SECURITY & DEFENCE QUARTERLY

Rethinking Cognition in Cognitive Warfare: A Functional–Cognitive Framework for Prediction and Control

Kyuichi Miyazaki

MIYAZAKIa3v@inet.aci.mod.go.jp

Air and Space Studies Institute, Japan Air Self-Defense Force, 2-2-1, Nakameguro, Meguro 153-0061, Tokyo, Japan

Abstract

Research on cognitive warfare often relies on broad claims about cognition, attitudes, narratives, trust, and behaviour while leaving unclear what type of claim is being made and what evidence would be needed to support it. This paper addresses that problem by presenting the functional–cognitive framework (FCF) as an operational specification framework for analysing, evaluating, and designing cognitive-warfare claims. This conceptual paper brings together the functional–cognitive distinction, learning-theory exemplars, cognitive-warfare doctrine, strategic-communication evaluation, and causal-inference reasoning. It integrates a worked example, clarifies the bridge between individual-level mechanisms and population-level indicators, and incorporates military conditions, such as time pressure; overload; confusion; stress; observe, orient, decide, and act (OODA)-loop disruption; and performance degradation. FCF is not presented as a novel theory of cognition. Instead, it is presented as a disciplined way to specify inputs, pathways, contexts, indicators, rival explanations, and missing links before strong claims are made. The paper shows how this reframing changes analytical questions, data requirements, countermeasure selection, and practical conclusions, especially when analysts move too quickly from aggregate indicators to claims about individual cognition or controllable mechanisms. FCF does not solve the problem of unobserved mechanisms, and it does not replace standard tools of causal inference or programme evaluation. Its contribution is narrower but useful: it helps analysts to state more precisely what has been observed, what has only been inferred, and what type of countermeasure or design recommendation can legitimately follow. In that sense, the framework supports both conceptual discipline and more credible military analysis.

Keywords:

functional analysis, cognitive warfare, prediction and control, functional–cognitive framework, influence operations

Article info
Received: 10 January 2026
Revised: 2 May 2026
Accepted: 10 June 2026
Available online: 26 July 2026

Citation: Miyazaki, K. (2026) 'Rethinking cognition in cognitive warfare: A functional–cognitive framework for prediction and control', *Security and Defence Quarterly*, 55(3). doi: 10.35467/sdq/224399.

Introduction

Cognitive warfare is often discussed through broad terms such as cognition, perception, belief, narrative, trust, susceptibility, resilience, and behavioural effect. These terms point to real concerns about how information environments shape judgement and action. The analytical problem is that they can appear to explain more than the evidence supports. A campaign may coincide with a change in opinion, sharing behaviour, hesitation, confusion, or decision tempo. The difficulty begins when that pattern is treated as evidence about what happened inside the target population, which mechanism produced the effect, and which countermeasure should follow (Buck, 2019, pp. 148–150; Cornish *et al.*, 2011, pp. 4–8).

This problem is especially important in cognitive warfare because the concept is used in more than one way. In some texts, cognitive warfare mainly refers to influence on beliefs, attitudes, or narratives. In others, especially those closer to North Atlantic Treaty Organization (NATO) usage, it also includes attempts to overload attention, degrade orientation, exploit stress, disrupt cohesion, or slow decision cycles under operational conditions. These concerns are related, but they are not the same. Unless the analysis states what type of claim is being made, description, explanation, prediction, and intervention can easily slide into one another (Claverie and du Cluzel, 2021, pp. 2–4; Masakowski and Blatny, 2023, pp. 1–8).

This paper starts from a narrow claim. The central issue is not whether cognition matters but how claims about cognition are written, inferred, and operationalised. A population-level shift in support, a fall in operational performance, or a rise in message engagement may be meaningful observations. None of them, however, automatically identifies a specific individual-level mechanism, and none of them by itself shows which intervention is likely to work in future conditions. If cognitive warfare is to function as an analytical category rather than a broad label, its claims must make their evidential status clear (Curran and Bauer, 2011, pp. 583–590; King, 1997, pp. 3–24; Robinson, 1950, pp. 351–357).

To address this problem, the paper applies the functional–cognitive framework (FCF) as an operational specification framework for cognitive-warfare analysis. FCF is not presented as a replacement theory of mind. Nor does it assume that internal pathways are fully observable in real operations. Its role is more practical. It helps to separate observable inputs, operational context, hypothesised pathways, observable outputs, rival explanations, and inferential gaps. This is useful because many debates about cognitive warfare are not only about whether an effect occurred. They are also about what type of claim the available evidence can support (Buck, 2019, pp. 148–150; Robinson, 1950, pp. 351–357; Yarkoni and Westfall, 2017, pp. 1100–1105).

The argument also connects cognitive warfare to a wider problem in strategic-communication research. Strategic communication often has to act under uncertainty while still reporting impact. In that setting, cautious statements about associations or aggregate change can be rewritten as claims about mechanism, within-person change, or reliable targeting. The same pressure appears in cognitive-warfare analysis, and it is often stronger because the field combines cognitive constructs, military urgency, and operational relevance. When these elements are combined, readers may assume a stronger claim than the evidence warrants unless the text prevents that shift (Buck, 2019, pp. 158–160; Hernán and Robins, 2020, pp. 1–22).

The paper therefore makes a specific contribution. It does not offer a full theory of persuasion, and it does not reject construct-based cognitive research. Instead, it proposes a way

to keep strong verbs, mechanism language, and practical recommendations aligned with the evidence available. This matters for research because conceptual inflation weakens cumulative knowledge. It also matters for practice because over-specified interpretations can harden into flawed targeting, misleading assessments, or brittle countermeasure plans (Buck, 2019, pp. 158–160; Pamment and Smith, 2022, pp. 7–15).

The paper first explains why cognitive-warfare claims often remain underspecified and where level shifts typically enter the argument. It then frames FCF as an operational specification framework rather than a novel theory. Next, it discusses associative learning and operant contingencies as two examples of mechanism language and clarifies the limits of importing individual-level learning concepts into security analysis. The paper then examines the bridge between individual-level mechanisms and population-level indicators, presents a stepwise operationalisation procedure, and applies the framework to a worked example. The final parts draw out implications for military analysis and defence practice, such as overload; confusion; time pressure; and observe, orient, decide, and act (OODA)-loop disruption (Cornish *et al.*, 2011, pp. 4–8, 28–31; Buck, 2019, pp. 150–160).

Why cognitive-warfare claims often remain underspecified

A recurring difficulty in cognitive-warfare analysis is that several tasks are bundled together. A paper may describe a campaign, interpret its potential meaning, suggest a psychological pathway, imply that the same pathway operated across the target population, and then recommend a countermeasure. Each move may sound reasonable in isolation. The problem is that each move changes the type of claim being made. If those shifts are left implicit, the conclusion appears firmer than the evidence justifies (Druckman and McGrath, 2019, pp. 111–119; Firebaugh, 2015, pp. 865–867; King, 1997, pp. 3–24; Kleinberg *et al.*, 2015, pp. 491–495; Molenaar, 2004, pp. 201–218; Robinson, 1950, pp. 351–357).

