American Journal of
Biomedical Science & Research

www.biomedgrid.com
ISSN: 2642-1747

Research Article

Serious Games - Discourse Conflict among Power Groups in the Information Environment and the Public's Information Fatigue and Recovery by Imperfect or Perfect Information State

Yasuko Kawahata*

Faculty of Sociology, Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo

***Corresponding author:** Yasuko Kawahata, Faculty of Sociology, Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima-ku, Tokyo, Japan.

To Cite This Article: *Yasuko Kawahata*. Serious Games - Discourse Conflict among Power Groups in the Information Environment and the Public's Information Fatigue and Recovery by Imperfect or Perfect Information State. Am J Biomed Sci & Res. 2025 26(3) AJBSR.MS.ID.003437,*
DOI: 10.34297/AJBSR.2025.26.003437

Received: February 26, 2025; **Published:** March 25, 2025

Abstract

The polarization of discourse in online spaces and the disorder caused by fragmented information have become increasingly conspicuous. In particular, conflicts among power clusters have escalated, and as AI and bots automate certain aspects of debate, the general public is experiencing severe information fatigue. Here, a mathematical model based on game theory is constructed to derive the optimal scenario for conflict resolution among power clusters, suppression of AI/bot influence, and recovery from information fatigue within the general public. Building on this model, we propose measures to stabilize society over the long term. Furthermore, as a framework for policy proposals derived from social problem simulations, an approach employing Serious Games is tentatively explored.

Introduction

Political and social debates in online spaces, as well as controversies stemming from the disclosure or nondisclosure of information, are intensifying in their degree of polarization alongside the escalation of conflict. In particular, confrontations among power clusters tend to become zero-sum games, making it difficult to converge opinions. Moreover, during election campaigns and social movements, there have been confirmed instances where AI and bots intervene in de- bates, automatically reinforcing or guiding specific discourse. As a result, those in the general public—who exist in an environment of incomplete information are inundated with an excessive volume of data, leading to information fatigue. This situation can drive individuals to withdraw from discussions or abandon the process of forming opinions.

In this discussion, we construct a mathematical model concerning the convergence of conflicts among power clusters, the suppression of AI/bot influence, and the recovery from information fatigue in the general public, ultimately deriving an optimal convergence probability. Specifically, we focus on the following three points:

1. Convergence Conditions for Conflict Among Power Clusters: We introduce a variable D that indicates the intensity of debate and analyze the effectiveness of platform interventions and dialog programs aimed at mitigating confrontations.

2. Suppression of AI/bot Influence: We formulate the strategies by which power clusters utilize AI/bots in the model and examine the efficacy of platform detection capabilities and stricter regulations.
3. Recovery from Information Fatigue Among the General Public: We investigate how variations in the volume of information affect the general public 's information fatigue and evaluate the impact of information-organizing and recovery programs (educational initiatives and media literacy enhancements).

In addition, this study goes beyond merely con- structing a mathematical model by integrating a Serious Games approach. Serious Games, as defined in [1,2], are designed not simply for entertainment purposes but also for educational, policy-making, and social problem simulation objectives. Prior research [3,4] has explored applications of Serious Games in fields such as economics, environmental issues, urban planning, and international politics. In this paper, online discourse polarization is incorporated into a game- theoretic framework, and player actions (power clusters, AI/bots, and the general public) are simulated to evaluate which types of interventions lead to the most optimal convergence. Such an approach can be applied to real-world policy proposals and holds additional significance as research on the applications of Serious Games. Taking the above three points into account, we undertake a mathematical analysis using game theory and propose a model to optimize the probability of debate convergence PS. On the basis of computational results, we discuss optimal methods of adjusting the information environment and examine measures to promote the stabilization of society.

Asymmetry Between Complete-Information and Incomplete-Information Games

Power clusters possess one-way media outlets and thus wield enormous influence in disseminating information. Consequently, they have complete information regarding each player 's payoff function and the strategies available. Although this enables strategic manipulation of information and the steering of debates, the general public is forced into a state of incomplete information, lacking the data necessary for informed decision-making. This asymmetry structurally exacerbates social fragmentation.

Examination of Social Risks Arising from Segmented Information Environments and Information Fatigue Among the General Public

The backdrop to advancing polarization in online spaces includes differences in the conditions surrounding each cluster 's game. Notably, while power clusters operate under conditions akin to complete- information games, the general public remains in a state of incomplete-information games. This asymmetry accelerates societal division. In addition, it triggers information fatigue among the public and severely undermines the social function of dialog.

Structural Factors Behind Information Fatigue

Information fatigue in the general public must be recognized as a systemic issue triggered by this asymmetry. Power clusters with access to complete information emit a massive volume of data, while the general public operating under incomplete information cannot adequately process it all, compounding their

fatigue. Especially when multiple power clusters are at odds with each other, the general public must grapple with conflicting information, leading to a drastic increase in cognitive load.

The Ambivalent Position of the Intermediate Layer and Its Impact on Information Fatigue

Those in the intermediate layer, who have substantial reach on interactive media such as social networking services, occupy a position between complete- information and incomplete-information game conditions. While they have the potential to serve as a bridge, relaying information between power clusters and the general public, they also risk functioning as amplifiers of misinformation. When the intermediate layer intensifies the spread of information, the processing burden on the general public grows, further exacerbating information fatigue.

Challenges in Information Fatigue and Its Recovery for the General Public

Information fatigue among the general public can- not be explained solely by the overabundance of data; qualitative issues also play a role. Under incomplete- information conditions, significant cognitive resources are required to assess both the authenticity of information and its relative importance, often triggering ongoing fatigue. Merely reducing the quantity of information is insufficient; improvement in the quality of the information environment is essential for alleviating this fatigue.

Mutual Reinforcement Between Structural Fragmentation and Fatigue

The structural asymmetry in the information environment and the public 's information fatigue exacerbates each other. Deeper fragmentation results in a heavier information-processing load, which advances fatigue, while growing fatigue hampers rational judgment and further intensifies fragmentation—a vicious cycle. Breaking this interlinked chain of effects demands simultaneous structural reform of the information environment and support for the recovery of the general public.

An Integrated Approach to Social Recovery

Achieving both a reduction in fragmentation and restoration from information fatigue necessitates an integrative strategy encompassing three aspects. First, platforms must assure the quality of information and manage its circulation appropriately. Second, the role of the intermediate layer in providing constructive mediation should be strengthened. Third, there must be a system that

supports the advancement of information literacy and recovery from fatigue among the general public. These initiatives should not be implemented in isolation; rather, a unified effort is needed to improve the structure of the information environment while aiding the public 's capacity for recovery. Narrowing the gap between complete-information and incomplete- information game conditions and fostering a more symmetrical information landscape are essential to preventing social division and ensuring the sustained resilience of the general public.

The Concept of Serious Games

The concept of Serious Games was first introduced by *CC, Abt* [1]. He argued that games are not purely recreational but can also serve educational, social, and political objectives, coining the term "Serious Games" Unlike conventional games intended solely for entertainment, such games are characterized by designs aimed at solving specific problems or facilitating learning. According to Zyda, Serious Games have evolved from virtual reality and visual simulation, incorporating elements of entertainment while promoting learning and behavioral change [2]. Zyda also highlights the importance of interactivity, goal-oriented design, and player engagement as core components of Serious Games.

Sawyer discusses the potential of Serious Games to function as simulations that inform the improvement of public policy [3]. His work demonstrates their utility in visualizing complex social issues—such as urban planning, crisis management, and environmental policy and enabling policymakers to experiment with various scenarios. Susi et al. enumerate several features of Serious Games, including immersion, inter active learning, and real-world applicability [4]. Their research shows examples in education, military training, and healthcare, noting that Serious Games in medical contexts are often leveraged for surgical simulation and rehabilitation support.

Applications of Serious Games

Serious Games are implemented in a wide range of fields. In the military domain, they are used for tactical training and to refine the decision-making skills of soldiers. Within healthcare, they aid in simulations for surgical training and tools for cognitive rehabilitation. The educational sector employs them to assist learning in history, mathematics, and science. For environmental issues, simulation games on sustainable city planning and climate change awareness are adopted. Zyda emphasizes that the success of Serious Games relies on realistic simulation, designs that sustain user motivation, and the use of data analytics to provide feed- back [2]. These elements reinforce their effectiveness not simply as learning instruments but also as practical methods for fostering behavioral changes.

Game Theory and Its Connection to Serious Games

Game theory offers a mathematical framework for analyzing rational decision-making and aligns well with Serious Games. In the context of political decision-making and social simulations, it is especially useful for modeling the strategic actions of various players. For instance, Sawyer 's research demonstrates that policymakers can test different strategies and their corresponding outcomes, thereby facilitating more in- formed decision-making [3].

In this work, the Serious Games approach is used to scrutinize the issue of polarized online discourse. We formulate three components the conflict among power clusters, the influence of AI/bots, and "information fatigue among the general public" and simulate how they interact. As indicated in the research by Susi et al., the design of Serious Games necessitates incentive structures that promote player behavioural change [4]. Our model introduces mechanisms by which players acquire accurate information and make rational judgments, with the objective of guiding discourse toward a healthier convergence.

Objective

The polarization of discourse and the fragmentation of information have both become pronounced, with conflicts among power clusters intensifying as a result. Under these conditions, AI/bots intervene in debates and manipulate information, leading to severe information fatigue among the general public. In response, the goal of this study is to develop a mathematical model using game theory that identifies the convergence conditions for power-cluster conflicts, minimizes AI/bot influence, and fosters the recovery of the general public from information fatigue. By employing a Serious Games approach aimed at stabilizing the information environment, we draw on a game-theoretic framework that goes beyond mere entertainment and serves educational, policy-making, and social problem simulation purposes. Within this model, the behavior of players (power clusters, AI/bots, and the general public) is simulated and discussed, focusing on the mechanisms driving polarization in online discourse and how each player's strategic choices shape debate outcomes.

Challenges

One reason for the increasing polarization in online spaces is that the game conditions differ among various clusters. Organizing these conditions by cluster, we can categorize them as follows:

Power Clusters (denoted A, B, and C)

They possess one-way media outlets and thus hold substantial power in disseminating information. As a result, they are often in a complete-information game† condition.

†A complete-information game refers to a situation in which, at the point of any player 's decision-making, every event that has occurred in the game so far is fully known to all players. This concept was formalized by von Neumann and morgennstern in 1944.

In a complete-information game, the following conditions are satisfied. These criteria comprise the defining features of "complete information" in game theory:

I. All players are fully aware of the set of possible choices available at every stage of the game.

II. Every player 's payoff function is common knowledge to all participants.

III. All players have a thorough understanding of the rules of the game.

Under complete-information conditions, analysis via backward induction is viable Zermelo (1913) demonstrated the existence of a winning strategy in finite complete-information games such as chess. This leads to the following key properties:

1) Any finite complete-information game has a pure-strategy equilibrium solution.

2) Such an equilibrium solution is uniquely determined as a subgame-perfect equilibrium derived through backward induction.

Classic examples of complete-information games include:

a. Chess

b. Shogi (Japanese chess)

Intermediate Layer

Utilizes interactive media such as social networking services, possesses significant communicative reach, and occupies a middle position between complete-information and incomplete-information game settings.

General Public

Although they represent the largest group, their collective voice is variable in influence, and they may either unite or oppose one another. This group often faces incomplete information game conditions.

Under this cluster structure, we take into account the following points:

I. The nature of the game may differ by cluster i.e., whether it constitutes a complete-information game or an incomplete-information game††.

II. Go

III. Nim

However, several limitations exist in the theory of complete- information games, largely due to practical constraints related to the complexity of real-world situations:

I. Computational complexity: Although an optimal solution may exist in theory, it is often very difficult to compute in practice.

II. Bounded human cognition: There can be significant discrepancies between the theoretically predicted behavior (assuming perfect rationality) and actual human actions.

††An incomplete-information game is one in which at least one player does not fully know certain attributes of other players, such as their payoff functions (preferences) or available strategies. This concept was systematically theorized by Harsanyi (1967–1968) and forms an important area of research in modern game theory.

Characteristic features of incomplete-information games include:

I. Information asymmetry: Players hold different sets of information.

II. Type space: A parameter (type) set is defined to represent each player 's characteristics.

III. Beliefs: Each player forms subjective probabilities regarding the types of other players.

Key concepts employed in the theoretical analysis of incomplete- information games are:

a. Bayesian game: A model that uses a probabilistic frame- work to represent incomplete-information settings.

b. Bayesian Nash equilibrium: A state in which each player 's strategy is an optimal response based on their beliefs.

c. Common prior belief: A shared probability distribution among all players before the game unfolds.

d. Representative examples of incomplete-information games include:

Power clusters tend to monopolize information, while the general public is prone to confusion stemming from its incompleteness.

Once the information fatigue of the general public surpasses a certain threshold, people disengage from discussions and can no longer make optimal decisions.

Here, we focus on the power clusters 'strategic decisions (concealment versus explanation) and analyze how the probability of being exposed and expected payoffs are related. Furthermore, we construct a mathematical model to optimize the probability of discourse convergence PS and also examine measures aimed at stabilizing society.

i. Auctions: Bidders do not know the valuation of other bidders.

ii. Labor markets: Employers do not have complete information about workers 'abilities.

iii. Insurance markets: Insurance providers cannot fully discern the risk profiles of the insured.

iv. Signalling games: The sender knows their own type, but the receiver does not.

Theoretical challenges in analyzing incomplete-information games involve:

i. Higher-order beliefs: There can be infinitely layered struc-

tures of beliefs about others 'beliefs.

ii. **Multiplicity of equilibria:** Multiple Bayesian Nash equilibria can exist.

iii. **Belief updating:** Establishing consistent rules for the evolution of beliefs over the course of the game.

The theory of incomplete-information games is applied to analyze various real-world economic and social phenomena, including:

i. Price formation mechanisms in markets.

ii. Optimal incentive design in contract theory.

iii. Information transmission and decision-making within organizations.

iv. Evaluating the impacts of information asymmetry in financial markets.

Practical implications of information asymmetry include:

i. **Adverse selection (Adverse selection):** High-quality goods or services may be driven out of the market.

ii. **Moral hazard (Moral hazard):** Unobservable actions may change after a contract is established.

iii. **Information rent (Information rent):** Additional gains acquired by those with superior information.

As society becomes increasingly information-driven, the relevance of incomplete-information games continues to grow. Beyond economics, this framework extends to a variety of fields, including political science, sociology, and business management.

Information Across Clusters Based on Serious Game Theory

In this section, we perform an analysis under the Serious Games framework, where different clusters operate under distinct game conditions. Notably, we assume that each cluster may be in either a complete- information or an incomplete-information scenario, and we consider payoff functions that incorporate factors such as information integrity and the avoidance of exploitative outcomes.

Classification of Clusters

We define the following three types of clusters:

I. **Power Clusters (A, B, C):** Groups wielding one way media with significant influence on information dissemination.

II. **Intermediate Layer:** Entities that utilize interactive media like SNS and hold considerable broad- casting power.

III. **General Public:** Although they represent the largest population, their individual influence is variable; some of them form alliances, while others come into conflict.

Additionally, the general public frequently operates under incomplete-information conditions and tends to follow or align with either power clusters or the inter- mediate layer. We analyse the impact that occurs when commonly accepted "truths" are later revealed to be "falsehoods."

Specification of Payoff Functions

I. Positive Components:

I. Integrity of the information environment

II. Practical utility of acquired knowledge

II. Negative Components:

1) Emergence of exploitative states

2) Confusion arising from misinformation

Taking these into account, we define utility in the following manner:

$$U = A - B$$

where A represents the positive returns, and B denotes the negative losses.

Power Cluster Strategic Choices: Concealment vs. Explanation

Each power cluster (A, B, C) is assumed to choose between two strategies: "concealment" or "explanation".

