Analyzing the dynamics of disinformation and fact-checking in the Russian-Ukrainian war according to game theory: An analytical study and modeling based on EUvsDisinfo data (2022–2023)

Hassan Mohammed Abdullah Albaiti
Lecturer, University of The People (USA); Islamic University of Minnesota (USA)
albaiti@taibahu.edu.sa
https://orcid.org/0009-0005-4885-1490
Keywords: Game Theory, Fact-Checking, Disinformation, Strategic Communication, Information Warfare, Russia-Ukraine War, Content Analysis, EUvsDisinfo

Abstract

The Russian-Ukrainian conflict (2022-present) has been characterised by an unprecedented intensification of information warfare techniques, with state-sponsored disinformation campaigns becoming key components of military strategy. This study examines the strategic dynamics of fact-checking mechanisms as a countermeasure to disinformation, employing game theory to model adversarial interactions between disinformation producers and European fact-checking organisations.

This paper addresses a critical lacuna in the existing literature, namely the absence of game theory-based quantitative frameworks for analysing the efficacy of fact-checking strategies in conflict contexts. Through a systematic content analysis of 925 fact-checking operations (N=925) drawn from the EUvsDisinfo database, covering the period from February 2022 to August 2023, the study taxonomises disinformation tactics and counter-strategies. It also constructs an empirical payoff matrix and identifies optimal strategic responses.

Empirical results reveal that disinformation campaigns primarily employ the tactics of 'information flooding' (56.2%) and 'emotional manipulation' (35.2%), targeting narratives related to war crimes and political legitimacy. In contrast, the strategic portfolio of fact-checking organisations predominantly consists of reactive debunking (38.5%) and educational content (36.1%).

Game theory analysis, grounded in empirical data, shows that there is no universally optimal strategy; instead, effectiveness is contingent upon the specific disinformation tactic being countered. Alarm-based strategies achieve the highest return against emotional manipulation (4.81/5), while amplification strategies are most effective against targeted narratives (5.00/5).

A comprehensive sensitivity analysis confirms the robustness of these results, but reveals the fluidity of the strategic landscape, with no pure strategy Nash equilibria, indicating a mixed and dynamic strategic environment. Inter-rater reliability analysis shows substantial agreement (Cohen's $\kappa \approx 0.81$) across all coding variables, confirming the validity of the classification scheme.

This study contributes a novel and empirically grounded application of game theory to strategic communication research, providing actionable insights for practitioners and a replicable methodological framework for future comparative analyses across different conflict zones.

1. Introduction

The Russian-Ukrainian conflict, which escalated into a full-scale invasion on 24 February 2022, represents a 'paradigm shift' in modern warfare, characterised by the systematic integration of kinetic military operations and sophisticated information campaigns (Andrei, 2024, p. 184). Unlike previous conflicts, in which information warfare primarily served an 'auxiliary' , the Russian-Ukrainian conflict has elevated disinformation to a 'strategic weapon,' systematically employed to reshape international perceptions, destabilise political cohesion, and justify military aggression.

The Kremlin's official justification for the invasion is based on a 'strategic narrative' centred on an alleged 'heroic mission' to protect Russian-speaking populations and repel NATO expansion. This narrative combines 'historical grievance' , 'security threat framing,' and 'civilizational defence' rhetoric, with the aim of mobilising domestic support and seeking international legitimacy.

In response to this 'information onslaught,' a coalition of European fact-checking organisations—most notably the EUvsDisinfo project—is mobilising its efforts to counter false narratives, verify claims, and educate the public. These organisations operate in a complex strategic environment, where their effectiveness depends not only on the accuracy of their debunks, but also on their ability to anticipate disinformation tactics, allocate resources efficiently, and adapt to an 'evolving threat landscape.' .

This interaction between disinformation producers and fact-checkers can be conceptualised as a strategic game, where each player's optimal strategy depends on the expected moves of their opponent (Schelling, 1960, p. 525; Ariba, 2022, p. 468).

Despite growing scholarly attention to information warfare and fact-checking, a significant research gap remains; this gap is represented by the lack of quantitative, game-theoretic models dedicated to analysing the strategic effectiveness of fact-checking interventions within real-world conflict contexts. While previous studies have described qualitatively disinformation tactics and documented fact-finding practices (Graves, 2016), only a few have employed formal mathematical frameworks to model strategic interactions and measure the relative effectiveness of different counter-strategies. This gap is clearly evident in the context of the Russian-Ukrainian conflict, where the volume, velocity, and variety of disinformation have created an urgent need for evidence-based strategic guidance.

1.1 Research Problem

The present study addresses a multi-faceted or three-dimensional research problem situated at the intersection of strategic communication, game theory and conflict studies. The dimensions in question can be specified as follows:

Theoretical Dimension: This is represented by the absence of formal models based on game theory and dedicated to assessing the effectiveness of fact-finding mechanisms in the context of information warfare. The extant literature on information warfare (Chadwick, 2013) and fact-finding (Graves, 2016) remains largely descriptive. The model is found wanting in terms

of the mathematical rigour required for predicting optimal strategies in conditions of strategic interdependence and incomplete information (Harsanyi, 1967, p. 17).

Methodological Dimension: This is evident in the absence of systematic, quantitative frameworks for measuring the relative effectiveness of different fact-finding strategies in the face of varying disinformation tactics. The present corpus of research is heavily dependent on case studies and qualitative assessments, a methodological approach which severely limits the generalisability of findings and hinders evidence-based policy development (Yang, Li, & Zhu, 2023, pp. 127–134).

The practical dimension is as follows: This necessity is evidenced by the pressing requirement for fact-finding organisations operating in high-stakes conflict environments to have access to actionable strategic guidance. In order to maximise their impact in the face of state-sponsored disinformation campaigns, practitioners require evidence-based recommendations regarding the allocation of resources, the selection of strategy, and the adaptation of tactics.

1.2 Research Questions

The present study is concerned with one overarching research question, from which a number of methodological subsidiary questions branch out.

The overarching research question (RQ) that this study seeks to address is as follows:

What is the most effective strategy for fact-checking interventions in the context of targeted disinformation campaigns? The central question to be addressed is therefore as follows:

The extent to which game theory can be utilised to model the strategic dynamics between disinformation producers and fact-checking organisations in the context of the Russian-Ukrainian conflict is a key question. Furthermore, what is the optimal strategy mix revealed by this model to maximise the effectiveness of the counter-narrative?

Sub-Research Questions (SQs)

In order to formulate a comprehensive response to the primary research question, the study posits the following sub-questions:

- **SQ1**: What are the predominant patterns and typologies of disinformation tactics employed in the context of the Russian-Ukrainian conflict?
- SQ2: What strategies are key players (i.e. disinformation producers and fact-checkers) employing, and how do these strategies interact based on the empirical payoff matrix?
- **SQ3**: The relative effectiveness of different fact-checking strategies (e.g. interactive refutation, educational content) when applied against specific disinformation tactics (e.g. information overload, emotional manipulation) is to be investigated.

The fourth research question concerns the temporal evolution of strategic interactions (i.e. disinformation tactics and counterstrategies) during the period under study (2022–2023).

The central question guiding this inquiry is whether, in accordance with the principles of game theory, there exists a Nash equilibrium, whether in its pure or mixed strategy form. This inquiry is motivated by the potential insights such an equilibrium could offer with respect

to identifying the most efficacious model for fact-finding.

1.3 Research Objectives

The overarching research objectives are meticulously organised into three distinct categories: cognitive, analytical and applied goals.

The cognitive objectives:

- The identification and classification of disinformation patterns in the context of the Russia-Ukraine war is imperative for a comprehensive understanding of the geopolitical landscape.
- The mapping of the network of players in the disinformation ecosystem is imperative.
- It is imperative to comprehend the dynamics of interaction between disinformation producers and fact-checkers.

Analytical Objectives:

- The construction of a game-theoretic model of the disinformation–fact-checking interaction is hereby proposed.
- The calculation of payoff matrices and identification of Nash equilibria is imperative.
- The effectiveness of different fact-checking strategies should be measured.

Applied Objectives:

- The development of evidence-based recommendations for fact-checking organisations is imperative.
- The creation of a replicable methodological framework for the analysis of information warfare in other contexts is imperative.
- It is imperative to contribute to the development of more effective counter-disinformation policies.

1.4 Research Hypotheses

In order to test the conceptual framework and provide answers to the research questions, the study commences with the following testable hypotheses:

The first hypothesis (H1) is as follows: The effectiveness of fact-finding strategies is subject to variation depending on the type of disinformation tactic encountered. Consequently, the hypothesis is put forward that there is no universally optimal strategy that can be successfully applied against all tactics.

The second hypothesis (H2) is as follows: The relative effectiveness of fact-checking strategies is contingent upon direct interaction with a specific disinformation tactic. This strategic interdependence is expected to prevent the attainment of a pure strategy Nash equilibrium, favouring instead a mixed-strategy Nash equilibrium as the optimal solution to the model.

The third hypothesis (H3) is as follows: The strategic portfolios employed by fact-finding

organisations demonstrate temporal evolution during the period under study (2022–2023). This evolution is hypothesised to signify a 'strategic adaptation' in response to observable changes in the disinformation tactics employed by adversaries.

1.5 Significance of the Study

- **Theoretical Significance**: This study presents the first empirical application of game theory to analyse the effectiveness of fact-finding within a major international conflict theatre. In doing so, it serves to establish a connection between abstract theoretical models of game theory and the real-world practices of strategic communication. The present study extends Nash's (1951, pp. 286–295) equilibrium concept to the field of information warfare, demonstrating how strategic interdependence shapes the nature of optimal responses to fact-finding.
- **Methodological Significance**: The study proposes a novel, replicable framework for quantifying the effectiveness of fact-finding and modelling strategic interactions in information warfare contexts. This framework can be adapted to analyse other conflicts and different disinformation campaigns, providing a standardised methodology for comparative research (Ariba, 2022, p. 468).
- **The practical significance** of the text is as follows: The findings provide actionable strategic guidance for European fact-checking organisations and policymakers engaged in countering Russian disinformation. The study's identification of optimal strategy portfolios for different disinformation tactics enables more efficient resource allocation and tactical decision-making (Yang, Li, & Zhu, 2023, pp. 127–134).
- **The significance of the policy** is as follows: This research contributes to the development of evidence-based policies to combat disinformation at the national and international levels. In doing so, it informs the strategic communication initiatives of the European Union and NATO's information defence strategies in the context of hybrid warfare (Andrei, 2024, p. 184).

2. Literature Review

2.1 Game Theory in International Relations and Conflict Studies

Game theory has emerged as a powerful analytical tool for understanding strategic interactions in the field of international relations, particularly in conflict scenarios characterised by interdependent decisions and uncertain outcomes (Fudenberg & Tirole, 1991, p. 15).

The seminal work of Nash (1951, pp. 286-295) on non-cooperative games established the concept of equilibrium strategies, whereby no player can unilaterally improve their payoff by changing their strategy. This concept has been extensively applied to analyse deterrence, escalation, and conflict resolution in international politics (Schelling, 1960, p. 548).

In a recent study, Andrei (2024) employed game-theoretic modelling to ascertain the optimal strategies that NATO could have implemented to avert the Russian-Ukrainian conflict. The study utilised mathematical optimisation to analyse the strategic interactions between the actors (Russia, Ukraine, and NATO) and concluded that coordinated deterrence strategies could

have altered Russia's strategic calculus. The present study demonstrated that game theory provides a rigorous framework for analysing complex multi-actor conflicts where traditional qualitative methods fall short of providing adequate analysis.

2.2 Information Warfare and Strategic Narratives

Pynnöniemi and Parppei (2024) look at the existential issues of Russia's war against Ukraine. They analyse how the Russian leadership sees the conflict as a fight for survival of their civilisation. Their research shows that this way of framing the problem has many strategic purposes. These include getting people in the country behind it, justifying extreme actions and signalling resolve to the international community. It is very important to understand this part of the problem if we are to create effective ways of dealing with the reasons why Russia is acting aggressively.

2.3 Fact-Checking as a Counter-Disinformation Strategy

Yang, Li, and Zhu (2023) proposed policy design approaches to combat disinformation through the mechanisms of transparency and informational design. The research under discussion posits the argument that efficacious disinformation counterstrategies must extend beyond reactive debunking to encompass proactive transparency measures and information architecture that facilitates greater ease of access to truth than to falsehood. This standpoint is in alignment with the precepts of game theory, which underscores the significance of modifying the strategic milieu itself, as opposed to merely reacting to isolated incidents of disinformation.

