologies.

Artificial Intelligence

Artificial Intelligence is a technology that is already widely used in civilian day to day life. It is for example used to identify patterns in consumer behaviour. But those applications and their impact are not strictly kept within civilian applications. As AI can and will change the behavioural influencing interventions in terms of tactics and capabilities, it also provides opportunities when it comes to measuring effects.

One such example is the use of Network Analysis to follow Target Audiences (TAs). As it allows for following the specific disseminators of information in order to pinpoint whether or not the behavioural influencing tactic reaches the intended TA. Network Analysis allows for identification of key influencers, communities, or nodes that play a significant role in disseminating information. A specific application would be Social Network Analysis, making use of online data. By examining the characteristics and behaviours of an individual's social network connections, assumptions can be made about their preferences. For instance, this technique has been applied to map sentiments on Twitter regarding Covid-19 related themes.[93] More specifically, Social Network Analysis was used to first determine the dominant topics, as well as whether sentiments were positive, neutral, or negative.[94]

Predictive Analytics uses AI, particularly machine learning, to develop forecasting or predictive models.[95] Machine learning is yet another branch under the AI umbrella, referring to the ability of computers to use data and algorithms in such a way that they independently learn and improve their own accuracy, thus imitating the way humans learn.[96] This in turn can be used for simulation purposes to estimate likelihood of effectiveness of influencing operations

91 Julia Hirschberg and Christopher D. Manning, 'Advances in Natural Language Processing', *Science* 349, no. 6245 (17 July 2015): 261–66, https://doi.org/10.1126/science.aaa8685.

92 Nicholas Carlini et al., 'Extracting Training Data from Large Language Models', *30th USENIX Security Symposium*, 2021, 2633–50.

93 Man Hung et al., 'Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence', *Journal of Medical Internet Research* 22, no. 8 (18 August 2020): e22590, https://doi.org/10.2196/22590.

94 Hung et al.

95 S.A. Khan and Reemiah Muneer, 'A Novel Thresholding for Prediction Analytics with Machine Learning Techniques', *International Journal of Computer Science and Network Security* 23, no. 1 (2023): 33–40, https://doi.org/10.22937/IJCSNS.2023.23.1.5.

96 M. I. Jordan and T. M. Mitchell, 'Machine Learning: Trends, Perspectives, and Prospects', *Science* 349, no. 6245 (17 July 2015): 255–60, https://doi.org/10.1126/science.aaa8415.

or forecast potential future behaviour or target audiences for planning purposes to determine intended goals of an operation.[97] One example is a study that provided description of future behaviour of ISIS using machine learning .[98] Predictive Analytics can make use of historical data, but also current data such as data generated from performing Social Network Analysis explained above.

Natural Language Processing

In a study done by the Artificial Intelligence Software Architectures and Algorithms Group of the Lincoln Laboratory at the Massachusetts Institute of Technology, Natural Language Processing was used to develop a framework to quantify the level of impact of influence operations undertaken by foreign hostile actors within a network.[99] The developed framework aids the identification of actors spreading disinformation, as well as the assessment of their efforts. This framework is used to automate detection and depiction of influencing operations done by hostile foreign actors. It is thus not in first instance a framework to use for information-based influencing operations carried out by one's own armed forces. However, as outlined in the main contributions, the use of the framework to detect, classify, discover and estimate the causal impact as well as performance of influencing operations, specifically done online, does provide opportunities for effect measurement of interventions carried out by the home armed forces.

> The developed framework aids the identification of actors spreading disinformation, as well as the assessment of their efforts.

Another, more general, application of NLP for effect measurement of influencing operations is Sentiment Analysis. Sentiment Analysis refers to the use of large bodies of text, mostly originating from media, to determine beliefs or emotions of the target audience.[100] Sentiment analysis can be used for carrying out Target Audience Analysis, for example to determine country image or perceptions of the TA towards a country or armed forces in particular.[101] Additionally, it can thus aid in measuring effectiveness of influencing operations by comparing the sentiments amongst the TA before and after the influencing operation took place.

Large Language Models

A study carried out by OpenAI, the company behind ChatGPT, Georgetown University's Centre for Security and Emerging Technology and the Stanford Internet Observatory, researched the impact of LLMs on influence operations.[102] The study focuses on three dimensions of influence operations, being actors, behaviours, content[103], and the misuse

97 Benjamin Jensen and Ryan Kendall, 'Waze for War: How the Army Can Integrate Artificial Intelligence', War on the Rocks, 2 September 2016, https://warontherocks.com/2016/09/waze-for-war-how-the-army-can-integrate-artificial-intelligence/.

98 Andrew Stanton et al., 'Mining for Causal Relationships: A Data-Driven Study of the Islamic State', in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15 (New York, NY, USA: Association for Computing Machinery, 2015), 2137–2146, https://doi.org/10.1145/2783258.2788591.

99 Steven T. Smith et al., 'Automatic Detection of Influential Actors in Disinformation Networks', *Proceedings of the National Academy of Sciences* 118, no. 4 (26 January 2021): e2011216118, https://doi.org/10.1073/pnas.2011216118.

100 Ganesh Kumar Wadhwani et al., 'Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia–Ukraine War', *SN Computer Science* 4, no. 4 (21 April 2023): 346, https://doi.org/10.1007/s42979-023-01790-5.

101 Seow Ting Lee, 'A Battle for Foreign Perceptions: Ukraine's Country Image in the 2022 War with Russia', *Place Branding and Public Diplomacy*, 28 November 2022, 1–14, https://doi.org/10.1057/s41254-022-00284-0.

102 Josh A. Goldstein et al., 'Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations', 2023.

103 Camille François, 'Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses' (Graphika and Berkman Klein Center for Internet & Society, 2019).

and threats of LLMs for disinformation.[104] However, the study does highlight opportunities of positive impact that LLMs can have on measuring effects of such operations, it also points out the general impact of LLMs on information-based influence. LLMs can potentially make influencing operations inherently more effective, and thus also a larger potential threat when employed by foreign actors, particularly due to the difficulty of identifying when they are employed during influencing operations.[105] Focusing on online influence operations, in order to determine their effectiveness, the focus is often on gathering metrics associated with engagement.[106] These metrics are then used as a proxy for behavioural change, for which a comparison group is needed to determine changes that can be attributed to the influencing operation. Here LLMs can help twofold. First, LLMs can be trained to identify trends and patterns,[107] which aids the recognition of changes in behaviour. Second, since LLMs have the capability to recreate human answers to questions, and thus responses, they could be used to compile a hypothetical comparison group to test measures of effectiveness.

Emerging technologies can aid data protection and privacy while collecting the data that is needed to carry out effect measurement.

Data protection technologies

Emerging technologies can aid data protection and privacy while collecting the data that is needed to carry out effect measurement. As discussed, measuring effects of information-based influencing operations requires data collection. However, the specific uses of AI, NLP, and LLMs require access to large datasets for training purposes of these technologies. Therefore, the question of privacy and data protection becomes increasingly more important. Additionally, data breaches have led to increased concern about the security of private data. These privacy concerns can limit the ways organisations can collect and use private data. A number of methods are being developed that would allow organisations to utilise personal data without compromising privacy.

An intermediary can be used that can provide usable data to organisations while simultaneously protecting the privacy of individuals providing the data. One way of doing so is to use a deep learning model like generative adversarial networks (GANs) to generate synthetic data that closely mimics the original dataset.[108]

An alternative to using intermediates is the emerging field of “remote data science”. Remote data science techniques enable data analysis without the need to collect data from individual’s devices. Techniques like federated learning allows for AI training remotely on individual devices without the need to pull data into a collective database.[109] Encryption techniques like homomorphic encryption makes it possible to analyse data while simultaneously keeping it encrypted the entire time.[110]

104 Goldstein et al., ‘Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations’.

105 Goldstein et al.

106 Christopher A. Bail et al., ‘Assessing the Russian Internet Research Agency’s Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017’, *Proceedings of the National Academy of Sciences* 117, no. 1 (7 January 2020): 243–50, https://doi.org/10.1073/pnas.1906420116.

107 Benjamin Jensen and Dan Tadross, ‘How Large-Language Models Can Revolutionize Military Planning’, War on the Rocks, 12 April 2023, https://warontherocks.com/2023/04/how-large-language-models-can-revolutionize-military-planning/.

108 Piyush Anand and Clarence Lee, ‘Using Deep Learning to Overcome Privacy and Scalability Issues in Customer Data Transfer’, *Marketing Science* 42, no. 1 (January 2023): 189–207, https://doi.org/10.1287/mksc.2022.1365.

109 Kim Martineau, ‘What Is Federated Learning?’, *IBM Research Blog* (blog), 9 February 2021, https://research.ibm.com/blog/what-is-federated-learning.

110 Xun Yi, Russell Paulet, and Elisa Bertino, *Homomorphic Encryption and Applications*, SpringerBriefs in Computer Science (Springer, 2014).

6. Conclusions

This paper aimed to provide an overview of methods to measure the effect of behavioural influencing operations in a military context, with a focus on the processes and different stages associated with effect measurement. In doing so, we've sought to explore a part of the behavioural influencing process, in a military context that remains underdeveloped in broader research. As such, this paper serves as a basis and steppingstone for further research into the topic that is effect measurement of behavioural influencing in the military context. Areas of further research include, broadening the research towards a range of different scenarios to aid the identification of potential best practices or the impact of environmental specific factors on effect measurement in order to test the robustness of the measuring methods.

As the research is scattered across different fields, it also provides opportunities for further research, if and when conducted in a multidisciplinary manner.

As became apparent throughout the paper, the topic of effect measurement in this particular context of military operations is spread out over a range of different disciplines, e.g. behavioural economics, marketing, intelligence studies, psychology, sociology, or war studies. By bridging these disciplines, in particular research methods from behavioural economics and marketing, with the military context and placing them in a single thought framework, this paper attempts to centre the research on this topic. But as the research is scattered across different fields, it also provides opportunities for further research, if and when conducted in a multidisciplinary manner. What became clear in bridging these disciplines, is that the same translating step is needed in the implementation and further research of these measures. So, in a similar fashion, the implementation of these methods within the military context can be aided when carried out by a multi-disciplinary team, as suggested by the needed expertise amongst the feasibility criteria.