Expected Payoffs for Power Clusters

The payoff function for a power cluster is expressed as follows:

$$EU_P = P(-b_1 S) + (1 - P)(-c_1 S)$$

where:

i. **P:** Probability of selecting concealment

ii. **1 – P:** Probability of opting for explanation

iii. $\mathbf{b_1}$**:** Penalty if concealment is exposed ($b_1 > c_1$)

iv. $\mathbf{c_1}$**:** Penalty incurred when offering an explanation

Comparison of Concealment and Explanation

1) Expected Payoff When Opting for Concealment:

$$EU_P \text{ (concealment)} = -b_1 S$$

2) Expected Payoff When Opting for Explanation:

$$EU_P \text{ (concealment)} = -c_1 S$$

Concealment is More Advantageous If:

EU_P (concealment) > EU_P (explanation)

i.e.,

$$-b_1 S > -c_1 S$$

$$b_1 S < c_1 S$$

$$S < \frac{c_1}{b_1}$$

When this holds, it is rational for the power cluster to choose concealment.

Simulation-Based Analysis

We assign parameters and compare the payoffs of concealment and explanation (Table 1).

Table 1: Parameters for the Payoff Function of Power Clusters.

Parameter	Value
b1	10
c1	5

Assumed Parameter Values Threshold Calculation

$$S^* = \frac{c_1}{b_1} = \frac{5}{10} = 0.5$$

Therefore, if S < 0.5, opting for concealment is more rational, whereas if S > 0.5, choosing explanation is preferable.

Discussion

a) S < 0.5 (Low Probability of Exposure):

Concealment yields a higher expected pay- off.

Concealment is thus the rational choice.

b) S > 0.5 (High Probability of Exposure):

Explanation offers a higher expected payoff.

Explanation is therefore the rational choice.

Using game-theoretic analysis, this study has identified the optimal threshold for when a power cluster decides between concealment and explanation. The result shows that if the probability S of being exposed exceeds 0.5, it becomes rational for the cluster to cease concealment and instead opt for explanation. By estimating the actual value of S, it is possible to conduct a more detailed investigation into the strategies each cluster is likely to adopt.

Overview of the Simulation

In this section, we compare the expected payoffs for a power cluster when choosing between the strategies of "concealment" and "explanation" We examine how the concealment expected payoff EU_p (concealment) and the explanation expected payoff EU_p (explanation) vary as functions of S, the prob- ability that concealment is exposed, and discuss the optimal strategy at the threshold S∗ = 0.5 (Figure 1).

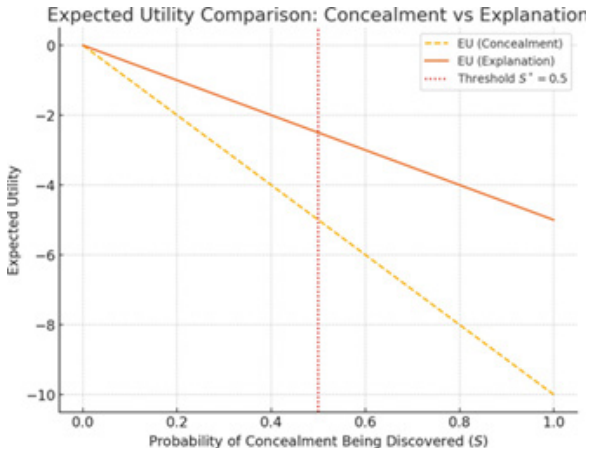

Figure 1: Expected Utility Comparison: Concealment vs Explanation.

From the simulation results, the following trends emerge:

(a) In the range S < 0.5, the expected payoff of concealment EU_p (concealment) is higher than that of explanation EU_p (explanation). Consequently, choosing concealment is more rational in this interval.

(b) In the range S > 0.5, the expected payoff of explanation EU_p (explanation) becomes higher. Therefore, it is more logical for the power cluster to opt for explanation rather than concealment.

(c) At the threshold S∗ = 0.5, the two expected payoffs are equal, marking the boundary at which the power cluster 's decision changes.

Interpretation

These findings indicate that the probability of concealment being revealed is a critical factor when a power cluster makes rational decisions. The following policy implications are derived: ultimately, when S exceeds 0.5, choosing explanation is more advantageous. This suggests that increasing the transparency of the in- formation environment influences the actions of power clusters, highlighting the effectiveness of early information interventions.

Derivation of the Optimal Payoff Convergence Rate for the General Public

Within the Serious Games framework described above, we now analyze the behavior choices available to the general public in or-

der to achieve an optimal pay- off. Specifically, we calculate how the general public might act so that its own payoff converges optimally, depending on whether the power cluster adopts con- cealment or explanation.

Definition of the General Public 's Payoff Function

We define the general public 's payoff function as follows:

$$U_G = A_G - B_G$$

where A_G signifies the positive gains and BG denotes the negative losses. An increase in AG derives from a healthier information environment and accurate learning, while an increase in BG corresponds to losses due to misinterpretation or confusion.

Factors Affecting the General Public 's Payoff

We introduce the following factors that influence the general public 's payoff function:

a) **I_P:** Reliability of the information provided by the power cluster ($0 \le IP \le 1$)

b) **I_C:** Reliability of the information supplied by the intermediate layer ($0 \le IC \le 1$)

c) **E_G:** Information literacy of the general public ($0 \le EG \le 1$)

Taking these into account, we model the general public 's payoff as follows:

$$A_G = \alpha I_P + \beta I_C + \gamma E_G$$

$$B_G = \delta(1 \,.\, I_P) + \varepsilon(1 \,.\, I_C) + \varsigma(1 \,.\, E_G)$$

Here, α, β, γ, δ, ϵ, ζ are constants representing the weight of each factor.

Deriving the Optimal Payoff Convergence Rate

To achieve the highest payoff, the general public must maximize UG:

$$U_G = (\alpha I_P + \beta I_C + \gamma E_G) - (\delta(1.I_P) + \in (1.I_C) + \zeta(1.E_G))$$

Rearranging the terms, we obtain:

Under this assumption, the general public 's payoff function becomes:

$$U_G = (\alpha + \delta) I_P + (\beta + \in) I_C + (\gamma + \zeta) E_G . (\delta + \in + \zeta)$$

We define P_S as the probability that the general public obtains appropriate information, and consider the function

$$P_S = f(I_P, I_C, E_G)$$

The optimal strategy that maximizes the general public 's payoff can be summarized in three points:

Avoid excessive dependence on power-cluster information

When I_P is low (i.e., when the power cluster is engaged in concealment), the general public needs to compensate for inaccuracies by increasing I_C or by enhancing E_G.

Make proactive use of information from the intermediate layer

Maximizing the contribution of I_C helps prevent misinformation and mitigates information fatigue.

Improve information literacy

Raising EG directly increases U_G and reduces the likelihood of confusion caused by inaccurate data.

Simulation Results

A simulation is performed under the above model, assigning particular values to each parameter.

Assumed Parameter Values

(Table 2) Under this assumption, the general public 's payoff function becomes:

Table 2: Parameters of the Payoff Function for the General Public.

Parameter	Value
α, δ	0.4
β, ϵ	0.3
γ, ζ	0.3

$$U_G = 0.8I_P + 0.6I_C + 0.6E_G - 0.9$$

Interpreting the Simulation Results

When $I_p < 0.5$, improving I_C and E_G is crucial.

I. If I_C is high, it can compensate for low IP, stabilizing the general public 's payoff.

II. Maintaining E_G at or above 0.5 minimizes the impact of information fatigue and facilitates more effective information gathering.

Based on this hypothesis, we derive a mathematical guideline for actions that enable the general public to achieve maximum payoff. The following conclusions can be drawn:

I. Even if I_P decreases, ensuring the reliability of I_C can preserve the integrity of information for the general public.

II. Improving the general public 's information literacy (EG) allows for better evaluation of power- cluster data, thereby reducing losses caused by misinformation.

III. The simulation shows that, even when I_P is low, appropriate adjustments to I_C and E_G can steadily enhance the general public 's payoff.

Medium- to Long-Term Risks of Concealment Strategies by Power Clusters

Here, we use a mathematical model to analyze the medium to long-term risks that arise if a power cluster opts for concealment.

Although concealment may be a rational choice in the short term, it can lead to inconsistencies in information over time, there by escalating risk. In particular, we consider losses stemming from repeated acts of concealment, decreased payoffs for other clusters, and the eventual detriment to the power cluster itself.

Power Cluster Concealment Strategies and Their Effects

Risk Accumulation Through Repeated Concealment

Once a power cluster initiates concealment, it may require additional obfuscation or manipulation to maintain consistency in the narrative. Repeated concealment entails the following risks:

1) Information inconsistencies emerge, leading to contradictions.

2) If concealment is uncovered, the resulting damage may increase exponentially.

Impact on Other Clusters

Continuing a concealment strategy reduces the pay- offs of other clusters. We consider this in terms of the intermediate layer and the general public.

a. Payoff for the Intermediate Layer (EU_M)

Restricted information flow reduces learning opportunities.

Fractured opinions undermine the reliability of disseminated data.

b. Payoff for the General Public (EU_G)

Information confusion makes it difficult to discern factual accuracy.

Heightened polarization destabilizes society.

Long-Term Losses for the Power Cluster

While concealment can circumvent risk in the short term, the exposure risk S grows over time. If concealment is discovered, the penalty b1 can be significantly larger than usual.

Mathematical Model of Long-Term Concealment Risks

Temporal Variation in the Probability of Exposure

We assume that the probability S(t) of uncovering concealment increases over time t. This relationship is modelled as follows:

$$S(t) = S_0 + rt$$

where

i. **S_0:** Initial likelihood of concealment being ex- posed.

ii. **r:** Rate at which the exposure risk grows as time progresses.

iii. **t:** Elapsed time (the period during which concealment continues).

Long-Term Expected Payoff

The long-term expected payoff for a power cluster that persists in concealment is defined as:

$$EU_P^{long} = -b_1 S(t) = -b_1 (S_0 + rt)$$

Here, b1 is the penalty if concealment is discovered. As time goes by, S(t) increases, thereby reducing the expected payoff.

Simulation: Risk Fluctuations Over Time

We calculate how the power cluster's payoff changes when the probability S(t) of being exposed grows over time, and identify the optimal moment to switch to explanation.

The following graph shows the variation in expected payoffs for concealment versus explanation as time advances:

a. Orange dashed line: Expected payoff if concealment is maintained (EU^{long})

b. Blue solid line: Expected payoff if explanation is adopted ($EU^{confess}$)

c. Red dotted line: Optimal point in time for switching to explanation (t^*)

a. Initial Phase (t < 4.5):

a. Concealment yields a higher expected pay- off, making it the more rational choice in the short run.

b. At Time (t = 4.5):

I. The expected payoffs for concealment and explanation invert.

II. This point is the optimal moment for switching to explanation, after which explanation becomes the more rational strategy.

Long-Term Consequences (t > 4.5)

I. Persisting in concealment amplifies risk to the extent that the expected payoff turns sharply negative.

II. Continued concealment over the long term could lead to the eventual collapse of the power cluster.

In summary, this analysis demonstrates that the risk the expected payoff from explanation EU confess of concealment increases over time for a power cluster. The conclusions are as follows:

I. Although concealment may be advantageous in the short run, as time passes, the probability of detection grows, and the expected payoff becomes less than what can be obtained by explanation.

II. If the power cluster is acting rationally, it will switch to explanation after a certain period.

III. Considering the payoffs of both the intermediate layer and the general public, excessive concealment can cause societal disruption and ultimately magnify losses in the medium to long term.

Simulation Overview

(Figure 2) This simulation compares the expected payoff of a

power cluster when it persists with concealment versus when it opts for explanation. Over time, the probability S(t) that concealment is uncovered increases, causing the expected payoff from concealment EU long to decline. Because the expected payoff from explanation EU confess remains constant, there exists some point in time at which explanation becomes more advantageous than concealment.

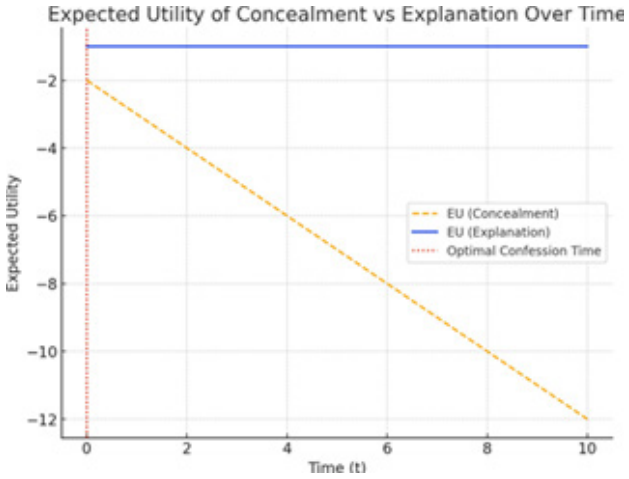

Figure 2: Expected Utility of Concealment vs. Explanation Over Time.

Based on the simulation results, the following points were noted:

a) The expected payoff from concealment EU long decreases over time, eventually dropping below once a certain time threshold has passed.

b) Under the parameters used for this simulation, the time t∗ at which the expected payoffs of concealment and explanation are equal was found to be t∗ = 0.0.

c) In other words, from the initial state onward, ex- planation already offers a higher expected payoff, leaving no window of time in which concealment is the rational choice.

The simulation thus reveals that any short-term gains from concealment rapidly diminish as the detection risk grows, making explanation the more favorable strategy in short order.

1. Concealment quickly becomes disadvantageous as the probability of detection rises.

Because the detection probability S(t) increases linearly, even a short duration significantly heightens risk.

Consequently, the benefits obtained by continuing concealment are extremely limited.

2. Explanation provides a more stable payoff.

Since explanation 's payoff is constant, there is no escalating penalty arising from in- creased risk.

Therefore, choosing explanation from the outset represents a more rational long-term strategy.

3. Employing concealment even as a short-term strategy is extremely risky.

Even over a brief period, the risk of being found out increases and drastically reduces concealment 's payoff.

A power cluster can evade such hazards and maintain social trust by opting for explanation early on.

While the above analysis treated S(t) as a simple linear function, more detailed modeling is possible by in- corporation additional factors:

I. Nonlinear growth in the probability of detection (e.g., exponential increases or threshold effects).

II. Gradual changes in concealment strategy (e.g., partial disclosure).

III. Reactions from the general public and the intermediate layer.

Overall, the simulation indicates that the expected payoff of concealment declines with time due to a growing risk of exposure, quickly rendering it less favorable than explanation. In particular, in the present model, choosing explanation from the outset emerges as the optimal strategy. This conclusion suggests that adopting concealment, even for short-term gains, poses very high risks and is not sustainable.

Deriving Action Recommendations to Maximize the General Public's Convergence Rate under Concealment

We now analyze the behavior of the general public that maximizes its payoff convergence rate when a power cluster employs concealment. Specifically, we use a mathematical model to deter-

mine which strategy yields the highest payoff for the general public, given that the probability of concealment being discovered increases over time.

Definition of the General Public 's Payoff Function

The general public 's payoff function is defined as follows:

$$U_G = A_G - B_G$$

where AG represents the positive payoff and BG is the negative loss. The general public 's payoff depends on the reliability of information provided by the power cluster, the accuracy of data from the intermediate layer, and each individual 's level of information literacy.

Factors Affecting the General Public 's Payoff

We identify the following factors that influence the general public 's payoff function:

a) **IP:** Reliability of information delivered by the power cluster ($0 \le IP \le 1$)

b) **IC:** Reliability of information provided by the intermediate layer ($0 \le IC \le 1$)

c) **EG:** Information literacy of the general public ($0 \le EG \le 1$)

Taking these into account, the general public 's pay- off is modeled as follows:

$$A_G = \alpha I_P + \beta I_C + \gamma E_G$$

$$B_G = \delta(1 - I_P) + \epsilon(1 - I_C) + \zeta(1 - E_G)$$

where α, β, γ, δ, ϵ, ζ are constants denoting the weight of each factor.