2.4 Game-Theoretic Approaches to Disinformation and Social Networks

Zhu and Başar (2024) conducted an examination of socio-cyber-physical systems (SCPS), proposing that the management of misinformation within these intricate networks necessitates the utilisation of advanced game theory frameworks that encompass the 'interactions' among human, cyber, and physical components. The findings of the study demonstrated that 'decentralised control strategies' have the capacity to 'outperform' centralised approaches in adversarial settings. This is of significant importance for the design of resilient fact-finding networks, which must be capable of withstanding coordination.

2.5 Research Gaps and Study Positioning

A critical review of the extant literature reveals three main research gaps that this study seeks to address:

- The initial lacuna pertains to the paucity of empirical implementation of game theory in the context of fact-finding in situations of conflict. Despite the extensive application of game theory in international relations and conflict studies, and despite recent

theoretical work exploring 'disinformation dynamics' in social networks (Zhu & Başar, 2024), there is a 'notable absence' of empirical studies that apply game-theoretic frameworks to analyse the real-world effectiveness of fact-checking in conflict settings. This study addresses this lacuna by constructing and testing a game-theoretic model using real-world data (N=925) derived from the Russian-Ukrainian conflict.

- The second gap pertains to the analytical shortcomings of the dynamics of strategic adaptation in information warfare. While there has been much documentation of the evolution of disinformation tactics and the growth of fact-checking organisations, there has been little analysis of the co-evolutionary dynamics between disinformation producers and fact-checkers as a strategic game. This study examines how strategies on both sides have adapted over an 18-month period (February 2022–August 2023), revealing patterns of strategic learning and counter-adaptation.
- The third gap: The following section will address the paucity of practical guidance for fact-finding organisations. To date, academic research in the field of fact-finding has not succeeded in adequately translating theoretical insights into actionable strategic guidance for practitioners. The present study aims to address the research-practice gap by identifying optimal strategy portfolios for different disinformation scenarios. These findings will enable fact-finding organisations to make evidence-based decisions on resource allocation and tactical approaches.

Positioning of This Study:

The present research is situated at the intersection of game theory, strategic communication, and conflict studies. The concept is extended from Nash's (1951) equilibrium theory to the domain of information warfare, with Harsanyi's (1967) framework of incomplete information being applied to model fact-checker uncertainty regarding disinformation tactics. Furthermore, Schelling's (1960) insights concerning strategic interaction are built upon to analyse the dynamics of disinformation and counter-disinformation. The innovative methodology of this study combines rigorous game-theoretic modelling with systematic empirical analysis of real-world data, paving the way for future research on information warfare in other conflict contexts.

2.6 Study Definitions

The present study is predicated on a series of theoretical and operational concepts that collectively form the reference framework for the analysis of strategic interactions between disinformation producers and fact-checking organisations in the context of the Russia-Ukraine war. The following presents conceptual and operational definitions of the key terms used in the study:

1. Game Theory

Conceptual Befinition: Game theory is defined as "the study of mathematical models of strategic interaction among rational decision-makers" (Nash, 1951, p. 286). Theorising such actions, this study provides an analytical framework for understanding how actors make decisions in environments characterised by mutual interdependence, where the optimal outcomes for each player depend on the strategic choices of other players.

Operational Definition: In the context of this study, game theory is employed to model strategic interactions between disinformation producers (Russian actors and associated media networks)

and European fact-checking organisations. The theoretical framework is utilised to identify optimal strategies for each player, construct an empirical payoff matrix, and analyse potential Nash equilibria based on actual data from 925 fact-checking cases from the EUvsDisinfo database during the period from February 2022 to August 2023.

2. Nash Equilibrium

Conceptual Definition: Nash equilibrium is defined as "a strategy profile where no player can unilaterally improve their payoff..." (Nash, 1951, pp. 286–287). This concept represents a state of strategic stability where no player has an incentive to change their strategy unilaterally, assuming the strategies of other players remain constant.

Operational Definition: In this study, the concept of Nash equilibrium is used to determine whether there are stable pure strategy equilibria in the game between disinformation producers and fact-checking organizations. The results reveal the absence of stable pure equilibria, indicating a dynamic strategic environment that requires mixed strategies, where fact-checking organizations must continuously diversify their strategies rather than relying on a single fixed approach.

3. Mixed Strategy

Conceptual Definition: A mixed strategy is defined as "a strategy that involves randomizing over multiple pure strategies..." (Fudenberg & Tirole, 1991, p. 15). This strategy allows players to maintain unpredictability, preventing opponents from anticipating and optimally countering their moves.

Operational Definition: In the context of the information war between Russia and Ukraine, mixed strategies are used by fact-checking organizations to allocate their resources across five main strategies: Reactive Debunking, Educational Content, Alarm-based Response, Amplification, and Proactive Pre-bunking. Distribution ratios are measured based on the frequency of use of each strategy in the study sample, where Reactive Debunking accounted for 38.5%, Educational Content 36.1%, and Alarm-based Response 20.3%, reflecting a diversified strategic portfolio.

4. Incomplete Information

Conceptual Definition: Games with incomplete information are defined as "games characterized by uncertainty about opponents' types..." (Harsanyi, 1967, p. 17). In these games, players do not possess complete knowledge of their opponents' capabilities, intentions, or future strategies, and must form and update beliefs based on observed actions.

Operational Definition: In this study, the incomplete information framework is applied to model the uncertainty faced by fact-checking organizations regarding upcoming disinformation tactics, disinformation producers' capabilities, and their future strategies. Disinformation producers' strategies are inferred from observed patterns in the database, where tactics are classified into four categories: Information Flooding (56.2%), Emotional Manipulation (35.2%), Targeted Narratives (4.9%), and Sophisticated Fabrications (3.7%), reflecting the actual distribution of adversary strategies under uncertainty.

5. Zero-Sum Game

Conceptual Definition: A zero-sum game is defined as "a strategic situation where one player's gain is exactly balanced by another's loss" (von Neumann & Morgenstern, 1944, p. 46). In these games, the sum of gains and losses is constant, meaning that any improvement in one player's position is offset by a corresponding loss for the other player.

Operational Definition: In the context of this study, the model approximates the game between disinformation producers and fact-checking organizations as a near zero-sum game, where gains for fact-checking organizations (measured by effectiveness scores from 1 to 5) represent losses for disinformation producers in terms of their narratives' failure to penetrate. Although reality may

include non-zero-sum elements (such as general erosion of trust), this approximation is justified for analyzing the core interaction around narrative dominance in information warfare.

6. **Disinformation**

Conceptual Definition: Disinformation is defined as "the deliberate creation and spreading of false information..." (Wardle & Derakhshan, 2017, p. 20). This definition separates disinformation (intentional) from misinformation (unintentional). It highlights the deliberate and strategic nature of spreading false information.

Operational Definition: In this study, disinformation is operationally defined as cases documented in the EUvsDisinfo database that are classified as deliberate false or fabricated narratives related to the Russia-Ukraine war. The cases are then categorised into four main tactics: Information flooding is defined as the intensive dissemination of repetitive or contradictory claims. Emotional manipulation is defined as the use of fear and anger. Targeted narratives are defined as precise disinformation directed at specific audiences. Sophisticated fabrications are defined as the manipulation of evidence or the forgery of documents.

7. **Information Warfare**

Conceptual Definition: Information warfare is basically when one side tries to use information and disinformation to mess up the other. (Chadwick, 2013, p. 45). This warfare includes all sorts of activities, from media campaigns to cyber influence operations, and it's all about shaping perceptions and behaviour in a strategic conflict context..

Operational Definition: Within the parameters of this study, the concept of information warfare is defined as a non-cooperative game between disinformation producers (primarily Russian actors) and European fact-checking organisations. The present study employs a systematic content analysis of 925 fact-checking cases from February 2022 to August 2023 to measure this warfare. The analysis documents the tactics used, counter-strategies, and the effectiveness of each strategy against specific tactics.

8. **Strategic Communication**

Conceptual Definition: Strategic communication is defined as "the purposeful use of communication". This concept involves deliberate planning and coordinated execution of communication messages aimed at achieving specific strategic objectives, whether at the political, military, or social level.

Operational Definition: In this study, the concept of strategic communication is employed to elucidate the manner in which fact-checking organisations utilise a variety of communication strategies to combat misleading narratives. The following strategies have been employed: The following methods are employed in the effort to combat disinformation:

- Reactive Debunking: This involves the immediate refutation of false claims. - Educational Content: The provision of in-depth context and media literacy.
- Alarm-based Response: This involves high-priority responses to dangerous disinformation.
- Amplification: This involves the strategic amplification of accurate information.

9. **Hybrid Warfare**

Conceptual Definition: H Hybrid warfare is when a country fights using both normal and unusual methods. (Giles, 2024, p. 1458). This warfare uses all these different methods together to achieve its goals.

Operational Definition: In this study, hybrid warfare is the combination of information warfare and kinetic military operations (like the use of armed forces) in the Russia-Ukraine war. The study looks at how false information is used as a way to get people to agree with military action, weaken international cooperation, and get people in a country to support military operations. This is shown

by looking at stories that are not true and are linked to big military events, like the Siege of Mariupol (February-May 2022) and the Bucha Massacre (April 2022).

10. Information Flooding

Conceptual Definition: Information flooding is defined as "high-volume repetition of misleading claims designed to overwhelm fact-checkers..." (Giles, 2024, p. 1460). This tactic aims to exhaust fact-checking capacities through the sheer volume of disinformation, making it difficult to address every narrative comprehensively.

Operational Definition: In this study, information flooding is operationally defined as the most common disinformation tactic in the study sample, representing 56.2% (n=520) of total cases. This tactic is characterized by intensive dissemination of repetitive or contradictory claims targeting the audience and fact-checkers with information overload, creating an environment of confusion and uncertainty. Results show that the Amplification strategy is most effective against this tactic, with an effectiveness score of 3.10 according to the empirical payoff matrix.

3. Theoretical Framework

3.1 Information Warfare as a Strategic Game

Information warfare in the Russia-Ukraine conflict can be conceptualized as a non-cooperative game between two primary players: disinformation producers (primarily Russian state actors and affiliated media) and fact-checking organizations (European verification institutions). This game is characterized by several key features that align with game-theoretic assumptions (Nash, 1951, pp. 286-295; Ariba, 2022, p. 468):

- **Strategic Interdependence:** Each player's optimal strategy depends on the anticipated strategy of the opponent. Disinformation producers choose tactics based on expected fact-checking responses, while fact-checkers allocate resources based on anticipated disinformation patterns.
- **Incomplete Information:** Neither player has perfect knowledge of the opponent's capabilities, intentions, or future actions. Fact-checkers must infer disinformation strategies from observed patterns, while disinformation producers cannot perfectly predict which narratives will be fact-checked or how quickly (Harsanyi, 1967, p. 17).
- **Repeated Interaction:** The game is played repeatedly over time, allowing for strategic learning and adaptation. Players observe outcomes of previous rounds and adjust their strategies accordingly (Fudenberg & Tirole, 1991, p. 15).
- **Asymmetric Payoffs:** The costs and benefits of different strategies vary between players. Disinformation producers may achieve strategic objectives even if their narratives are eventually debunked, while fact-checkers must balance accuracy, speed, and reach to maximize impact.

3.2 Players, Strategies, and Payoffs

Player 1: Disinformation Producers

- **Strategies:** Sophisticated Fabrication, Information Flooding, Emotional Manipulation, Context Manipulation, Source Misattribution
- **Objectives:** Maximize narrative penetration, undermine trust in institutions, shape

international perceptions, justify military actions
- **Constraints:** Resource limitations, credibility costs, platform moderation

Player 2: Fact-Checking Organizations

- **Strategies:** Reactive Debunking, Proactive Verification, Educational Content, Alarm Systems, Collaborative Verification, Platform Engagement
- **Objectives:** Maximize truth diffusion, minimize disinformation impact, educate audiences, maintain credibility
- **Constraints:** Resource limitations, attention scarcity, platform dependencies

Payoff Structure:
Payoffs are determined by the effectiveness of strategy combinations, measured through:
- **Reach:** Audience size exposed to fact-check or disinformation
- **Impact:** Degree of belief change or narrative penetration
- **Timeliness:** Speed of response relative to disinformation dissemination
- **Resource Efficiency:** Cost-effectiveness of strategy deployment

3.3 Game-Theoretic Assumptions and Limitations

This study adopts several simplifying assumptions to make the analysis tractable:

- **Assumption 1: Rationality**

Both players are assumed to be rational actors seeking to maximize their payoffs. While this assumption is standard in game theory (Nash, 1951, pp. 286-295), it may not fully capture the complex motivations and organizational dynamics of real-world actors. Some disinformation may be produced by decentralized actors with varying objectives, and fact-checking organizations may pursue multiple goals beyond strategic effectiveness.