This difficulty becomes more acute in security contexts because the underlying processes are often not directly observable. Analysts rarely know exactly who was exposed to which cue, how often exposure occurred, what competing influences were present, or whether the relevant change concerned belief, emotion, compliance, hesitation, confusion, fear, fatigue, or tactical decision quality. Even if data exists, it is often aggregate, *post hoc*, or operationally noisy. Under these conditions, a cognitive label can stand in for missing specification. A claim about attention, appraisal, trust, or bias may then appear to be both an explanation and a recommendation, even when the data supports neither role on their own (Curran and Bauer, 2011, pp. 583–590; King, 1997, pp. 3–24; Robinson, 1950, pp. 351–357). Psychometric work on construct validity makes the same point in measurement terms: indicators should not be treated as direct readouts of an underlying attribute unless the assumptions linking measure and construct are specified and defended (Borsboom *et al.*, 2004, pp. 1061–1063; Strauss and Smith, 2009, pp. 1–4).

Replacing cognitive vocabulary with functional language does not solve this problem by itself. The same overreach occurs if functional pathways are written as if they had been directly observed when they have only been inferred. The question is therefore not whether cognition should be rejected. It is what type of claim is being made, what evidence would be needed for that claim, and where the argument moves from observation to inference (Cronbach and Meehl, 1955, pp. 281–302; Flake and Fried, 2020, pp. 456–465).

Three common moves that create slippage

The first common move is the shift from aggregate difference to individual response. A report may note that one segment showed lower support, higher susceptibility, or greater engagement than another, and then write as if identifiable individuals in that segment can now be targeted or predicted. This is the classic ecological or aggregation problem in a military setting: relationships observed across groups or populations do not automatically become rules for persons (Curran and Bauer, 2011, pp. 583–590; King, 1997, pp. 3–24; Robinson, 1950, pp. 351–357).

The second common move is the shift from plausible mechanism to demonstrated process. Cognitive-warfare analysis often draws on psychologically plausible constructs, such as salience, fear, trust, or bias. Such language can be useful as a hypothesis. The problem arises when the analysis moves from plausibility to certainty: because a mechanism sounds theoretically coherent; it is written as if it had been observed operating in the target population under field conditions (Cronbach and Meehl, 1955, pp. 281–302; Flake and Fried, 2020, pp. 456–465).

The third common move is the shift from *post hoc* coherence to causal or operational conclusion. A campaign may coincide with public hesitation, greater narrative reach, degraded morale, or slower decision processes. That pattern may matter. Yet a coherent story about those observations is not equivalent to evidence that the campaign produced them, and it is not equivalent to a validated rule for future intervention (Cornish *et al.*, 2011, pp. 4–8, 28–31; UK Ministry of Defence and NATO, 2019, pp. xvii–xviii, 1-1–1-5).

Why military contexts intensify the problem

Military writing can intensify this problem because it often needs decisive verbs: deter, reassure, disrupt, degrade, harden, shape, and compel. These verbs are useful because they connect communication activity to operational objectives. They can also hide uncertainty, especially when a product is written under time pressure and moves through an organisation that rewards clarity and actionability (NATO Standardization Office, 2023, pp. 1-1–1-6; UK Ministry of Defence and NATO, 2019, pp. xvii–xviii, 1-1–1-5).

Cognitive warfare creates a second pressure because it is not limited to persuasion. It also concerns confusion, overload, stress exploitation, cohesion erosion, and decision degradation. These are real military concerns, but they are difficult to infer cleanly from available indicators. A fall in performance, for example, may reflect fatigue, poor training, equipment problems, organizational friction, or adversary action. If the analysis calls the outcome cognitive warfare without specifying the route by which the effect supposedly occurred, the concept risks of becoming a catch-all label rather than an analytical aid.

Recent NATO work makes this broader operational emphasis explicit. The NATO chief scientist research report describes cognitive warfare as part of the fight for cognitive superiority. It notes that adversaries exploit vulnerabilities in the OODA decision cycle to target cognitive weaknesses. It also presents cognitive warfare as involving influence, degradation of adversary capabilities, strengthening of resilience, and improvement of human and technological cognition (NATO Science and Technology Organization, 2025, pp. 4, 6, 8–10).

Scope of the argument

The argument is deliberately limited. It does not claim that strong causal claims are impossible, that cognitive variables are unusable, or that aggregate indicators are uninformative. It claims that different type of evidence support different types of conclusions. Cognitive-warfare analysis becomes methodologically weak when these distinctions are hidden. The next part responds to that problem by treating FCF as a framework for specification discipline rather than as a stand-alone explanatory theory.

FCF as an operational specification framework

Functional–cognitive framework is used here as an operational specification framework for cognitive-warfare analysis (Figure 1). At its most basic level, it distinguishes between a cognitive description of what is thought to matter and a functional specification of what can be stated about inputs, conditions, observable responding, and consequences. That distinction is not a new discovery. Its value lies in application. In cognitive-warfare writing, it helps to separate what has been observed from what has only been inferred and what remains unknown (De Houwer, 2011, pp. 202–209; De Houwer *et al.*, 2013, pp. 252–256; Hernán and Robins, 2020, pp. 1–22; Hughes *et al.*, 2016, pp. 4–14; Shmueli, 2010, pp. 289–310; Yarkoni and Westfall, 2017, pp. 1100–1105).

This framing also narrows the originality claim of the paper. FCF is not presented as a novel theory of cognition or as a replacement for psychology, learning theory, or strategic communication. The contribution is a domain application: it brings the functional–cognitive distinction to a field in which claims about effect, mechanism, and intervention often travel further than the evidence allows. The framework is intended to sharpen analysis, not to displace neighbouring research traditions.

This operational emphasis also clarifies the relationship between FCF and adjacent methods. FCF is not a substitute for causal inference, programme evaluation, psychometrics, or computational analysis. It sits earlier in the workflow. It asks the analyst to specify the type of claim being made, the implied pathway, the required indicators, and the missing links before strong causal or operational verbs are allowed to stand. This role is modest, but useful, because many weaknesses in cognitive-warfare analysis enter before formal evaluation begins (Shmueli, 2010, pp. 289–310; Yarkoni and Westfall, 2017, pp. 1100–1105).