Deriving the Optimal Payoff Convergence Rate

To maximize the general public 's payoff, we need to maximize UG:

$$U_G = (\alpha I_P + \beta I_C + \gamma E_G) - (\delta(1 - I_P) + \epsilon(1 - I_C) + \zeta(1 - E_G))$$

Reorganizing:

$$U_G = (\alpha + \delta) I_P + (\beta + \epsilon) I_C + (\gamma + \zeta) E_G - (\delta + \epsilon + \zeta)$$

We define the general public 's optimal payoff con- vergence rate PS under conditions that maximize UG. We regard PS as the probability that the general public obtains accurate information:

$$P_S = f(I_P, I_C, E_G)$$

The optimal strategy for the general public to maximize its payoff converges on three key points:

i. Limit reliance on information from the power cluster.

When IP is low (i.e., when the power cluster is engaging in concealment), the general public must bolster the accuracy of information through improvements in IC or EG.

ii. Actively utilize information from the intermediate layer.

Maximizing the contribution of IC mitigates the influence of inaccurate data and reduces information fatigue.

iii. Elevate information literacy.

Increasing EG directly enhances UG and lowers the likelihood of misinterpretation losses. We conduct a simulation under this model, assuming specific parameter values.

Assumed Parameters

(Table 3) Under these assumptions, the general public's payoff function becomes:

$$U_G = 0.8 I_P + 0.6 I_C + 0.6 E_G - 0.9$$

Table 3: Parameters for the General Public 's Payoff Function.

Parameter	Value
α, δ	0.4
β, ϵ	0.3
γ, ζ	0.3

Interpretation of the Simulation Results

a. If IP < 0.5, improving IC or EG becomes indispensable.

b. A high IC value can compensate for a low IP, stabilizing the general public 's payoff.

c. Maintaining EG at or above 0.5 mitigates the impact of information fatigue and facilitates sound information gathering.

From these considerations, we derive behavioral guidelines that enable the general public to maximize its payoff. The principal findings are:

a) Even if IP decreases, ensuring a high IC preserves the integrity of information for the general public.

b) Strengthening the public 's information literacy (E_G) makes it easier to evaluate power-cluster content accurately, thereby reducing losses incurred by misunderstanding.

c) The simulation indicates that even when IP is low, strategic adjustments to IC and EG can boost the general public 's payoff in a stable manner.

The Effect of Offering Compensatory Payoffs by Power Clusters

We now turn to the scenario in which a power cluster chooses a concealment strategy, resulting in reduced payoffs for the intermediate layer and the general public. Specifically, we examine the "compensatory payoff" strategy, in which the power cluster offers additional benefits to these groups. Using a game theoretic model, we investigate how their behaviors change when such rewards are provided.

Compensatory Payoffs Offered by Power Clusters

By providing compensatory payoffs (R) to the intermediate lay-

er and the general public, a power cluster can alter the system in the following ways:

i. Intermediate Layer

a) Receives an incentive to collaborate with the power cluster rather than disseminate inde- pendent information.

b) However, if the compensatory payoff is in- adequate, it may still attempt its own distri- bution of information.

ii. General Public

Despite reduced trust in the cluster 's data, they may align with the power cluster 's viewpoint if economic or social advantages are sufficiently large.

Nevertheless, in the long run, degraded in- formation accuracy might spur further polarization of opinion.

New Payoff Functions

i. Power Cluster 's Payoff

Taking compensatory payoffs into account, the pay- off function for the power cluster is defined as:

$$EU_P = P(-b_1S)+(1-P)(-c_1S)-R_MM -R_GG$$

where

a) RM: Compensatory payoff to the intermediate layer

b) RG: Compensatory payoff to the general public

c) M: Degree of cooperation by the intermediate layer

d) G: Degree of support from the general public

ii. Intermediate Layer 's Payoff

We define the payoff function of the intermediate layer as:

$$EU_M = M(a_2L - b_2C) + (1 -M)(-d_2I) + RM$$

where

a) L: Gains from collaborating with the power cluster

b) C: Cost of independently disseminating information

c) I: Societal disruption caused by information manipulation

iii. General Public 's Payoff

The payoff function of the general public is ex- pressed as:

$$EU_G = G(a_3V - b_3C) + (1 - G)(-e_3I) + RG$$

where

i. V: Societal benefits obtained from accepting the power cluster 's information

ii. C: Cost of gathering and evaluating information

iii. I: Societal chaos caused by data manipulation

Simulation: The Influence of Compensatory Payoffs

We calculate the expected payoffs EU_P , EU_M , EU_G for each cluster under varying values of RM and RG. This allows us to identify the range of values that might be optimal.

i. Orange dashed line: Power cluster 's expected payoff (EU_P)

ii. Blue solid line: Intermediate layer 's expected payoff (EU_M)

iii. Green dotted line: General public's expected pay- off (EU_G)

Power Cluster (EU_P) Payoff

a) The cluster 's payoff diminishes as RM and RG rise.

b) In other words, there is a cost to maintaining concealment via compensatory payments.

c) Excessive compensation may drive the power cluster 's payoff into negative territory, rendering the concealment strategy unfeasible.

Intermediate Layer (EU_M) Payoff

i. Increasing RM improves the intermediate layer 's payoff.

ii. If RM surpasses a certain threshold, the intermediate layer gains a stronger incentive to support the concealment strategy instead of distributing its own information.

General Public (EU_G) Payoff

i. Raising RG boosts the general public 's pay- off.

ii. However, if RG is too low, confusion persists, and public trust in the information environment diminishes.

Here, we analyze how the actions of the intermediate layer and the general public shift when the power cluster offers compensatory payoffs. We draw the following conclusions:

i. Although the power cluster can provide some compensation to preserve short-term gains, such payments represent an added cost that makes long-term maintenance of concealment challenging.

ii. Determining an appropriate balance of RM and RG is crucial for sustaining the concealment strategy.

iii. Excessively high compensatory payoffs can push the power cluster 's payoff into negative values, eventually making explanation a more logical choice.

Simulation Overview

We compute the expected payoffs of the power cluster (EU_P), the intermediate layer (EU_M), and the general public (EU_G) when compensatory payoffs RM and RG are provided by the power cluster. By examining how these payoffs vary, we consider an optimal range for these compensations (Figure 3).

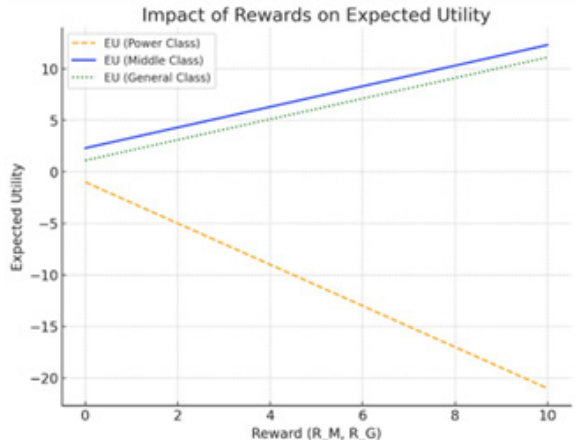

Figure 3: Impact of Rewards on Expected Utility.

Payoff for the Power Cluster (EU_P):

i. As RM and RG grow, the power cluster 's payoff decreases.

ii. Particularly large compensatory payments drive the cluster 's payoff into negative territory, making sustained concealment untenable.

Payoff for the Intermediate Layer (EU_M):

i. Increasing RM raises the intermediate layer' s payoff.

ii. Once RM exceeds a certain level, the incentive to support the concealment strategy rather than disseminate information independently intensifies.

Payoff for the General Public (EU_G):

i. RG enhances the public's payoff as it grows.

ii. However, if the compensatory payoff is in- sufficient, confusion remains and the credibility of information deteriorates.

i. A suitable setting of compensatory payoffs is essential:

i. While raising RM and RG increases the pay- offs for the intermediate layer and the general public, it reduces the payoff for the power cluster.

ii. By appropriately adjusting RM and RG, one can balance the cluster 's profits against the interests of other groups.

ii. The sustainability limits of the concealment strategy:

i. Paying excessively high compensatory pay- offs erodes the power cluster 's own pay- off, making it unfeasible to maintain concealment over the long run.

ii. Beyond a certain threshold, the power cluster finds it more rational to switch to an ex- planation strategy.

iii. Differentiating the responses of the intermediate layer and the general public:

i. The intermediate layer (EUM) is highly sensitive to changes in RM; when rewards are substantial, it is more likely to cooperate than to distribute information on its own.

ii. For the general public (EUG), larger RG boosts payoffs, but inadequate compensatory payoffs can precipitate social polarization.

Although this analysis is based on a simple expected- payoff model, the following factors could be incorporated for a more realistic approach:

a) A dynamic model in which the power cluster 's information manipulation evolves over time.

b) Nonlinear factors influencing how the intermediate layer and the general public respond to new information.

c) Social simulations integrating the effects of social networking platforms and media.

These results confirm that although a power cluster may enhance the payoffs of the intermediate layer and the general public by providing compensatory rewards, its own payoff declines accordingly. In particular, exorbitant compensation can negatively affect the cluster 's interests, rendering long-term concealment infeasible. Therefore, if a power cluster employs a concealment strategy, it must carefully weigh short-term profits against long-term costs. Future research should investigate how best to determine the optimal level of compensatory payoffs in dynamic settings, including policy recommendations that account for the broader flow of information within society.

Deriving Action Suggestions for the General Public to Maximize Convergence Rate

Using a mathematical model, we examine the optimal behavioral strategy for the general public when a power cluster employs a concealment strategy yet offers compensatory payoffs. Specifically, we explore how these rewards alter the actions of each cluster.

Compensatory Payoffs Provided by the Power Cluster

By granting compensatory payoffs (R_M, R_G) to the intermediate

layer and the general public, a power cluster induces the following changes:

Intermediate Layer

i. Gains an incentive to cooperate with the power cluster rather than disseminate information independently.

ii. However, if the payoff is insufficient, the intermediate layer may attempt its own distribution of information.

General Public

i. Although trust in the power cluster may be diminished, people may follow the cluster' s position if they obtain economic or social benefits.

ii. In the long run, however, the accuracy of information may decline, and polarization of opinion could intensify.

Defining New Utility Functions

a) Power Cluster 's Utility

Taking compensatory payoffs into account, the power cluster 's utility function is defined as follows:

$$EU_P = P(-b_1S)+(1-P)(-c_1S)-R_MM -R_GG$$

where

I. R_M: Compensatory payoff to the intermediate layer

II. R_G: Compensatory payoff to the general public

III. M: Degree of cooperation from the intermediate layer

IV. G: Extent to which the general public follows the power cluster

b) Intermediate Layer 's Utility

The intermediate layer 's utility function is given by:

$$EU_M = M(a_2L - b_2C) + (1 -M)(-d_2I) + R_M$$

where

i. L: Gains from collaborating with the power cluster

ii. C: Cost associated with independently distributing information

iii. I: Societal confusion caused by manipulative in- formation

c) General Public 's Utility

The general public 's utility function is expressed as:

$$EU_G = G(a_3V - b_3C) + (1 - G)(-e_3I) + R_G$$

where

I. V: Social benefits gained from adopting the power cluster 's information

II. C: Cost of information gathering and evaluation

III. I: Societal turmoil arising from manipulated data

Simulation: Effect of Compensatory Payoffs

We vary the compensatory payoffs (R_M, R_G) and calculate the expected utility (EU_P, EU_M, EU_G) for each cluster to identify which values might be optimal.

I. Orange dashed line: Power cluster's expected utility (EU_P)

II. Blue solid line: Intermediate layer's expected utility (EU_M)

III. Green dotted line: General public 's expected utility (EU_G)

Power Cluster (EU_P)

i. Increasing RM or RG reduces the power cluster 's utility.

ii. Therefore, there is a cost to maintaining concealment.

iii. Excessive compensatory payoffs can drive the power cluster 's utility below zero, undermining the concealment strategy.

Intermediate Layer (EU_M)

i. Raising RM curbs independent dissemination, increasing the incentive to collaborate with the power cluster.

ii. If RM is too small, the intermediate layer may opt to share information on its own.

General Public (EU_G)

I. A higher RG helps reduce confusion and enhances trust.

II. If the payoff is insufficiently large, public trust diminishes, fostering social polarization.

We analysed how providing compensatory payoffs to the intermediate layer and the general public changes each cluster 's decision-making. The conclusions reached are:

I. Although the power cluster can offer such rewards to maintain short-term utility, the corresponding costs make prolonged concealment challenging in the long run.

II. Finding an appropriate balance of RM and RG is crucial to sustaining the concealment strategy.

III. Excessively generous compensatory payoffs can push the power cluster 's own utility into negative territory, making an explanation strategy more rational.

Scenario After Information Leakage from the Power Cluster to the Intermediate Layer

When internal data from the power cluster leaks to the intermediate layer, three key developments occur:

I. Intermediate Layer 's Action Choice: "Dissemination" vs. "Cautious Alignment"

II. Formation of Trust Among the General Public

III. Internal Conflicts Within the Power Cluster

In this model, each cluster 's decision-making process is formalized mathematically and examined through simulation.

Action Model for the Intermediate Layer

The intermediate layer selects between disseminating the information and aligning itself more cautiously with the power cluster.

Dissemination Strategy

$$EU_M(\text{dissemination}) = a_2L - b_2C + R_M - d_PP$$

I. a_2L: Value derived from circulating information

II. b_2C: Cost of social turmoil

III. R_M: Reward from the power cluster

IV. d_P P: Pressure or potential retaliation from the power cluster

Cautious Alignment Strategy

$$EU_M(\text{alignment}) = -d_2I + R_M + e_PP$$

I. $-d_2I$: Negative impact on data integrity due to withholding information

II. e_P P: Protection from the power cluster

Threshold for Action Selection

The condition for choosing dissemination is:

$$a_2L + d_2I > b_2C + (d_P + e_P)P - R_M$$

Summary

I. A higher power cluster pressure P inclines the intermediate layer toward caution.

II. Greater data integrity I increases the likelihood of dissemination.

III. A larger reward RM from the power cluster curbs dissemination.

Trust Formation Model for the General Public

Whether the general public believes the information depends on its quality I and the degree of dissemination M.

$$EU_G = G(a_3V - b_3C) + (1 - G)(-e_3I)\,b$$

Determination of Trust

$$G > \frac{-e_3I}{a_3V - b_3C + e_3I}$$

Discussion

I. As I (information integrity) increases, the public' s trust G also rises.

II. When social disruption C is large, G declines.

Internal Conflict Model of the Power Cluster

Once information leakage is discovered, internal strife arises within the power cluster, leading to reduced influence.

Conflict Costs

$$EU_P = -bl_S - f_P\,T$$

1. **f_P T:** Cost of internal discord

2. **T:** Intensity of the internal dispute

Discussion

a) Severe internal conflict undermines the power cluster 's capacity to disseminate information.

b) Infighting reduces media dominance, thus making independent dissemination by the intermediate layer more feasible.

Simulation: Changes in Expected Utility for Each Cluster

We perform simulations by assuming specific parameter values.

Assumed Parameters:

Table 4

Table 4: Parameter Settings for the Simulation.

Parameter	Value
d_p	2 (Suppression effect of power cluster pressure on the intermediate layer)
e_p	1 (Protective effect provided by the power cluster)
e_3	2 (General public 's sensitivity to information integrity)
I	0.8 (Initial level of information integrity)
f_P	3 (Cost of internal conflict within the power cluster)

a) Power Cluster 's Utility: Rises in pressure P and internal conflict T reduce its utility.

b) Intermediate Layer 's Choice: Absent extreme pressure, dissemination is frequently chosen.

c) General Public 's Level of Trust: Escalating social turmoil C decreases trust G.

d) The intermediate layer generally Favors disseminating information.

e) As internal conflict within the power cluster deepens, the ability to control information declines.

f) The public 's trust depends heavily on the inter- mediate layer 's level of dissemination.