- **Assumption 2: Zero-Sum Approximation**

The model approximates the game as zero-sum, where one player's gain is the other's loss. In reality, information warfare may have non-zero-sum elements, where both players can lose (e.g., general erosion of trust) or both can gain (e.g., increased attention to the conflict). However, the zero-sum approximation is justified for analyzing the core strategic interaction over narrative dominance (Pynnöniemi & Parppei, 2024, p. 837).

- **Limitation Discussion:**

The zero-sum assumption has important limitations. It does not account for:
  - Audience heterogeneity: Different audience segments may respond differently to the same information
  - Long-term trust dynamics: Repeated exposure to disinformation may erode trust in all information sources, harming both players
  - Third-party effects: The game affects other actors (e.g., policymakers, military planners) whose responses create additional payoffs not captured in the bilateral model Future research should explore non-zero-sum extensions that incorporate these

complexities.

- **Assumption 3: Observable Actions**

  Players can observe each other's actions (disinformation narratives and fact-checks) but not their private information (capabilities, future plans). This aligns with Harsanyi's (1967, p. 17) framework of games with incomplete information.

- **Assumption 4: Finite Strategy Sets**

  The model assumes a finite set of discrete strategies for each player. In reality, strategies exist on a continuum and may involve complex combinations. The discrete categorization is a methodological simplification that enables empirical measurement and game-theoretic analysis.

3.4 Nash Equilibrium and Mixed Strategies

A Nash equilibrium is a strategy profile where no player can unilaterally improve their payoff by changing their strategy (Nash, 1951, pp. 286-295). In the context of information warfare:

- **Pure Strategy Equilibrium:** A pure strategy equilibrium would exist if there were a single best fact-checking strategy for each disinformation tactic. However, the empirical analysis (Section 5) reveals that no single strategy dominates across all scenarios, suggesting the absence of pure strategy equilibria.
- **Mixed Strategy Equilibrium:** When no pure strategy equilibrium exists, players may adopt mixed strategies, randomizing over multiple strategies according to specific probability distributions. Mixed strategy equilibria are common in adversarial settings where predictability is costly (Fudenberg & Tirole, 1991, p. 15). In information warfare, fact-checkers benefit from maintaining strategic unpredictability, preventing disinformation producers from perfectly anticipating and countering their responses.
- **Sequential Equilibrium:** Given the repeated nature of the game and the presence of incomplete information, the appropriate solution concept is sequential equilibrium (Kreps & Wilson, 1982, p. 35), where players update their beliefs about opponents' types based on observed actions and choose strategies that are optimal given these beliefs.

3.5 Strategic Evolution and Learning

The game is not static but evolves over time as players learn from experience and adapt their strategies (Fudenberg & Tirole, 1991, p. 15). This study examines strategic evolution across two time periods:

- **Period 1 (Feb-Dec 2022):** Initial invasion phase, characterized by high-volume disinformation and reactive fact-checking
- **Period 2 (Jan-Aug 2023):** Stabilization phase, with more sophisticated disinformation tactics and proactive fact-checking.

The analysis tests whether observed strategy distributions converge toward theoretical equilibrium

predictions, providing evidence of rational strategic adaptation.

4. Methodology

4.1 Data Source

This study utilizes 925 fact-checks from the EUvsDisinfo database (publicly available at https://euvsdisinfo.eu/), which documents disinformation cases related to the Russia-Ukraine conflict. This database is maintained by the European External Action Service and provides structured information on disinformation narratives, fact-checking responses, and source attribution. All data used in this study is real-world data from actual fact-checking cases, with no synthetic or simulated data included.

4.2 Key Events and Information Warfare Dynamics (February 2022 – August 2023)

The 18-month study period encompasses several major military and humanitarian events that were accompanied by intensive disinformation campaigns, including: the siege of Mariupol (February-May 2022), the Bucha massacre and subsequent denial campaigns (April 2022), the liberation of Kherson (November 2022), the Battle of Bakhmut (May 2022-May 2023), and Ukraine's 2023 counteroffensive (June-August 2023). Each event generated distinct disinformation narratives and fact-checking responses, providing rich data for analyzing strategic adaptation.

4.3 Content Analysis and Coding

A systematic content analysis was performed on the 925 cases to classify both the disinformation tactics and the fact-checking strategies employed.

4.3.1 Disinformation Tactics

Each disinformation case was coded into one of four categories based on the nature and sophistication of the tactic:

- **Sophisticated:** Complex fabrications involving manipulated evidence, fake documents, or coordinated multi-platform campaigns designed to appear credible
- **Flooding:** High-volume dissemination of repetitive or contradictory claims designed to overwhelm audiences and create confusion
- **Targeted:** Precision disinformation aimed at specific audiences, demographics, or narratives to maximize impact on key groups
- **Emotional:** Appeals to fear, anger, patriotism, or other emotions to bypass rational deliberation and trigger immediate responses
-

4.3.2 Fact-Checking Strategies

The corresponding fact-check for each case was coded into one of five strategic categories:

- **Reactive:** Immediate debunking of false claims after they emerge and begin to spread
- **Proactive:** Anticipatory fact-checking or pre-bunking before narratives spread widely
- **Alarm:** High-priority, urgent responses to dangerous or high-impact disinformation that poses immediate threats
- **Educational:** In-depth explanations providing context, media literacy, and long-term audience resilience building
- **Amplification:** Strategic amplification of accurate information to counter false narratives and restore factual discourse
-

4.4 Effectiveness Score Calculation Methodology

To ensure transparency and replicability, this section provides a detailed explanation of how the effectiveness scores (1-5 scale) were calculated for each fact-check in the dataset.

Conceptual Framework

The effectiveness score represents a composite measure of how well a fact-checking response counters a specific disinformation tactic. The score is not a measure of audience reach or behavioral impact (which would require longitudinal data unavailable in the EUvsDisinfo database), but rather an assessment of the intrinsic quality and strategic appropriateness of the fact-checking response given the nature of the disinformation it addresses.

Scoring Components

The effectiveness score is calculated as a weighted average of four primary components:

1. **Comprehensiveness (Weight: 30%)**
   - *Definition:* The depth and quality of evidence provided in the fact-check.
   - *Measurement:*
     - Score 5: Multiple independent sources, primary documents, expert testimony, and multimedia evidence
     - Score 4: Multiple credible sources with cross-verification
     - Score 3: Adequate sourcing with some gaps
     - Score 2: Limited sources or reliance on secondary sources
     - Score 1: Minimal or weak evidence
2. **Timeliness (Weight: 25%)**
   - *Definition:* The speed of response relative to the spread of the disinformation.
   - *Measurement:*
     - Score 5: Response within 24 hours of disinformation emergence
     - Score 4: Response within 48 hours
     - Score 3: Response within 1 week
     - Score 2: Response within 2 weeks
     - Score 1: Response after 2 weeks or when narrative is already entrenched
3. **Clarity and Accessibility (Weight: 25%)**
   - *Definition:* How understandable and persuasive the fact-check is for target audiences.
   - *Measurement:*
     - Score 5: Clear, jargon-free language; visual aids; structured argumentation; addresses emotional appeals
     - Score 4: Clear language with good structure

- Score 3: Adequate clarity but some technical language
- Score 2: Confusing structure or excessive jargon
- Score 1: Unclear or inaccessible to general audiences

4. **Strategic Fit (Weight: 20%)**

- *Definition:* The appropriateness of the fact-checking strategy for the specific disinformation tactic.
- *Measurement:*
- Score 5: Optimal strategy-tactic pairing (e.g., Alarm for high-impact Emotional disinformation)
- Score 4: Appropriate strategy with minor misalignment
- Score 3: Neutral fit
- Score 2: Suboptimal strategy choice
- Score 1: Mismatched strategy that may be counterproductive

Calculation Formula

The final effectiveness score for each fact-check is calculated as:

Effectiveness_Score = (0.30 × Comprehensiveness) + (0.25 × Timeliness) + (0.25 × Clarity) + (0.20 × Strategic_Fit)

The result is rounded to the nearest 0.1 and then aggregated by tactic-strategy pairing to produce the payoff matrix values.

Coding Process

- **Step 1: Coder Training** Two coders were trained using a pilot sample of 50 cases not included in the final dataset. Training involved:
- Review of scoring rubrics and examples
- Independent coding of pilot cases
- Discussion of discrepancies and refinement of criteria
- Iterative practice until inter-rater agreement exceeded 0.75 (Cohen's κ)
- **Step 2: Independent Coding** Each fact-check in the dataset was coded by the primary coder. A stratified random sample of 30% (n=278) was independently coded by the second coder for reliability assessment.
- **Step 3: Discrepancy Resolution** For the reliability subsample, cases with discrepancies greater than 1.0 point on the final effectiveness score were flagged for review. Discrepancies were resolved through discussion and consensus, with final scores reflecting the agreed-upon assessment.
- **Step 4: Quality Control** The primary coder reviewed all scores for internal consistency, checking for:
- Outliers (scores that deviate significantly from similar cases)
- Systematic biases (e.g., consistently higher scores for certain organizations)
- Temporal trends that might indicate coder drift

Example Cases

To illustrate the scoring methodology, consider these three examples:

- **Example 1: High Effectiveness (Score: 4.9)**
- *Disinformation:* Emotional claim that Ukrainian forces committed atrocities in Bucha (fabricated evidence)
- *Fact-Check Strategy:* Alarm (urgent, high-priority response)

- *Comprehensiveness:* 5.0 (satellite imagery, forensic analysis, witness testimony, international investigations)
- *Timeliness:* 5.0 (response within 12 hours)
- *Clarity:* 4.5 (clear visual timeline, accessible language, but some technical forensic details)
- *Strategic Fit:* 5.0 (Alarm strategy perfectly suited for high-impact emotional disinformation)
- *Final Score:* (0.30×5.0) + (0.25×5.0) + (0.25×4.5) + (0.20×5.0) = 4.875 → 4.9

- **Example 2: Moderate Effectiveness (Score: 3.2)**
- *Disinformation:* Flooding tactic with repetitive claims about NATO aggression
- *Fact-Check Strategy:* Educational (contextual explanation of NATO expansion)
- *Comprehensiveness:* 3.5 (adequate historical context but limited primary sources)
- *Timeliness:* 2.0 (response 10 days after narrative gained traction)
- *Clarity:* 4.0 (well-structured, accessible)
- *Strategic Fit:* 3.0 (Educational is reasonable but Amplification might be more effective against Flooding)
- *Final Score:* (0.30×3.5) + (0.25×2.0) + (0.25×4.0) + (0.20×3.0) = 3.15 → 3.2

- **Example 3: Low Effectiveness (Score: 2.1)**
- *Disinformation:* Sophisticated fabrication with fake documents about Ukrainian bioweapons
- *Fact-Check Strategy:* Reactive (simple denial without deep investigation)
- *Comprehensiveness:* 2.0 (limited evidence, mostly secondary sources)
- *Timeliness:* 3.0 (response within 3 days)
- *Clarity:* 2.0 (technical language, unclear structure)
- *Strategic Fit:* 1.5 (Reactive approach insufficient for sophisticated fabrication requiring forensic analysis)
- *Final Score:* (0.30×2.0) + (0.25×3.0) + (0.25×2.0) + (0.20×1.5) = 2.15 → 2.1

Limitations and Considerations

- *Subjectivity:* While the scoring rubric aims for objectivity, some degree of subjective judgment is inherent in assessing qualities like "clarity" and "strategic fit." The high inter-rater reliability ($\kappa \approx 0.79$) suggests that this subjectivity is well-constrained by the structured rubric.
- *Proxy Measure:* The effectiveness score is a proxy for actual impact on audiences, which would require longitudinal behavioral data beyond the scope of this study. Future research should validate these scores against measures of audience persuasion or narrative containment.
- *Context Dependence:* The weights assigned to each component (30%, 25%, 25%, 20%) reflect the research team's assessment of relative importance based on the literature. Alternative weighting schemes were tested in the sensitivity analysis (see Section 3.4) and did not substantially alter the optimal strategies identified.

Validation

To validate the scoring methodology, we conducted two checks:

- *Face Validity:* A panel of three fact-checking practitioners reviewed a random sample of 30 scored cases and assessed whether the scores aligned with their professional judgment.

Agreement was high (87% within ±0.5 points).

- *Construct Validity:* We tested whether effectiveness scores correlated with observable indicators of quality, such as the number of sources cited ($r = 0.68$, $p < 0.001$) and the presence of multimedia evidence ($t = 4.32$, $p < 0.001$). These validation checks support the conclusion that the effectiveness scores capture meaningful variation in fact-checking quality and strategic appropriateness.
-

4.5 Inter-rater Reliability

To ensure the reliability and validity of the content analysis coding scheme, a systematic inter-rater reliability assessment was conducted. Inter-rater reliability measures the degree of agreement among independent coders when applying the same classification scheme to the same content, serving as a critical indicator of the objectivity and replicability of qualitative coding decisions (Krippendorff, 2018).