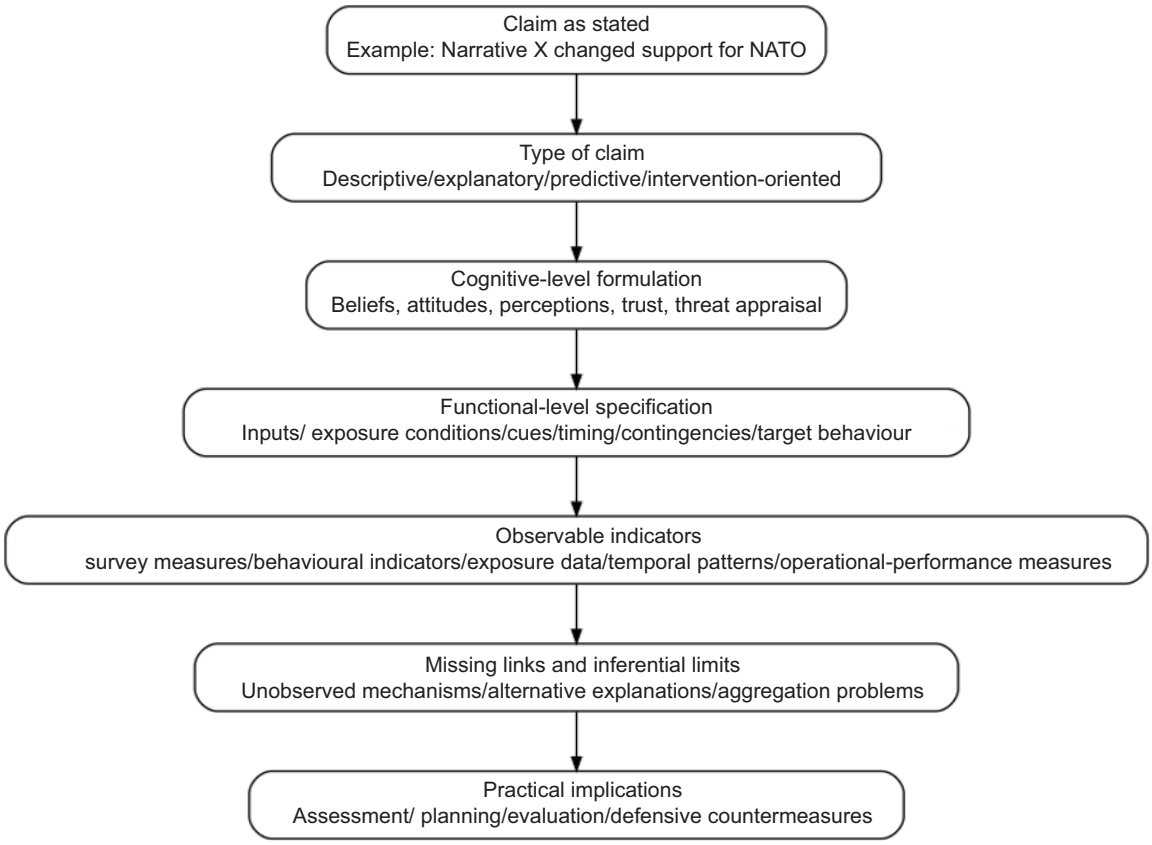
This position is consistent with strategic-communication evaluation work. Macnamara and Gregory (2018, pp. 469–470, 474–476) argue that evaluation is not only an end-stage measurement. It can also expose objectives, inputs, contexts, and underlying logic. The authors also criticise narrow message-tracking approaches because such approaches miss the adaptive, participatory, and context-sensitive character of strategic communication. Corman and Dooley (2008, pp. 2–4, 15) similarly warn against control-oriented searches for a single best message in complex communication environments. Together, these points support the role assigned to FCF here: it does not replace evaluation, but it makes its assumptions explicit before effects are claimed.

What the framework does

Used in this narrower sense, FCF performs four tasks. Firstly, it clarifies the target of analysis by distinguishing description, explanation, prediction, and intervention recommendation. Secondly, it separates cognitive labels from the inputs, contexts, and outputs that

http://doi.org/10.35467/sdq/224399

Figure 1. Functional–Cognitive Framework for the Analysis of Cognitive Warfare Claims. This figure summarises the overall analytical logic of the Functional–Cognitive Framework, moving from the claim as stated to claim type, cognitive-level formulation, functional-level specification, observable indicators, inferential limits, and practical implications.



would have to be observed if the claim were to be credible. Thirdly, it makes inferential gaps visible. Finally, it helps to align recommendations with the strength of the evidence rather than with the appeal of the narrative (Breiman, 2001, pp. 199–231; Hernán and Robins, 2020, pp. 1–22).

What the framework does not do

Functional–cognitive framework does not provide privileged access to internal mechanisms. It does not eliminate uncertainty about what occurred within a person or a unit. It also does not guarantee that a detailed specification can always be tested in operational settings. In many cases, using the framework leads to a more bounded conclusion rather than a stronger one. That is not a weakness. It is one of its safeguards (Curran and Bauer, 2011, pp. 583–590; Hamaker, 2012, pp. 43–61).

Two mechanism families: Associative learning and operant contingencies

Here, associative learning and operant contingencies are treated together because they illustrate two complementary ways of specifying mechanism language. The point is not that they are the only relevant pathways in cognitive warfare. The point is that they

show how mechanism claims can be written more clearly without pretending that real operations offer laboratory-level observability. From a behaviour-analytical perspective, relevant task is to identify the environmental variables functionally related to responding. Applied behaviour analysis defines itself as a scientific approach for discovering the environmental variables that reliably influence socially significant behaviour and for developing a technology of behavioural change; within that approach, the three-term contingency of antecedent, behaviour, and consequences is treated as the basic unit of operant analysis (Cooper *et al.*, 2020, pp. 57–58).

Associative learning is relevant when patterned responding emerges through repeated pairings, salience, emotional tagging, or cue generalisation (Rescorla and Wagner, 1972, pp. 64–99). In cognitive-warfare contexts, this may appear in claims that repeated co-occurrence between a symbol and a threat image changes how the symbol is processed, or that repeated narrative pairing shifts the emotional meaning of a policy object. Operant contingencies are relevant when social reinforcement, punishment, incentives, reputational costs, or repeated feedback stabilise or suppress public or organisational behaviour. A disinformation ecosystem, for example, may not merely persuade. It may reward amplification, punish dissent, or create uncertainty that changes what actors are willing to say or do in public. Here, behaviour-analytical work is useful because it insists that behaviour occurs in an environmental context and that analysts specify the antecedent and consequent events relevant to the pattern under study (Cooper *et al.*, 2020, pp. 45, 57–58; Staal, 2004, pp. 5–17).

The acquisition/performance distinction is also treated here as a standard learning-theory distinction, not as a new contribution. The distinction matters in security settings because a disposition, association, or behavioural tendency may have been acquired but still be expressed only under particular conditions. Fatigue, time pressure, fear, command climate, social surveillance, or reward structure may affect whether a response appears. A pattern that looks like an attitude change may therefore be a performance issue, and a performance failure may be misread as persuasion (Claverie and du Cluzel, 2021, pp. 2–4; Staal, 2004, pp. 31–45).

Why these exemplars remain useful

These examples are useful because they force more explicit writing. A paper that invokes associative learning should specify the repeated pairings, the expected temporal pattern, and the output that would count as relevant evidence. A paper that invokes operant contingencies should specify the reinforcing or punishing conditions, the unit of behaviour, the probable feedback loops, and the conditions under which the pattern should weaken or persist. The same applies to motivational conditions. If the proposed pathway depends on social approval, reassurance, relief, or avoidance of reputational cost, the paper should specify what currently makes those consequences valuable. Once these questions are asked, some mechanism claims will remain plausible, while others will be revealed as underspecified metaphors (Cooper *et al.*, 2020, pp. 412–418; Michael, 1993, pp. 191–193).

From persuasion to degradation

The same logic applies when the target outcome is impairment rather than persuasion. Cognitive warfare is often discussed as influence on minds, but military analysis also worries about overload, confusion, attentional fragmentation, premature closure, and degraded decision quality under stress. These effects fit the NATO emphasis on disrupting

orientation and decision cycles, not only changing opinions. Under such conditions, the analytical question is not simply what message was believed. It is whether the information environment made observation, orientation, decision, or action slower, less accurate, or more brittle. FCF helps by asking the analyst to specify the route from input to degraded performance instead of treating every decline as a general cognitive effect (Claverie and du Cluzel, 2021, pp. 2–4; Masakowski and Blatny, 2023, pp. 1–8; Staal, 2004, pp. 31–45).