Scenario After Information Leakage from the Power Cluster to the Intermediate Layer

Here, we calculate each cluster 's expected utility as power-cluster pressure P and internal discord T vary, analyzing the resulting outcomes.

Expected Utility for the Intermediate Layer:

We compare expected utility under the "dissemination" and "cautious alignment" strategies. Our results show that as the

power cluster 's pressure P grows, the expected utility of dissemination drops while the utility of alignment relatively increases. This shift reflects the intermediate layer's recalibrated incentive structure under strict power cluster control.

$$EU_M(\text{dissemination}) = 2_I - 2_P$$

$$EU_M(\text{alignment}) = 2_I + 2_P$$

The above indicates that the superiority of dissemination hinges largely on the level of information integrity I. Specifically, while high I boost the utility of dissemination, rising P may make cautious alignment the more rational decision.

Expected Utility for the Power Cluster:

The power cluster 's utility depends on internal conflict T. If the dispute intensifies, its overall control weakens, diminishing its utility:

$$EU_P = -f_P T$$

These results imply that unless the power cluster manages internal friction, it risks losing its influence in the long run (Figure 4).

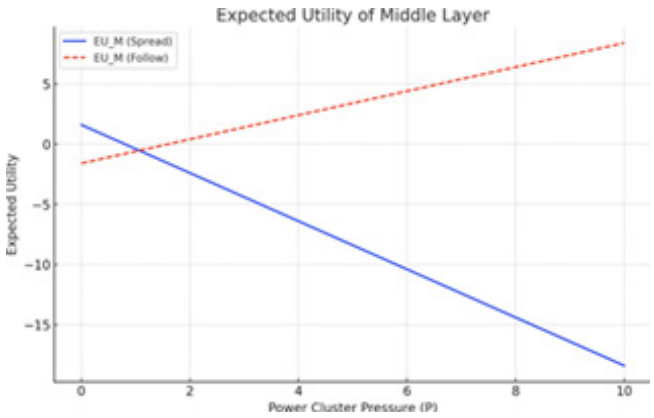

Figure 4: Intermediate Layer 's Expected Utility (EUM) vs. Power Cluster 's Pressure (P).

In sum, our analysis of how the power cluster's pressure and internal discord affect both the intermediate layer 's choice of strategy and the public 's trust leads to the following conclusions:

a) If information integrity is high, the intermediate layer tends to disseminate it. However, stronger power-cluster pressure makes cautious alignment more appealing (Figure 5).

b) As internal conflict escalates within the power cluster, its capacity to maintain control erodes, compromising its influence.

c) The general public 's trust depends on the degree of information sharing by the intermediate layer; intensifying social upheaval decreases trust.

Figure 5: Power Cluster 's Expected Utility (EUP) vs. Internal Conflict (T).

Action Suggestions for Maximizing the General Public 's Payoff Convergence Rate

Finally, we determine how the general public can attain optimal utility by examining the relationship be- tween information integrity I and trust G in more detail.

Deriving the Optimal Level of Trust

The general public 's expected utility is represented as follows:

$$EU_G = G(a_3V - b_3C) + (1 - G)(-e_3I)$$

To identify the condition under which the public chooses to trust, we solve for the optimal value of G:

$$G(a_3V - b_3C) > (1 - G)(-e_3I)$$

Rewriting:

$$G > \frac{-e_3I}{a_3V - b_3C + e_3I}$$

This equation shows how trust G varies as a function of information integrity I.

We perform a simulation under the following assumptions:

From the simulation, the following points emerge (Table 5).

Table 5: Parameter Settings for the Simulation.

Parameter	Value
d_P	1.5 (Value of information)
b_3	1.0 (Cost of societal unrest)
e_3	2.0 (Sensitivity to the reliability of information)

a) When the level of information integrity I is high, the public 's trust G tends to increase correspondingly.

b) If societal confusion intensifies, trust declines, curbing the public 's willingness to accept information.

c) Thus, elevating information integrity can boost the general public 's payoff convergence rate.

Optimal Responses by the Power Cluster and Strategies of the Intermediate Layer

Three Choices:

After information leakage, the power cluster can choose among the following three responses:

a) Pressure Strategy (Intensify P)

a) Strengthen pressure on the intermediate layer and the general public to suppress further dissemination.

b) However, greater pressure increases social unrest (C), reducing the general public 's payoff.

b) Adjustment Strategy (Partially Endorse the Leaked Information + Provide Compensation to the General Public)

a) Collaborate with the intermediate layer, formally acknowledging portions of the leaked information while offering a reward (RG) to the general public.

b) In this way, it may be possible to limit the scope of dissemination by the intermediate layer.

c) Split Strategy (Shift Responsibility for the Leakage to Another Faction)

I. Pin blame on a specific faction within the power cluster in an effort to regain trust.

II. However, this approach raises internal conflict (T) and entails considerable long-term risk.

Utility Functions for Each Strategy

I. Pressure Strategy

$$EU_P = -b_1S - d_PP - f_P\ T$$

Higher pressure means an increased value of P, but it also intensifies the conflict T.

II. Adjustment Strategy

$$EU_P = -b_1S + e_PM - R_GG$$

Setting an appropriate RG to mitigate social unrest (C) is crucial.

III. Split Strategy

$$EU_P = -b_1S - f_P\ T + gP\ (1 - T)$$

If factional rifts deepen, T may become too large to manage, potentially resulting in a loss of control.

I. In the short term, the "Split Strategy" has an advantage: If internal conflict (T) is still low, the payoff is higher.

II. Once T > 0.5, the "Adjustment Strategy" is superior: As conflict escalates, the power cluster 's dominance weakens, making an adjustment approach more effective.

III. The Pressure Strategy is not optimal: Applying strong pressure risks provoking a complete backlash from the intermediate layer.

Optimal Strategy for the Intermediate Layer Two Options

The intermediate layer decides whether to disseminate (M = 1) or align (M = 0).

EU_M (dissemination) > EU_M (alignment)

$$a_2L - b_2C > -d_2I + e_PP$$

Rewriting,

M = 1 (disseminate) if $a_2L+d_2I > b_2C+e_PP$

I. In most cases, dissemination is optimal: Dissemination consistently offers a higher payoff for the intermediate layer (EUM), providing little incentive to align.

II. **If the power cluster adopts an Adjustment Strategy, alignment may occur:** If an appropriate reward (RM) is offered, the incentive to follow the cluster grows.

Responses by the Power Cluster

a) In the short run, the Split Strategy yields greater benefit, but excessive internal conflict can lead to failure.

b) Over the long term, the Adjustment Strategy is optimal, partially disclosing information and co- operating with the intermediate layer.

c) The Pressure Strategy is of limited effective- ness and may even accelerate dissemination.

Responses by the Intermediate Layer

a) Generally, the Dissemination Strategy is more advantageous, so even under pressure, sharing information tends to yield higher returns.

b) Nevertheless, if the power cluster offers sufficient reward (RM), the intermediate layer may opt to align.

Optimal Response by the Power Cluster and the Strategy of the Intermediate Layer

Below we compare the expected payoffs from the three strategies (Pressure, Adjustment, Split) that the power cluster can deploy and determine which is the most logical in each scenario.

Comparing the Power Cluster 's Expected Payoffs

We define each strategy 's expected payoff as follows:

Pressure Strategy:

$$EU_P = -b1 - d_P T - f_P T$$

Adjustment Strategy:

$$EU_P = -b_1 + e_P (1 - T) - R_G T$$

Split Strategy:

$$EU_P = -b_1 - f_P T + gP (1 - T)$$

where

a) b1: Base cost to the power cluster

b) d_P: Suppressive cost from applying pressure

c) f_P: Cost arising from internal disputes

d) g_P: Gains from assigning blame externally

e) e_P: Benefits of cooperating with the intermediate layer

f) RG: Compensation to the general public

(Figure 6) Simulation results yield the following conclusions:

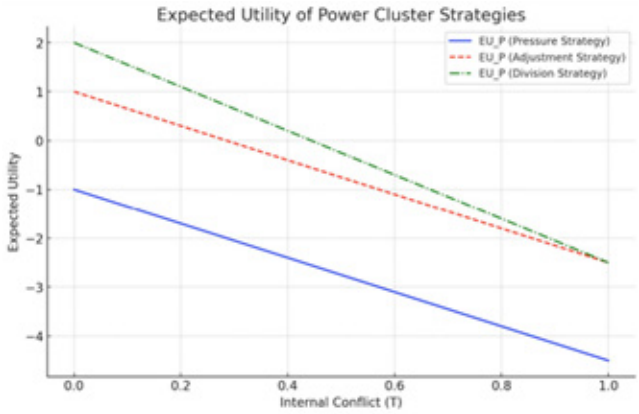

Figure 6: Power Cluster 's Expected Payoff (EU_P) by Strategy vs. Internal Conflict (T).

a) Short-term superiority of the Split Strategy: When internal conflict (T) is small, this strategy yields the highest payoff for the power cluster.

b) Once T > 0.5, the Adjustment Strategy becomes more advantageous: If factional strife grows too intense, the cluster 's control diminishes and high-pressure tactics are no longer viable, making adjustment more favorable for curbing social instability.

c) The Pressure Strategy has limited utility: Even if the cluster escalates pressure, rising resistance from the intermediate layer and the public makes long-term gains unlikely.

The power cluster must select the most suitable strategy according to the circumstances. Although the Split Strategy yields immediate gains, persistent internal conflict amplifies longer-term risk, making a shift toward the Adjustment Strategy necessary.

Suggestions for Maximizing the General Public 's Convergence Rate

We now examine the relationship between the power cluster 's strategy and the general public 's behavior to derive a formu-

la-based analysis and conduct simulation-based discussion, aiming to determine the actions that yield the highest payoff for the public.

Consensus Formation Model for the General Public

Whether or not the general public trusts the information depends on both the intermediate layer 's dissemination (M) and the power cluster 's influence (P).

$$EU_G = G(a_3V - b_3C) + R_G$$

The condition under which the general public accepts the information is expressed as follows:

$$G > \frac{-e_3I}{a_3V - b_3C + e_3I}$$

I. a3: Benefits from accepting the information

II. b3: Cost of social unrest

III. e3: Degree to which information integrity matters

IV. I: Initial reliability of information

V. V: Consistency of the data

VI. G: General public 's level of trust

VII. If RG < 2.6, the general public's expected payoff (EUG) remains negative, and they do not accept the information.

VIII. Offering a payoff of RG ≥ 2.6 persuades the public to accept it.

IX. Reducing social turbulence (C) and improving consistency (V) can bolster trust.

X. When RG ≥ 2.6, it is possible to secure the public 's acceptance.

XI. Increasing transparency while implementing measures to quell unrest is vital.

XII. Enhancing the consistency of the information (V) is essential for maintaining public trust.

Transition from the Split Strategy to the Adjustment Strategy and an Uncontrollable Scenario

Simulation findings indicate that once internal conflicts within the power cluster (T) exceed 1.5, control collapses.

Points of Losing Control

a) The split strategy remains sustainable as long as T ≤ 1.0.

b) Once T > 1.5, intensifying internal discord undermines the power cluster 's dominance.

c) Therefore, the shift to an adjustment strategy must begin by at least T = 1.0.

Strategies to Prevent Societal Breakdown and Turmoil in Cases of Divergent Language Spheres or Information Pathways

As the power cluster switches to the adjustment strategy, confusion can emerge from differences in language spheres and communication channels. We define the degree of acceptance of information as A, which declines the greater the linguistic or cultural gap.

$$A = A_0 - k_L L - k_C C$$

a) A_0**:** Initial level of acceptance

b) **L:** Extent of linguistic barriers (larger values indicate bigger differences)

c) **C:** Disparities in information pathways

d) k_L, k_C**:** Coefficients capturing how language and channel divergences reduce acceptance

Avoiding Social Turmoil

According to the simulation, larger values of L (language differences) and C (pathway differences) diminish acceptance (A), triggering unrest in society.

Conditions Under Which Social Turmoil Arises

a) When L + C > 4.0, the acceptance level A drops below 0.2, fuelling social destabilization.

b) Substantial linguistic and pathway differences make the power cluster 's adjustment strategy more prone to failure.

Strategies for Preventing Turmoil

a) Prepare a multilingual media plan in advance to maintain consistency of information.

b) Regulate information channels to minimize contradictions among different sources.

c) Provide the intermediate layer with accurate data to avert misunderstandings.

Discussion on Maximizing Overall Social Stability

The general public is affected by both (1) internal conflicts in the power cluster and (2) variation in language spheres and communication pathways.

Defining Social Stability S

We define S (the overall stability of society) with the following model:

$$S = S_0 - f_T T - f_A A - f_C C$$

a) S_0**:** Initial level of social stability

b) f_T**:** Influence of internal cluster conflicts (T) on social unrest

c) f_A**:** Extent to which diminishing acceptance (A) affects stability

d) f_C**:** Impact of diverging communication pathways on society

Conditions for Maximizing Stability

Society remains stable if

$$S > S_{min}$$

In other words,

$$S_0 - f_T T - f_A A - f_C C > S_{min}$$

Strategies for Maximizing Overall Social Stability

Simulation findings show that social stability (S) falls below 3.0 once society becomes precarious.

Conditions Leading to Instability

a) When T > 1.5, severe internal conflicts within the power cluster reduce stability.

b) If A < 0.3, divergences in language or information channels lower acceptance, causing instability.

c) When C > 3.0, social disorder progresses due to conflicting information pathways.

Strategies for Bolstering Social Stability

a) Contain Internal Conflicts in the Power Cluster

a) Shift to an adjustment strategy while T ≤ to prevent further splits.

b) Control internal strife and achieve organizational consensus before reaching a point of no return.

b) Multilingual Approaches and Information Control

I. Introduce a multilingual media strategy so that A ≥ 0.5 can be maintained.

II. Strive to limit channel-based discrepancies and boost acceptance.

c) Appropriately Set Compensation (R_G) for the General Public

I. Offer at least $R_G \geq 2.6$ so that the general public is more inclined to accept the information.

Detailed Design of the Adjustment Strategy and Social Stability: Transition from the Split Strategy to an Uncontrollable Scenario

Simulation results also confirm that once the power cluster 's internal conflict (T) exceeds 1.5, it spirals out of control.

Key Factor in Losing Control

a) The split strategy can persist while T ≤ 1.0.

b) Once T > 1.5, intensifying factional disputes destroy the power cluster 's grip on power.

c) Thus, shifting to the adjustment strategy must happen by or before T = 1.0.

Discussion on Maximizing Overall Social Stability

The general public experiences direct repercussions from (1) the power cluster 's internal conflicts and (2) disparities in language and information pathways (Figure 7).

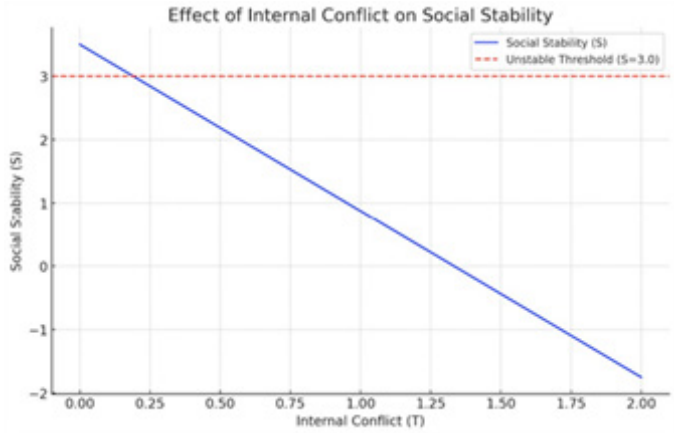

Figure 7: Relationship Between Internal Conflict (T) and Social Stability (S).