Besides the two main findings addressing the whole of the topic, regarding the under researched facets and scattered across different disciplines, this paper also brought forward three more in depth findings. The first, being that when carrying out effect measurements, one needs to start with the end, i.e. with the objective in mind. Which points to the idea that in the pre-intervention stage of setting up the behavioural influencing operation, one should already focus on the post-intervention phase, or intended end-state. Since without the identification of, on the one hand, the baseline of the current situation that you would want to impact, and on the other hand a clear grasp of what that intended impact (ideally) looks like, it is near impossible to establish the actual impact itself. Therefore, integration of the planning for effects measurement into the core of the military decision-making process is crucial to carrying out the measurement procedures.

Secondly, conditions and tools for measuring remain highly context dependent and close to nongeneralizable. Examples of this are high versus low intensity conflict, the available resources, and most of all the time factor. As such, a toolbox for effect measurement of behavioural influencing operations will not consist of a myriad of tools instead of a 'one-size fits all' instrument. This indicates the necessity of a wide research base into the possible utility of different measurement tools and the need to invest in knowledge and skill to be able to select and utilize the appropriate tooling for specific interventions.

Lastly, the foreshadowing of the potential impact of emerging technologies extends from the capabilities of behavioural influencing onto the possibilities of their measurement as well.

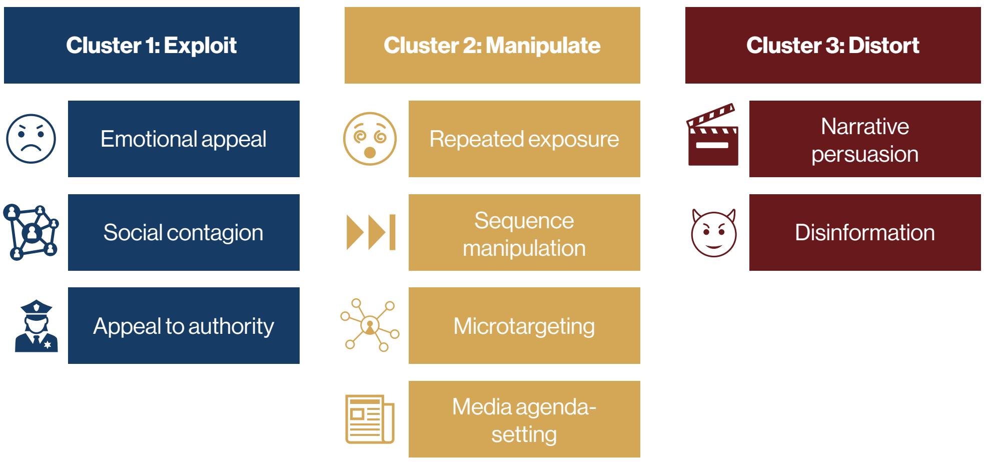
Although predictions regarding the impact of these technologies remains difficult, as they rarely develop in a linear fashion, building a solid knowledge base on the possibilities and possible utility of measuring tools is a solid requirement for the further operationalisation of BI in the future. Alas, as their potential uses, risks, constraints and with those their possibilities are yet to be determined and given the speed of developments, when it comes to the development and implementation of for example AI or LLMs, one is essentially attempting to build the plane while already flying it. Furthermore, their impact will not only stay confined to either the civilian or military context, but rather looks like it will, and is already, creating spillovers into both. That requires the close cooperation across many different fields of research to be able to stay in tune with the potential of these emerging technological innovations ever more important.

Annex

A. Information-based Behavioural Influencing Tactics

The following behavioural influencing tactics all utilise information, with the aim to change the perceptions, attitudes and eventually behaviour of the target audience. In doing so they target the inherent weaknesses of human cognition in the information environment, for military purposes. The nine described influencing tactics all find their footing in behavioural sciences to explain the underlying cognitive mechanisms. These clusters are not mutually exclusive, and the boundaries between them are not always clearly defined. They are often used in concert as they can be mutually reinforcing. As the purpose of this paper is not the implementation of these tactics, but rather the assessment of their behavioural effects, the overview of these tactics are kept rather brief.[111] There are three clusters of influencing tactics, together accounting to nine different tactics of behavioural influencing. This allocation, and consequently this section, is based upon a previous report of HCSS.[112] For the purpose of this paper, the variations in effectiveness between the different influencing tactics will not be discussed. This is, however, an area that can benefit from further research.

Figure 10. Behavioural Influencing Tactics



111 For a more detailed account of the underlying mechanisms and behavioural processes that underpin the different tactics please consult: Lotje Boswinkel et al., 'Weapons of Mass Influence: Shaping Attitudes, Perceptions and Behaviours in Today's Information Warfare' (The Hague: The Hague Centre for Strategic Studies, April 2022), https://hcss.nl/wp-content/uploads/2022/04/Weapons-of-Mass-Influence-Information-Warfare-HCSS-2022-V2.pdf.

112 For more a more in detail account of the behavioural influencing tactics, please consult: Boswinkel et al.

Cluster 1: Exploit

The first cluster encloses influencing tactics that aim to exploit the reality and how this is observed, and reacted to by the target audience. These tactics thus built on what already exists in the information environment and its three dimensions. Within this cluster there are three specific tactics that can be used. The first tactic is called emotional appeal, which targets human emotion to create certain reactions. The second tactic is social contagion, which utilises individual's tendency to copy each other's behaviour. Lastly, appeal to authority is related to the second tactic as within society there often exist certain key figures that are regarded as more influential than others. Such key figures can be compelled to display certain behaviour in order to provide an example of 'preferred' behaviour and so guide the behaviour of the target audience.

Cluster 2: Manipulate

As with Cluster 1, Cluster 2 also uses the reality that is already present. However, now this reality is not only exploited, but the tactics also manipulate information by rearranging and overwhelming the target audience with information, targeting a specific audience, or prioritising certain bits of information over others. There are four different tactics identified under this cluster. Repeated exposure refers to the persuasive effects of information that is frequently presented to the TA on a consistent basis. Secondly, and related, individuals are more susceptible to the information that is first presented to them. Thus by manipulating the sequence in which information is ranked, the TA can be influenced. Another tactic of manipulation is microtargeting, which describes the way that information is sought out and directed at specific audiences so that they are more susceptible to it. The last tactic under this cluster is media agenda setting. This practice makes use of mass media to shape public opinion in a favourable way. As mass media is still often the main medium of communication between policymaking and the public, it determines the public agenda.

Cluster 3: Distort

The remaining two tactics under Cluster 3 do not use the existing reality by exploiting or manipulating it. Rather, they create a whole new reality by distorting the existing one. This can be done by narrative persuasion, which taps into how individuals use narratives to make sense of the world and find solutions. Altering these narratives in turn guide behavioural change as they sway the audience in taking a certain course of action. Lastly, there is disinformation as a tactic. Disinformation is the act of deliberately spreading manipulated and false information with the intention of deceiving or manipulating the audience. Thus, it renders the individual unable to distinguish between what is fact and what is fiction. Disinformation is often used as an umbrella term for all previously mentioned tactics, as it in part relies on their underlying mechanisms. However, it is treated as a separate influencing tactic, as it has distinct effects on the decision-making process of the individual.

B. Extended description of measuring methods

This section of the annex provides a more in-depth explanation of the measuring and data collection methods presented earlier. However more in-depth compared to the in-text compilation previously provided, this does remain an introduction into the possibilities of measuring and does not provide an all-encompassing overview. For that we refer to the provided references and additional reading materials.

B.1 Experimental methods

Experimental methods, especially RCTs are generally understood to be the 'gold standard' for providing the strongest evidence for causality.[113] In experimental methods, the target audience is arbitrarily divided into two or more groups. With a sufficiently large sample size, and correct randomisation, this should lead to the groups being as similar as possible. This makes the control group a valid counterfactual to the group being exposed to the intervention, and thus any changes observed between the group should be due to the intervention.[114] Some of the groups are exposed to the intervention while the others are not. This allows one to observe how the TA reacts when they are exposed to the intervention as well as the counterfactual situation where they are not exposed to the intervention. Any possible external factors that might influence their behaviour should impact the groups in a similar way.[115] While experimental methods are a great way to measure the impact of influencing operations, implementing them does come with some challenges. Randomization can be difficult to achieve especially in the field, as the groups need to be similar to each other, only a part of the TA needs to be exposed to the treatment, and there shouldn't be any spillover effects.[116]

Randomised Controlled Trial (RCT)

A prime example of an experimental method is a randomised control trial (RCT). It is widely used in the medical field as it produces a reliable measurement of causality when implemented correctly. An RCT follows a similar structure as the one mentioned above, and it is a reliable method to measure if an intervention has an effect and how large that effect is.[117]

A/B testing

While RCTs are mainly focused on figuring out if an intervention has an effect, A/B testing is better used for figuring out which intervention works best.[118] Similar to RCTs, in A/B testing the TA is divided into two or more groups. The different groups are then exposed to different interventions, or different versions of the same intervention.[119] By comparing the different groups

113 Patten and Newhart, *Understanding Research Methods: An Overview of the Essentials*.

114 Esther Duflo, 'Randomized Controlled Trials, Development Economics and Policy Making in Developing Countries', 2016.

115 Hariton and Locascio, 'Randomised Controlled Trials—the Gold Standard for Effectiveness Research'.

116 A. D. Nichol et al., 'Challenging Issues in Randomised Controlled Trials', *Injury* 41 Suppl 1 (July 2010): S20-23, https://doi.org/10.1016/j.injury.2010.03.033.

117 Hariton and Locascio, 'Randomised Controlled Trials—the Gold Standard for Effectiveness Research'.

118 Roy Bendor Cohen, 'RCT vs. A/B Testing: A Quest to Answer "Why"', 27 April 2021, https://www.q-bt.co/blog-posts/rct-vs-a-b-testing-a-quest-to-answer-why.