The problem is therefore operational, not only semantic. A force may remain committed to a mission while becoming slower, more hesitant, or more brittle in how it orients, prioritises, and acts. The NATO chief scientist report links cognitive warfare to this operational dimension by discussing propaganda, deception, interference, and manipulation as ways in which adversaries may alter behaviour and decision-making across military and civilian populations in the digital information environment (NATO Science and Technology Organization, 2025, pp. 6–10).

Read through this OODA-focused lens, a cognitive-warfare claim should specify more than whether a target believed a message. It should also state which part of the decision cycle is claimed as affected. An operation may distort observation by flooding the environment with contradictory cues, impair orientation by making the situation harder to interpret, delay decision by increasing uncertainty about costs and consequences, or disrupt action by weakening coordination and confidence. FCF helps to analyse such claims by asking the analyst to state relevant input, the decision-cycle function allegedly affected, the observable performance indicator, and the rival explanations that could also account for slower or poorer performance.

From individual-level mechanisms to population-level indicators

In many security contexts, the analyst cannot directly observe how many cue–outcome pairings a specific subject experienced, how a specific individual updated a response over time, or whether a moment of hesitation reflected persuasion, fatigue, fear, overload, or task complexity. The problem is even sharper when the claimed target is performance degradation rather than persuasion. A drop in accuracy or decision speed may be consistent with adversary influence, but it does not identify a particular mechanism on its own (Curran and Bauer, 2011, pp. 583–590; King, 1997, pp. 3–24; Staal, 2004, pp. 31–45).

Functional–cognitive framework therefore makes a narrower claim. It does not require complete direct observation of individual pathways. It requires the analyst to state the inferential steps between inputs, hypothesized pathways, observable indicators, and conclusions. In this sense, the framework is more honest than broad cognitive shorthand. It does not remove the missing data; it identifies the missing links before strong conclusions are drawn (Hernán and Robins, 2020, pp. 1–22).

Why direct observation is often unavailable

In laboratory or tightly controlled settings, an analyst may be able to manipulate cues, monitor acquisition across trials, or estimate within-subject change under repeated exposure. In cognitive-warfare settings, such control is rarely available. Exposure is uneven, populations are heterogeneous, channels overlap, and adversaries, media, institutions, and local events interact over time. Military conditions add further noise: tempo, fatigue,

stress, uncertainty, and physical danger can alter both exposure and responding (Staal, 2004, pp. 5–17; UK Ministry of Defence and NATO, 2019, pp. xvii–xviii, 1-1–1-5).

For that reason, a direct one-to-one mapping from learning-theory language to field evidence is usually impossible. A paper that assumes such mapping without qualification risks of replacing one untestable vocabulary with another. Individual-level mechanism language is not forbidden, but its evidential status must be stated honestly (Curran and Bauer, 2011, pp. 583–590; Hamaker, 2012, pp. 43–61).

What aggregate indicators can and cannot show

This point does not mean that aggregate indicators should be dismissed. In practice, population-level indicators are often the most relevant data available. Surveys, digital trace measures, behavioural tallies, uptake data, voting patterns, refusal rates, compliance patterns, information-sharing data, or performance metrics may each show that something changed across a population, segment, or unit (Pamment and Smith, 2022, pp. 7–15; Stone *et al.*, 2022).

At the same time, aggregate indicators do not identify by themselves the route by which change occurred. A shift in support for a policy may be compatible with narrative persuasion, elite cueing, fear amplification, local events, selection effects, or ordinary political fluctuation. A drop in operational performance may be compatible with information overload, poor coordination, sleep loss, equipment problems, or command friction. Aggregate indicators show that a pattern exists; they do not automatically settle the mechanism (Hernán and Robins, 2020, pp. 1–22; King, 1997, pp. 3–24; Robinson, 1950, pp. 351–357).

Recent disinformation research illustrates why this distinction matters. In a 19-country survey on Russian narratives justifying the invasion of Ukraine, Zilinsky *et al.* (2024, pp. 965–967, 981–982) find that a pre-existing conspiratorial outlook predicts endorsement of war-related disinformation more consistently than media diets, while the association of specific platforms with belief is weaker and varies across countries. That pattern does not make exposure irrelevant, but it shows why platform-level association alone is an insufficient basis for strong claims about mechanism, susceptibility, or targeting.

The inferential bridge

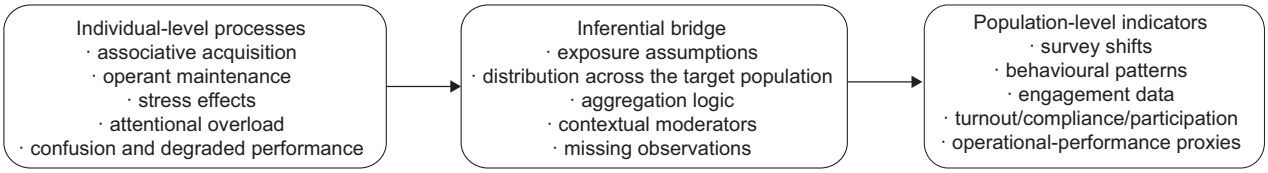
The inferential bridge addresses this problem. Between individual-level mechanism language and population-level indicators, lies a chain of conditional propositions: what was delivered, to whom, under what operational conditions, through what plausible route, and with what expected pattern in the data. The bridge also asks what remains unknown. Were the people who changed actually exposed? Was exposure concentrated in the subgroup that shifted? Did the timing fit the proposed route? Did rival explanations move in the same period? (Curran and Bauer, 2011, pp. 583–590; King, 1997, pp. 3–24; Robinson, 1950, pp. 351–357).

That's why effect claims should remain bounded in geopolitical analysis. Gerrits (2019, pp. 3–5) argues that disinformation in international relations often exploits the existing divisions and concerns rather than creating them from scratch, and that its security impact should not be overstated even when it is disruptive and politically salient. That caution is compatible with the present argument: analysts should neither dismiss disinformation nor treat aggregate movement as automatic proof of deep cognitive conversion.

A practical way to think about the bridge is to treat it as a chain rather than a leap. The analyst first specifies the input or influence environment. The analyst then specifies the conditions that make a particular route plausible. The analyst then states the expected pattern in observable indicators. Only after that sequence is stated should the paper decide how strong its conclusion could be (Hernán and Robins, 2020, pp. 1–22).

The inferential bridge between individual-level processes and population-level indicators is illustrated in Figure 2.

Figure 2. Bridging Individual-Level Mechanisms and Population-Level Indicators. This figure illustrates the inferential bridge between individual-level processes, the assumptions required to aggregate them, and population-level indicators. FCF does not eliminate missing data; it makes the inferential steps and missing links more explicit.