Defining Social Stability S

We represent societal stability S through the model:

$$S = S_0 - f_T T - f_A A - f_C C$$

a) S_0 = 5.0: The initial level of social stability

b) f_T = 2.0: Influence of internal cluster strife

c) f_A = 1.5: Sensitivity to decreased acceptance

d) f_C = 1.0: Effect of diverging communication pathways

Simulation Results on Deteriorating Stability

a. When T > 1.5, mounting internal cluster conflicts diminish stability.

b. If A < 0.3, lower acceptance, driven by differences in language or channels, destabilizes society.

c. For C > 3.0, contradictory pathways incite social unrest.

Strategies to Maximize Stability

a) Keep Internal Conflicts in the Power Cluster Under Control

Move to the adjustment strategy by T ≤ 1.0 to forestall further fragmentation.

Manage internal struggles and reach a consensus before losing control.

b) Multilingual Solutions and Information Regulation

I. Implement multilingual measures so that A ≥ 0.5 can be maintained.

II. Reduce communication-path conflicts and strengthen acceptance.

c) Set Appropriate Compensation (R_G) for the General Public

I. Provide at least RG ≥ 2.6 so that the public is receptive to the information.

II. Shift to the adjustment strategy before T > 1.5.

III. Employ a multilingual, multi-channel control pol- icy to maintain acceptance A.

IV. Offer suitable returns (RG ≥ 2.6) to the general public, achieving consensus.

Optimal Combination of Complete vs. Incomplete Information Games and Cooperative vs. Non- cooperative Games: Defining the Social Stability Model

Assume that social stability S depends on the following factors:

$$S = S_0 - f_T T - f_A(1 - A) - f_C C$$

where

I. S0: Initial (maximum) level of social stability

II. fT: Degree to which internal conflict (T) induces social unrest

III. fA: Impact on stability arising from a decrease in acceptance (A)

IV. fC: Effect of divergent communication pathways

V. T: Internal conflict within the power cluster

VI. A: Acceptance level among the general public

VII. C: Confusion stemming from channel discrepancies

Calculating the Proportions of Complete- Information and Incomplete-Information Games

Denote the proportion of complete-information games by p and of incomplete-information games by 1 − p. We posit that a higher share of complete- information games increase the transparency of data, boosting social stability:

Complete = p, Incomplete = 1 − p

We then define

$$T = (1 - p)T_0$$

and

$$A = A_0 + k_p p$$

Calculating the Proportions of Cooperative and Non-cooperative Games

Let q be the share of cooperative games, and 1 − q

the share of non-cooperative ones:

Coop = q, NonCoop = 1 − q

A higher proportion of cooperative games reduces social schisms, thereby improving stability:

$$C = (1 - q)C_0$$

Deriving Social Stability S

By substituting these definitions into our model for social stability:

$$S = S_0 - f_T (1-p)T_0 - f_A(1-(A_0+k_p p)) - f_C(1-q)C_0$$

Rewriting,

$$S = S_0 - f_T T_0 + f_{Tp} T_0$$

$$- f_A + f_A A_0 + f_A k_p p$$

$$- f_C C_0 + f_C q C_0$$

Threshold for Stability and Deriving Optimal Pro- portions

Society remains stable if

$$S \geq S_{min}$$

which, after rearranging, becomes

$$(f_T T0 + f_A k_p)p + f_{Cq} C_0 \geq S_{min} + S_0 + f_T T_0 + f_A \cdot f_A A_0 + f_C C$$

Using this condition, we solve for optimal values of p (the share of complete-information games) and q (the share of cooperative games).

Results and Discussion

The following optimal proportions emerged from our calculations:

(Table 6) From these results, we draw the following conclusions:

Table 6: Optimal Proportions for Complete- Information and Cooperative Games.

Item	Value
Proportion of Complete- Information Games (p)	0.17 (17.3%)
Proportion of Cooperative Games (q)	0.69 (68.8%)

I. An optimal share of complete-information games is around 17%.

II. An optimal share of cooperative games is around 69%.

III. Up to 30% of non-cooperative games is tolerable before the risk of destabilization rises significantly.

IV. Strengthening cooperative relationships con- tributes more to societal stability than merely increasing information disclosure.

Thus, the following prerequisites are key to maximizing social stability:

I. Maintaining about 17–20% of games in a complete-information setting

II. Keeping about 65–70% of games in a cooperative mode

III. Restricting the non-cooperative share to under 30% to enhance social stability (Figure 8).

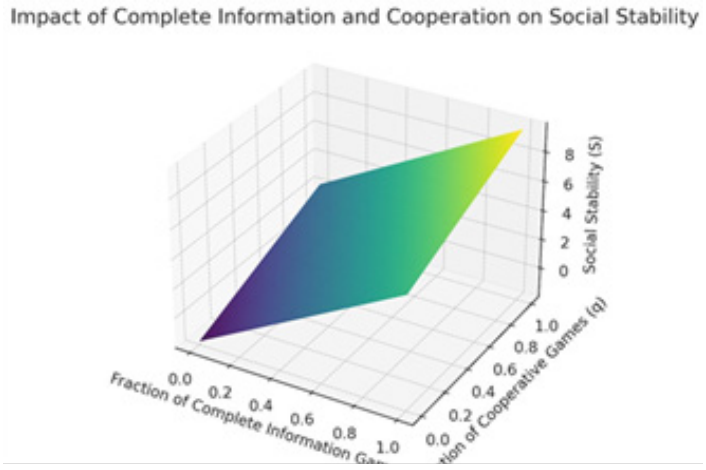

Figure 8: Impact of Complete Information and Cooperation on Social Stability.

Changes in Social Stability in Response to Complete-Information and Cooperative Games

Simulation outcomes demonstrate how fluctuations in the share of complete-information games (p) and the share of cooperative games (q) affect social stability (S).

Influence of Complete-Information and Cooperative Games

Analysis of these graphs indicates the following:

I. Effect of the Proportion p of Complete-Information Games

I. Increasing the share of complete-information games generally enhances S, since transparency mitigates internal strife (T).

II. However, even a very high p cannot drive social stability above a certain threshold if the proportion of cooperative games is too low.

II. Effect of the Proportion q of Cooperative Games

a. As the share of cooperative games increases, social stability S rises sharply.

b. This is because confusion in communication channels (C) subsides, and acceptance by the general public grows.

c. Conversely, even if p is high, low q can limit improvements in stability.

Interaction Between Complete-Information and Cooperative Games

a. When both p and q increase simultaneously, the highest stability levels are attained.

b. This results from transparency and cooperation operating in tandem, minimizing both internal conflict (T) and confusion in information path- ways (C).

Conditions for Exceeding the Threshold $S_{min} = 5$

Evaluating how social stability compares to the lower limit $S_{min} = 5$ reveals:

$$(f_T T0 + f_A k_p)p + f_{Cq} C_0 \geq S_{min} \cdot S_0 + f_T T_0 + f_A \cdot f_A A_0 + f_C C_0$$

Calculations show that the following must be satisfied for society to remain stable:

a. At least 17% of games must be complete information

b. At least 65% of games must be cooperative

Thus, not only is transparency important strengthening cooperative relationships is likewise vital for societal stability.

Policy Recommendations for Maximizing Social Stability Based on Simulation Results

Based on the above results, we will summarize the arguments for maximizing social stability.

I. Enhance Transparency without Insisting on Full Disclosure

i. Set the proportion of complete- information games at 17–20% to maintain a moderate level of information openness.

ii. Over-disclosing information can trigger con- fusion; therefore, careful information management is essential.

II. Prioritize the Promotion of Cooperative Games

i. Raise the share of cooperative games to 65–70%, limiting non-cooperative games to no more than 30%.

ii. Facilitate social dialog, policy coordination, and decision-making that encourages civic participation.

III. Tolerate Some Non-Cooperative Games While Preventing Their Excess

i. Maintain a degree of competitive principles but cap the portion of non-cooperative games at a maximum of 30% to avert societal fragmentation.

IV. Manage Diversity in Media and Communication Channels

i. Discrepancies in information pathways pose the gravest risk to social stability.

ii. Strengthen coordination mechanisms among diverse information sources to reduce misinformation and confrontations.

iii. Setting the proportion of complete-information games at about 17–20% is recommended.

iv. A 65–70% share of cooperative games is crucial for maximizing social stability.

v. Promoting deeper social cooperation is more vital than expanding information disclosure.

vi. Pre-empting confusion in communication path- ways helps elevate acceptance A and thus fosters long-term stability.

Building on these findings, further fine-tuning of pol- icy design and additional simulations can enhance the optimization of social stability.

Optimal Convergence Rate for the General Public s Payoff

Taking the above discussion into account, we mathematically organize how the optimal mix of complete and incomplete information games, as well as cooperative and non-cooperative games, can be structured in a way that considers social stability.

Social Stability Model

Social stability S is defined as follows:

$$S = S_0 - f_T T - f_A(1 - A) - f_C C$$

where

1. S_0: The initial (maximum) level of social stability
2. f_T: Extent to which internal conflict (T) amplifies social unrest
3. f_A: Extent to which a decline in acceptance (A) undermines stability
4. f_C: Impact of diverging communication path- ways on society
5. **T:** Internal power-cluster conflict
6. **A:** The public 's acceptance level of information
7. **C:** Turmoil arising from dissimilar information pathways

Effects of Complete vs. Incomplete Information Games

We consider the proportions of complete- information games (p) and incomplete-information games (1 – p):

$$T = (1 - p)T_0$$

Because an increase in complete-information games raises acceptance A,

$$A = A_0 + k_p p$$

where k_p denotes the degree to which the proportion of complete-information games augments acceptance.

Effects of Cooperative vs. Non-Cooperative Games

We also account for the share of cooperative games

(q) and non-cooperative games (1 – q):

$$C = (1 - q)C_0$$

A higher share of cooperative games reduces confusion C arising from communication pathways.

Deriving Social Stability S

By substituting these relationships into our funda- mental equation for social stability:

$$S = S_0 - f_T (1 - p)T_0 - f_A(1 - (A_0 + k_p p)) - f_C(1 -$$

$$= S_0 - f_T T0 + f_{TP} T_0 - f_A + f_A A_0 + f_A k_p p - f_C C_0 + f_C q C_0$$

Rewriting:

$$S = S_0 - f_T T_0 - f_A + f_A A_0 - f_C C_0 + (f_T T_0 + f_A k_p)p + f_C q C_0.$$

Threshold of Social Stability

For society to remain stable, we require

$$S \geq S_{min}$$

Therefore,

$$(f_T T_0 + f_A k_p)p + f_C q C_0 \geq S_{min} - S0 + f_T T_0 + f_A - f_A A_0 + f_C C_0.$$

Optimal Proportions for Complete-Information and Cooperative Games

$$(f_T T_0 + f_A k_p)p + f_C q C_0 \geq S_{min} - S0 + f_T T_0 + f_A - f_A A_0 + f_C C_0.$$

When running numerical simulations with this equation, we arrive at the following optimal proportions:

Thus, certain conditions are crucial for maximizing social stability:

I. The share of complete-information games from 17% to 20% is appropriate.

II. A cooperative-game ratio of about 65% to 70% is also suitable.

III. Keeping the proportion of non-cooperative games

Bayesian Game Analysis of Social Stability: Basic Model Definition

Here, we examine how the general public (G) infers the power cluster 's (P) reliability, along with how that inference influences social stability, all within a Bayesian-game framework.

Setting of Players

There are two players:

I. **Power Cluster (P):** Provides information to society, but its reliability is uncertain.

II. **General Public (G):** Estimates P 's reliability and decides whether to trust or distrust.

The power cluster is assumed to have two types:

I. **Honest Type (TG):** Delivers beneficial information to society.

II. **Dishonest Type (TB):** Engages in concealment or manipulation of data.

The general public believes with probability θ that P

is honest, and with probability 1 – θ that it is dishonest:

$$P(T_G) = \theta, P(T_B) = 1 - \theta.$$

The general public 's action choices are:

I. **Trust:** Accept information from the power cluster.

II. **Distrust:** Disregard the power cluster and make independent assessments.

Expected Utility of the General Public

The general public 's expected utility is computed separately for each type of power cluster:

$$EU_G(\text{Trust}) = \theta U_G(TG) + (1 . \theta)U_G(T_B), EU_G(\text{Distrust}) = U_G(D).$$

The public will choose to trust if

$$\theta(U_G(T_G) - U_G(T_B)) \geq U_G(D) . U_G(T_B).$$

Defining θ^*:

$$\theta^* = \frac{U_G(D) - U_G(T_B)}{U_G(T_G) - U_G(T_B)}.$$

Therefore, if $\theta \geq \theta^*$, the general public trusts the power cluster.

The Power Cluster 's Strategy Choices

The power cluster decides whether to provide honest information (Honest) or to conceal (Hide):

$$EU_P(\text{Honest}) = \theta U_P(T_G) + (1 . \theta)U_P(T_B), EU_P(\text{Hide}) = U_P(H).$$

Honest disclosure is chosen if

$$\theta(U_P(T_G) - U_P(T_B)) \geq U_P(H) . U_P(T_B).$$

Defining θ_p^* :

$$\theta_p^* = \frac{U_P(H) - U_P(TB)}{U_P(TG) - U_P(TB)}$$

If $\theta \geq \theta \geq \theta_p^*$, the power cluster acts honestly.

Bayesian Equilibrium

In a Bayesian game, we identify a point at which the general public 's threshold for trust, θ∗, coincides with the power cluster 's threshold for honesty, θ_p^* :

$$\theta^* = \theta_p^*$$

When this condition holds, the system reaches an equilibrium in which both players 'strategies stabilize.

Effects on Social Stability

Social stability S can be modeled as a function of the probability θ that the power cluster is honest:

$$S = S_0 + k_\theta(\theta - \theta^*).$$

Society remains stable if

$$S \geq S_{min},$$

i.e.,

$$\theta \geq \theta^* + \frac{Smin - S0}{k\theta}.$$

Key variables are as follows:

I. S_0: Base level of social stability (set to 5 here)

II. k_θ: Degree to which honesty affects stability (k_θ = 10 in this simulation)

III. **θ*:** Trust threshold for the general public

IV. **θ:** Probability of the power cluster being honest

V. **S_{min}:** Minimum threshold for stable societal conditions (S_{min} = 3 here)

Relationship Between Social Stability and Honesty

From the figure, we note the following:

1. When the power cluster 's honesty probability θ exceeds θ∗, social stability S increases.
2. If θ is below θ∗, stability falls short of S0 and can sink below S_{min}, destabilizing society.
3. 3While higher θ further stabilizes society, marginal returns diminish above a certain point.

Points Where Social Stability Unravels

I. The red dashed line at S_{min} = 3 indicates the lowest permissible stability level.

II. If θ < θ∗, S might fall under S_{min}.

III. Consequently, public trust breaks down, acceptance wanes, and societal unease is likely to escalate.

Measures to Safeguard Social Stability

From these simulation results, the following policies appear effective in boosting social stability:

Enhancing the Power Cluster 's Honesty

I. Increase information transparency.

II. Establish independent oversight mechanisms to improve perceived trustworthiness.

III. Strengthen incentives for accurate information sharing with the public.

Raising the General Public 's Trust Level

I. Expand media literacy education to pre-empt mis information.

II. Boost the public's capacity to vet reliable sources.

Lowering the Threshold θ*

I. Improve confidence in the power cluster 's credibility, such that the general public accepts lower explicit proofs of honesty.

In conclusion, the power cluster 's honesty probability θ surpassing the threshold θ∗ is a necessary condition for maintaining social stability. When honesty is insufficient, public distrust grows, destabilizing society. Future research may consider real-world factors (policy interventions, the impact of information dissemination, etc.) to refine this model further.

Convergence Conditions for "Flaming" Debates in Online Spaces

In online spaces, the power cluster and its affiliated intermediate layer may engage in heated arguments "(flaming") with the general public. Possible tactics to end such debates include consensus-building, terminating discussions, or deleting posts. This study formalizes the convergence model mathematically to identify the conditions under which flaming subsides.