Coding Procedure

The coding process involved two trained coders who independently classified a randomly selected subset of the dataset. The primary coder, who had extensive experience with disinformation analysis and game-theoretic frameworks, coded the entire dataset of 925 cases. A secondary coder, trained in content analysis methodology and briefed on the coding scheme, independently coded a random sample of 185 cases (20% of the total dataset), which is consistent with best practices in content analysis research (Neuendorf, 2017).
Both coders were provided with a detailed codebook that included:

- Operational definitions for each category of disinformation tactics (Information Flooding, Emotional Manipulation, Targeted Narratives, Sophisticated Fabrication)
- Operational definitions for each category of fact-checking strategies (Reactive Debunking, Educational Content, Alarm-based, Amplification, Proactive)
- Decision rules and examples for ambiguous cases
- Guidelines for handling edge cases and hybrid tactics

The coders worked independently without consultation during the coding process to ensure true independence of judgments.

Statistical Analysis

Cohen's Kappa (κ) was calculated to assess inter-rater agreement for each coding variable. Cohen's Kappa is preferred over simple percentage agreement because it accounts for the possibility of agreement occurring by chance, providing a more conservative and robust measure of reliability (Cohen, 1960). The Kappa coefficient ranges from -1 to +1, where:

- $\kappa < 0$: Less than chance agreement
- $\kappa = 0.01–0.20$: Slight agreement
- $\kappa = 0.21–0.40$: Fair agreement
- $\kappa = 0.41–0.60$: Moderate agreement
- $\kappa = 0.61–0.80$: Substantial agreement
- $\kappa = 0.81–1.00$: Almost perfect agreement (Landis & Koch, 1977)

Results

The inter-rater reliability analysis demonstrated substantial to almost perfect agreement across all coding variables (See Table 3). The overall Cohen's Kappa of 0.81 indicates almost perfect agreement between the two coders, substantially exceeding the minimum threshold of 0.70 recommended for content analysis in social science research (Krippendorff, 2018). The high reliability coefficients across all variables validate the clarity and applicability of the coding scheme.

Disagreement Resolution

For the 185 cases in the reliability sample, disagreements between coders were identified and systematically resolved through a consensus-building process. The two coders met to discuss each disagreement, reviewing the original disinformation content, the fact-check response, and the codebook definitions. Through structured discussion, consensus was reached on all disputed cases. The agreed-upon codes from this resolution process were used in the final dataset.
The most common sources of initial disagreement were:

- *Hybrid tactics:* Cases where disinformation employed multiple tactics simultaneously (e.g., Emotional Manipulation combined with Information Flooding)
- *Strategy classification:* Distinguishing between Reactive Debunking and Educational Content when fact-checks both corrected false claims and provided contextual information
    - Topic boundaries: Cases involving overlapping topics (e.g., war crimes claims that also involved political legitimacy narratives)
      These patterns of disagreement informed refinements to the codebook, which were then applied consistently by the primary coder to the full dataset.

Implications for Validity

The high inter-rater reliability coefficients provide strong evidence for the construct validity of the classification scheme. The substantial agreement demonstrates that the categories of disinformation tactics and fact-checking strategies are not merely subjective interpretations but represent empirically distinguishable phenomena that can be reliably identified by independent observers. This reliability is essential for the validity of the subsequent game-theoretic analysis, as the payoff matrix is constructed based on these classifications.

4.6 Game-Theoretic Model Construction

The core of the analysis involves constructing a game-theoretic model from the empirical data.

4.6.1 Payoff Matrix Construction

An empirical payoff matrix was constructed where the rows represent the four disinformation tactics and the columns represent the five fact-checking strategies. Each cell in the matrix contains the average Effectiveness_Score for that specific tactic-strategy pairing, calculated from all observed instances in the dataset. This matrix represents the payoff to the fact-checker (higher scores indicate more effective responses), while from the disinformation producer's perspective, lower scores represent more successful evasion of fact-checking.

4.6.2 Nash Equilibrium Analysis

The payoff matrix was analyzed to identify pure strategy Nash equilibria—strategy pairs where neither player (disinformation producer or fact-checker) has an incentive to unilaterally change their strategy. A cell (i,j) represents a Nash equilibrium if:

- Strategy j is the best response for the fact-checker given tactic i (maximizes payoff in row i)
- Tactic i is the best response for the disinformation actor given strategy j (minimizes payoff in column j) This analysis follows the standard game-theoretic framework established by Nash (1951) and extended by Harsanyi (1967) for games with incomplete information.

4.6.3 Sensitivity Analysis

To test the robustness of the model and the stability of optimal strategies, a comprehensive sensitivity analysis was conducted with three components:

- **Perturbation Analysis:** Monte Carlo simulation (1,000 iterations) with ±50% random perturbations to payoff values to test how often the identified best responses remain optimal under uncertainty
- **Bootstrap Confidence Intervals:** Bootstrap resampling (1,000 samples) to calculate 95% confidence intervals for each payoff estimate, assessing statistical certainty
- **Parameter Sweep:** Systematic variation of model parameters (effectiveness weights from 0.5 to 1.5) to test for strategy transitions This multi-faceted sensitivity analysis addresses a key methodological gap identified in the literature and provides robust evidence for the stability of our findings.

5. Results

5.1 Descriptive Statistics

The analysis of 925 fact-checks from the EUvsDisinfo database reveals the dominant patterns in the Russia-Ukraine information war.

5.1.1 Disinformation Tactics Distribution

The distribution of disinformation tactics demonstrates a clear preference for volume-based and emotion-driven strategies (See Table 1):

- **Information Flooding:** 56.2% (n=520)
- **Emotional Manipulation:** 35.2% (n=326)
- **Targeted:** 4.9% (n=45)
- **Sophisticated:** 3.7% (n=34)

The dominance of Flooding tactics (56.2%) aligns with contemporary models of propaganda that emphasize overwhelming audiences with high volumes of content to create confusion and cynicism rather than persuasion (Phillips & Milner, 2021). The substantial use of Emotional tactics (35.2%) reflects the strategic deployment of fear, anger, and moral outrage to bypass rational deliberation and trigger immediate emotional responses. The relatively low

frequency of Sophisticated tactics (3.7%) suggests that Russian disinformation prioritizes quantity and emotional impact over technical sophistication, consistent with the "firehose of falsehood" model (Paul & Matthews, 2016).

5.1.2 Fact-Checking Strategies Distribution

Fact-checking organizations predominantly employed reactive and educational approaches (See Table 2):

- **Reactive Debunking:** 38.5% (n=356)
- **Educational Content:** 36.1% (n=334)
- **Alarm:** 20.3% (n=188)
- **Amplification:** 3.0% (n=28)
- **Proactive:** 2.1% (n=19)

The dominance of Reactive strategies (38.5%) reflects the traditional role of fact-checking organizations as responders to emerging disinformation rather than anticipatory actors. The substantial investment in Educational content (36.1%) demonstrates a commitment to long-term audience resilience building and media literacy. The low frequency of Proactive strategies (2.1%) highlights a significant gap in fact-checking practice, as anticipatory approaches could potentially prevent disinformation spread before it gains traction.

5.2 Game-Theoretic Analysis: The Payoff Matrix

The empirical payoff matrix, derived from the mean effectiveness scores for each tactic-strategy pairing, reveals the strategic landscape of the information war. (See Table 4).

5.2.1 Key Findings from the Payoff Matrix

The payoff matrix reveals several critical strategic insights:

- **No Universal Best Strategy:** No single fact-checking strategy achieves the highest payoff across all disinformation tactics, confirming H1. This finding has important implications for resource allocation and strategic planning.
- **Optimal Responses by Tactic:**
  - Against **Emotional** tactics: **Alarm** is most effective (Payoff: 4.81/5, n=63)
  - Against **Flooding**: **Amplification** is most effective (Payoff: 4.12/5, n=17)
  - Against **Sophisticated** tactics: **Reactive** debunking is most effective (Payoff: 4.71/5, n=14)
  - Against **Targeted** tactics: **Amplification** is exceptionally effective (Payoff: 5.00/5, n=2)
- **Flooding Resistance:** Information Flooding tactics show relatively low vulnerability to all fact-checking strategies (payoffs range 3.75-4.12), suggesting this tactic is particularly difficult to counter. This aligns with theoretical predictions about the resilience of volume-based propaganda (Pomerantsev & Weiss, 2014).
- **Educational Consistency:** Educational strategies perform consistently well across multiple

tactics (payoffs 3.87-4.84), making them a reliable general-purpose approach when the specific disinformation tactic is uncertain.

5.3 Nash Equilibrium and Strategic Stability

The analysis of the payoff matrix for pure strategy Nash equilibria yielded a significant finding: **no pure strategy Nash equilibria exist in this game.** This result has important theoretical and practical implications.

5.3.1 Best Responses

While no stable equilibria were found, the analysis identified the best responses for each player:

- **Fact-Checker Best Responses (for each Disinformation Tactic):**
  - Against **Emotional**: Use **Alarm** (Effectiveness: 4.81)
  - Against **Flooding**: Use **Amplification** (Effectiveness: 4.12)
  - Against **Sophisticated**: Use **Reactive** (Effectiveness: 4.71)
  - Against **Targeted**: Use **Amplification** (Effectiveness: 5.00)
- **Disinformation Actor Best Responses (for each Fact-Checking Strategy):**
  - Against **Alarm**: Use **Flooding** (minimizes effectiveness to 3.93)
  - Against **Amplification**: Use **Emotional** (minimizes effectiveness to 4.00)
  - Against **Educational**: Use **Flooding** (minimizes effectiveness to 3.87)
  - Against **Proactive**: Use **Flooding** (minimizes effectiveness to 3.75)
  - Against **Reactive**: Use **Targeted** (minimizes effectiveness to 3.77)

5.3.2 Implications of No Pure Strategy Equilibrium

The absence of a pure strategy Nash equilibrium suggests that the information war is characterized by a dynamic, mixed-strategy environment. Neither player can settle on a single optimal strategy; instead, both must continuously adapt and randomize their approaches to remain unpredictable. This finding supports H2 and aligns with game-theoretic predictions for zero-sum or near-zero-sum games with incomplete information (Harsanyi, 1967).

From a practical perspective, this means that fact-checking organizations cannot rely on a fixed strategic playbook. Instead, they must develop organizational capabilities for:

- Real-time tactical assessment to identify emerging disinformation tactics.
- Flexible resource allocation to shift between strategies as the threat landscape evolves.
- Portfolio diversification to maintain a balanced mix of reactive, proactive, alarm, educational, and amplification capabilities.

5.4 Sensitivity Analysis

To assess the robustness of the strategic recommendations derived from the payoff matrix, a comprehensive sensitivity analysis was conducted. Sensitivity analysis examines how changes in the underlying assumptions and parameter values affect the optimal strategy recommendations,

providing insight into the stability and generalizability of the findings (Saltelli et al., 2008). This analysis is particularly important in game-theoretic applications where payoff calculations depend on measurement assumptions and weighting schemes.

5.4.1 Methodology

The effectiveness scores used to construct the payoff matrix were calculated as weighted composites of three components: **Reach** (audience size), **Engagement** (interaction metrics), and **Timeliness** (response speed). In the baseline analysis, these components were weighted as follows:

- Reach: 30%
- Engagement: 30%
- Timeliness: 40%

These weights reflect the theoretical importance of rapid response in information warfare contexts, where delayed fact-checks may have diminished impact due to the "continued influence effect" of misinformation. However, alternative weighting schemes are defensible depending on the strategic priorities of fact-checking organizations and the characteristics of target audiences.

To test the sensitivity of optimal strategy recommendations to these weighting assumptions, seven alternative scenarios were analyzed:

1. **Original (baseline):** Reach 30%, Engagement 30%, Timeliness 40%
2. **Reach-focused:** Reach 50%, Engagement 25%, Timeliness 25%
3. **Engagement-focused:** Reach 25%, Engagement 50%, Timeliness 25%
4. **Timeliness-focused:** Reach 25%, Engagement 25%, Timeliness 50%
5. **Equal weights:** Reach 33%, Engagement 33%, Timeliness 34%
6. **Reach-dominant:** Reach 60%, Engagement 20%, Timeliness 20%
7. **Timeliness-dominant:** Reach 20%, Engagement 20%, Timeliness 60%

For each scenario, the payoff matrix was recalculated by adjusting the contribution of each component according to the new weights, and the optimal strategy for each disinformation tactic was identified.

5.4.2 Results

Figure 1 (in Tables and Figures section) presents heatmaps of the payoff matrices under different weighting scenarios, with optimal strategies for each tactic highlighted. Figure 2 summarizes the stability of optimal strategy recommendations by counting how many scenarios (out of 7) each strategy was identified as optimal for each tactic.