Implications for design and evaluation

Once this bridge is stated openly, aggregate indicators can be used more productively. They become one layer of evidence in a structured argument, rather than a shortcut to internal mechanism claims. This posture is also more compatible with causal-inference and evaluation logic. The analyst can define the estimate more clearly, specify what comparison matters, and distinguish explanation, process inference, and prediction instead of treating them as interchangeable (Hernán and Robins, 2020, pp. 1–22; Shmueli, 2010, pp. 289–310).

This shift also affects evaluation design. Instead of asking simply whether a campaign "changed minds," analysts can ask more specific questions: what type of change is being claimed; what comparison supports it; what data would reduce uncertainty; and what practical decision depends on the answer? In that sense, FCF is best understood as a front-end specification tool that makes later evaluation more disciplined (Moore *et al.*, 2015, p. h1258; Shmueli, 2010, pp. 289–310; Yarkoni and Westfall, 2017, pp. 1100–1105).

Operationalising cognitive-warfare claims

The framework can be used as a stepwise analytical procedure. The aim is to turn broad claims into a series of visible decisions rather than leaving the inferential structure implicit. Although the steps are presented sequentially, the process is iterative. New data may force the analyst to redefine the claim, narrow the population, change the preferred pathway, or soften the recommendation (Hernán and Robins, 2020, pp. 1–22; W.K. Kellogg Foundation, 2004, pp. 1–9).

Clarify the type of claim

The first step is to identify the type of claim being made. At minimum, the analyst should distinguish descriptive, explanatory, predictive, and intervention-oriented claims. Many cognitive-warfare texts begin as description and end as intervention recommendation

without marking the transition. This first step prevents that drift (Shmueli, 2010, pp. 289–310; Yarkoni and Westfall, 2017, pp. 1100–1105).

A useful practical test is sentence-level. If the wording implies stable susceptibility, mechanism, or reliable targeting, the paper should ask whether the available design supports that level of inference. If it does not, the sentence should be rewritten until the wording matches the evidence (Buck, 2019, pp. 148–160; Yarkoni and Westfall, 2017, pp. 1100–1105).

Specify the target population and operational context

The second step is to define relevant population and context. The analyst should identify who is claimed to be affected, under what operational conditions, during what time window, and in relation to which decision or behaviour. In military settings, context is part of the pathway, not background. Time pressure, fatigue, uncertainty, and command relationships can change how the same information affects performance (NATO Standardization Office, 2023, pp. 1-1–1-6; Staal, 2004, pp. 5–17; UK Ministry of Defence and NATO, 2019, pp. xvii–xviii, 1-1–1-5).

A message that produces hesitation in a low-pressure public discussion may not function in the same way during a crisis. Conversely, a cue that is weak in ordinary conditions may become powerful under acute threat or overload. Operational context therefore has to be stated as part of the claim, not as an afterthought (Staal, 2004, pp. 31–45).

Specify inputs, cues, and hypothesised pathways

The third step is to specify relevant inputs. Inputs may include narratives, symbols, emotional cues, framing devices, repeated social feedback, or engineered ambiguity. Once the inputs are specified, the analyst can state one or more hypothesised pathways. One pathway may be mainly associative, another may depend more on reinforcement or punishment, and another may involve overload or stress-induced performance degradation. Behaviour analysis is especially useful here because it distinguishes antecedent cues that indicate when a consequence is likely to follow a response from motivational conditions that alter how valuable that consequence is at that moment. Conflating these functions makes influence claims harder to test (Claverie and du Cluzel, 2021, pp. 2–4; Cooper *et al.*, 2020, pp. 412–418).

The important point is that the pathway remains a hypothesis until the data supports more. The framework encourages more explicit hypotheses; it does not permit the analyst to write them as observed facts (Cronbach and Meehl, 1955, pp. 281–302; Flake and Fried, 2020, pp. 456–465).

This requirement is especially important in digitally mediated campaigns. Howard *et al.* (2023, pp. 47–49) describe computational propaganda as a socio-technical phenomenon involving algorithms, automation, and human curation, and they argue that its study requires mixed methods rather than only technical or message-level inspection. Mustafa *et al.* (2025, pp. 45–47, 59–60) similarly emphasise that computational propaganda develops through interactions among actors, drivers, and means of diffusion over time. Feedback loops and structural conditions shape what becomes visible in the data.

Identify observable indicators and expected patterns

The fourth step is to identify what can be observed if the hypothesised pathway is at least partly correct. Observable indicators may include changes in survey responses, shifts in digital behaviour, repeated refusals, altered decision timing, degraded task performance, or changes in public or organisational language. Expected patterns should also be made explicit. If the pathway is plausible, what should appear first, in which subgroup, and with what timing? (Pamment and Smith, 2022, pp. 15–24; Stone *et al.*, 2022).

This question matters because an underspecified claim can often accommodate almost any outcome after the fact. A more explicit expected-pattern statement makes the analysis more falsifiable and reduces the temptation to fit a mechanism to whatever happened (Breiman, 2001, pp. 199–231; Hernán and Robins, 2020, pp. 1–22).

Identify missing links and rival explanations

The fifth step is to make missing links explicit. These may include unknown exposure rates, inability to observe private cognition, poorly measured baseline conditions, or lack of a credible comparison group. Rival explanations should also be considered systematically. A narrative campaign may coincide with economic shock, leadership turnover, military loss, domestic scandal, or independent media dynamics. If those alternatives are not addressed, the wording should remain bounded (Hernán and Robins, 2020, pp. 1–22).

This step also shows why the procedure is a framework rather than a list. The missing-link step directly constrains the strength of the conclusion and therefore shapes the decision that follows.

Integrate ethics and governance into operationalisation

The sixth step is to integrate ethics and governance into the analysis rather than leaving them to a short concluding paragraph. Cognitive-warfare research often concerns vulnerable populations, democratic legitimacy, coercive conditions, or military necessity. These are not external issues. They help to determine appropriate recommendations, permissible interventions, and how much uncertainty is acceptable before action is taken (Organization for Economic Cooperation and Development [OECD], 2022, pp. 19–20; Pamment and Smith, 2022, pp. 22–24; UK Ministry of Defence and NATO, 2019, pp. xvii–xviii, 1-1–1-5).

Ethics also affects method. It requires the analyst to ask whether a proposed intervention is not only conceptually plausible but also operationally and normatively defensible under the stated conditions. This burden is consistent with guidance on behaviourally informed intervention, which treats proportionality, transparency, and safeguards as parts of responsible design (Nuffield Council on Bioethics, 2007, pp. 3–4; OECD, 2019, pp. 144–145; OECD, 2022, pp. 19–20).

Table 1 condenses the procedure into a short audit checklist. Its purpose is not to replace substantive analysis but to force explicit answers to the minimum questions that must be addressed before a cognitively phrased influence claim can be treated as evaluable.

Table 1. Audit Checklist for Operationalising Cognitive-Warfare Claims. This table summarises the operational checklist used to specify outcome indicators, exposure units, functional pathways, alternative explanations, and falsification checks in the Functional–Cognitive Framework.