Basic Model for Discussion Convergence

We define the following players in an online flaming debate:

I. Intermediate Layer (M): Individuals who sup- port the power cluster 's perspective.

II. General Public (G): Individuals who oppose the power cluster.

III. Third-Party Moderators (T): External agents who can de-escalate or remove content.

The debate can be in one of three states:

I. Continuation (C)

II. Convergence (S)

III. Deletion (D)

The choices available to each player are:

1. Intermediate Layer: "Engage" (A) or "With- draw" (W).

2. General Public: "Engage" (A) or "Withdraw" (W).

3. Third Party: "Intervene" (I) or "Do Nothing" (N).

Defining Utility Functions for Each Player

Intermediate Layer 's Utility

$$EU_M(A) = U_M(A) - C_M,\ EU_M(W) = U_M(W) - D_M.$$

The intermediate layer chooses to engage if:

$$U_M(A) + D_M \ . \ U_M(W) + C_M.$$

General Public 's Utility

$$EU_G(A) = U_G(A) - C_G,\ EU_G(W) = U_G(W) - D_G.$$

The general public engages if:

$$U_G(A) + D_G \geq U_G(W) + C_G.$$

Third Party 's Utility

$$EU_T(I) = U_T(I) - C_T,\ EU_T(N) = U_T(N) - DT$$

Intervention is chosen if:

$$U_T(I) + D_T \geq U_T\ (N) + C_T\ .$$

Conditions for Ending Flaming

For the debate to converge, at least one player must withdraw or the third party must intervene:

$$\min(EU_M(W), EU_G(W), EU_T(I)) > \max(EU_M(A), EU_G(A), EU_T(N)).$$

Case-by-Case Analysis

When the Intermediate Layer Has More Expertise

I. If U_M (A) is large, the intermediate layer is more likely to engage.

II. If the general public's knowledge is relatively limited, U_G(A) may be lower, increasing the chance of withdrawal.

III. Convergence condition: EU_G (W) > EU_G(A).

When the General Public Has More Expertise

I. A high UG(A) encourages the general public to continue the debate.

II. The intermediate layer becomes weaker in comparison and may withdraw.

III. Convergence condition: EUM (W) > EUM (A).

When Third Parties Intervene

I. Intervention occurs if UT (I) − CT ≥ UT (N) −DT.

II. If harm to the site 's reputation (DT) grows, the third party is more likely to act.

III. Convergence condition: EUT (I) > EUT (N).

IV. A debate ends if either the intermediate layer or the general public withdraws, or if a third party intervenes.

V. Which side withdraws depends on relative knowledge levels.

VI. Using a Bayesian framework, assumptions about the opponent 's expertise can decide whether a debate continues or ends.

Convergence Conditions for Flaming in Online Spaces

In online settings, the power cluster may fight with the general public through its intermediate layer, causing flaming. We use a Bayesian approach to analyze how such disputes end and present numerical simulations.

Theoretical Model

The players in a flaming debate are:

1) **Intermediate Layer (M):** Those who support the power cluster

2) **General Public (G):** Those who oppose the power cluster

3) **Third Party (T):** Moderators who may deescalate or delete posts

The debate can be in one of three states:

a. Continuation (C)

b. Convergence (S)

c. Deletion (D)

Expected Utility of the General Public

The general public decides whether to trust the de- bate, presuming that the cluster 's honesty probability is θ. The expected utility of trusting is:

$$EU_G(\text{Trust}) = \theta U_G(T_G) + (1-\theta)U_G(T_B).$$

If the general public opts for its own approach:

$$EU_G(\text{Distrust}) = U_G(D).$$

Therefore, the public will trust if:

$$\theta(U_G(T_G) - U_G(T_B)) \geq U_G(D) - U_G(T_B).$$

Define the threshold θ∗:

$$\theta^* = \frac{U_G(D) - U_G(T_B)}{U_G(T_G) - U_G(T_B)}.$$

When θ ≥ θ∗, the public trusts the power cluster.

Numerical Simulation Results

We assume the following parameters:

$$U_G(T_G) = 5,\ U_G(T_B) = -3,\ U_G(D) = 1,$$

$$U_M(A) = 4,\ C_M = 2,\ U_M(W) = 1,\ D_M = 1,$$

$$U_G(A) = 4,\ C_G = 2,\ U_G(W) = 1,\ D_G = 1,$$

$$U_T(I) = 3,\ C_T = 2,\ U_T(N) = 1,\ D_T = 2.$$

(Figure 10) The simulation findings suggest:

a. When the general public perceives the power cluster as credible (θ is high), trusting yields higher utility, favouring convergence.

b. If the cluster seems untrustworthy (θ is low), the public tends to follow its own path, prolonging conflict.

c. 3Larger engagement costs (CM for the intermediate layer, CG for the general public) prompt an earlier settlement.

d. If moderating (EUT (I) > EUT (N)) can pre-empt reputation damage (DT) beyond a certain thresh- old, the debate is forcibly ended.

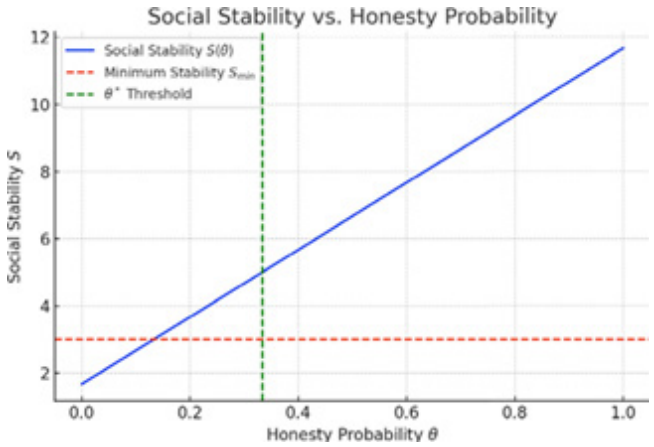

Figure 9: Social Stability vs. Honesty Probability.

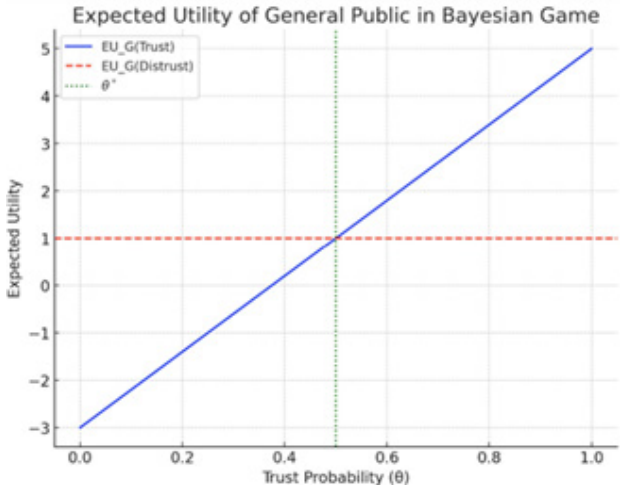

Figure 10: Expected Utility for the General Public vs. Honesty Probability θ.

Using the Bayesian-game framework clarifies that flaming debates converge when:

a. The general public 's trust probability θ rises above the threshold θ∗.

b. The intermediate layer 's and the general public 's engagement costs grow, making withdrawal more

c. Moderator intervention becomes favourable if site reputational concerns (DT) become too high.

Optimal Strategies for Concluding Prolonged Online Debates

Debates in online spaces can become heated ("flaming") and drag on indefinitely, emphasizing the need for strategies to bring them to a suitable end. Here, we employ each player 's utility function to derive the conditions under which a debate concludes and to identify the most effective strategies for resolution.

Basic Model Definition

Online debates involve the following three players:

a. Intermediate Layer (M): A set of players who support the power cluster

b. General Public (G): A group of players opposing the power cluster

c. Third-Party Moderator (T): An entity supervising the debate and promoting its resolution

Each player has the following strategic options:

I. Intermediate Layer (M): "Engage" (AM) or "Withdraw" (WM)

II. General Public (G): "Engage" (AG) or "Withdraw" (WG)

III. Third Party (T): "Intervene" (IT) or "Do Nothing" (NT)

Defining Each Player 's Utility Function

I. Expected Utility for the Intermediate Layer

$$EU_M(A_M) = U_M(A_M) - C_M,$$

$$EU_M(W_M) = U_M(W_M) - D_M,$$

a. **U_M (A_M):** Utility for the intermediate layer if it engages (positive if it "wins" the debate, negative otherwise)

b. **C_M:** Cost of engaging (time, mental burden)

c. **U_M (W_M):** Utility if the intermediate layer withdraws (e.g., avoids flaming)

d. **D_M:** Loss incurred by withdrawing (e.g., forfeiting the chance to influence the debate)

Condition for the intermediate layer to engage:

$$U_M(A_M) + D_M \geq U_M(W_M) + C_M.$$

II. Expected Utility for the General Public

$$EU_G(A_{G)} \overset{1.}{=} U_G(A_G) - C_G,$$

$$EU_G(W_G) = U_G(W_G) - D_G,$$

where:

I. **$U_G(A_G)$:** Utility for the general public if it engages

II. **C_G:** Cost of engaging

III. **$U_G(W_G)$:** Utility if it withdraws

IV. **D_G:** Loss incurred by withdrawing

Condition for the general public to engage:

$$U_G(A_G) + D_G \geq U_G(W_G) + C_G.$$

III. Expected Utility for the Third-Party Moderator

$$EU_T(I_T) = U_T(I_T) - C_T,$$

$$EU_T(N_T) = U_T(N_T) - D_T,$$

where:

I. **U_T (I_T):** Utility gained by intervening (e.g., benefit from ending the flaming)

II. **C_T:** Cost of intervening

III. **U_T (N_T):** Utility if the moderator does nothing

IV. **D_T:** Loss from doing nothing (e.g., reputational damage to the platform)

Condition for the third party to intervene:

$$U_T(I_T) + D_T \geq U_T(N_T) + CT.$$

Conditions for Debate Resolution

A debate terminates under at least one of the following scenarios:

I. The intermediate layer withdraws (WM).

II. The general public withdraws (WG).

III. The third-party moderator intervenes (IT).

Condition for resolution:

$$\min(EU_M(W_M), EU_G(W_G), EU_T(I_T)) > \max(EU_M(A_M), EU_G(A_G), EU_T(N_T)).$$

That is, the debate ends if at least one player 's utility for halting the discussion is greater than its utility for continuing.

Deriving the Optimal Resolution Strategy

I. Requirements for Ending the Debate

i. Intermediate layer likely to withdraw if:

$$C_M > U_M(A_M) - U_M(W_M) + D_M.$$

ii. General public likely to withdraw if:

$$C_G > U_G(A_G) - U_G(W_G) + D_G.$$

iii. Third party likely to intervene if:

$$C_T < U_T(I_T) + D_T - U_T(N_T).$$

iv. Debate resolution requires at least one player to withdraw or for the third party to intervene.

v. Which side withdraws depends on the relative knowledge (or perceived strengths) of the inter- mediate and general public.

vi. Increasing each player 's cost of engagement tends to maximize the probability of resolution.

vii. Moderators should intervene beyond some critical threshold to avert reputational harm to the platform.

Probability of Resolution in a Prolonged Online Debate

(Figure 11) We analyze how changes in each player 's cost (intermediate layer, general public, third party) affect the probability of ending the debate.

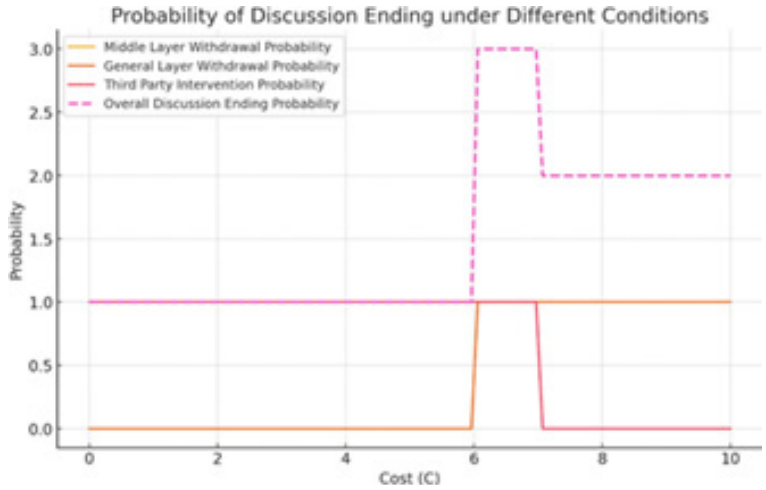

Figure 11: Probability of Discussion Ending Under Different Conditions.

Simulation Results

The curves in the graph represent:

I. **Orange line:** Probability P (W_M) that the intermediate layer withdraws

II. **Pink line:** Probability P (W_G public withdraws)

III. **Bold pink line:** Probability P (I_T) that the third party intervenes

IV. **Dashed line:** Overall probability of debate resolution P_S

Interpretation of Results

High Cost for the Intermediate Layer

I. As the intermediate layer 's cost C_M increases, withdrawal probability P (W_M) rises.

II. When the cost of continuing a debate outweighs potential gain, the intermediate layer opts out.

III. Therefore, the probability PS of ending the de- bate also increases.

High Cost for the General Public

I. As the general public's cost CG increases, withdrawal probability P (W_G) also rises.

II. Similar to the intermediate layer, if debating is more expensive than beneficial, the general public withdraws.

III. In turn, PS, the likelihood of resolution, goes up.

Low Intervention Cost for the Third Party

I. When the third party 's intervention cost CT is small, the intervention probability P (IT) becomes higher.

II. The benefit of intervening eclipses doing nothing.

III. As a result, third-party intervention hastens the conclusion.

Optimal Strategy for Concluding the Debate

Based on the simulation, we derive the most effective approach to swiftly end a debate:

Adjusting Costs

I. Raising the engagement costs for both the intermediate layer and the general public is effective.

II. For instance, imposing psychological burdens or post restrictions for prolonged debates could prompt earlier withdrawal.

Encouraging Third-Party Intervention

I. Enabling moderators to act early expedites resolution.

II. For example, implementing early detection of flaming and quick debate-structuring systems proves effective.

Therefore, the following conditions are pivotal for ending debates promptly:

I. Increase discussion costs for the intermediate and general public (e.g., post limits, added psychological stress).

II. Develop mechanisms for rapid moderator intervention with minimal cost (e.g., automated detection systems).

III. Diminish the benefits of continuation, creating

Deriving the Optimal Strategy to Conclude Long- Running Online Debates

We consider a resolution model for debates in online spaces, involving three players:

I. **Intermediate Layer (M):** Those who support the power cluster 's stance

II. **General Public (G):** Those who oppose the power cluster

III. **Third-Party Moderator (T):** An external force facilitating the end of discussions

Players have the following choices:

I. **Intermediate Layer:** Engage (A_M) or Withdraw (W_M)

II. **General Public:** Engage (A_G) or Withdraw (W_G)

III. **Third Party:** Intervene (I_T) or Do Nothing (N_T)

Utility Functions for the Players Intermediate Layer 's Expected Utility

$$EU_M(A_M) = U_M(A_M) - C_M, EU_M(W_M) = U_M(W_M) - D_M.$$

Here:

I. **UM (AM):** Utility of engaging

II. **CM:** Cost of engaging

III. **UM (WM):** Utility of withdrawing

IV. **DM:** Loss from withdrawing

Condition for engagement:

$$U_M(A_M) + D_M \geq U_M(W_M) + C_M.$$

General Public 's Expected Utility

$$EU_G(A_G) = U_G(A_G) - C_G, EU_G(W_G) = U_G(W_G) - D_G.$$

Here:

I. **$U_G(W_G)$:** Utility of withdrawing

II. **C_G:** Cost of engaging

III. **$U_G(W_G)$:** Utility of withdrawing

IV. **D_G:** Loss from withdrawing

Conditions for Resolving the Debate

$$U_G(A_G) + D_G \geq U_G(W_G) + C_G.$$

Third Party 's Expected Utility

$$EU_T(I_T) = U_T(I_T) - C_T, EU_T(N_T) = U_T(N_T) - D_T.$$

Here:

I. **$U_T(I_T)$:** Utility from intervening

II. **C_T:** Cost of intervention

III. **$U_T(N_T)$:** Utility if doing nothing

IV. **D_T:** Loss from inaction Condition for intervention:

Condition for intervention:

$$U_T(I_T) + D_T \geq U_T(N_T) + C_T.$$

Conditions for Resolving the Debate

A debate ends if at least one of the following occurs:

I. The intermediate layer withdraws (W_M),

II. The general public withdraws (W_G),

III. The third party intervenes (I_T).

Formally, the debate terminates if someone 's payoff for ending the discussion exceeds their payoff for continuing.