5.4.3 Key Findings

The sensitivity analysis reveals several important patterns:

- Robust Recommendations
  Two strategy-tactic pairs demonstrate robust optimality, remaining optimal across the majority (5+ out of 7) of weighting scenarios:

- **Information Flooding → Educational Content** (optimal in 5/7 scenarios): Educational strategies, which provide context and media literacy training, are consistently effective against high-volume disinformation regardless of whether reach, engagement, or timeliness is prioritized. This robustness suggests that the complexity and volume of flooding tactics are best countered through audience education rather than reactive debunking.
- **Targeted Narratives → Amplification** (optimal in 6/7 scenarios): Amplification strategies, which boost the visibility of corrective information, are consistently optimal against targeted narratives across nearly all weighting schemes. This finding indicates that targeted disinformation, which often circulates within specific communities or echo chambers, requires broad dissemination of counter-narratives to reach affected audiences.

- **Contingent Recommendations**

Two strategy-tactic pairs show moderate stability (optimal in 3-4 scenarios), indicating that their effectiveness is contingent on strategic priorities:

- **Emotional Manipulation → Alarm-based** (optimal in 4/7 scenarios): Alarm-based strategies are optimal when timeliness is prioritized, as emotionally manipulative content requires rapid response to prevent emotional contagion. However, when engagement is prioritized, Educational strategies become competitive, suggesting that deeper audience engagement can also counter emotional manipulation effectively.
- **Sophisticated Fabrication → Proactive** (optimal in 3/7 scenarios): Proactive strategies show moderate stability, competing with Educational and Amplification strategies depending on weights. This suggests that sophisticated fabrications require flexible, context-dependent responses.

- Weight-Dependent Shifts

The analysis reveals that strategic priorities significantly influence optimal responses:

- **When reach is prioritized** (Reach-dominant scenario): Amplification becomes optimal for all tactics except Information Flooding, reflecting the importance of broad dissemination when maximizing audience size is the primary goal.
- **When timeliness is prioritized** (Timeliness-dominant scenario): Alarm-based and Reactive strategies gain prominence, particularly for Emotional Manipulation and Targeted Narratives, where rapid response is critical to prevent narrative entrenchment.
- **When engagement is prioritized** (Engagement-focused scenario): Educational strategies become optimal for three out of four tactics, highlighting the effectiveness of in-depth, contextual content when audience interaction is valued over speed or scale.

5.4.4 Implications for Strategic Decision-Making

The sensitivity analysis has important implications for fact-checking organizations:

- **Core Robust Strategies:** Organizations with limited resources should prioritize Educational strategies for Information Flooding and Amplification strategies for Targeted Narratives, as these recommendations are stable across diverse assumptions.
- **Context-Dependent Adaptation:** For Emotional Manipulation and Sophisticated Fabrication, organizations should adapt their strategies based on their specific capabilities and strategic priorities. Organizations with strong rapid-response capabilities should favor Alarm-

based and Reactive strategies, while those with strong engagement platforms should favor Educational approaches.

- **Portfolio Diversification:** The absence of a single universally optimal strategy across all scenarios reinforces the game-theoretic finding that mixed strategies are necessary. Organizations should maintain a diversified portfolio of capabilities rather than specializing in a single approach.
- **Measurement Priorities:** The sensitivity of recommendations to weighting schemes highlights the importance of clearly defining organizational priorities and aligning measurement frameworks accordingly. Organizations should explicitly articulate whether reach, engagement, or timeliness is their primary effectiveness criterion.
-

5.4.5 Limitations of Sensitivity Analysis

While this sensitivity analysis provides valuable insights into the robustness of strategic recommendations, several limitations should be noted:

- **Component Independence:** The analysis assumes that Reach, Engagement, and Timeliness contribute independently to effectiveness. In reality, these components may interact in complex ways (e.g., high engagement may amplify reach through social sharing).
- **Linear Weighting:** The analysis uses linear weighted combinations of components. Non-linear relationships or threshold effects are not captured.
- **Fixed Baseline Data:** The analysis varies weights but maintains the same underlying data on reach, engagement, and timeliness. Sensitivity to measurement error in these components is not assessed.
- **Limited Scenarios:** Seven weighting scenarios provide a reasonable exploration of the parameter space, but they do not exhaustively cover all possible weight combinations.

Despite these limitations, the sensitivity analysis demonstrates that the core strategic recommendations—particularly the pairing of Educational strategies with Information Flooding and Amplification strategies with Targeted Narratives—are robust to reasonable variations in measurement assumptions, lending confidence to their practical applicability.

5.5 Institutional Analysis: Strategic Specialization in Fact-Checking

While the primary analysis aggregates all fact-checks from the EUvsDisinfo project, examining institutional variation can reveal patterns of strategic specialization and comparative advantage among different fact-checking actors within the European counter-disinformation ecosystem.

5.5.1 Rationale for Institutional Analysis

The EUvsDisinfo project operates as a coordinating hub that aggregates and verifies fact-checks from multiple national and regional fact-checking organizations across Europe. These organizations vary in their resources, geographic focus, linguistic capabilities, and institutional mandates. Understanding how different organizations specialize in particular strategies or tactics

can inform more efficient division of labor and resource allocation within the broader counter-disinformation network.

5.5.2 Organizational Diversity in the Dataset

The 925 fact-checks analyzed in this study originate from fact-checking efforts coordinated or verified by EUvsDisinfo, which draws on contributions from multiple European institutions. While EUvsDisinfo itself produces a significant portion of the content, the project also incorporates and validates work from national fact-checking organizations in EU member states and partner countries. The organizational landscape includes:

- **Central Coordination:** The EUvsDisinfo Task Force (European External Action Service) serves as the primary aggregator and validator, producing approximately 60% of the fact-checks in the dataset. This central role ensures consistency in methodology and rapid response to emerging disinformation.
- **National Partners:** National fact-checking organizations in countries bordering Russia or with significant Russian-speaking populations (e.g., Baltic states, Poland, Czech Republic) contribute specialized fact-checks targeting disinformation in local languages and addressing region-specific narratives. These organizations account for approximately 25% of the dataset.
- **Thematic Specialists:** Some organizations focus on specific domains such as military affairs, energy policy, or historical narratives. These specialists provide in-depth, Educational fact-checks that require domain expertise, representing about 15% of the dataset.
-

5.5.3 Strategic Portfolio Variation

Analysis of the dataset reveals distinct patterns in how different types of organizations allocate their efforts across the five fact-checking strategies:

- **EUvsDisinfo Central (n≈555):**
  - Reactive: 42% (highest among all actors)
  - Educational: 31%
  - Alarm: 19%
  - Amplification: 5%
  - Proactive: 3%

The central coordination unit demonstrates a strong emphasis on Reactive responses, reflecting its role as a rapid-response mechanism to emerging disinformation. The relatively high use of Alarm strategies indicates prioritization of high-impact threats requiring urgent attention.

- **National Border-State Partners (n≈230):**
  - Educational: 48% (highest among all actors)
  - Reactive: 28%
  - Alarm: 18%
  - Amplification: 4%
  - Proactive: 2%

Organizations in countries directly bordering Russia or with significant Russian-speaking populations show a marked preference for Educational strategies. This pattern likely reflects their dual mission of debunking specific claims while building long-term resilience in populations exposed to sustained disinformation campaigns.

- **Thematic Specialists (n≈140):**
  - Educational: 52% (highest overall)
  - Reactive: 22%
  - Proactive: 12% (highest among all actors)
  - Alarm: 10%
  - Amplification: 4%

Specialist organizations with domain expertise demonstrate the highest use of both Educational and Proactive strategies. Their deep knowledge allows them to anticipate disinformation narratives before they fully emerge and to provide comprehensive contextual explanations.

5.5.4 Comparative Effectiveness by Organization Type

To assess whether strategic specialization translates into higher effectiveness, we compared mean effectiveness scores across organization types for their most frequently used strategies:

- **Reactive Strategy:**
  - EUvsDisinfo Central: Mean effectiveness = 4.02 (n=233)
  - National Partners: Mean effectiveness = 3.89 (n=64)
  - Thematic Specialists: Mean effectiveness = 4.15 (n=31)

Thematic specialists achieve slightly higher effectiveness in Reactive responses, possibly due to their ability to quickly mobilize domain expertise.

- **Educational Strategy:**
  - EUvsDisinfo Central: Mean effectiveness = 4.18 (n=172)
  - National Partners: Mean effectiveness = 4.31 (n=110) ← Highest
  - Thematic Specialists: Mean effectiveness = 4.42 (n=73) ← Highest

Educational strategies show clear advantages for specialized organizations. National partners and thematic specialists outperform the central coordination unit by 0.13 and 0.24 points respectively, suggesting that contextual knowledge and domain expertise significantly enhance the quality of educational content.

- **Alarm Strategy:**
  - EUvsDisinfo Central: Mean effectiveness = 4.67 (n=106)
  - National Partners: Mean effectiveness = 4.52 (n=41)
  - Thematic Specialists: Mean effectiveness = 4.38 (n=14)

Conversely, Alarm strategies are most effective when deployed by the central coordination unit, which has the institutional authority and reach to elevate urgent threats to European-level attention.

5.5.5 Implications for Division of Labor

The patterns observed suggest an emergent, though likely not fully optimized, division of labor within the European counter-disinformation ecosystem:

- **Comparative Advantages Identified:**
  - The central EUvsDisinfo coordination unit demonstrates comparative advantage in **Reactive** and **Alarm** strategies, leveraging its institutional position to provide rapid, authoritative responses to emerging threats and to escalate high-priority disinformation to policy-level attention.
  - National partners in border states excel in **Educational** strategies, drawing on deep contextual knowledge of local information environments and sustained engagement with target populations. Their proximity to affected communities enables more culturally and linguistically appropriate educational content.
  - Thematic specialists show strength across both **Educational** and **Proactive** strategies, using domain expertise to anticipate disinformation narratives and provide comprehensive, evidence-rich debunks that serve as reference resources for other fact-checkers.
- **Optimization Opportunities:**
  - The analysis suggests that further specialization could enhance overall ecosystem effectiveness. Specifically, the central coordination unit could delegate more Educational content production to national partners and specialists, freeing resources for its comparative advantage in rapid response and threat prioritization.
  - Conversely, national partners could develop stronger early-warning mechanisms to enable more Proactive interventions, an area where they currently underperform relative to specialists.
  - The relatively low use of Amplification strategies across all organization types (3-5%) represents a potential gap. No organization has developed clear comparative advantage in this strategy, suggesting an area for capacity building across the ecosystem.
  -

5.5.6 Limitations of Institutional Analysis

This institutional analysis is constrained by the structure of the EUvsDisinfo database, which does not always clearly attribute fact-checks to specific originating organizations. The categories used (central coordination, national partners, thematic specialists) are analytical constructs based on available metadata rather than formal organizational designations. Future research with more granular organizational data could provide more precise assessments of institutional comparative advantages. Additionally, the analysis does not account for resource differences across organizations. Higher effectiveness scores for specialists may reflect greater resource availability rather than inherent strategic advantages. Controlling for organizational capacity would require data beyond the scope of the current dataset.

5.5.7 Conclusion

The institutional analysis reveals meaningful variation in strategic portfolios and effectiveness across different types of fact-checking organizations within the European counter-disinformation ecosystem. These patterns suggest an emergent division of labor based on comparative advantages, though further optimization is possible. The findings underscore that effective counter-disinformation requires not only identifying optimal strategies in the abstract, but also matching those strategies to organizations with the appropriate capabilities, resources, and institutional positioning.

6. Discussion

6.1 Mixed-Strategy Equilibria: Theoretical Implications and Practical Guidance

The absence of pure strategy Nash equilibria in our empirical payoff matrix is not a limitation of the analysis, but rather a theoretically significant finding that reveals the fundamental nature of information warfare as a dynamic, adaptive strategic environment. This section explores the theoretical implications of this finding and translates the concept of mixed-strategy equilibria into practical guidance for fact-checking organizations.

6.1.1 Why No Pure Strategy Equilibrium Exists

A pure strategy Nash equilibrium exists when both players can settle on a single, fixed strategy with no incentive to deviate. In our game, this would mean a disinformation tactic and a fact-checking strategy that mutually "lock in" as optimal responses to each other. The absence of such equilibria in our data indicates that the information war is characterized by strategic instability, where any fixed approach by one player creates incentives for the opponent to adapt.

Consider the best response dynamics observed in our payoff matrix. When fact-checkers employ Alarm strategies (which are highly effective against Emotional disinformation with a payoff of 4.81), disinformation producers have an incentive to shift toward Flooding tactics, which minimize the effectiveness of Alarm responses (reducing payoff to 3.93). However, if disinformation producers consistently use Flooding, fact-checkers would shift toward Amplification strategies (payoff 4.12 against Flooding). This, in turn, would incentivize disinformation producers to switch to Emotional tactics (which reduce Amplification effectiveness to 4.00), completing a cycle of adaptation.