119 Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020).

together it can be measured which intervention has had the greatest effect. A/B testing is popular especially in the online environment, as it is easy to randomly present slightly modified versions of interventions to different users.[120] It is also possible to use specialised algorithms to efficiently select the best treatment arm if many interventions are possible, such as in mega-studies of nudges, and the goal is to select the most effective one.[121]

B.2 Quasi-experimental methods

Quasi-experimental methods can be used when randomisation is not feasible. Quasi-experimental methods have the same idea as experimental methods: compare the treatment group to its counterfactual to obtain the average treatment effect. However, where quasi-experimental methods differ from experimental methods is that in the former the subjects are not assigned to treatment and control groups by random.[122] Instead, the control group is deliberately chosen so that it resembles the intervention group as closely as possible. As the groups are chosen deliberately, some assumptions are needed in order to gain a reliable measurement of causality. These assumptions are also addressed in more detail in the explanations below.

Pre-Post analysis

A pre-post analysis is one of the simplest ways to measure the impact of a BI operation on behavioural change. In a behavioural influencing context, this consists of observing the TA before and after the intervention and comparing the two to each other.[123] An example of a kinetic warfare equivalent would be observing an area before and after a missile strike to determine the damage inflicted.

Compared to experimental and other quasi-experimental methods, a pre-post analysis is easier to implement as there is no division into groups and the intervention can be implemented to the whole target audience. The simplicity of the analysis makes it a useful tool when the resources and the access to the target audience is limited, which is usually the case in a military context.

However, the pre-post analysis relies on the assumption that there are no outside factors affecting the target audience while the intervention is implemented.[124] This may hold true in the context of a missile strike where the physical environment and time period observed is very limited. However, the information space is quickly changing and volatile especially in the military context. As a result, the measurements obtained include different levels of outside noise from concurrent events and therefore the measurements usually produce only rough estimates. The trade-off between accuracy and the number of resources used should be taken into consideration when considering the pre-post analysis as a measurement method.

120 Ron Kohavi and Roger Longbotham, 'Online Controlled Experiments and A/B Testing', in *Encyclopedia of Machine Learning and Data Mining*, 2017, 922–29, https://doi.org/10.1007/978-1-4899-7687-1_891.

121 Marco Gregori, 'Adaptive Sampling and Hypotheses Testing to Identify the Best Nudge in Factorial Experiments'.

122 Rossi, Lipsey, and Freeman, *Evaluation*.

123 Samuel J. Stratton, 'Quasi-Experimental Design (Pre-Test and Post-Test Studies) in Prehospital and Disaster Research', *Prehospital and Disaster Medicine* 34, no. 6 (December 2019): 573–74, https://doi.org/10.1017/S1049023X19005053.

124 David J. Torgerson and Carole J. Torgerson, 'The Limitations of Before and After Designs', in *Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*, ed. David J. Torgerson and Carole J. Torgerson (London: Palgrave Macmillan UK, 2008), 9–16, https://doi.org/10.1057/9780230583993_2.

Difference-in-difference

Difference-in-difference is a quasi-experimental method that compares two similar groups to each other with one group being exposed to the intervention and the other one not.[125] However, unlike experimental methods, DID relaxes the requirement of randomising the target audience into groups. Instead the division into the intervention and control group is done in a non-randomised way. Even though the allocation to treatment and control groups are not random, the aim of the allocation is to make the two groups as similar as possible. There exist multiple options to make control and treatment groups as similar as possible:

- **Matching:**[126] Matching can happen either at an individual level or at an aggregate level. At the individual, two individuals need to be identified that are as similar to each other as possible and then allocate one randomly to the control group and one to the treatment group. While matching individuals it is important to keep in mind the inverse relationship between similarity and sample size. The more similar the subject pair needs to be, the harder it is to find a match to an individual subject. At the aggregate level individuals are not matched to each other, but individuals are allocated in a way to make the aggregate groups as similar as possible e.g., having a similar average age and gender division.
- **Propensity score matching:**[127] In propensity score matching, the first step is to identify all the different characteristics that affect an individual's probability of exhibiting the behaviour of interest. Second, subjects are ranked based on their propensity to exhibit the behaviour of interest based on their individual characteristics. Finally, subjects need to be matched to each other based on their propensity score and allocate them to different groups. Matched subjects can differ based on individual characteristics as long as their propensity score is similar. Alternatively to matching, individuals can also be inversely weighted based on their propensity score, thus the whole sample can be used and more weight is put on individuals who are less likely to be exposed to the intervention.[128]
- **Synthetic control methods:**[129] In using synthetic control methods, a synthetic control group is build that closely matches the treatment group. The synthetic control group is constructed by using a weighted average of similar available control units. The weights are selected in a way so that the weighted average follows closely the trendline of the treatment group before the treatment is implemented. Lastly, the assumption is made that the trendlines would continue as similar in the absence of the treatment.

For a control group to be a valid counterfactual to the treatment group some assumptions are needed, the main ones being the parallel trends assumption and the Stable Unit Treatment Value Assumption (SUTVA).[130] The parallel trends assumption states that in the absence of treatment the difference between the intervention group and the control group is constant.

125 Steve Bradley and Colin Green, *The Economics of Education A Comprehensive Overview*, 2nd ed. (Academic Press, 2020).

126 The World Bank, 'Dimewiki Matching', accessed 12 September 2023, https://dimewiki.worldbank.org/Matching.

127 The World Bank, 'Dimewiki Propensity Score Matching', accessed 12 September 2023, https://dimewiki.worldbank.org/Propensity_Score_Matching.

128 Nicholas C. Chesnaye et al., 'An Introduction to Inverse Probability of Treatment Weighting in Observational Research', *Clinical Kidney Journal* 15, no. 1 (January 2022): 14–20, https://doi.org/10.1093/ckj/sfab158.

129 Alberto Abadie, Alexis Diamond, and Jens Hainmueller, 'Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program', *Journal of the American Statistical Association* 105, no. 490 (1 June 2010): 493–505, https://doi.org/10.1198/jasa.2009.ap08746.

130 "Columbia University Mailman School of Public Health, 'Difference-in-Difference Estimation', Columbia University Mailman School of Public Health, 3 August 2016, https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation.

The SUTVA assumption requires that the composition of the two groups stays stable throughout the implementation and that there are no spill overs across groups.

To measure the effect of the intervention, two types of differences are important, the difference in the intervention group before and after the intervention, as well as the difference between the intervention and the control group. A DID combines these two differences and measures how the intervention group changes before and after the intervention and compares it to how the control group changes before and after the intervention.[131]

Regression discontinuity

Under the right circumstances regression discontinuity can be considered a good way to measure the impact of a treatment. It can be used when dealing with a variable that is continuous but has a distinct and impactful cutoff point where a treatment is administered.[132] This cutoff point can divide target audiences into two groups, those who are over the cut-off point and receive the treatment, and those who are under and do not. Generally, these two groups are fundamentally different from each other and not comparable. However, target audiences close to either side of the cut-off point are usually quite similar and therefore can be compared to each other. By analysing these border cases, you are able to identify the effect of the treatment.

In order for a setup to be favourable for a regression discontinuity, the treatment should be assigned based on some observable variable or index.[133] There also should be a distinct cut off point for the treatment where the discontinuity is expected to be found. This cut off point should be arbitrary, to make sure the target audiences on both sides of the point are similar to each other. The same cut off point cannot be used to determine the efficiency of multiple treatments. For example, in the 1950s the German army introduced compulsory military service for all males born after July 1st 1937. By comparing individuals born just before and just after this cut-off point it can be possible to measure the impact of military service on labour production.[134]

Regression discontinuity is a great analytical method to use when the circumstances allow for it. It can be utilised to obtain unbiased effects of the treatment near the cut-off point. It utilises an allocation system that is popular in the design of social policies which increases the number of natural experiments available. Although regression discontinuity can be a useful method it also faces some limitations.[135] Only using the observations at the cut-off point limits the number of usable observations compared to the entire sample available. It also limits external validity as the differences are tied to a specific cut off point. Results might differ when choosing a different cut off point. In addition a usually shaped distribution curve can lead to inaccurate results.

131 Paul J. Gertler, Sebastian Martinez, and Patrick Premand, *Impact Evaluation in Practice, Second Edition* (Washington, DC: World Bank, 2016).

132 Gertler, Martinez, and Premand.

133 The World Bank, 'Dimewiki Regression Discontinuity', n.d., https://dimewiki.worldbank.org/Regression_Discontinuity#:~:text=Regression%20Discontinuity%20Design%20(RDD)%20is,who%20is%20eligible%20to%20participate.

134 Thomas K. Bauer et al., 'Evaluating the Labor-Market Effects of Compulsory Military Service', *European Economic Review* 56, no. 4 (1 May 2012): 814–29, https://doi.org/10.1016/j.euroecorev.2012.02.002.

135 The World Bank, 'Dimewiki Regression Discontinuity'.

Instrumental variables (IV)

IV is a quasi-experimental method that utilises an outside source of variation as a proxy for being exposed to the intervention.[136] It is used when the explanatory variable used is dependent on another variable affecting the model, creating a bias. The bias can be caused by an omitted variable, an error in the measurement of the explanatory variable or a simultaneous relationship between the explanatory and the dependent variable.[137] By using an instrumental variable as a proxy for the independent variable, it is possible to get rid of the bias and capture only the effect of the independent variable. However, it should be noted this does not give us an average treatment effect but a local average treatment effect, meaning the average treatment effect for only those individuals who are also subject to the instrument.

The most important thing when using an instrumental variable's estimation is to pick an instrument that is valid.[138] For an instrument to be a good proxy for an independent variable, it should be highly correlated with the independent variable in question. This is to ensure that the instrument actually has a strong effect on the independent variable and that this effect is measurable. The instrumental variable should also not be correlated with the independent variable itself to ensure it only captures the effect of that variable. The lack of a link between the instrumental variable and variable of interest is not always clear and there should be clear arguments why the two are not connected. An example of a good instrumental variable would be an individual's draft number as a proxy for going to war. A lot of different factors affect whether an individual goes to war or not, and going to war will have an impact on a number of aspects of an individual's life. However, a draft number is usually picked randomly and therefore is not correlated with any other characteristics of an individual but is highly correlated with the probability if an individual goes to war or not.[139]

B.3 Non-experimental methods

Non-experimental methods differ from experimental and quasi-experimental methods in the sense that there is no interest in manipulating an isolated variable, but instead observe and measure how different variables operate in their natural habitat. This approach can be useful when the operation focuses on a single variable instead of a relationship between two variables, on a non-causal relationship between two variables, or on a causal relationship but the independent variable cannot be manipulated. It is also useful when the topic of interest is broad and exploratory.