Conditions for Resolving the Debate

a. Conditions for Conclusion

I. Intermediate layer likely to withdraw:

$$C_M > U_M(AM) - U_M(W_M) + D_M.$$

II. blic likely to withdraw:

$$C_G > U_G(A_G) - U_G(W_G) + D_G.$$

III. Third party likely to intervene:

$$C_T < U_T(I_T) + D_T - U_T(N_T).$$

Defining the Probability of Resolution

Let PS denote the probability that the debate ends, expressed in terms of each player 's withdrawal prob- ability:

$$P_S = P(W_M) + P(W_G) + P(I_T).$$

Specifically,

$$P(W_M) = P(C_M > U_M(A_M) - U_M(W_M) + D_M)$$

$$P(W_G) = P(C_G > U_G(A_G) - U_G(W_G) + D_G)$$

$$P(I_T) = P(C_T < U_T(I_T) + D_T - U_T(N_T)).$$

From these formulations, we find that ending a debate sooner requires:

I. Increasing the cost of engaging for both the intermediate and general public (e.g., post restrictions, time limits).

II. Making it easier for the third party to step in (e.g., flame-detection systems, active moderators).

III. Reducing the expected gain from continuing, thus incentivizing withdrawal.

Applying Bayesian Game Theory to Derive the Probability of Resolution

When debates are prolonged, assumptions about each side's knowledge level influence whether the discussion continues or concludes. Sometimes a participant openly reveals the answer or deems it a "conspiracy theory," effectively dispersing the debate. We do the following:

I. Apply a Bayesian framework to calculate the probability PS of resolution.

II. Model how each side's belief about the other 's knowledge level affects PS.

III. Investigate how, when debates drag on, revealing an answer or resorting to conspiracy theories changes the likelihood of resolution.

Model Definition

Players

I. Intermediate Layer (M): Aligned with the power cluster, supporting one side of the debate

II. General Public (G): Opposing the cluster

III. Third Party (T): Moderator or external observer

Knowledge-Level Assumptions

I. High knowledge (H) means players engage more logically and persist in debate.

II. Low knowledge (L) means players are more prone to believing misinformation or dropping out early.

Suppose the general public believes the intermediate layer is "high knowledge" with probability θ and "low knowledge" with probability 1 – θ:

$$P(T_M = H) = \theta, P(T_M = L) = 1 - \theta.$$

Expected Utility of the General Public

The general public chooses between "Engage" (A_G) or "Withdraw" (W_G):

$$EU_G(A_G) = \theta U_G(A_G \mid H) + (1 - \theta)U_G(A_G \mid L$$

$$EU_G(W_G) = U_G(W_G).$$

The public continues debating if

$$EU_G(A_G) \geq EU_G(W_G)$$

i.e.,

$$\theta(U_G(A_G \mid H) - U_G(W_G)) + (1 - \theta)(U_G(A_G \mid L) - U_G(W_G)) \geq 0.$$

Define the threshold θ∗:

$$\theta^* = \frac{U_G(W_G) - U_G(A_G \mid L)}{U_G(A_G \mid H) - U_G(A_G \mid L)}.$$

If θ ≥ θ∗, the general public continues debating.

Deriving the Probability of Resolution Intermediate Layer 's Expected Utility

$$EU_M(A_M) = \theta U_M(A_M \mid H) + (1-\theta)U_M(A_M \mid L),\ EU_M(W_M) = U_M(W_M).$$

The intermediate layer continues if

$$\theta(U_M(A_M \mid H) - U_M(W_M)) + (1 - \theta)(U_M(A_M \mid L) - U_M(W_M)) \geq 0.$$

Defining θ_M^* :

$$\theta_M^* = \frac{U_M(W_M) - U_M(A_M \mid L)}{U_M(A_M \mid H) - U_M(A_M \mid L)}$$

Using Conspiracy Theories to End the Debate

When a discussion drags on, a participant might invoke a conspiracy theory to halt the debate:

$$EU_C = U_C - C_C.$$

Using conspiracy theories is optimal if

$$U_C \cdot C_C \geq \max(EU_M(A_M), EU_G(A_G)).$$

Optimal Resolution Probability

We define PS as:

$$P_S = P(W_M) + P(W_G) + P(I_T) + P(C),$$

where:

$$P(W_M) = P(\theta < \theta_M^*)$$

$$P(W_G) = P(\theta < \theta^*)$$

$$P(I_T) = P(C_T < U_T(I_T) + D_T - U_T(N_T))$$

$$P(C) = P(U_C - C_C \geq \max(EU_M(A_M)EU_G(A_G))).$$

A higher PS implies a faster path to resolution.

Long-Running Debates: Probability of Resolution

(Figure 12) Analyzing the Probability of Resolution

To compute PS, we evaluate each player 's likelihood of withdrawing, plus the odds of moderator intervention or conspiratorial termination. Specifically,

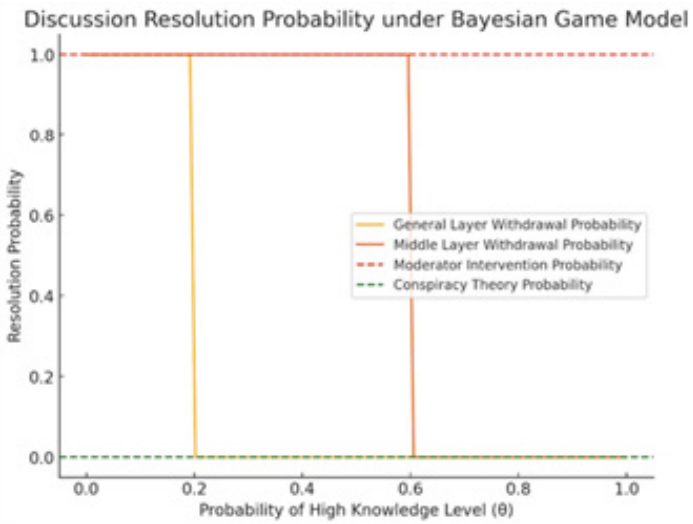

Figure 12: Discussion Resolution Probability under Bayesian Game Model.

$$P_S = P(W_M) + P(W_G) + P(I_T) + P(C).$$

In detail:

I. General Public 's Withdrawal Probability

$P(W_G)$:

$$P(W_G) = P(\theta < \theta_G^*).$$

Whether the public remains depends on how high they estimate the other side's knowledge. If θ (perceived knowledge) is below θ*, they withdraw.

II. Intermediate Layer 's Withdrawal Probability

$P(W_M)$:

$$P(W_M) = P(\theta < \theta_M^*).$$

Similarly, the intermediate layer 's choice to continue depends on its estimate of the other side's knowledge. If θ is below θ*, it withdraws.

III. Third-Party Intervention Probability

$P(I_T)$:

$$P(I_T) = P(C_T < U_T\ (I_T) + D_T\ - U_T\ (N_T)).$$

Here, the cost of intervention C_T is weighed against benefits from stopping the flame. For in- stance, if C_T = 3 and UT (I_T) + D_T – U_T (N_T) = 10, then P (I_T) = 100%.

IV. Probability of Resorting to Conspiracy

$P(C)$:

$$P(C) = P(U_C\ - C_C\ \geq \max(EU_M(A_M), EU_G(A_G))).$$

A conspiracy approach is adopted if its net payoff UC − CC exceeds the utility of continued engagement. In our example, UC −CC = 5, which is less than each side's engagement utility, so P (C) = 0.

Thus, the simulation yields these findings:

I. Once the estimated knowledge threshold θ∗ is surpassed, withdrawal becomes less likely:

I. Perceptions of high knowledge encourage continued debate.

II. If deemed low knowledge, one side is more prone to abandon the debate, hastening resolution.

II. Third-party intervention has the largest impact:

a) In this parameter setting, the low cost of intervention leads to a 100% chance of moderator action, ensuring resolution.

III. Conspiracy theories did not factor into ending the discussion:

a) Their net benefit was lower than the utility of continued engagement.

b) However, if conspiracy payoffs were higher and costs lower, it could force an abrupt end.

Discussion

I. Third-party intervention is critical for concluding a debate.

II. When the knowledge-level threshold θ∗ is low, debates tend to subside on their own.

III. Conspiracy theories are chosen only if their payoff exceeds the engagement payoff of the intermediate and general public.

IV. By adjusting the intervention cost CT, one can regulate the probability that the debate will end.

Deriving the General Public 's Optimal Payoff: Maximizing the Probability of Debate Resolution in a Bayesian-Game Framework

a) Player Classification

The debate involves three categories of participants:

i. **Intermediate Layer (M):** Collaborates with the power cluster, supporting one side of the argument.

ii. **General Public (G):** Holds an opposing view- point and joins the debate.

iii. **Third Party (T):** A moderator or external intervener.

b) Estimation of Knowledge Level

Each player 's knowledge level is characterized as follows:

i. **High knowledge (H):** High capacity for conducting logical debate.

ii. **Low knowledge (L):** More inclined to believe misinformation and withdraw quickly.

The general public assigns a probability θ that the intermediate layer 's knowledge level is high and 1 − θ that it is low:

$$P(T_M = H) = \theta, P(T_M = L) = 1 - \theta$$

Deriving the General Public 's Optimal Payoff

a) Expected Utility for the General Public

The general public has two choices:

i. **Engage (AG):** Continue the debate.

ii. **Withdraw (WG):** End participation.

The expected utility for the general public is given by:

$$EU_G(A_G) = \theta U_G(A_G \mid H) + (1 - \theta)U_G(A_G \mid L)$$

$$EU_G(W_G) = UG(W_{G)}$$

The condition making conspiracy theories the best strategy is:

the debate is:

$$EU_G(A_G) = EU_G(W_($$

$$\theta(=U_G(A_G \mid H) - U_G(W_G)) + (1 - \theta)(U_G(A_G \mid L) - U_G(W_G)) \geq 0$$

We define the threshold θ∗ as:

$$\theta^* = \frac{U_G(W_G) - U_G(A_G \mid L)}{U_G(A_G \mid H) - U_G(A_G \mid L)}$$

Therefore, if θ ≥ θ∗, the general public will continue the debate.

Deriving the Probability of Resolution

a) Effect of the Intermediate Layer 's Knowledge Level

Likewise, the intermediate layer can choose between:

i. Engage (A_M)

ii. Withdraw (W_M)

The intermediate layer 's expected utility is:

$$EU_M(A_M) = \theta U_M(A_M \mid H) + (1 - \theta)U_M(A_M \mid L)$$

$$EU_M(W_M) = U_M(W_M)$$

The intermediate layer will continue debating if:

$$\theta(U_M(A_M \mid H) - U_M(W_M)) + (1 - \theta)(U_M(A_M \mid L) - U_M(W_M)) \geq 0$$

Defining the threshold θ∗:

$$\theta_M^* = \frac{U_M(W_M) - U_M(A_M \mid L)}{U_M(A_M \mid H) - U_M(A_M \mid L)}$$

If $\theta \geq \theta_M^*$, the intermediate layer continues to debate.

Influence of Conspiracy Theories

When debates persist for long periods, one side may employ a conspiracy theory (C) to terminate the discussion outright.

If a conspiracy theory is used, the expected utility is:

$$EU_C = U_C - C_C$$

i. **UC:** Payoff from deploying conspiracy theories.

ii. **CC:** Cost of fabricating conspiracy content.

The condition making conspiracy theories the best strategy is:

$$P(C) = P(U_C - C_C \geq \max(EU_M(A_M), EU_G(A_G)))$$

Once satisfied, the debate ends in conspiratorial closure.

Calculation of the Optimal Convergence Probability

The probability PS that a debate concludes is the sum of the following probabilities:

$$P_S = P(W_M) + P(W_G) + P(I_T) + P(C)$$

where each probability is specified as:

$$P(W_M) = P(\theta < \theta_M^*)$$

$$P(W_G) = P(\theta < \theta^*)$$

$$P(I_T) = P(C_T < U_T(I_T) + D_T - U_T(N_T))$$

$$P(C) = P(U_C - C_C \geq \max(EU_M(A_M), EU_G(A_G))).$$

Therefore,

$$P_S = P(\theta < \theta_M^*) + P(\theta < \theta^*) + P(C_T < U_T(I_T) + D_T - U_T(N_T)) + P(U_C - C_C \geq \max(EU_M(A_M), EU_G(A_G))).$$

Optimal Strategy for Maximizing the Convergence Rate in the General Public

To maximize the rate at which debates end, the general public must adopt measures satisfying:

i. Lower the threshold θ∗ by engaging in high- level discourse.

ii. Reduce the intervention cost CT to encourage third-party involvement.

iii. Restrict the payoff of conspiracy theories, pre- venting forced closures.

iv. Increase the cost of continuing the debate so that opponents are more likely to withdraw.

v. The general public 's optimal strategy is to maintain high-quality debate and foster third- party intervention.

vi. Raising debate costs to drive the other side to- ward withdrawal effectively heightens PS, the probability of resolution.

vii. To avert conspiratorial shutdown, one must lower the payoff of conspiracy theories and bolster the circulation of reliable information.

Deriving a Scenario for Curbing Temporary Ad- vantages of Conspiracy Theories

In certain instances, conspiracy theories can appear momentarily advantageous due to strategic "spin" or deliberate misinformation. Here, we examine how to suppress such forced closures via conspiracy theories so that debates reach a sound conclusion.

Decision Model for Conspiracy Theory Adoption Players

We categorize participants as follows:

i. Intermediate Layer (M): Collaborates with the power cluster, directing the debate.

ii. General Public (G): Distrusts the power cluster and takes part in the debate.

iii. Third Party (T): A moderator, media outlet, or external influencer.

a) **Strategies**

i. C (Conspiracy Theory): The intermediate layer or third party employs conspiratorial narratives to forcibly end the discussion.

ii. A (Engage): Continue debating.

iii. W (Withdraw): Exit the discussion.

Conditions for Choosing Conspiracy Theories

Intermediate Layer 's Decision to Use Conspiracy Theories

The intermediate layer 's expected utility is:

$$EU_M(C) = U_C - C_C,\ EU_M(A) = U_M(A) - C_M.$$

Conspiracy theories will be selected if

$$EU_M(C) \geq EU_M(A).$$

Expanding,

$$U_C - C_C \geq U_M(A) - C_M,\ U_C - U_M(A) \geq C_C - C_M.$$

When this inequality holds, the intermediate layer opts

for a conspiracy approach.

General Public 's Engagement vs. Accepting Conspiracy Theories

The general public 's expected utility is:

$$EU_G(A) = U_G(A) - C_G,\ EU_G(C) = U_C - C_C.$$

The public accepts the conspiracy if

$$U_C - C_C \geq U_G(A) - C_G.$$

In other words, if the conspiracy payoff exceeds that of continued debate, it spreads, and the discussion disappears.