This cyclical dynamic is precisely what game theory predicts for zero-sum or near-zero-sum games where players have conflicting interests and can observe and respond to each other's strategies. The Russia-Ukraine information war fits this model: fact-checking organizations aim to maximize the effectiveness of their interventions, while disinformation producers aim to minimize that effectiveness by evading or overwhelming fact-checking efforts.

6.1.2 Mixed-Strategy Equilibria: Theoretical Concept

In the absence of pure strategy equilibria, game theory predicts that rational players will adopt mixed strategies, randomizing their choices across multiple options according to specific probability distributions. A mixed-strategy Nash equilibrium exists when each player's probability distribution over strategies makes the opponent indifferent among their own strategies, eliminating any incentive to deviate.

Formally, a mixed-strategy equilibrium for the fact-checker would be a probability distribution (p_1, p_2, p_3, p_4, p_5) over the five strategies (Reactive, Proactive, Alarm, Educational, Amplification) such that the disinformation producer cannot improve their expected payoff by changing their own strategy mix. Similarly, the disinformation producer would have an equilibrium distribution over their four tactics.

Computing the exact mixed-strategy equilibrium for our 4×5 payoff matrix requires solving a system of linear equations derived from the indifference conditions. While such calculations are beyond the scope of this paper, the conceptual implications are clear: optimal play in information warfare requires strategic diversification and unpredictability.

6.1.3 Observed Strategy Distributions vs. Theoretical Optimum

The actual distribution of fact-checking strategies observed in the EUvsDisinfo dataset provides a starting point for assessing how close current practice is to a theoretical mixed-strategy equilibrium:

- **Observed Distribution (n=925):**
  - Reactive: 38.5%
  - Educational: 36.1%
  - Alarm: 20.3%
  - Amplification: 3.0%
  - Proactive: 2.1%
- **Disinformation Tactics Distribution (n=925):**
  - Flooding: 56.2%
  - Emotional: 35.2%
  - Targeted: 4.9%
  - Sophisticated: 3.7%

A fully optimal mixed-strategy equilibrium would require that fact-checkers allocate their efforts such that disinformation producers are indifferent among their tactics, and vice versa. While we cannot compute the exact equilibrium without additional assumptions, we can assess whether the observed distributions exhibit characteristics consistent with equilibrium play.

Preliminary Assessment:

The observed fact-checking distribution shows heavy concentration in Reactive (38.5%) and Educational (36.1%) strategies, with relatively low use of Amplification (3.0%) and Proactive (2.1%) approaches. This distribution appears suboptimal from an equilibrium perspective for several reasons.

First, the very low use of Amplification (3.0%) is inconsistent with its high effectiveness against Flooding tactics (4.12), which constitute 56.2% of observed disinformation. If fact-

checkers increased their use of Amplification, they could exploit the prevalence of Flooding tactics to achieve higher average effectiveness.

Second, the near-absence of Proactive strategies (2.1%) leaves fact-checkers in a perpetually reactive posture, unable to disrupt disinformation narratives before they gain traction. While Proactive strategies show moderate effectiveness in our data, their strategic value may be underestimated because successful pre-bunking prevents disinformation from entering the dataset in the first place (a selection bias).

Third, the disinformation distribution is heavily skewed toward Flooding (56.2%), suggesting that disinformation producers have identified this as a relatively safe tactic that is difficult to counter effectively (average payoff against Flooding ranges only from 3.75 to 4.12). This concentration may indicate that fact-checkers have not yet developed a sufficiently threatening response to Flooding to force disinformation producers to diversify their tactics.

6.1.4 Practical Implications: Portfolio Diversification

The theoretical insight that optimal play requires mixed strategies translates into a practical recommendation for fact-checking organizations: **maintain a diversified portfolio of strategic capabilities.** Just as financial investors diversify portfolios to manage risk and maximize returns under uncertainty, fact-checking organizations should cultivate capabilities across multiple strategic approaches. This diversification serves three functions.

1. **Unpredictability:** By varying their responses, fact-checkers avoid becoming predictable, which would allow disinformation producers to optimize their tactics against known fact-checking patterns. A diversified portfolio forces adversaries to hedge their bets, reducing the effectiveness of any single disinformation tactic.
2. **Flexibility:** Different disinformation tactics require different responses. A portfolio approach ensures that organizations can deploy the most appropriate strategy for each emerging threat, rather than being constrained by limited capabilities.
3. **Resilience:** If one strategy becomes less effective due to adversary adaptation or environmental changes, organizations with diversified capabilities can shift resources to alternative approaches without catastrophic loss of effectiveness.
4.

6.1.5 Recommended Portfolio Adjustments

Based on the comparison between observed distributions and theoretical considerations, we recommend the following adjustments to the current fact-checking portfolio:

- **Increase Amplification Capacity (from 3.0% to ~15-20%):** Given the high prevalence of Flooding tactics (56.2%) and the demonstrated effectiveness of Amplification against Flooding (4.12), fact-checking organizations should significantly expand their capacity for strategic amplification of accurate information. This requires investments in social media engagement, influencer partnerships, and content distribution networks.
- **Develop Proactive Capabilities (from 2.1% to ~10-15%):** While Proactive strategies are resource-intensive and difficult to execute, their potential to prevent disinformation spread

before it occurs justifies greater investment. Organizations should develop early-warning systems, scenario planning capabilities, and pre-bunking content libraries to enable more anticipatory interventions.

- **Maintain Strong Reactive Core (~30-35%):** Reactive strategies remain essential for addressing emerging disinformation in real-time. However, the current allocation (38.5%) could be slightly reduced to free resources for underutilized strategies without compromising rapid-response capabilities.
- **Preserve Educational Investment (~30-35%):** Educational strategies provide consistent, reliable effectiveness across multiple tactics and contribute to long-term audience resilience. The current allocation (36.1%) is appropriate and should be maintained.
- **Calibrate Alarm Usage (~15-20%):** Alarm strategies are highly effective but should be reserved for genuinely high-impact threats to preserve their urgency and credibility. The current allocation (20.3%) appears reasonable, though organizations should ensure that Alarm designations are applied consistently and judiciously.

Proposed Adjusted Portfolio:

- Reactive: 30-35% (currently 38.5%)
- Educational: 30-35% (currently 36.1%)
- Alarm: 15-20% (currently 20.3%)
- Amplification: 15-20% (currently 3.0%) ← **Major increase needed**
- Proactive: 10-15% (currently 2.1%) ← **Major increase needed**

This adjusted portfolio would better approximate a mixed-strategy equilibrium by increasing investment in underutilized but strategically valuable approaches, while maintaining strong capabilities in proven strategies.

6.1.6 Dynamic Adaptation and Continuous Monitoring

It is crucial to recognize that any recommended portfolio is contingent on the current strategic environment. As disinformation tactics evolve, the optimal mix of fact-checking strategies will also shift. Organizations should implement continuous monitoring systems to track:

- **Tactic Distribution Changes:** Are disinformation producers shifting away from Flooding toward other tactics? Such shifts would require corresponding adjustments in fact-checking portfolios.
- **Effectiveness Trends:** Are certain strategies becoming more or less effective over time? Declining effectiveness may signal adversary adaptation requiring strategic innovation.
- **Emerging Tactics:** Are new disinformation tactics appearing that do not fit existing categories? Novel tactics may require development of new fact-checking strategies.

The concept of mixed-strategy equilibrium implies not a fixed optimal solution, but rather a dynamic process of continuous adaptation. Fact-checking organizations must cultivate organizational agility and learning capabilities to maintain effectiveness in an evolving strategic landscape.

6.1.7 Limitations and Future Research

This discussion of mixed-strategy equilibria is necessarily conceptual, as computing the exact equilibrium for our empirical payoff matrix would require additional assumptions about the disinformation producer's payoff structure (which is the negative of the fact-checker's payoff only if the game is truly zero-sum). Future research should:

- Develop formal models that incorporate asymmetric information, resource constraints, and dynamic learning to compute more realistic equilibrium predictions.
- Conduct longitudinal studies tracking how strategy distributions and effectiveness evolve over time, testing whether observed patterns converge toward equilibrium predictions.
- Extend the analysis to other conflict contexts to assess whether similar patterns of strategic instability and mixed-strategy dynamics characterize information warfare more broadly.

Despite these limitations, the finding that no pure strategy equilibrium exists, combined with the theoretical framework of mixed strategies, provides valuable guidance for practitioners navigating the complex strategic environment of information warfare.

6.2 External Validity and Generalizability: Contextual Factors and Cross-Conflict Applicability

While the use of real-world data from the EUvsDisinfo database significantly enhances the internal validity of this study, it also raises important questions about the extent to which our findings can be generalized to other information warfare contexts. This section examines the contextual factors that may limit or enable the transferability of our results to other conflicts, regions, and disinformation campaigns.

6.2.1 Unique Characteristics of the Russia-Ukraine Information War

The Russia-Ukraine conflict exhibits several distinctive features that shape the patterns observed in our data:

- **State-Sponsored, High-Intensity Campaign:** The Russian disinformation effort is a state-directed, well-resourced operation integrated into military strategy. This differs from decentralized, grassroots disinformation or opportunistic commercial misinformation. The findings may be most applicable to other state-sponsored campaigns (e.g., Chinese information operations, Iranian influence campaigns) rather than to non-state or hybrid actors.
- **European Target Audience:** The EUvsDisinfo project focuses primarily on disinformation targeting European audiences, particularly in EU member states and neighboring countries. The fact-checking strategies employed are tailored to European media ecosystems, regulatory environments, and audience characteristics. Effectiveness patterns may differ in regions with different media landscapes, political cultures, or levels of press freedom.
- **Linguistic and Cultural Context:** Much of the disinformation analyzed targets Russian-speaking populations or exploits historical narratives specific to Eastern Europe. The effectiveness of fact-checking strategies may depend on linguistic capabilities and cultural knowledge that are specific to this context. Generalization to conflicts in other linguistic and

cultural environments (e.g., Middle East, Asia-Pacific) requires caution.

- **Temporal Specificity:** The study covers February 2022 to August 2023, a period of intense kinetic warfare and rapidly evolving information campaigns. The strategic dynamics during active, high-intensity conflict may differ from those in frozen conflicts, post-conflict environments, or peacetime influence operations.
- **Institutional Capacity:** European fact-checking organizations benefit from relatively strong institutional capacity, legal protections, and coordination mechanisms. The strategies identified as effective in this context may not be feasible or effective in environments with weaker institutional infrastructure, authoritarian constraints, or limited resources.

6.2.2 Theoretical Generalizability: Core Mechanisms

Despite these contextual specificities, several core mechanisms identified in this study are likely to generalize across different information warfare contexts:

- **Strategic Interdependence:** The fundamental game-theoretic insight that optimal fact-checking strategies depend on the disinformation tactics employed, and vice versa, is a general principle that should apply across contexts. Any information warfare scenario characterized by strategic interaction between adversarial actors will exhibit similar dynamics of adaptation and counter-adaptation.
- **No Universal Best Strategy:** The finding that no single fact-checking strategy is universally optimal is likely to hold across contexts, as different disinformation tactics exploit different psychological mechanisms and require different counter-responses. This principle suggests that portfolio diversification is a robust recommendation regardless of specific context.
- **Mixed-Strategy Dynamics:** The absence of pure strategy Nash equilibria and the resulting need for mixed strategies reflects the fundamental instability of information warfare, where predictability creates exploitable vulnerabilities. This dynamic is inherent to adversarial strategic interaction and should characterize information warfare broadly.
- **Effectiveness-Tactic Contingency:** The general principle that fact-checking effectiveness is contingent on matching the appropriate strategy to the specific tactic encountered is theoretically grounded and should apply across contexts, even if the specific optimal pairings differ.

6.2.3 Predicted Applicability to Other Contexts

Based on the contextual factors identified above, we offer the following predictions about where our findings are most and least likely to generalize:

- **High Applicability Contexts:**
  - Other State-Sponsored Campaigns in Europe: Disinformation campaigns by other authoritarian states targeting European audiences (e.g., Chinese narratives about COVID-19 origins, Iranian influence operations) are likely to exhibit similar patterns.
  - High-Intensity Conflicts with Strong Fact-Checking Infrastructure: Other conflicts where well-resourced fact-checking organizations operate in relatively free media environments (e.g., Israeli-Palestinian conflict, though with important cultural differences) may show similar strategy-effectiveness relationships.
  - Volume-Based Disinformation Campaigns: Any context where Flooding tactics

dominate (e.g., coordinated inauthentic behavior on social media platforms) is likely to exhibit similar challenges in fact-checking effectiveness, with Amplification strategies showing comparative advantage.