Audit question	Prompt
1. Outcome indicator	What observable response will be used as the outcome (e.g., rate, latency, persistence, or choice), and what pattern of change would count as meaningful over time?
2. Exposure unit	What exactly counts as an exposure event, what cue class is repeated, and how is exposure distributed across time and the target population?
3. Functional pathway	What is the proposed functional link from exposure to outcome—does the claim concern induction (cue–outcome relations driving acquisition) or maintenance (post-behaviour consequences sustaining responding), and what boundary conditions are assumed?
4. Alternative explanation	What is the most credible confounding pathway (e.g., self-selection, algorithmic amplification, pre-existing tendencies), and how will it be handled in design or interpretation?
5. Falsification check	What observation would weaken the claim—for example, which indicator should change first if the mechanism is operating, and what pattern would be inconsistent with the proposed pathway?

Worked example

The worked example shows how FCF changes the analytical workflow. The example is realistic rather than fully empirical. It illustrates how the framework changes the questions asked, the evidence required, and the conclusions justified (Pamment and Smith, 2022, pp. 7–15).

Consider a claim of the type often made in policy discussion: "Russian narrative X weakened public support for NATO reinforcement in a Baltic population." In conventional cognitive framing, this sentence may quickly lead to a discussion of fear, trust erosion, threat perception, or receptivity to disinformation. Those are plausible avenues. The problem is that the sentence bundles together several claims: that the narrative reached the relevant population, that it produced a change rather than merely coinciding with one, that the change occurred through a specific pathway, and that the analyst now knows what to counter (Claverie and du Cluzel, 2021, pp. 2–4; Pamment and Smith, 2022, pp. 7–15).

Reframing the claim

Under FCF, the first move is to rewrite the claim more precisely. The question is no longer whether narrative X "changed minds" in general. Instead, the paper asks whether exposure to narrative X, under specified conditions, is associated with observable changes in stated support, behavioural indicators, or decision-relevant attitudes, and whether the proposed pathway is persuasive, stress-inducing, confusion-enhancing, or socially reinforced (Hernán and Robins, 2020, pp. 1–22).

This reframing changes the required evidence. The analyst would need to specify the channels through which narrative X reached the population, the time window in which change is claimed, the subgroup or unit to which the claim applies, and the observable indicator on which the conclusion rests.

Mapping a plausible pathway

One plausible pathway might involve repeated pairing between NATO reinforcement and images of escalation, local insecurity, or loss of control. Another might involve repeated social reinforcement within online communities, rewarding expressions of scepticism, and punishing contrary views. The third might not be persuasion in the narrow sense at all. It might involve confusion, attention capture, or decision fatigue during periods of high informational noise. In still other cases, the crucial variable may be motivational rather than persuasive: uncertainty, threat, or social exposure may temporarily increase the value of reassurance, conformity, or conflict avoidance. That could change what people publicly say or do without supporting a strong inference about durable belief change (Claverie and du Cluzel, 2021, pp. 2–4; Cooper *et al.*, 2020, pp. 412–418; Staal, 2004, pp. 31–45).

The framework does not require the analyst to prove that every exposed individual followed the same route. It requires the paper to state which route is being proposed, why that route is plausible under the circumstances, and what pattern would be expected in the available indicators if the route mattered (Hernán and Robins, 2020, pp. 1–22).

Assessing the evidence and its limits

Suppose the available evidence shows that support for reinforcement declined modestly during the campaign period, that engagement with the narrative increased in some online spaces, and that elite messaging was inconsistent. That pattern may be important but it does not by itself identify the mechanism. The decline may reflect the campaign, local political dynamics, independent media framing, economic anxieties, or broader war fatigue. The framework therefore does not allow the paper to conclude simply that narrative X caused the decline (Hernán and Robins, 2020, pp. 1–22; Pamment and Smith, 2022, pp. 7–15).

Functional–cognitive framework changes the conclusion as much as the analysis. Instead of stating that narrative X weakened support through mechanism M, the analyst may conclude that the campaign plausibly contributed to an environment in which support declined, especially in subgroups exposed under particular conditions. The evidence, however, does not yet isolate the mechanism or justify an individual-level targeting rule (Shmueli, 2010, pp. 289–310; Yarkoni and Westfall, 2017, pp. 1100–1105).

This bounded conclusion is also more consistent with the current empirical findings on Russian disinformation. Zilinsky *et al.* (2024, pp. 966–970, 981–982) show that endorsement of narratives justifying the invasion is not evenly distributed across the public, but is concentrated among respondents with stronger conspiratorial predispositions. That does not rule out exposure effects, but it cautions against treating platform reach as broad cognitive conversion or assuming that the same mechanism operates across all subgroups.

Practical implications of reframing

This reframing also changes countermeasure planning. A conventional conclusion about "belief change" might suggest more persuasive messaging in response. FCF may instead indicate that the first requirement is better mapping of exposure, clearer identification of the subgroup in which change occurred, or interventions aimed at reducing overload, inconsistency, or uncertainty rather than merely adding new counter-narratives (Pamment and Smith, 2022, pp. 15–24).

Table 2. Conventional Cognitive Framing and FCF-Based Reframing of an Influence Claim. This table compares a conventional cognitive framing with an FCF-based reframing of a plausible influence claim, showing differences in unit of analysis, assumed mechanism, required data, analytical question, and practical implication.

Analytical element	Conventional cognitive framing	FCF-based reframing
Initial claim	Narrative X changed public support for NATO	Narrative X, under conditions A–C, altered exposure patterns, cue salience, and response tendencies relevant to support-related behaviour.
Unit of analysis	Population attitude	Inputs, exposure conditions, behavioural outputs, and inferential links
Assumed mechanism	Belief change/persuasion	Specify associative cues, operant contingencies, stressors, overload, or contextual triggers
Required data	Survey before/after	Exposure data, timing, repetition, subgroup differences, behavioural indicators, and alternative explanations
Key analytical question	Did attitudes change?	What was delivered, to whom, under what conditions, through what observable pathway, and with what measurable output?
Main limitation and practical implication	Mechanism often underspecified; possible influence effect	Mechanism still partly unobserved, but inferential gaps are made explicit; more precise basis for assessment, intervention design, and countermeasure planning

The framework is useful not because it produces a stronger answer but because it produces a more discriminating one. Table 2 provides a side-by-side comparison between conventional cognitive framing and FCF-based reframing.

Implications for military analysis and defence practice

The military relevance of this approach lies in the fact that cognitive warfare is not confined to persuasion. It may also involve overload, confusion, ambiguity, stress exploitation, cohesion erosion, and degraded decision quality under time pressure. This is where OODA-loop disruption is especially important. A force does not need to be convinced in the strong sense for its orientation, coordination, or timing to be impaired. Analysis that focuses only on explicit belief change may therefore miss operationally important pathways (Claverie and du Cluzel, 2021, pp. 2–4; Masakowski and Blatny, 2023, pp. 1–8; Staal, 2004, pp. 31–45).

The framework supports four practical tasks. Firstly, it can improve analytical writing by requiring staff products and research papers to state whether they are describing an observed pattern, inferring a pathway, estimating an effect, or recommending an intervention. Secondly, it can improve collection planning by identifying what type of data would reduce uncertainty. Thirdly, it can improve evaluation by showing when aggregate effects must be treated as bounded findings rather than as proof of mechanism. Finally, it can support defence planning by distinguishing between responses aimed at persuasion, responses aimed at resilience under overload, and responses aimed at organisational decision discipline (Buck, 2019, pp. 150–160; Hernán and Robins, 2020, pp. 1–22; Stone *et al.*, 2022).