Causal Loop Diagrams (CLD) & computational System Dynamics Model (SDM)

Causal loop diagrams are illustrations that use boxes, connections and feedback loops to visually represent variables and their interactions within a system.[140] Figure 11 is a simplified example of a CLD illustrating how BI operations targeting hostile troops can affect overall military success. CLDs can be a useful tool in decision making to gain further insight on how different systems work and how manipulating a particular variable will affect the system overall.

136 Gertler, Martinez, and Premand, *Impact Evaluation in Practice, Second Edition*.

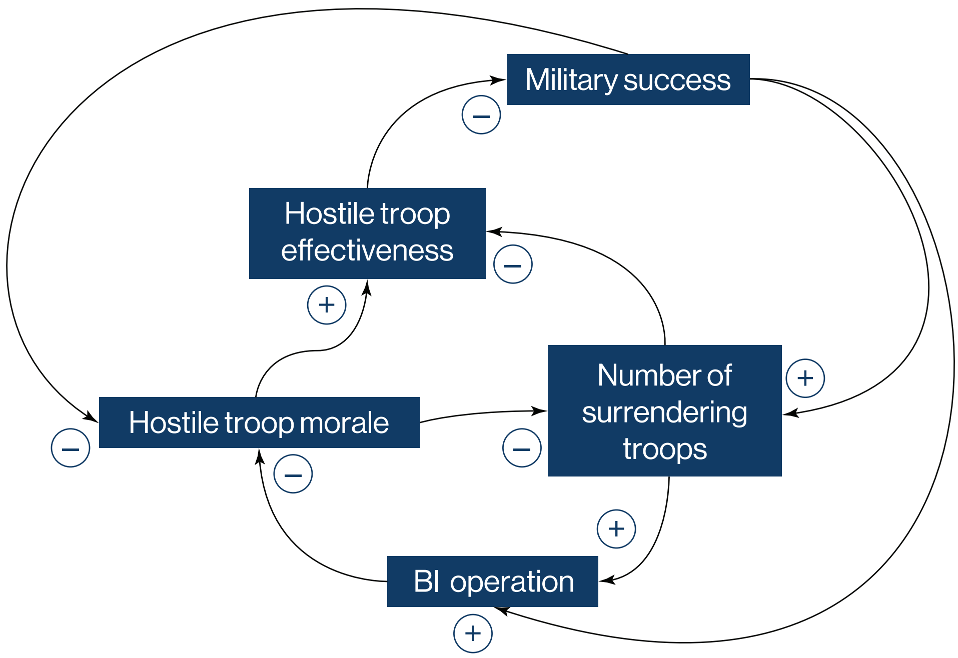
137 Columbia University Mailman School of Public Health, 'Instrumental Variables', Columbia University Mailman School of Public Health, 4 August 2016, https://www.publichealth.columbia.edu/research/population-health-methods/instrumental-variables.

138 Columbia University Mailman School of Public Health.

139 Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin, 'Identification of Causal Effects Using Instrumental Variables', *Journal of the American Statistical Association* 91, no. 434 (1996): 444–55, https://doi.org/10.2307/2291629.

140 Pete Barbrook-Johnson and Alexandra S. Penn, *Systems Mapping: How to Build and Use Causal Models of Systems*, 1st ed. (Palgrave Macmillan Cham, 2022).

Figure 11. Example of CLD in the military context



By combining CLDs with sufficient data and relevant formulas they can be turned into computational System Dynamics Models (SDM).[141] While a CLD only visually represents the causal connections between different variables, a SDM turns these visual illustrations into measurable effects. Indeed, while a CLD explain that BI operations can affect overall military success, a SDM can calculate exactly how much. A SDM can simulate how manipulating a particular variable will affect the overall system. These simulations can be used as support when making decisions; whether that is planning a BI operation, choosing the optimal aggression level to minimise the number of insurgents,[142] or the most cost-effective cyber security strategy.[143]

Agent Based Models (ABM)

ABMs are computerised models used to map the relationships and interactions between individuals, things, environments and places. At the simplest, it consists of multiple individual agents and the interactions between them.[144] ABMs are used to model social networks and behaviour patterns that emerge with inter-agent interactions.[145] They can be divided into classical models and AI based models. In classical models, the choice options of individual agents are based on clear and simple rules. For example, an agent will change their opinion if they are exposed to an X amount of persuasive literature on the topic. However, in AI based models the choices of agents are not restricted by predetermined rules, but instead the agents are constantly learning by interacting with other agents in order to reach their goals more efficiently.

141 Loes Crielaard et al., 'Refining the Causal Loop Diagram: A Tutorial for Maximizing the Contribution of Domain Expertise in Computational System Dynamics Modeling', *Psychological Methods*, 12 May 2022, https://doi.org/10.1037/met0000484.

142 J. Morrison, Daniel Goldsmith, and Michael Siegel, 'Dynamic Complexity in Military Planning: A Role for System Dynamics', 1 January 2008.

143 Derek L. Nazareth and Jae Choi, 'A System Dynamics Model for Information Security Management', *Information & Management* 52, no. 1 (1 January 2015): 123–34, https://doi.org/10.1016/j.im.2014.10.009.

144 Eric Bonabeau, 'Agent-Based Modeling: Methods and Techniques for Simulating Human Systems', *Proceedings of the National Academy of Sciences of the United States of America* 99, no. 10 (2002): 7280–87.

145 C M Macal and M J North, 'Tutorial on Agent-Based Modelling and Simulation', *Journal of Simulation* 4, no. 3 (1 September 2010): 151–62, https://doi.org/10.1057/jos.2010.3.

Agent based models can be used as effective tools in decision making. For example in the informational environment, ABMs have been used to identify which individuals to target in order to spread information most efficiently,[146] while in a military context, ABMs have been used to simulate troop movement.[147] As with most models, there is always a trade-off between how detailed the model needs to be and how easy it is to build and implement.

Structural econometric models

Structural econometric models utilise economic theory to try and develop mathematical statements about the way observable "explanatory" variables affect observable "exogenous" variables.[148] As theory alone is rarely enough to accomplish this, structural models include statistical assumptions to compensate the un-observables in the models. Together, the theory and assumptions create a model that is able to rationalise and estimate all the possible observable outcomes.[149] They can be used in decision making by modelling previous behaviour or by simulating all the different outcomes of a decision. One area of structural models is dynamic games, where individual actors make decisions to achieve competing goals. This has been previously utilised for example in marketing, by finding the optimal marketing frequency in a competitive market.[150] Similar methods could also be utilised in a military context to simulate the effect of different BI strategies. While using structural models can be a useful tool for decision makers, it is important to understand the context in which it can and should be used. As structural models rely heavily on theory and assumptions, it is important that these are based on sound, reliable and valid empirical evidence.

Time series models

The change in beliefs, knowledge and behaviour is rarely instantaneous but a gradual process. To model and predict this gradual process, one can use time series models. Time series models are statistical models that utilise chronological data collected over time to make future predictions, analyse trends or identify anomalies.[151] There are a number of different time series available that can be used based on the specific context. Time series models can be used to perform a trend analysis to identify whether an intervention has an impact on the overall trend of the behaviour of interest, for example the number of enlisting soldiers.[152] Other time series models, like persistence modelling, can be used to disentangle the short- and long-term effects of interventions[153] or spillover effects across channels.[154] As the data is ordered chronologically, it is likely that earlier observations have predictive power over future observations, which can make predictive models useful.[155]

146 Sinan Aral and Paramveer S. Dhillon, 'Social Influence Maximization under Empirical Influence Models', *Nature Human Behaviour* 2, no. 6 (June 2018): 375–82, https://doi.org/10.1038/s41562-018-0346-z.

147 James Moffat, Smith Josephine, and Witty Susan, 'Emergent Behaviour: Theory and Experimentation Using the MANA Model', *Journal of Applied Mathematics and Decision Sciences* 2006 (2 October 2006), https://doi.org/10.1155/JAMDS/2006/54846.

148 Peter C. Reiss and Frank A. Wolak, 'Chapter 64 Structural Econometric Modeling: Rationales and Examples from Industrial Organization', in *Handbook of Econometrics*, ed. James J. Heckman and Edward E. Leamer, vol. 6 (Elsevier, 2007), 4277–4415, https://doi.org/10.1016/S1573-4412(07)06064-3.

149 Reiss and Wolak.

150 Jean-Pierre Dubé, Günter Hitsch, and Puneet Manchanda, 'An Empirical Model of Advertising Dynamics', *Quantitative Marketing and Economics (QME)* 3, no. 2 (2005): 107–44.

151 Berend Wierenga and Ralf van der Lans, *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science (Springer Cham, 2017).

152 US Army, 'FM 3-05.301 Psychological Operations Tactics, Techniques, and Procedures', 2003.

153 Wierenga and van der Lans, *Handbook of Marketing Decision Models*.

154 Isaac M. Dinner, Harald J. Heerde Van, and Scott A. Neslin, 'Driving Online and Offline Sales: The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising', *Journal of Marketing Research* 51, no. 5 (1 October 2014): 527–45, https://doi.org/10.1509/jmr.11.0466.

155 Wierenga and van der Lans, *Handbook of Marketing Decision Models*.

C. Extended description of data collection methods

C.1 Self-report measures

Implicit methods

Implicit Association Test (IAT)

IAT is a test where the participant is given two concepts (e.g. countries) and two attributes (e.g. friendly & unfriendly) that are attached to two response keys (e.g. left and right arrow on a keyboard). One concept and one attribute are assigned to one response key and the combination of concepts, attributes and response keys are changed. If a participant associates a concept and attribute sharing a response key to each other, the response time should be faster.[156] Thus one can measure the response times between different concepts and attribute combos to elicit implicit beliefs a target audience holds.