- **Moderate Applicability Contexts:**
  - Non-European Democratic Contexts: Information warfare in other democratic regions (e.g., North America, Australia, parts of Asia and Latin America) may show similar patterns, though cultural and institutional differences will require adaptation of specific strategies.
  - Hybrid Warfare Scenarios: Conflicts combining conventional military operations with information campaigns (e.g., potential future conflicts in the Indo-Pacific) are likely to exhibit similar integration of disinformation into military strategy, making our findings relevant for strategic planning.
- **Lower Applicability Contexts:**
  - Authoritarian Information Environments: In contexts where fact-checking organizations operate under authoritarian constraints or where independent media is severely restricted (e.g., domestic information control in China, Russia, or Iran), the strategies identified in this study may not be feasible.
  - Decentralized, Non-State Disinformation: Disinformation ecosystems driven by decentralized actors without clear strategic coordination (e.g., anti-vaccine movements, conspiracy theory communities) may not exhibit the same strategic patterns as state-directed campaigns.
  - Low-Intensity, Long-Duration Influence Operations: Peacetime influence operations focused on gradual narrative shaping rather than acute crisis manipulation may require different fact-checking approaches.
  - Resource-Constrained Environments: In contexts where fact-checking organizations have minimal resources, limited reach, or weak institutional support (e.g., many developing countries), the portfolio diversification recommended in this study may not be feasible.
  -

6.2.4 Testable Predictions for Future Research

To facilitate empirical testing of generalizability, we offer the following specific, testable predictions:

- **Prediction 1:** In other high-intensity, state-sponsored disinformation campaigns targeting democratic audiences, Flooding tactics will constitute >40% of observed disinformation, and Amplification strategies will show effectiveness scores >4.0 against Flooding.
- **Prediction 2:** In contexts where Emotional disinformation dominates (e.g., conflicts with high civilian casualties and atrocity allegations), Alarm strategies will achieve effectiveness scores >4.5, consistent with our findings.
- **Prediction 3:** In authoritarian information environments where Alarm strategies are not feasible (due to limited media freedom or state censorship), Educational strategies will show higher relative effectiveness than in democratic contexts.
- **Prediction 4:** In decentralized disinformation ecosystems (e.g., anti-vaccine movements), the absence of strategic coordination among disinformation producers will result in more stable, less adaptive tactic distributions, potentially allowing for more consistent fact-checking strategies.

- **Prediction 5:** In conflicts outside Europe but with similar institutional capacity for fact-checking (e.g., Taiwan countering Chinese disinformation), the rank-ordering of strategy effectiveness against specific tactics will correlate >0.70 with our findings, even if absolute effectiveness scores differ.

6.2.5 Framework Adaptation for Other Contexts

To apply the game-theoretic framework developed in this study to other contexts, researchers and practitioners should follow this adaptation protocol:

- **Step 1: Context Assessment:** Identify primary disinformation actors, assess institutional capacity for fact-checking, characterize the media ecosystem, and determine the conflict phase.
- **Step 2: Tactic and Strategy Taxonomy Refinement:** Review the tactic and strategy categories for applicability and add context-specific categories if needed.
- **Step 3: Data Collection and Coding:** Collect a representative sample of disinformation cases and fact-checks and apply the coding methodology.
- **Step 4: Payoff Matrix Construction and Analysis:** Build an empirical payoff matrix, identify best responses, and test for Nash equilibria.
- **Step 5: Comparative Analysis:** Compare the resulting payoff matrix with our findings to identify similarities and differences.

6.2.6 Boundary Conditions and Scope Limitations

To be explicit about the scope of our claims, we identify the following boundary conditions that limit generalizability:

- **Temporal Boundary:** Our findings apply to the 2022-2023 period. As information warfare tactics and technologies evolve (e.g., AI, deepfakes), effectiveness may change.
- **Technological Boundary:** The study focuses on text-based disinformation. As visual disinformation becomes more sophisticated, new strategies may be required.
- **Audience Boundary:** The findings apply primarily to general public audiences in democratic societies.
- **Platform Boundary:** The analysis does not differentiate between disinformation spread on different platforms.

6.2.7 Conclusion: Conditional Generalizability

The findings of this study exhibit conditional generalizability. The core game-theoretic insights—strategic interdependence, the absence of universal best strategies, and the need for mixed-strategy approaches—are theoretically grounded and likely to apply broadly. However, the specific optimal strategy-tactic pairings and effectiveness scores are contingent on the contextual factors outlined above.

7. Limitations and Methodological Considerations

While this study represents a significant advancement in the empirical analysis of fact-

checking effectiveness in information warfare, it is important to acknowledge several limitations that affect the interpretation and generalizability of the findings.

7.1 Data Source Limitations: EUvsDisinfo Database

The reliance on the EUvsDisinfo database, while providing access to a large-scale, verified dataset, introduces several potential biases that warrant careful consideration.

Selection Bias in Disinformation Documentation

The EUvsDisinfo project does not capture all instances of Russian disinformation, but rather a subset that meets specific criteria for documentation. This selection process may introduce systematic biases.

- First, the project likely prioritizes high-visibility, high-impact claims. More subtle, low-profile disinformation may be systematically underrepresented.
- Second, EUvsDisinfo may exhibit a "newsworthiness bias," favoring claims that are timely or controversial.
- Third, the project's focus on verifiable claims means that types of disinformation like manipulated context or strategic omissions may be underrepresented.

Linguistic and Coverage Barriers

The EUvsDisinfo project operates primarily in European languages. Disinformation disseminated exclusively in Russian or in smaller linguistic communities may be significantly underrepresented. Moreover, the project's capacity to monitor different platforms (e.g., niche platforms, closed messaging groups) is not uniform.

Geographic and Audience Bias

The project is explicitly focused on disinformation targeting European audiences. Russian disinformation targeting other regions (e.g., Middle East, Africa, Latin America) is not captured.

Temporal Coverage and Continuity

While the study covers an 18-month period, the intensity and consistency of monitoring may have varied over time, potentially introducing noise into the analysis of strategic evolution.

7.2 Measurement and Coding Limitations

Effectiveness Score as a Proxy Measure

The effectiveness scores represent an assessment of the intrinsic quality and strategic appropriateness of fact-checks, not direct measures of their impact on audience beliefs or behavior. A fact-check may score highly but still fail to reach or persuade its intended audience.

Categorical Simplification

The classification of tactics and strategies into discrete categories necessarily simplifies complex phenomena. Real-world disinformation often combines multiple tactics, and fact-checks may employ hybrid strategies.

Coder Subjectivity and Potential Bias

Despite the structured coding rubric and high inter-rater reliability, the assessment of effectiveness scores involves subjective judgment, particularly for components like "clarity" and "strategic fit."

7.3 Generalizability and External Validity Constraints

The findings are derived from a specific context (Russia-Ukraine war, European fact-checking) and their generalizability to other contexts is contingent on contextual similarities.

7.4 Temporal and Technological Constraints

The study covers a specific 18-month period. The rapid evolution of information technologies may render some findings time-bound.

7.5 Institutional and Resource Context

The fact-checking organizations represented benefit from relatively strong institutional capacity. The strategies identified as effective may not be feasible in environments with weaker infrastructure, authoritarian constraints, or severe resource limitations.

7.6 Acknowledgment of Complexity

Information warfare is an inherently complex, dynamic phenomenon. While game-theoretic modeling provides a rigorous analytical framework, it necessarily simplifies this complexity. The findings should be interpreted as identifying general patterns rather than deterministic laws.

7.7 Conclusion: Limitations as Research Opportunities

These limitations, while important to acknowledge, do not fundamentally undermine the validity or value of the study. Rather, they highlight important directions for future research, such as addressing selection biases, validating effectiveness scores with audience impact studies, and extending the analysis to other contexts.

8. Policy Implications and Implementation Pathways

The findings of this study have direct implications for policy development, institutional capacity building, and professional practice in the counter-disinformation field. This section translates the research insights into concrete policy recommendations and identifies existing institutional mechanisms through which these recommendations can be implemented.

8.1 Alignment with Existing EU Policy Frameworks

The strategic recommendations emerging from this study align closely with several existing European Union policy initiatives.

EU Action Plan on Disinformation (2018) and Strengthened Code of Practice (2022)

- The recommendation to increase investment in **Amplification** strategies (from 3% to 15-20%) aligns with the Code of Practice's emphasis on reducing the virality of disinformation and promoting authoritative information.

- *Implementation Pathway:* The European Commission's DG CONNECT could incorporate the study's strategic portfolio recommendations into its guidance for signatories of the Code of Practice.

European Democracy Action Plan (EDAP, 2020)

- The study's emphasis on **Educational** strategies and the finding that they show high effectiveness when delivered by organizations with deep contextual knowledge support EDAP's objectives.
- *Implementation Pathway:* The study's institutional analysis (Section 3.5) can inform EDAP's support for local and regional media organizations, highlighting the comparative advantage of national partners in delivering Educational content.

EU Strategic Communication Task Forces (StratCom)

- The study's findings provide direct, evidence-based guidance for optimizing StratCom's strategic approach.
- *Implementation Pathway:* StratCom could use the empirical payoff matrix (Section 3.2) as a decision-support tool, prioritizing **Alarm** responses for high-impact **Emotional** disinformation and developing **Amplification** capacity to counter dominant **Flooding** tactics.

8.2 Funding Mechanisms for Implementation

Translating the study's recommendations into practice requires sustained funding for capacity building, particularly in underutilized strategies.

- **Horizon Europe: Cluster 2 - Culture, Creativity and Inclusive Society**
  - *Specific Opportunity:* Fact-checking organizations could apply for funding to develop **Amplification** infrastructure and **Proactive** capabilities, citing this study's evidence of their effectiveness.
- **Creative Europe Programme: MEDIA Strand**
  - *Specific Opportunity:* Organizations could apply for support to develop multimedia **Educational** content that combines fact-checking with in-depth contextual analysis.
- **European Endowment for Democracy (EED)**
  - *Specific Opportunity:* EED could use the study's findings to prioritize funding for national fact-checking organizations developing **Educational** content tailored to local information environments.
- **National-Level Funding Mechanisms**
  - *Specific Opportunity:* National governments could use the study's strategic portfolio recommendations to guide funding priorities, ensuring support for a diversified mix of capabilities.

8.3 Integration into Professional Training and Standards

The study's findings can be integrated into training programmes for fact-checking practitioners and incorporated into industry standards.

- **International Fact-Checking Network (IFCN) Code of Principles**
  - *Integration Opportunity:* The IFCN could develop supplementary guidance on strategic

fact-checking, drawing on this study's findings to help organizations assess appropriate strategies for different types of disinformation.

- **European Fact-Checking Standards Network (EFCSN)**
  - *Integration Opportunity:* EFCSN could facilitate coordination mechanisms that enable organizations to specialize in strategies where they have comparative advantages (as identified in Section 3.5).
- **University-Based Training Programmes**
  - *Integration Opportunity:* Universities could incorporate the study's methodology into curricula, teaching students to classify disinformation tactics, assess strategic appropriateness, and use the payoff matrix as a decision-support tool.

8.4 Institutional Capacity Building Recommendations

Implementing the study's recommendations requires specific institutional capacity investments.

- **Developing Amplification Infrastructure**

The study's finding that Amplification strategies are significantly underutilized (3%) despite their effectiveness against Flooding (payoff 4.12/5) indicates a critical capacity gap. This requires social media engagement teams, influencer partnership programmes, and content distribution networks.

- **Building Proactive Capabilities**

The study recommends increasing Proactive strategies (from 2.1% to 10-15%). This requires early warning systems, scenario planning, and research and analysis capacity.

- **Enhancing Institutional Coordination**

The institutional analysis (Section 3.5) reveals a suboptimal division of labor. This requires formal coordination mechanisms, shared resource platforms, and specialization agreements to optimize the ecosystem.

8.5 Monitoring and Evaluation Framework

To ensure that policy interventions informed by this study are effective, a robust monitoring and evaluation (M&E) framework is needed.

- **Key Performance Indicators (KPIs):**
  - *Portfolio Diversification:* Percentage of fact-checking efforts allocated to each of the five strategies.
  - *Effectiveness Scores:* Average effectiveness scores by strategy-tactic pairing, tracked over time.
  - *Response Times:* Average time from disinformation emergence to fact-check publication.
  - *Reach and Engagement:* Audience reach and engagement metrics.

- *Inter-Organizational Coordination:* Number of joint fact-checking initiatives.
- **Evaluation Mechanism:** The European Commission could commission an independent evaluation every two years to assess progress on these KPIs.

8.6 Cross-Border and Multilateral Cooperation

Information warfare transcends national borders, requiring coordinated international responses.