These distinctions matter under operational conditions. A communication environment that impairs orientation through contradiction, saturation, and ambiguity may require different countermeasures from one that shifts attitudes through repeated associative pairings. A

population-level drop in support may also call for different action depending on whether the relevant route is narrative persuasion, reputational signalling, fear of escalation, or confusion under conditions of crisis (Claverie and du Cluzel, 2021, pp. 2–4; Staal, 2004, pp. 31–45).

In an operational setting, such differences may appear in command posts, staff cells, or the public-support environments exposed to ISR saturation, contradictory reporting, tempo pressure, fatigue, and uncertainty about source credibility. Under such conditions, the immediate question may not be whether an adversary persuaded a target in the strong sense. It may be whether the information environment degraded orientation, slowed prioritisation, or increased brittleness in collective decision-making. This is closer to the OODA-focused concern highlighted in a recent NATO framing (NATO Science and Technology Organization, 2025, p. 6).

This distinction also gives FCF a defensive use. If the claim concerns observation, the analyst should ask whether the problem is cue flooding, source ambiguity, deception, or missing information. If the claim concerns orientation, the focus shifts to sense-making, shared mental models, and the ability to distinguish relevant from irrelevant information. If the claim concerns decision, the relevant indicators may include delay, premature closure, or inconsistent prioritisation. If the claim concerns action, the concern may be coordination failure, loss of confidence, or breakdown in execution. In each case, the countermeasure differs. Filtering and source validation, decision-support tools, rehearsed decision rules, communication discipline, or training for performance under overload may be relevant but only after the affected function and the evidence for that claim have been specified (NATO Science and Technology Organization, 2025, p. 6; Staal, 2004, pp. 31–45; UK Ministry of Defence and NATO, 2019, pp. xvii–xviii, 1-1–1-5).

A further practical implication concerns training and doctrine. When cognitive-warfare claims are underspecified, doctrine may encourage a false sense of controllability: the idea that once a cognitive variable is named, a lever has been identified. A more disciplined specification encourages the opposite habit. It asks planners and analysts to state what is known about exposure, timing, pathway, context, and observability before translating a diagnosis into a design recommendation (Cornish *et al.*, 2011, pp. 28–31, 40–41; UK Ministry of Defence and NATO, 2019, pp. xvii–xviii, 1-1–1-5).

The framework is therefore not a demand for methodological maximalism in all cases. Military analysis often has to proceed with incomplete evidence. The practical rule is narrower: when evidence is incomplete, wording and recommendations must remain bounded. Strong claims about mechanism or intervention should be reserved for cases in which the design supports them (Hernán and Robins, 2020, pp. 1–22; Yarkoni and Westfall, 2017, pp. 1100–1105).

Bounded inference should not be mistaken for complacency. Gerrits (2019, pp. 3–5, 8–10) argues that disinformation should not be overstated as a security variable, but he also notes that it is disruptive and worsens an already strained international environment by exploiting the existing divisions and concerns. In defence analysis, the implication is to avoid both inflation and dismissal: effect claims should guide collection and countermeasure planning but remain tied to the evidence available.

Conclusions

This paper has argued that the problem in cognitive-warfare analysis is not simply that cognition is difficult to define or measure. The deeper problem is that

cognitive-warfare claims often move too easily across levels of analysis and too quickly from observed pattern to explanatory and operational conclusion. In such circumstances, broad cognitive language can hide uncertainty rather than solve it (Buck, 2019, pp. 148–150; Robinson, 1950, pp. 351–357; Yarkoni and Westfall, 2017, pp. 1100–1105).

Therefore, FCF is presented as an operational specification framework rather than as a replacement theory of cognition. Its contribution is to make the inferential structure of a claim more explicit: what was observed, what pathway is being proposed, what context is required, what indicator is being used, what rival explanations remain, and what type of recommendation can legitimately follow. This contribution is methodological and practical rather than metaphysical (Hernán and Robins, 2020, pp. 1–22).

The framework also has clear limits. It does not guarantee direct access to individual pathways, especially in security environments where exposure and responding are only partly visible. It does not remove the need for stronger designs, better measurement, or careful causal reasoning. Nor does it replace standard tools of evaluation. What it offers is a disciplined way to prevent broad mechanism language and strong recommendation language from outrunning the evidence (Breiman, 2001, pp. 199–231; Hernán and Robins, 2020, pp. 1–22).

Used in this narrower way, FCF can help reframe cognitive-warfare analysis away from broad psychological assertion and towards explicit specification, bounded inference, and more credible practical guidance (Buck, 2019, pp. 150–160; Shmueli, 2010, pp. 289–310).

Three practical recommendations follow. Firstly, analysts should distinguish descriptive, explanatory, predictive, and intervention claims before using operational verbs, such as disrupt, degrade, harden, or compel. Secondly, cognitive-warfare assessments should document the inferential bridge between individual-level mechanism language and the population-level indicators actually available. Thirdly, military and strategic-communication planners should treat overload, confusion, and time pressure as pathways to be specified and measured, not simply assumed, when discussing cognitive effects in operational environments.

Funding

This research received no external funding.

Data Availability Statement

Data sharing is not applicable to this article, as no datasets were generated or analysed during the current study.

Disclosure Statement

No potential conflict of interest was reported by the author. The author read and agreed to the published version of the manuscript.

References

Borsboom, D., Mellenbergh, G.J. and van Heerden, J. (2004) 'The concept of validity', *Psychological Review*, 111(4), pp. 1061–1071. doi: 10.1037/0033-295X.111.4.1061.

Breiman, L. (2001) 'Statistical modeling: The two cultures', *Statistical Science*, 16(3), pp. 199–231. doi: 10.1214/ss/1009213726.

Buck, G. (2019) 'The elephant in the room: Measurement of effect', *Defence Strategic Communications*, 7, pp. 147–160. doi: 10.30966/2018.RIGA.7.6.

Claverie, B. and du Cluzel, F. (2021) *The cognitive warfare concept*. NATO ACT Innovation Hub. Available at: https://innovationhub-act.org/wp-content/uploads/2023/12/CW-article-Claverie-du-Cluzel-final_0.pdf (Accessed: 26 June 2026).

Cooper, J.O., Heron, T.E. and Heward, W.L. (2020) *Applied behavior analysis*. Global 3rd edn. Harlow: Pearson Education.

Corman, S.R. and Dooley, K.J. (2008) *Strategic communication on a rugged landscape: Principles for finding the right message*. Report No. 0801. Tempe, AZ: Consortium for Strategic Communication, Arizona State University. Available at: https://csc.asu.edu/sites/g/files/litvpz846/files/white-papers/121.pdf (Accessed: 26 June 2026).

Cornish, P., Lindley-French, J. and Yorke, C. (2011) *Strategic communications and national strategy*. London: Chatham House.

Cronbach, L.J. and Meehl, P.E. (1955) 'Construct validity in psychological tests', *Psychological Bulletin*, 52(4), pp. 281–302. doi: 10.1037/h0040957.

Curran, P.J. and Bauer, D.J. (2011) 'The disaggregation of within-person and between-person effects in longitudinal models of change', *Annual Review of Psychology*, 62, pp. 583–619. doi: 10.1146/annurev.psych.093008.100356.