While an IAT can be easily carried out with a mass audience through the internet and used to measure implicit constructs that can be hard to measure with explicit methods, for example sensitive topics such as racial stereotypes,[157] there is some discussion questioning the validity of the measure.[158] An individual's results may change based on context and over time. Therefore, it is useful to use other measurement methods together with the IAT. It is also noteworthy from a behaviour influencing point of view to acknowledge that implicit beliefs do not automatically translate into behaviour. Even if an individual prefers one group over another does not mean they will present discriminatory behaviour in real life.

Affect Misattribution Procedure (AMP)

AMP is a priming method where an individual is first given a stimulus related to the topic of interest, and then the individual is asked to express their opinion of an ambiguous figure. As individuals tend to misattribute their feelings toward the stimulus to the ambiguous figure, it can be used to measure the implicit emotions and beliefs of individuals.[159] The literature of AMP has shown that the method produces valid and reliable results about individuals' implicit beliefs.[160]

156 A. G. Greenwald, D. E. McGhee, and J. L. Schwartz, 'Measuring Individual Differences in Implicit Cognition: The Implicit Association Test', *Journal of Personality and Social Psychology* 74, no. 6 (June 1998): 1464–80, https://doi.org/10.1037//0022-3514.74.6.1464.

157 Anthony G. Greenwald, Mahzarin R. Banaji, and Brian A. Nosek, 'Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects', *Journal of Personality and Social Psychology* 108, no. 4 (April 2015): 553–61, https://doi.org/10.1037/pspa0000016.

158 Ulrich Schimmack, 'Invalid Claims About the Validity of Implicit Association Tests by Prisoners of the Implicit Social-Cognition Paradigm', *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 16, no. 2 (March 2021): 435–42, https://doi.org/10.1177/1745691621991860.

159 B. Keith Payne et al., 'An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement', *Journal of Personality and Social Psychology* 89, no. 3 (September 2005): 277–93, https://doi.org/10.1037/0022-3514.89.3.277.

160 Keith Payne and Kristjen Lundberg, 'The Affect Misattribution Procedure: Ten Years of Evidence on Reliability, Validity, and Mechanisms', *Social and Personality Psychology Compass* 8, no. 12 (2014): 672–86, https://doi.org/10.1111/spc3.12148.."

Explicit methods

Issues to consider with self-reported data

Privacy concerns

Some individuals might be reluctant to report their actual knowledge, beliefs, emotions, or behaviour about sensitive topics due to privacy concerns or social image concerns. Different questioning techniques have been developed to mitigate this issue.[161] The privacy concerns are not just a limiting factor when it comes to individual respondents. There can be legal regulations, like the GDPR implemented by the EU, that limit the collection and storage of personal data.

Open-ended vs. closed-ended questions

It should be considered when to ask open-ended or closed-ended questions. Closed-ended questions have a distinct set of predefined responses which makes them easy to administer, analyse and compare. However, they offer limited information compared to open-ended questions. Open- ended questions may offer richer insights. While these answers can be harder to analyse and compare, tools like Natural Language Processing (NLP) models can be utilised to extract underlying topics, emotions, sentiment and, possible evolution over time.[162]

Outcome prediction

In some cases the outcome cannot be measured but it might be possible to still predict it using specific models.[163] For instance, if the link between personality and social media likes is known based on a specific sample of respondents that provided both personality ratings and gave access to their social media data, it is possible to accurately predict the personality of other people for whom we only have social media data.[164] This information can then be used to develop ads that are tailored to an individual's inferred personality. Similarly, it is possible try to use a sample of diaspora respondents to predict opinions or behaviour in countries where certain topics cannot be discussed or where behaviour is illegal, and respondents are unwilling to disclose such information. Naturally, diaspora respondents may be quite different from respondents who live inside such countries, so this type of sample selection has to be addressed in a statistical model that is used for inference.

Poll

A poll is a simple question about a topic of interest used to measure the emotions, beliefs and knowledge of a target audience. The question in a poll can be either close or open ended.

161 These include randomised response, see Warner, 'Randomized Response'. , list experiments, see Miller, 'A New Survey Technique for Studying Deviant Behavior'., Crosswise models, see Yu, Tian, and Tang, 'Two New Models for Survey Sampling with Sensitive Characteristic'., and Bayesian truth serum, see Prelec, 'A Bayesian Truth Serum for Subjective Data'.

162 These models include for example simple dictionary methods like LIWC and unsupervised learning models like Latent Dirichlet Allocation, see Blei, Ng, and Jordan, 'Latent Dirichlet Allocation'.

163 Leo Breiman, 'Random Forests', *Machine Learning* 45, no. 1 (1 October 2001): 5–32, https://doi.org/10.1023/A:1010933404324; Christopher Winship and Robert D. Mare, 'Models for Sample Selection Bias', *Annual Review of Sociology* 18 (1992): 327–50.

164 Michal Kosinski, David Stillwell, and Thore Graepel, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior', *Proceedings of the National Academy of Sciences* 110, no. 15 (9 April 2013): 5802–5, https://doi.org/10.1073/pnas.1218772110.

While a poll might be inferior to a multi question survey when it comes to the amount of information obtained, it is easier to administer to a large target audience to gain general insight about the topic of interest.

Survey

A survey is a series of close or open-ended questions used to measure the emotions, beliefs and knowledge of a target audience as well as past, future and hypothetical behaviour (through vignettes). Surveys are a straightforward way to gain insight about a target audience. If possible, it is best to use surveys together with other data collection methods to gain as much insight as possible and account for possible biases. A good example of this can be special "representative" survey panels where participants fill out surveys regularly, while e.g. their online behaviour, and background characteristics are also observed.

Interview

An interview is a constructed conversation where an individual is asked a series of questions. Interviews are usually conducted verbally either face-to-face or over the phone which introduces a social aspect into the data collection and allows for richer data with the inclusion of audio and vision. As the interview is an interactive process it allows the interviewer to probe further into interesting topics discovered during the interview and specify any ambiguities. However, increased contact with the interviewer can lead to an increased risk of socially desirable responding. Interviews are commonly used if the sample size available is small or as an exploratory tool when the potential answers to the questions are still unclear.

Focus group

A focus group is a group interview with multiple interviewees focusing on a specific topic of interest. As a focus group is conducted with multiple people simultaneously, it allows for more efficient data collection compared to individual interviews. It also allows one to gain further insight on how social interaction affects an individual's answers and increased social interaction can lead to more creative and richer answers. However, the group set up also comes with its difficulties, as social interaction within the group can lead to socially desirable responding, groupthink and confirmation bias.

C.2 **Online data**

Online search data & web browsing behaviour

This type of data refers to the tracking of general web-based metrics like the frequency of specific search terms or the number of visits on a specific website. These can be used to measure the online behaviour of target audiences. These metrics can be very useful as they represent actual behaviour and data collection can be computerised.

Social media data

This refers to the analysis of a target audience's social media activity, networks and metadata. Social network analysis is useful because people tend to connect with others who share similar preferences and interests. By examining the characteristics and behaviours of an individual's social network connections, we can make assumptions about their preferences.

Individuals often join social network groups or communities based on shared interests. These tools can be used together or separately to uncover patterns in beliefs and behaviours across different networks over time. Target audiences can be analysed as a group or the analysis can focus on individuals separately.

C.3 Multimedia Data

Audio and text analysis

From audio and text data it is possible to analyse both content and linguistics. From audio data it is additionally possible to analyse vocal cues. Through NLP, linguistics and vocal cues can be used to identify individuals, sentiment, deception and characteristics, while content itself can be analysed to recognise patterns and intent. [165]

Image and video analysis

Image and video refer to both material found online as well as imagery obtained straight from sensors like surveillance cameras. Facial recognition programs can be used to identify individuals and their emotional state.[166] Object recognition can be used to recognise the context of the image and categorisation.[167] It is also possible to identify locations from images,[168] and the usage of open source intelligence and geolocated images have been used in the war in Ukraine to get a glimpse through the fog of war.[169] Video and images can even be used to identify behaviour.[170] As images are heavy with data, AI and machine learning models can be utilised to make the analysis process more efficient.[171]

C.4 Sensor data

Physiological measures

Physiological measures refer to the measurement of voluntary or more commonly involuntary bodily functions. These measures can be obtained by sensors and used to gain insight into the mind of an individual. Involuntary bodily functions like blood pressure, pupil dilation and galvanic skin response can be used to measure the emotions of an individual or to detect deception. While it may not be possible to identify specific emotions through physiological

165 IBM, 'What Is Natural Language Processing? | IBM', accessed 12 September 2023, https://www.ibm.com/topics/natural-language-processing.

166 European Data Protection Supervisor, 'TechDispatch #1/2021 - Facial Emotion Recognition | European Data Protection Supervisor', 23 August 2023, https://edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-12021-facial-emotion-recognition.

167 Mitts, Phillips, and Walter, 'Studying the Impact of ISIS Propaganda Campaigns'.

168 Jianxiong Xiao et al., 'Basic Level Scene Understanding: From Labels to Structure and Beyond', *MIT Web Domain*, November 2012, https://dspace.mit.edu/handle/1721.1/90941.

169 The Economist, 'Open-Source Intelligence Is Piercing the Fog of War in Ukraine', *The Economist*, 13 January 2023, https://www.economist.com/interactive/international/2023/01/13/open-source-intelligence-is-piercing-the-fog-of-war-in-ukraine.

170 A. B. Klimenko and I. S. Korovin, 'A Technique of Deviant Behaviour Detection in Video Surveillance Systems Based on Complex Behaviour Analysis', *Journal of Physics: Conference Series* 1661, no. 1 (November 2020): 012050, https://doi.org/10.1088/1742-6596/1661/1/012050.