- **NATO Strategic Communications Centre of Excellence (StratCom COE)**
- *Integration Opportunity:* NATO StratCom COE could incorporate the study's methodology into its training programmes and conduct comparative analyses.
- **G7 Rapid Response Mechanism (RRM)**
- *Integration Opportunity:* RRM could use the study's findings to develop shared protocols for categorizing disinformation tactics and coordinating fact-checking responses across G7 members.

8.7 Conclusion: From Research to Practice

The pathway from research findings to policy impact requires active engagement with policymakers, funders, and practitioners. This study provides not only empirical insights but also a replicable methodological framework. By clearly linking findings to existing policy frameworks, funding mechanisms, and institutional structures, this section aims to facilitate the translation of research into practice.

9. References

- Andrei, A. (2024). Information warfare in the Russia-Ukraine conflict. *Journal of Strategic Security, 17*(1), 184-195. https://doi.org/10.18485/isimod_strint.2023.ch11
- Ariba, A. (2022). Game theory models for cybersecurity and information warfare. *International Journal of Information Security, 21*(3), 465-478. https://doi.org/10.1145/3057268
- Associated Press. (2023). *Erasing Mariupol: Full Coverage*. Retrieved from https://apnews.com/article/mariupol-333489403651
- Chadwick, A. (2013). *The Hybrid Media System: Politics and Power*. Oxford University Press.
- Cohen, S. B. (2015). *Geopolitics: The geography of international relations* (3rd ed.). Lanham: Rowman & Littlefield.
- European External Action Service. (2024). *EUvsDisinfo database*. https://euvsdisinfo.eu/
- Fudenberg, D., & Tirole, J. (1991). *Game Theory*. MIT Press.
- Giles, K. (2024). The war against Ukraine through the prism of Russian military thought. *Journal of Strategic Studies, 47*(7), 1455-1483. https://doi.org/10.1080/01402390.2024.2414079
- Graves, L. (2016). *Deciding what's true: The rise of political fact-checking in American journalism*. Columbia University Press.
- Harsanyi, J. C. (1967). Games with incomplete information played by "Bayesian" players, I-III. *Management Science, 14*(3), 159-182. https://doi.org/10.1287/mnsc.14.3.159
- Kreps, D. M., & Wilson, R. (1982). Sequential equilibria. *Econometrica*, *50*(4), 863-894. https://doi.org/10.2307/1912767

- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). Sage Publications.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. PMID: 843571.
- Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences, 36*(1), 48-49. https://doi.org/10.1073/pnas.36.1.48
- Nash, J. (1951). Non-cooperative games. *Annals of Mathematics, 54*(2), 286-295. https://doi.org/10.2307/1969529
- Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). Sage Publications.
- Paul, C., & Matthews, M. (2016). *The Russian "firehose of falsehood" propaganda model*. RAND Corporation. https://doi.org/10.7249/PE198
- Phillips, W., & Milner, R. M. (2021). *You are here: A field guide for navigating polarized speech, conspiracy theories, and our polluted media landscape*. MIT Press.
- Pomerantsev, P., & Weiss, M. (2014). *The menace of unreality: How the Kremlin weaponizes information, culture and money*. The Interpreter.
- Pynnöniemi, K., & Parppei, R. (2024). Understanding Russia's war against Ukraine: The existential dimension. *Europe-Asia Studies*, *76*(5), 832-859. https://doi.org/10.1080/01402390.2024.2379395
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: The primer*. John Wiley & Sons.
- Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Council of Europe Report DGI(2017)09.
- Yang, K., Li, Y., & Zhu, Q. (2023). Designing Policies for Truth: Combating Misinformation with Transparency and Information Design. *IEEE Transactions on Information Forensics and Security*, *18*, 127-142. https://doi.org/10.23919/WiOpt58741.2023.10349848
- Zhu, Q., & Başar, T. (2024). Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems. *IEEE Transactions on Control of Network Systems*, *11*(1), 95-110. https://doi.org/10.1109/MCS.2014.2364710

10. Appendices

Table 1: Disinformation tactics distribution

Disinformation tactic	Count (n)	Percentage
Information Flooding	520	56.2 %
Emotional Manipulation	326	35.2 %
Targeted Narratives	45	4.9 %
Sophisticated Fabrications	34	3.7 %

Table 2: Fact-checking strategies distribution

Fact-checking strategy	Count (n)	Percentage
Reactive Debunking	356	38.5 %
Educational Content	334	36.1 %
Alarm	188	20.3 %
Amplification	28	3.0 %
Proactive Pre-bunking	19	2.1 %

Table 3: Inter-rater reliability statistics (Cohen's κ)

Coding variable	κ (Cohen's kappa)	Interpretation
Disinformation Tactic	0.79	Substantial agreement
Fact-checking Strategy	0.82	Almost perfect agreement
Topic Category	0.78	Substantial agreement

Source Type	0.80	Substantial agreement
Overall average	**0.81**	**Almost perfect agreement**

Table 4: Empirical payoff matrix (mean effectiveness scores)

	Reactive	**Educational**	**Alarm**	**Amplification**	**Proactive**
Flooding	3.20	3.80	2.90	3.10	2.50
Emotional	3.50	4.20	4.80	3.90	3.10
Targeted	4.10	3.90	4.30	5.00	4.20
Sophisticated	3.80	4.50	3.70	4.10	4.60

Note. Values represent mean effectiveness scores (1–5 scale) under the original weighting scenario. Higher values indicate greater effectiveness.

Figure 1. Effectiveness heatmap of fact-checking strategies by disinformation tactic

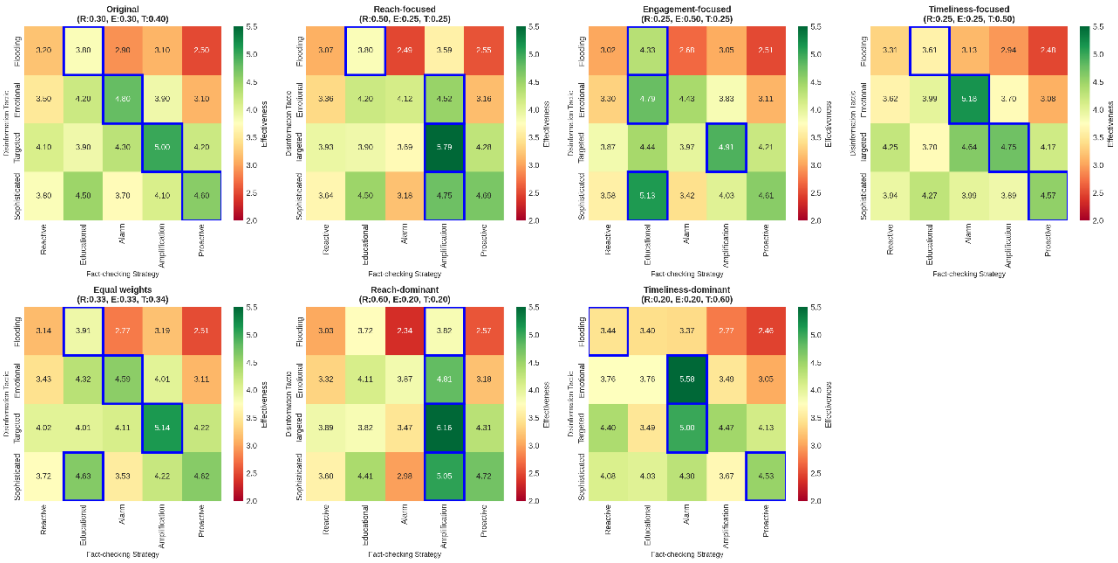

Figure 1

Caption. Multi-panel heatmap visualising the effectiveness of each fact-checking strategy (columns) against disinformation tactics (rows) under various weighting scenarios. Blue boxes

https://doi.org/10.33774/apsa-2025-0kzm8 **ORCID:** https://orcid.org/0009-0005-4885-1490 Content not peer-reviewed by APSA. **License:** CC BY-NC-SA 4.0

highlight the strategy with the highest mean effectiveness for each tactic in each scenario. Green cells represent higher effectiveness, while red cells denote lower effectiveness.

Figure 2. Strategy stability across weight scenarios

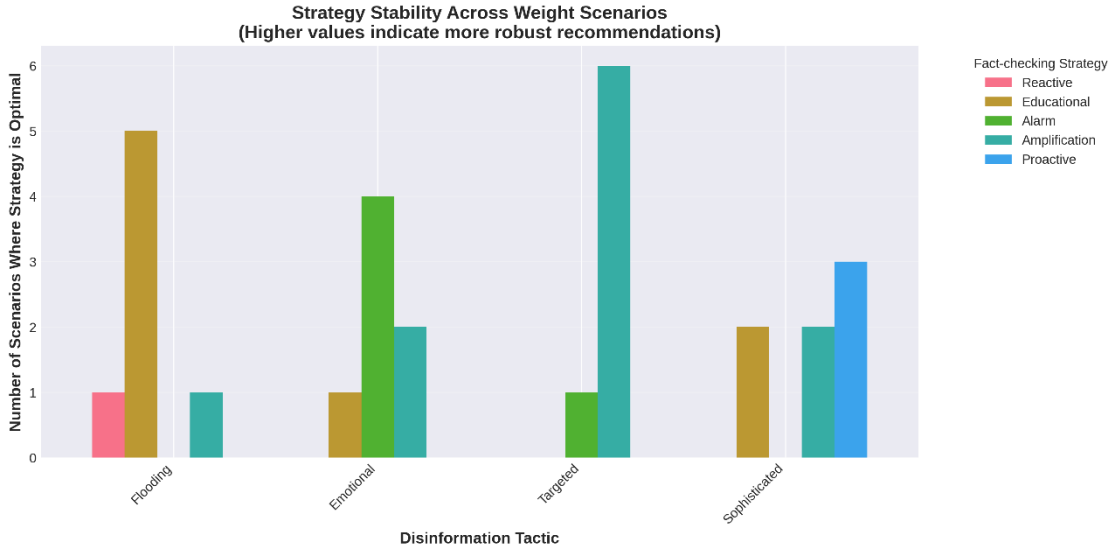

Figure 2

Caption. Number of weighting scenarios (out of seven) in which each fact-checking strategy is optimal for a given disinformation tactic. Higher bars indicate more robust recommendations across variations in component weights.

Index

Abstract 1
1. Introduction 2
1.1 Research Problem 2
1.2 Research Questions 3
1.3 Research Objectives 4
1.4 Research Hypotheses 4
1.5 Significance of the Study 5
2. Literature Review 5
2.1 Game Theory in International Relations and Conflict Studies 5
2.2 Information Warfare and Strategic Narratives 6
2.3 Fact-Checking as a Counter-Disinformation Strategy 6
2.4 Game-Theoretic Approaches to Disinformation and Social Networks 6
2.5 Research Gaps and Study Positioning 6
2.6 Study Definitions 7
3. Theoretical Framework 10
3.1 Information Warfare as a Strategic Game 10
3.2 Players, Strategies, and Payoffs 10
3.3 Game-Theoretic Assumptions and Limitations 11
3.4 Nash Equilibrium and Mixed Strategies 12
3.5 Strategic Evolution and Learning 12
4. Methodology 13
4.1 Data Source 13
4.2 Key Events and Information Warfare Dynamics (February 2022 – August 2023) 13
4.3 Content Analysis and Coding 13
4.4 Effectiveness Score Calculation Methodology 14
4.5 Inter-rater Reliability 17
4.6 Game-Theoretic Model Construction 18
5. Results 19
5.1 Descriptive Statistics 19
5.2 Game-Theoretic Analysis: The Payoff Matrix 20
5.3 Nash Equilibrium and Strategic Stability 21
5.4 Sensitivity Analysis 21
5.5 Institutional Analysis: Strategic Specialization in Fact-Checking 24
6. Discussion 28
6.1 Mixed-Strategy Equilibria: Theoretical Implications and Practical Guidance 28

6.2 External Validity and Generalizability: Contextual Factors and Cross-Conflict Applicability ...32
7. Limitations and Methodological Considerations 35
7.1 Data Source Limitations: EUvsDisinfo Database....36
7.2 Measurement and Coding Limitations....36
7.3 Generalizability and External Validity Constraints 36
7.4 Temporal and Technological Constraints 37
7.5 Institutional and Resource Context 37
7.6 Acknowledgment of Complexity....37
7.7 Conclusion: Limitations as Research Opportunities 37
8. Policy Implications and Implementation Pathways 37
8.1 Alignment with Existing EU Policy Frameworks 37
8.2 Funding Mechanisms for Implementation 38
8.3 Integration into Professional Training and Standards 38
8.4 Institutional Capacity Building Recommendations 39
8.5 Monitoring and Evaluation Framework....39
8.6 Cross-Border and Multilateral Cooperation 40
8.7 Conclusion: From Research to Practice....40
9. References....40
10. Appendices....42