De Houwer, J. (2011) 'Why the cognitive approach in psychology would profit from a functional approach and vice versa', *Perspectives on Psychological Science*, 6(2), pp. 202–209. doi: 10.1177/1745691611400238.

De Houwer, J., Gawronski, B. and Barnes-Holmes, D. (2013) 'A functional–cognitive framework for attitude research', *European Review of Social Psychology*, 24(1), pp. 252–287. doi: 10.1080/10463283.2014.892320.

Druckman, J.N. and McGrath, M.C. (2019) 'The evidence for motivated reasoning in climate change preference formation', *Nature Climate Change*, 9(2), pp. 111–119. doi: 10.1038/s41558-018-0360-1.

Firebaugh, G. (2015) 'Ecological fallacy, statistics of', in Wright, J.D. (ed.) *International encyclopedia of the social & behavioral sciences*. 2nd edn. Oxford: Elsevier, pp. 865–867.

Flake, J.K. and Fried, E.I. (2020) 'Measurement schmeasurement: Questionable measurement practices and how to avoid them', *Advances in Methods and Practices in Psychological Science*, 3(4), pp. 456–465. doi: 10.1177/2515245920952393.

Gerrits, A.W.M. (2019) 'Disinformation in international relations: How important is it?', *Security and Human Rights*, 29(1–4), pp. 3–23. doi: 10.1163/18750230-02901007.

Hamaker, E.L. (2012) 'Why researchers should think "within-person": A paradigmatic rationale', in Mehl, M.R. and Conner, T.S. (eds.) *Handbook of research methods for studying daily life*. New York, NY: Guilford Press, pp. 43–61.

Hernán, M.A. and Robins, J.M. (2020) *Causal inference: What if*. Boca Raton, FL: Chapman & Hall.

Howard, P.N., Lin, F. and Tuzov, V. (2023) 'Computational propaganda: Concepts, methods, and challenges', *Communication and the Public*, 8(2), pp. 47–53. doi: 10.1177/20570473231185996.

Hughes, S., De Houwer, J. and Perugini, M. (2016) 'The functional–cognitive framework for psychological research: Controversies and resolutions', *International Journal of Psychology*, 51(1), pp. 4–14. doi: 10.1002/ijop.12239.

King, G. (1997) *A solution to the ecological inference problem: reconstructing individual behavior from aggregate data*. Princeton, NJ: Princeton University Press.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Obermeyer, Z. (2015) 'Prediction policy problems', *American Economic Review: Papers & Proceedings*, 105(5), pp. 491–495. doi: 10.1257/aer.p20151023.

Macnamara, J. and Gregory, A. (2018) 'Expanding evaluation to progress strategic communication: Beyond message tracking to open listening', *International Journal of Strategic Communication*, 12(4), pp. 469–486. doi: 10.1080/1553118X.2018.1450255.

Masakowski, Y.R. and Blatny, J.M. (eds.) (2023) *Mitigating and responding to cognitive warfare*. STO Technical Report TR-HFM-ET-356. Brussels: NATO Science and Technology Organization.

Michael, J. (1993) 'Establishing operations', *The Behavior Analyst*, 16(2), pp. 191–206. doi: 10.1007/BF03392623.

Molenaar, P.C.M. (2004) 'A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever', *Measurement: Interdisciplinary Research and Perspectives*, 2(4), pp. 201–218. doi: 10.1207/s15366359mea0204_1.

Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., et al. (2015) 'Process evaluation of complex interventions: Medical Research Council guidance', *British Medical Journal*, 350, p. h1258. doi: 10.1136/bmj.h1258.

Mustafa, H., Luczak-Roesch, M. and Johnstone, D. (2025) 'Conceptualizing the evolving nature of computational propaganda: A systematic literature review', *Annals of the International Communication Association*, 49(1), pp. 45–60. doi: 10.1093/anncom/wlaf001.

NATO Science and Technology Organization (2025) *Cognitive warfare: NATO chief scientist research report*. Brussels: NATO Science and Technology Organization.

NATO Standardization Office (2023) *Allied joint doctrine for operations (AJP-3)*, Edn C, Version 1. Brussels: NATO Standardization Office.

Nuffield Council on Bioethics (2007) *Public health: ethical issues*. Available at: https://www.nuffieldbioethics.org/wp-content/uploads/Public-Health-short-guide-1.pdf (Accessed: 26 June 2026).

Organization for Economic Cooperation and Development (OECD) (2019) *Tools and ethics for applied behavioural insights: The basic toolkit*. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/06/tools-and-ethics-for-applied-behavioural-insights-the-basic-toolkit_bbbaaa7a/9ea76a8f-en.pdf (Accessed: 26 June 2026).

Organization for Economic Cooperation and Development (OECD) (2022) *Good practice principles for ethical behavioural science in public policy*. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/10/good-practice-principles-for-ethical-behavioural-science-in-public-policy_8be8043a/e19a9be9-en.pdf (Accessed: 26 June 2026).

Pamment, J. and Smith, V. (2022) *Attributing information influence operations: identifying those responsible for malicious behaviour online*. Riga: NATO Strategic Communications Centre of Excellence.

Rescorla, R.A. and Wagner, A.R. (1972) 'A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement', in Black, A.H. and Prokasy, W.F. (eds.) *Classical conditioning II: current research and theory*. New York, NY: Appleton-Century-Crofts, pp. 64–99.

Robinson, W.S. (1950) 'Ecological correlations and the behavior of individuals', *American Sociological Review*, 15(3), pp. 351–357. doi: 10.2307/2087176.

Shmueli, G. (2010) 'To explain or to predict?', *Statistical Science*, 25(3), pp. 289–310. doi: 10.1214/10-STS330.

Staal, M.A. (2004) *Stress, cognition, and human performance: a literature review and conceptual framework*. NASA technical memorandum 212824. Moffett Field, CA: Ames Research Center.

Stone, J.A., Flanders, K.J. and Can, S.H. (2022) 'Strategic communication? Measurement and evaluation of Twitter use among municipal governments', *Government Information Quarterly*, 39(4), Article 101755. doi: 10.1016/j.giq.2022.101755.

Strauss, M.E. and Smith, G.T. (2009) 'Construct validity: advances in theory and methodology', *Annual Review of Clinical Psychology*, 5, pp. 1–25. doi: 10.1146/annurev.clinpsy.032408.153639.

UK Ministry of Defence and NATO (2019) *Allied joint doctrine for strategic communications (AJP-10)*, Edn A, Version 1. London: Ministry of Defence.

W.K. Kellogg Foundation (2004) *Logic model development guide*. Battle Creek, MI: W.K. Kellogg Foundation.

Yarkoni, T. and Westfall, J. (2017) 'Choosing prediction over explanation in psychology: Lessons from machine learning', *Perspectives on Psychological Science*, 12(6), pp. 1100–1122. doi: 10.1177/1745691617693393.

Zilinsky, J., Theocharis, Y., Pradel, F., Tulin, M., de Vreese, C., Aalberg, T., *et al.* (2024) 'Justifying an invasion: When is disinformation successful?', *Political Communication*, 41(6), pp. 965–986. doi: 10.1080/10584609.2024.2352483.