171 Kate Macri, 'Army Is Modernizing Sensors for Data-Driven Decision-Making', 4 March 2022, https://governmentciomedia.com/army-modernizing-sensors-data-driven-decision-making.

measures, they can be used to identify emotional valence and the level of arousal.[172] In addition to emotions, attention and level of engagement can be measured through eye movement and brainwave activity. By tracking eye movements, it is possible to see what an individual pays attention to.[173] By analysing the brain activity of an individual through fMRI or EEG, it is possible to identify emotional valence, arousal and engagement.[174] As the measurement of physiological functions requires physical sensors being placed on or near the individual being measured, it limits their usability in the field.

C.5 Other data

Movement data

Movement data refers to information about the location or movement of individuals. This information can be obtained through open-source intelligence, satellite imagery, GPS-tracking or other ways. In some cases, the effect of behavioural influencing operation can be observed as changes in the movement of target audiences. Ukraine launched a wide information campaign in the summer of 2022 about a counteroffensive in Kherson. This led to the Russians to move troops from the Kharkiv region to Kherson and Ukraine successfully attacking the weakened defences of Kharkiv.[175]

Financial data

Financial data refers to all money flowing in, out, and within an organisation. Financial transactions are a good way to understand the relationships and connections of an organisation. It can provide information about the organisation structure, what the organisation does and who they are supported by. One objective of a BI operation could be to influence the relationships within a TA or between the TA and some external entity. Changes in financial structures and flows can be used to measure the impact of these operation.

172 Iris B. Mauss and Michael D. Robinson, 'Measures of Emotion: A Review', *Cognition & Emotion* 23, no. 2 (1 February 2009): 209–37, https://doi.org/10.1080/02699930802204677.

173 Mauss and Robinson.

174 Giovanni Vecchiato et al., 'Changes in Brain Activity during the Observation of TV Commercials by Using EEG, GSR and HR Measurements', *Brain Topography* 23, no. 2 (June 2010): 165–79, https://doi.org/10.1007/s10548-009-0127-0.

175 Santelises, 'The Ukrainian Kharkiv Counter-Offensive and Information Operations'.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 'Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program'. *Journal of the American Statistical Association* 105, no. 490 (1 June 2010): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

Anand, Piyush, and Clarence Lee. 'Using Deep Learning to Overcome Privacy and Scalability Issues in Customer Data Transfer'. *Marketing Science* 42, no. 1 (January 2023): 189–207. https://doi.org/10.1287/mksc.2022.1365.

Anderson, Edward G. 'A Proof-of-Concept Model for Evaluating Insurgency Management Policies Using the System Dynamics Methodology'. *Strategic Insights* 6, no. 5 (2007).

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 'Identification of Causal Effects Using Instrumental Variables'. *Journal of the American Statistical Association* 91, no. 434 (1996): 444–55. https://doi.org/10.2307/2291629.

Aral, Sinan, and Paramveer S. Dhillon. 'Social Influence Maximization under Empirical Influence Models'. *Nature Human Behaviour* 2, no. 6 (June 2018): 375–82. https://doi.org/10.1038/s41562-018-0346-z.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 'Exposure to Opposing Views on Social Media Can Increase Political Polarization'. *Proceedings of the National Academy of Sciences* 115, no. 37 (11 September 2018): 9216–21. https://doi.org/10.1073/pnas.1804840115.

Bail, Christopher A., Brian Guay, Emily Maloney, Aidan Combs, D. Sunshine Hillygus, Friedolin Merhout, Deen Freelon, and Alexander Volfovsky. 'Assessing the Russian Internet Research Agency's Impact on the Political Attitudes and Behaviors of American Twitter Users in Late 2017'. *Proceedings of the National Academy of Sciences* 117, no. 1 (7 January 2020): 243–50. https://doi.org/10.1073/pnas.1906420116.

Barbrook-Johnson, Pete, and Alexandra S. Penn. *Systems Mapping: How to Build and Use Causal Models of Systems*. 1st ed. Palgrave Macmillan Cham, 2022.

Bateman, Jon, Elonnia Hickok, Laura Courchesne, Isra Thange, and Jacob N. Shapiro. 'Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research'. Carnegie Endowment for International Peace, 28 June 2021. https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-re-search-pub-84824.

Bauer, Thomas K., Stefan Bender, Alfredo R. Paloyo, and Christoph M. Schmidt. 'Evaluating the Labor-Market Effects of Compulsory Military Service'. *European Economic Review* 56, no. 4 (1 May 2012): 814–29. https://doi.org/10.1016/j.euroecorev.2012.02.002.

Blei, David M, Andrew Ng, and Michael Jordan. 'Latent Dirichlet Allocation'. *Journal of Machine Learning Research* 3 (2003): 993–1022.

Bonabeau, Eric. 'Agent-Based Modeling: Methods and Techniques for Simulating Human Systems'. *Proceedings of the National Academy of Sciences of the United States of America* 99, no. 10 (2002): 7280–87.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 'A 61-Million-Person Experiment in Social Influence and Political Mobilization'. *Nature* 489, no. 7415 (September 2012): 295–98. https://doi.org/10.1038/nature11421.

Boswinkel, Lotje, Neill Bo Finlayson, John Michaelis, and Michel Rademaker. 'Weapons of Mass Influence: Shaping Attitudes, Perceptions and Behaviours in Today's Information Warfare'. The Hague: The Hague Centre for Strategic Studies, April 2022. https://hcss.nl/wp-content/uploads/2022/04/Weapons-of-Mass-Influence-Information-Warfare-HCSS-2022-V2.pdf.

Bradley, Steve, and Colin Green. *The Economics of Education A Comprehensive Overview*. 2nd ed. Academic Press, 2020.

Branscomb, Lewis M., Ash Carter, John A. Alic, and Gerald Epstein. *Beyond Spinoff: Military and Commercial Technologies in a Changing World*. Harvard Business Publishing, 1922.

Breiman, Leo. 'Random Forests'. *Machine Learning* 45, no. 1 (1 October 2001): 5–32. https://doi.org/10.1023/A:1010933404324.

Carlini, Nicholas, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, et al. 'Extracting Training Data from Large Language Models'. *30th USENIX Security Symposium*, 2021, 2633–50.

Castells, Manuel. *The Rise of the Network Society*. Manuel Castells, 2010.

Chesnaye, Nicholas C., Vianda S. Stel, Giovanni Tripepi, Friedo W. Dekker, Edouard L. Fu, Carmine Zoccali, and Kitty J. Jager. 'An Introduction to Inverse Probability of Treatment Weighting in Observational Research'. *Clinical Kidney Journal* 15, no. 1 (January 2022): 14–20. https://doi.org/10.1093/ckj/sfab158.

Clausewitz, Carl von. *On War*. Princeton University Press, 1989.

Cohen, Roy Bendor. 'RCT vs. A/B Testing: A Quest to Answer "Why"', 27 April 2021. https://www.q-bt.co/blog-posts/rct-vs-a-b-testing-a-quest-to-answer-why.

Columbia University Mailman School of Public Health. 'Difference-in-Difference Estimation'. Columbia University Mailman School of Public Health, 3 August 2016. https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation.

———. 'Instrumental Variables'. Columbia University Mailman School of Public Health, 4 August 2016. https://www.publichealth.columbia.edu/research/population-health-methods/instrumental-variables.

Cook, Thomas D., and Donald T. Campbell. 'The Causal Assumptions of Quasi-Experimental Practice: The Origins of Quasi-Experimental Practice'. *Synthese* 68, no. 1 (1986): 141–80.

Crielaard, Loes, Jeroen F. Uleman, Bas D. L. Châtel, Sacha Epskamp, Peter M. A. Sloot, and Rick Quax. 'Refining the Causal Loop Diagram: A Tutorial for Maximizing the Contribution of Domain Expertise in Computational System Dynamics Modeling'. *Psychological Methods*, 12 May 2022. https://doi.org/10.1037/met0000484.

Cuttler, Carrie. 'Overview of Non-Experimental Research'. In *Research Methods in Psychology*. KPU, 2019. https://kpu.pressbooks.pub/psychmethods4e/chapter/overview-of-non-experimental-research/.

Diehl, James G, and Charles E Sloan. 'Battle Damage Assessment: The Ground Truth'. *Joint Force Quarterly*, no. 37 (2005): 59–64.

Dinner, Isaac M., Harald J. Heerde Van, and Scott A. Neslin. 'Driving Online and Offline Sales: The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising'. *Journal of Marketing Research* 51, no. 5 (1 October 2014): 527–45. https://doi.org/10.1509/jmr.11.0466.

Dubé, Jean-Pierre, Günter Hitsch, and Puneet Manchanda. 'An Empirical Model of Advertising Dynamics'. *Quantitative Marketing and Economics (QME)* 3, no. 2 (2005): 107–44.

Ducheine, P.A.L., Jelle van Haaster, and Richard van Harskamp. 'Manoeuvring and Generating Effects in the Information Environment'. Amsterdam Center for International Law, 2017.

Duflo, Esther. 'Randomized Controlled Trials, Development Economics and Policy Making in Developing Countries', 2016.

Dunlap, Charles J. 'The Military-Industrial Complex'. *Daedalus* 140, no. 3 (2011): 135–47.

Dylan, Huw, David V. Gioe, and Joe Littell. 'The Kherson Ruse: Ukraine and the Art of Military Deception'. Modern War Institute, 10 December 2022. https://mwi.westpoint.edu/the-kherson-ruse-ukraine-and-the-art-of-military-deception/.

Enomoto, Carl E., and Kiana Douglas. 'Do Internet Searches for Islamist Propaganda Precede or Follow Islamist Terrorist Attacks?' *Economics and Sociology* 12, no. 1 (2018): 233–47. https://doi.org/10.14254/2071- 789X.2019/12-1/13.

European Data Protection Supervisor. 'TechDispatch #1/2021 - Facial Emotion Recognition | European Data Protection Supervisor', 23 August 2023. https://edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-12021-facial-emotion-recognition.

Fishstein, Paul, and Andrew Wilder. 'Winning Hearts and Minds? Examining the Relationship between Aid and Security in Afghanistan'. Feinstein International Center, 2012. https://fic.tufts.edu/assets/WinningHearts-Final.pdf.

François, Camille. 'Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses'. Graphika and Berkman Klein Center for Internet & Society, 2019.

Gertler, Paul J., Sebastian Martinez, and Patrick Premand. *Impact Evaluation in Practice, Second Edition*. Washington, DC: World Bank, 2016.

Gilbert, Emily. 'Military Geoeconomics: Money, Finance and War'. In *A Research Agenda for Military Geographies*, 100–114. Edward Elgar Publishing, 2019. https://www.elgaronline.com/display/edcoll/9781786438867/9781786438867.00014.xml.

Goldstein, Josh A., Girish Sastry, Micah Musser, Renee DiResta, Gentzel Matthew, and Katerina Sedova. 'Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations', 2023.

Greenwald, A. G., D. E. McGhee, and J. L. Schwartz. 'Measuring Individual Differences in Implicit Cognition: The Implicit Association Test'. *Journal of Personality and Social Psychology* 74, no. 6 (June 1998): 1464–80. https://doi.org/10.1037//0022-3514.74.6.1464.

Greenwald, Anthony G., Mahzarin R. Banaji, and Brian A. Nosek. 'Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects'. *Journal of Personality and Social Psychology* 108, no. 4 (April 2015): 553–61. https://doi.org/10.1037/pspa0000016.

Gregori, Marco. 'Adaptive Sampling and Hypotheses Testing to Identify the Best Nudge in Factorial Experiments'. Seminar, 29 May 2022.

Grylls, George. 'Ukraine Is Outflanking Russia with Ammunition from Big Tech', 12 September 2023, sec. news. https://www.thetimes.co.uk/article/ukraine-is-outflanking-russia-with-ammunition-from-big-tech-lxp6sv3qz.

Hariton, Eduardo, and Joseph J. Locascio. 'Randomised Controlled Trials—the Gold Standard for Effectiveness Research'. *BJOG : An International Journal of Obstetrics and Gynaecology* 125, no. 13 (December 2018): 1716. https://doi.org/10.1111/1471-0528.15199.

Hawkins, Richard. 'Marianna Mazzucato The Entrepreneurial State: Debunking Public vs Private Sector Myths'. *Science and Public Policy* 42, no. 1 (1 February 2015): 143–45. https://doi.org/10.1093/scipol/scu071.

Heidarysafa, Mojtaba, Sara Dalpe, Shelby Kiefner, Erfan Pakdamanian, Inki Kim, Donald E. Brown, and Janet I. Warren. 'Exploring the Experiential Impact of Online Propaganda Using Eye-Gaze and Pupil Dilation : A Comparison across Three Ideological Groups'. U.S. Department of Justice Office of Justice Programs, 2019.

Hesketh, Roger. *Fortitude: The D-Day Deception Campaign*. First Edition. Woodstock, N.Y: Abrams Press, 2000.

Hirschberg, Julia, and Christopher D. Manning. 'Advances in Natural Language Processing'. *Science* 349, no. 6245 (17 July 2015): 261–66. https://doi.org/10.1126/science.aaa8685.

Homer, and trans. Robert Fagles. *The Iliad*. New York: Penguin Classics, 1998.

Hung, Man, Evelyn Lauren, Eric S. Hon, Wendy C. Birmingham, Julie Xu, Sharon Su, Shirley D. Hon, Jungweon Park, Peter Dang, and Martin S. Lipsky. 'Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence'. *Journal of Medical Internet Research* 22, no. 8 (18 August 2020): e22590. https://doi.org/10.2196/22590.

IBM. 'What Is Natural Language Processing? | IBM'. Accessed 12 September 2023. https://www.ibm.com/topics/natural-language-processing.

Iven, Markus, Laura Jasper, and Michel Rademaker. 'Cognitive Effects in Combined Arms: A Case Study of the Division 2025'. The Hague Centre for Strategic Studies, 10 February 2023. https://hcss.nl/report/cognitive-effects-in-combined-arms-a-case-study-of-the-division-2025/.

Jensen, Benjamin, and Ryan Kendall. 'Waze for War: How the Army Can Integrate Artificial Intelligence'. War on the Rocks, 2 September 2016. https://warontherocks.com/2016/09/waze-for-war-how-the-army-can-integrate-artificial-intelligence/.

Jensen, Benjamin, and Dan Tadross. 'How Large-Language Models Can Revolutionize Military Planning'. War on the Rocks, 12 April 2023. https://warontherocks.com/2023/04/how-large-language-models-can-revolutionize-military-planning/.

Jones, Jeffrey Bryant, and Jack N. Summe. *Psychological Operations in Desert Shield, Desert Storm and Urban Freedom*. AUSA Institute of Land Warfare, 1997.

Jordan, M. I., and T. M. Mitchell. 'Machine Learning: Trends, Perspectives, and Prospects'. *Science* 349, no. 6245 (17 July 2015): 255–60. https://doi.org/10.1126/science.aaa8415.

Kawasaki, Charlie. '6 Ways AI Can Make Sense of Sensor Data in 2020'. C4ISRNet, 14 February 2020. https://www.c4isrnet.com/thought-leadership/2020/02/14/6-ways-ai-can-make-sense-of-sensor-data-in-2020/.

Khan, S.A., and Reemiah Muneer. 'A Novel Thresholding for Prediction Analytics with Machine Learning Techniques'. *International Journal of Computer Science and Network Security* 23, no. 1 (2023): 33–40. https://doi.org/10.22937/IJCSNS.2023.23.1.5.

Kinder, Tabby. 'Silicon Valley Chiefs Urge Pentagon Procurement Overhaul'. *Financial Times*, 26 June 2023, sec. Tech start-ups. https://www.ft.com/content/45da39f2-4e05-46f1-96f4-813fbba79b16.

Klimenko, A. B., and I. S. Korovin. 'A Technique of Deviant Behaviour Detection in Video Surveillance Systems Based on Complex Behaviour Analysis'. *Journal of Physics: Conference Series* 1661, no. 1 (November 2020): 012050. https://doi.org/10.1088/1742-6596/1661/1/012050.

Kohavi, Ron, and Roger Longbotham. 'Online Controlled Experiments and A/B Testing'. In *Encyclopedia of Machine Learning and Data Mining*, 922–29, 2017. https://doi.org/10.1007/978-1-4899-7687-1_891.

Kohavi, Ron, Diane Tang, and Ya Xu. *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press, 2020.

Kosinski, Michal, David Stillwell, and Thore Graepel. 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior'. *Proceedings of the National Academy of Sciences* 110, no. 15 (9 April 2013): 5802–5. https://doi.org/10.1073/pnas.1218772110.

Larson, Eric V., Richard E. Darilek, Daniel Gibran, Brian Nichiporuk, Amy Richardson, Lowell H. Schwartz, and Cathryn Quantic Thurston. 'Foundations of Effective Influence Operations: A Framework for Enhancing Army Capabilities'. RAND Corporation, 27 May 2009. https://www.rand.org/pubs/monographs/MG654.html.

Lee, Seow Ting. 'A Battle for Foreign Perceptions: Ukraine's Country Image in the 2022 War with Russia'. *Place Branding and Public Diplomacy*, 28 November 2022, 1–14. https://doi.org/10.1057/s41254-022-00284-0.

Lynn, William J. 'The End of the Military-Industrial Complex: How the Pentagon Is Adapting to Globalization'. *Foreign Affairs* 93, no. 6 (2014): 104–10.

Macal, C M, and M J North. 'Tutorial on Agent-Based Modelling and Simulation'. *Journal of Simulation* 4, no. 3 (1 September 2010): 151–62. https://doi.org/10.1057/jos.2010.3.

Macri, Kate. 'Army Is Modernizing Sensors for Data-Driven Decision-Making', 4 March 2022. https://governmentciomedia.com/army-modernizing-sensors-data-driven-decision-making.

Martineau, Kim. 'What Is Federated Learning?' *IBM Research Blog* (blog), 9 February 2021. https://research.ibm.com/blog/what-is-federated-learning.

Mauss, Iris B., and Michael D. Robinson. 'Measures of Emotion: A Review'. *Cognition & Emotion* 23, no. 2 (1 February 2009): 209–37. https://doi.org/10.1080/02699930802204677.

Mazarr, Michael J., Ryan Bauer, Abigail Casey, Sarah Heintz, and Luke J. Matthews. 'The Emerging Risk of Virtual Societal Warfare: Social Manipulation in a Changing Information Environment'. RAND Corporation, 9 October 2019. https://www.rand.org/pubs/research_reports/RR2714.html.

Merrin, William, and Andrew Hoskins. 'Tweet Fast and Kill Things: Digital War'. *Digital War* 1, no. 1–3 (2020): 184–93.

Miller, Judith Droitcour. 'A New Survey Technique for Studying Deviant Behavior'. Dissertation, George Washington University, 1984.

Ministry of Defence. 'Doctrine Note 19/04 on Information Manoeuvre'. British Ministry of Defence, 2019.

Mitts, Tamar, Gregoire Phillips, and Barbara F. Walter. 'Studying the Impact of ISIS Propaganda Campaigns'. *The Journal of Politics* 84, no. 2 (April 2022): 1220–25. https://doi.org/10.1086/716281.

Moffat, James, Smith Josephine, and Witty Susan. 'Emergent Behaviour: Theory and Experimentation Using the MANA Model'. *Journal of Applied Mathematics and Decision Sciences* 2006 (2 October 2006). https://doi.org/10.1155/JAMDS/2006/54846.

Morrison, J., Daniel Goldsmith, and Michael Siegel. 'Dynamic Complexity in Military Planning: A Role for System Dynamics', 1 January 2008.

Munoz, Arturo. *U.S. Military Information Operations in Afghanistan: Effectiveness of Psychological Operations 2001-2010*. RAND Corporation, 2012. https://www.jstor.org/stable/10.7249/mg1060mcia.

Nadkarni, Prakash M., Lucila Ohno-Machado, and Wendy W. Chapman. 'Natural Language Processing: An Introduction'. *Journal of the American Medical Informatics Association: JAMIA* 18, no. 5 (2011): 544–51. https://doi.org/10.1136/amiajnl-2011-000464.

NATO. 'About DIANA'. About Diana. Accessed 12 September 2023. https://www.diana.nato.int/about-diana.html.

———. 'NATO Allied Joint Doctrine (AJP-01)'. UK Government, 2010. https://www.gov.uk/government/publications/ajp-01-d-allied-joint-doctrine.

———. 'NATO and Afghanistan'. NATO, 2022. https://www.nato.int/cps/en/natohq/topics_8189.htm.

———. 'NATO Approves 2023 Strategic Direction for New Innovation Accelerator'. NATO, 10 December 2022. https://www.nato.int/cps/en/natohq/news_210393.htm.

Nato Standardization Office. 'AJP-3.10.1 Allied Joint Doctrine for Psychological Operations Edition B Version 1'. NATO, 2014. https://www.gov.uk/government/publications/ajp-3101-allied-joint-doctrine-for-psychological-operations.

Nazareth, Derek L., and Jae Choi. 'A System Dynamics Model for Information Security Management'. *Information & Management* 52, no. 1 (1 January 2015): 123–34. https://doi.org/10.1016/j.im.2014.10.009.

Nichol, A. D., M. Bailey, D. J. Cooper, POLAR, and EPO Investigators. 'Challenging Issues in Randomised Controlled Trials'. *Injury* 41 Suppl 1 (July 2010): S20-23. https://doi.org/10.1016/j.injury.2010.03.033.

Osinga, Frans. *Science, Strategy and War: The Strategic Theory of John Boyd*. 1st ed. Routledge, 2007.

Ostenberg, Alan Paul, and Fedora Maria Harmon Baquer. 'A Model for Battle Damage Assessment in Command and Control Warfare'. Master's Thesis, Naval Postgraduate School, 1994.

Patten, Mildred L, and Michelle Newhart. *Understanding Research Methods: An Overview of the Essentials*. 10th ed. Routledge, 2018.

Paul, Christopher. 'Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Worked Example'. RAND Corporation, 22 March 2017. https://www.rand.org/pubs/research_reports/RR809z4.html.

Payne, B. Keith, Clara Michelle Cheng, Olesya Govorun, and Brandon D. Stewart. 'An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement'. *Journal of Personality and Social Psychology* 89, no. 3 (September 2005): 277–93. https://doi.org/10.1037/0022-3514.89.3.277.

Payne, Keith, and Kristjen Lundberg. 'The Affect Misattribution Procedure: Ten Years of Evidence on Reliability, Validity, and Mechanisms'. *Social and Personality Psychology Compass* 8, no. 12 (2014): 672–86. https://doi.org/10.1111/spc3.12148.

Peisakhin, Leonid, and Arturas Rozenas. 'Electoral Effects of Biased Media: Russian Television in Ukraine'. *American Journal of Political Science* 62, no. 3 (2018): 535–50.

Peterson, Becky. 'Palantir Grabbed Project Maven Defense Contract after Google Left the Program: Sources'. Business Insider, 10 December 2019. https://www.businessinsider.com/palantir-took-over-from-google-on-project-maven-2019-12.

Pijpers, Peter B.M.J., and P.A.L. Ducheine. '"If You Have a Hammer..." Reshaping the Armed Forces' Discourse on Information Maneuver'. Amsterdam Center for International Law, 1 November 2021.

Prelec, Dražen. 'A Bayesian Truth Serum for Subjective Data'. *Science* 306, no. 5695 (15 October 2004): 462–66. https://doi.org/10.1126/science.1102081.

Reiss, Peter C., and Frank A. Wolak. 'Chapter 64 Structural Econometric Modeling: Rationales and Examples from Industrial Organization'. In *Handbook of Econometrics*, edited by James J. Heckman and Edward E. Leamer, 6:4277–4415. Elsevier, 2007. https://doi.org/10.1016/S1573-4412(07)06064-3.

Rid, Thomas. *Active Measures: The Secret History of Disinformation and Political Warfare*. Illustrated edition. New York: Farrar, Straus and Giroux, 2020.

Robins, Richard W., R. Chris Fraley, and Robert F. Krueger. *Handbook of Research Methods in Personality Psychology*. Handbook of Research Methods in Personality Psychology. New York, NY, US: The Guilford Press, 2007.

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. *Evaluation: A Systematic Approach*. 7th ed. SAGE Publications, 2003.

Santelises, Aaron. 'The Ukrainian Kharkiv Counter-Offensive and Information Operations'. The Cove, 2 November 2022. https://cove.army.gov.au/article/ukrainian-kharkiv-counter-offensive-and-information-operations.

Schimmack, Ulrich. 'Invalid Claims About the Validity of Implicit Association Tests by Prisoners of the Implicit Social-Cognition Paradigm'. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 16, no. 2 (March 2021): 435–42. https://doi.org/10.1177/1745691621991860.

Seese, Gregory S, Rafael E. Linera, and Erinn McQuagge. 'Effects-Based Psychological Operations Measures of Effectiveness: Measuring Change and Impact'. In *What Do Others Think and How Do We Know What They Are Thinking?* Strategic Multilayer Assessment. Department of Defense, 2018.

Smith, Steven T., Edward K. Kao, Erika D. Mackin, Danelle C. Shah, Olga Simek, and Donald B. Rubin. 'Automatic Detection of Influential Actors in Disinformation Networks'. *Proceedings of the National Academy of Sciences* 118, no. 4 (26 January 2021): e2011216118. https://doi.org/10.1073/pnas.2011216118.

Stanton, Andrew, Amanda Thart, Ashish Jain, Priyank Vyas, Arpan Chatterjee, and Paulo Shakarian. 'Mining for Causal Relationships: A Data-Driven Study of the Islamic State'. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2137–2146. KDD '15. New York, NY, USA: Association for Computing Machinery, 2015. https://doi.org/10.1145/2783258.2788591.

Stratton, Samuel J. 'Quasi-Experimental Design (Pre-Test and Post-Test Studies) in Prehospital and Disaster Research'. *Prehospital and Disaster Medicine* 34, no. 6 (December 2019): 573–74. https://doi.org/10.1017/S1049023X19005053.

Tatham, Steve. 'Target Audience Analysis'. *Three Swords Magazine* 28 (2015): 50–53.

The Economist. 'Open-Source Intelligence Is Piercing the Fog of War in Ukraine'. *The Economist*, 13 January 2023. https://www.economist.com/interactive/international/2023/01/13/open-source-intelligence-is-piercing-the-fog-of-war-in-ukraine.

The World Bank. 'Dimewiki Matching'. Accessed 12 September 2023. https://dimewiki.worldbank.org/Matching.

———. 'Dimewiki Propensity Score Matching'. Accessed 12 September 2023. https://dimewiki.worldbank.org/Propensity_Score_Matching.

———. 'Dimewiki Regression Discontinuity', n.d. https://dimewiki.worldbank.org/Regression_Discontinuity#:~:text=Regression%20Discontinuity%20Design%20(RDD)%20is,who%20is%20eligible%20to%20participate.

Torgerson, David J., and Carole J. Torgerson. 'The Limitations of Before and After Designs'. In *Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*, edited by David J. Torgerson and Carole J. Torgerson, 9–16. London: Palgrave Macmillan UK, 2008. https://doi.org/10.1057/9780230583993_2.

US Army. 'FM 3-05.301 Psychological Operations Tactics, Techniques, and Procedures', 2003.

Vecchiato, Giovanni, Laura Astolfi, Fabrizio De Vico Fallani, Febo Cincotti, Donatella Mattia, Serenella Salinari, Ramon Soranzo, and Fabio Babiloni. 'Changes in Brain Activity during the Observation of TV Commercials by Using EEG, GSR and HR Measurements'. *Brain Topography* 23, no. 2 (June 2010): 165–79. https://doi.org/10.1007/s10548-009-0127-0.

Verbruggen, Maaike. 'The Role of Civilian Innovation in the Development of Lethal Autonomous Weapon Systems'. *Global Policy* 10, no. 3 (2019): 338–42. https://doi.org/10.1111/1758-5899.12663.

Wadhwani, Ganesh Kumar, Pankaj Kumar Varshney, Anjali Gupta, and Shrawan Kumar. 'Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia–Ukraine War'. *SN Computer Science* 4, no. 4 (21 April 2023): 346. https://doi.org/10.1007/s42979-023-01790-5.

Wakabayashi, Daisuke, and Scott Shane. 'Google Will Not Renew Pentagon Contract That Upset Employees'. *The New York Times*, 1 June 2018, sec. Technology. https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html.

Wanless, Alicia, and Michael Berk. 'The Changing Nature of Propaganda: Coming to Terms with Influence in Conflict'. In *The World Information War*, 63–80. Routledge, 2021. https://doi.org/10.4324/9781003046905-7.

Wanless, Alicia, and James Pamment. 'How Do You Define a Problem Like Influence?' *Journal of Information Warfare* 18, no. 3 (2019): 1–14.

Warner, Stanley L. 'Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias'. *Journal of the American Statistical Association* 60, no. 309 (2012): 63–69. https://doi.org/10.1080/01621459.1965.10480775.

Wierenga, Berend, and Ralf van der Lans. *Handbook of Marketing Decision Models*. International Series in Operations Research & Management Science. Springer Cham, 2017.

Winship, Christopher, and Robert D. Mare. 'Models for Sample Selection Bias'. *Annual Review of Sociology* 18 (1992): 327–50.

Xiao, Jianxiong, Bryan C. Russell, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. 'Basic Level Scene Understanding: From Labels to Structure and Beyond'. *MIT Web Domain*, November 2012. https://dspace.mit.edu/handle/1721.1/90941.

Yi, Xun, Russell Paulet, and Elisa Bertino. *Homomorphic Encryption and Applications*. SpringerBriefs in Computer Science. Springer, 2014.

Yu, Jun-Wu, Guo-Liang Tian, and Man-Lai Tang. 'Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis'. *Metrika* 67, no. 3 (1 April 2008): 251–63. https://doi.org/10.1007/s00184-007-0131-x.

The Hague Centre
for Strategic Studies

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies
Email: info@hcss.nl
Website: www.hcss.nl