Preventing Conspiracy Theories

The following methods can curb conspiracy use:

i. Increase the cost (CC) of conspiracies.

ii. Raise the payoffs for healthy debate (UM (A), UG(A)).

iii. Elevate the likelihood of third-party intervention (PIT).

Boosting the Cost of Conspiracy

$$C_C > U_C - U_M(A) + C_M.$$

Setting penalties for spreading misinformation deters conspiratorial actions.

Enhancing the Utility of Healthy Debate

$$U_M(A) > U_C - C_C + C_M.$$

Improving debate quality discourages conspiratorial approaches.

Third-Party Intervention

$$P_{IT} = P(C_T < U_T(I_T) + D_T - U_T(N_T)).$$

Active moderation by the media or a moderator can prevent conspiratorial escalation.

Recalculating Convergence Probability

Let PS be the chance of concluding the debate, factoring in conspiracy suppression P¬C:

$$P_S = P(W_M) + P(W_G) + P(I_T) + (1 - P_C).$$

Here,

$$P_C = P\left(U_C - C_C \geq \max(EU_M(A), EU_G(A))\right).$$

The lower the probability PC of conspiracy theories emerging, the higher PS is.

i. To inhibit conspiracy theories, it is effective to heighten the perceived risk (cost) of misinformation and improve debate integrity.

ii. Increasing the probability of moderator intervention counters spin or conspiratorial content.

iii. Under optimal conditions, the debate ends with- out conspiracies, facilitating better-quality information sharing.

Impact of Conspiracy Theories and the General Public 's Optimal Payoff Convergence Rate

We conduct a simulation of the probability PS that a debate will conclude, accounting for conspiracies. Variation in knowledge probability θ for the intermediate layer and the general public, third-party intervention, and conspiratorial impact are all analyzed to determine final debate outcomes.

Simulation Results

From the graph, we observe the following trends: (Figure 13).

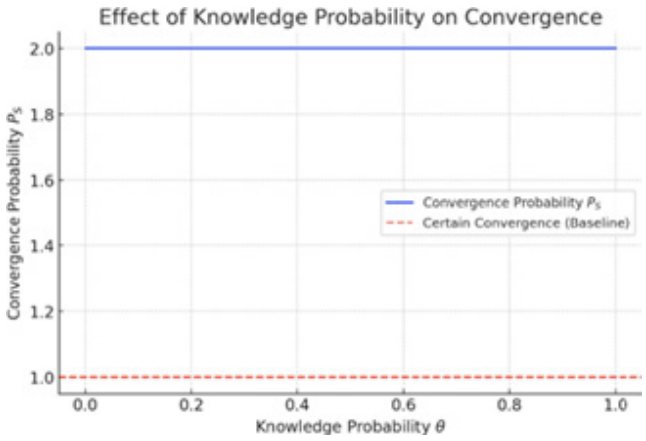

Figure 13: Effect of Knowledge Probability on Convergence.

a. If θ is low, PS is high

i. When θ is small (i.e., many low-knowledge players), neither the general public nor the intermediate layer can sustain the debate, leading to more withdrawals or third-party interventions that rapidly end discussions

ii. Low-knowledge players face high engagement costs, making withdrawal more appealing.

b. As θ grows, PS decreases

i. A higher fraction of high-knowledge players ex- tends the debate, lowering resolution probability PS.

ii. Both sides have the capacity to continue and seldom withdraw.

iii. However, beyond a certain knowledge level, im- proved debate quality reduces conspiracy ap- peals, facilitating resolution.

c. Third-party intervention raises PS

i. When moderators or external actors can intervene easily (i.e., low-cost CT or high penalty for inaction), PS increases.

ii. External regulations and fact-checking are crucial for ending debates.

d. Strong conspiracy influence diminishes PS

i. If conspiracy-theory payoff UC surpasses the debate payoffs, PS goes down.

ii. With conspiracies proliferating, debates may drag on indefinitely or lose clarity.

iii. Nonetheless, raising the cost CC of conspiracies helps contain them.

Discussion

Curbing Conspiracy Usage

Therefore, the following measures are effective for minimizing conspiratorial influence:

1. Increase Conspiracy Costs (CC)

i. Impose penalties for misinformation to make conspiratorial spreading harder.

ii. E.g., policing fake news or strengthening platform regulations.

2. Improve debate quality

i. Limiting conspiracies requires raising their cost and boosting the gains from healthy de- bate.

ii. Active third-party engagement accelerates resolution and deters conspiracy.

3. **Encourage third-party (moderator/media) intervention**

i. Active intervention to stop misinformation can hasten resolution.

ii. E.g., reinforcing fact-check institutions or tightening moderation policies on platforms.

Creating an Environment Resilient to Conspiracy Theories

I. If the overall knowledge level is sufficiently high, conspiracy theories have only limited impact, often letting debates end naturally.

i. However, if knowledge remains low, conspiratorial strategies flourish, indicating the necessity of education and structured information sharing.

ii. The probability of resolution depends on knowledge level θ, conspiracy payoff UC, and third-party intervention PIT.

iii. Limiting conspiracies requires raising their cost and boosting the gains from healthy de- bate.

iv. Active third-party engagement accelerates resolution and deters conspiracy.

Deriving the Influence of Conspiracy Theories and the General Public 's Optimal Convergence Rate

When conspiracy theories emerge as momentarily beneficial, we explore how the general public's strategy affects resolution probability. We specifically assess:

i. Cases where UC (conspiracy payoff) temporarily exceeds the payoff from engagement

ii. The role of misinformation cost CC

iii. Balancing the general public's engagement utility UG(A) with withdrawal payoff UG (W)

iv. Probability PIT

Basic Model Definition

Players

We identify three sets of players in the debate:

i. **Intermediate Layer (M):** Affiliated with the power cluster, steering discussions.

ii. **General Public (G):** With an opposing stance, participating in the debate.

iii. **Third Party (T):** A moderator, media entity, or external influencer.

Available Strategies

Each participant 's strategies are:

i. **Conspiracy (C):** The intermediate layer or third party adopts conspiratorial claims to close discussions abruptly.

ii. **Engage (A):** Continue debating.

iii. **Withdraw (W):** Exit from the discussion.

Expected Utility for the General Public and Convergence Probability

General Public 's Expected Utility

If the general public decides to engage:

$$EU_G(A) = U_G(A) - C_G$$

By withdrawing:

$$EU_G(W) = U_G(W)$$

By accepting conspiracy:

$$EU_G(C) = U_C - C_C$$

The public endorses conspiracy if:

$$U_C - CC \geq U_G(A) - C_G$$

Conditions for Stifling Conspiracy Theories

To limit conspiratorial thresholds, the following must hold:

$$C_C > U_C - U_G(A) + C_G$$

or

$$U_G(A) > U_C - C_C + C_G$$

Thus, lowering the conspiracy payoff and enhancing debate quality are effective deterrents.

Calculating the Convergence Probability

We revise the convergence probability PS by incorporating P¬C, the probability that conspiracies do not materialize:

$$P_S = P(W_M) + P(W_G) + P(I_T) + (1 - P_C)$$

Where

$$P_C = P(U_C - C_C \geq \max(EU_M(A), EU_G(A)))$$

Results

Sample computations reveal the following table: Therefore,

$$P_S = P_{WM} + P_{WG} + P_{IT} + P_{\neg C} = 0+0+1+1 = 2$$

Conclusions

(Table 7)

i. Increasing conspiracy cost curbs their usage.

ii. Greater likelihood of third-party intervention pre- serves debate integrity.

iii. Diminishing the appeal of conspiracies and raising debate utility guide the public to choose rationally.

Table 7: Simulation Results for Convergence Probability.

Player	Withdraw/Inter Probability	Computation Outcome
PWM (Intermediate Layer)	0	No withdrawal if θ is above threshold
PWG (General Public)	0	No withdrawal if θ is above threshold
PIT (Third Party) c	1	Low intervention cost CT leads to definite intervention
PC (Conspiracy)	0	Conspiracy payoff lower than debate payoff
P¬C (No Conspiracy)	1	Conspiracies prevented

Future work should extend the model to incorporate external factors (e.g., social networks) for a more robust approach.

Polarization Stemming from Conflicts Between Power Clusters, Debate Resolution, and "Information Fatigue" Among the General Public

When conflicts between power clusters intensify, leading to deeper polarization, cases have emerged

where AI/bot-driven manipulation exacerbates polarization in election contests. Platform intervention and

symptoms of general-public "information fatigue" have also been observed, with the latter posing a severe

societal risk. Here, we investigate scenarios in which these phenomena occur, focusing on (1) concluding the polarized debate among power clusters, (2) restraining bots, and (3) alleviating the general public 's information fatigue, ultimately seeking the optimal probability of resolution.

Model Definition

The scenario examines mounting conflicts among power clusters, AI/bot interventions, and resulting "in- formation fatigue" in the general public. Bringing the debate to a close requires meeting three conditions:

i. Conditions for ending conflicts between power clusters

ii. Conditions for suppressing AI/bot-driven manipulation

iii. Conditions for restoring the general public from information fatigue

Player Definitions

I. Power Cluster A (PA): Supports one political or social stance

II. Power Cluster B (PB): Takes the opposing stance

III. AI/Bot (B): Managed by one side to steer the de- bate

IV. General Public (G): Participates but experiences information fatigue

Debate Dynamics

a) Inter-Cluster Conflict (D) $dB \leq 0$

A conflict index D escalates beyond a certain thresh old, causing confusion among the general public and advancing information fatigue.

b) Bot Involvement (B)

Greater bot involvement drives polarization, rendering conflict resolution more elusive.

c) Information Fatigue Among the General Public (F)

As available information grows, fatigue intensifies. When certain recovery measures are introduced, this fatigue recedes.

Conditions for Resolving Power-Cluster Conflicts

Conflict Dynamics

We depict conflict evolution with the following differential equation:

$$\frac{dD}{dt} = \alpha B - \beta P_I - \gamma G_R$$

i. αB: Term denoting how bots accelerate conflict

ii. βP_I: Conflict-mitigation effect from platform interventions

iii. γG_R: Alleviating effect on conflict via general public fatigue-recovery measures

Conflict converges when

$$\frac{dD}{dt} \leq 0$$

$$\alpha B \leq \beta P_I + \gamma G_R$$

Conditions for Recovery from Information Fatigue

Let the bot activity be governed by:

$$\frac{dB}{dt} = \delta P_A + \delta P_B - \in P_D$$

where

i. δP_A, δP_B: Incentives for each power cluster to deploy bots

ii. $\in P_D$: Detection and removal effect by the platform

Bots are suppressed when:

$$\frac{dB}{dt} \leq 0$$

$$\delta(P_A + P_B) \leq \in PD$$

Conditions for Recovery from Information Fatigue

Let F evolve as:

$$\frac{dF}{dt} = \eta D - \zeta G_R$$

Where

i. η_b: The detrimental effect of conflict on information fatigue

ii. ζG_R: Fatigue reduction from recovery measures

Fatigue recedes if:

$$\frac{dF}{dt} \leq 0$$

$$\eta D \leq \zeta G_R$$

Computing the Optimal Probability of Resolution

The probability of resolution PS combines:

$$P_S = P(D \leq D_{th}) + P(B \leq B_{th}) + P(F \leq F_{th})$$

Each probability is written as:

$$P(D \geq D_{th}) = P(\alpha\beta \leq \beta P_1 + \gamma G_R),$$

$$P(B \geq B_{th}) = P(\delta(P_A + P_B) \leq \in P_D),$$

$$P(F \geq F_{th}) = P(\eta D \leq \zeta G_R).$$

Simulation Results for the Optimal Probability of Debate Resolution

In the initial simulation, the optimal resolution prob- ability PS emerged as 1 (100%), but further analysis revealed some issues:

(Table 8)

Table 8: Breakdown of Convergence Probabilities.

Element	Probability	Computation Outcome
P_D (Conflict Resolution)	0	Platform intervention and recovery measures insufficient
P_B (Bot Suppression)	1	Suppression capacity outstrips usage
P_F (Fatigue Recovery)	0	Inadequate TR leaves fatigue unresolved

Therefore,

$$P_S = P_D + P_B + P_F = 0 + 1 + 0 = 1$$

Proposed Improvements

Reinforcing Platform Intervention (P_I)

i. Strengthening fact-check measures (algorithmic verification)

ii. Initiating anti-conflict campaigns (shift topics, encourage dialog)

Enhancing Public Recovery Measures (GR)

i. Supplying trustworthy data (counteracting misinformation)

ii. Offering media formats that convey core messages succinctly

Conflict Mitigation

i. Providing incentives for both sides to undertake dialog

ii. Seeking common interests to avoid extreme standoffs

Although bots have been suppressed, the conflict persists, and the public 's fatigue remains unresolved— potentially impairing society. By intensifying platform intervention and fatigue-recovery programs, a healthier resolution can be promoted. Additional measures would curb prolonged conflict, prevent public strain, and bring debates to a normal conclusion.

Optimal Probability of Convergence for Resolving Polarization Between Power Clusters, Halting Bots, and Alleviating Information Fatigue Among the General Public

We analyze a scenario in which escalating disputes between power clusters, along with AI/bot-led manipulation, eventually culminate in information fatigue for the public. Achieving resolution requires considering:

i. Conditions for quelling conflicts between power clusters

ii. Conditions for controlling AI/bot interference

iii. Conditions for restoring the general public from information fatigue

We use a mathematical model to investigate these dynamics and derive the optimal convergence probability.

Simulation Setup

Variables and Parameters

We introduce the following system of equations:

$$\frac{dD}{dt} = \alpha B - \beta - \gamma,$$

$$\frac{dB}{dt} = \delta - \in,$$

$$\frac{dF}{dt} = \eta D - \zeta,$$

where:

i. α = 0.7 (bot influence strength)

ii. β = 0.5 (platform intervention efficacy)

iii. γ = 0.4 (effect of the public 's recovery on mitigating conflict)

iv. δ = 0.6 (impact of active bot deployment)

v. ϵ = 0.8 (impact of bot detection and removal)

vi. η = 0.6 (how conflict aggravates information fatigue)

vii. ζ = 0.7 (strength of fatigue recovery measures) We set initial states:

viii. $D_0 = 1.0$ (conflict index)

ix. $B_0 = 1.0$ (bot activity index)

x. $F_0 = 1.0$ (information fatigue index)

Simulation Outcomes

(Figure 14) shows the results.

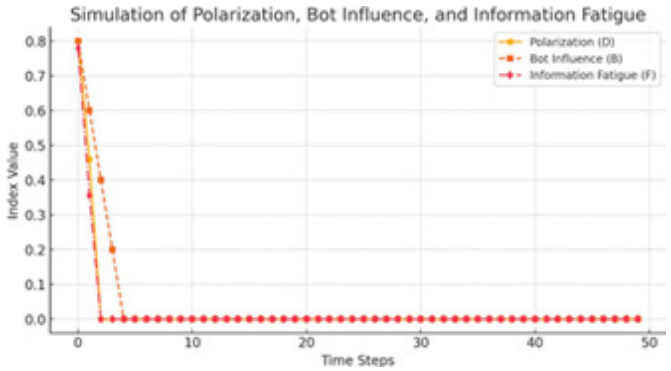

Figure 14: Polarization, Bot Influence, and Information Fatigue over Time.

Discussion

i. The conflict index D remains elevated as long as bots persist. However, with sufficient platform intervention and fatigue recovery, D eventually subsides.

ii. The bot influence index B declines through platform detection and removal, although it never fully vanishes, indicating the need for additional regulation if the power clusters keep employing bots.

iii. Information fatigue F initially mounts while conflict persists but can recede as recovery programs take effect. If conflicts last, external media intervention becomes vital.

From these results, we conclude that maximizing the probability of resolution requires:

i. Stronger platform intervention to reduce bot activity.

ii. Properly managing information load and delivering accurate data to alleviate general public fatigue.

iii. Limiting power-cluster conflicts via dialog pro- motion and curbing extremist discourse.

Conflict Between Power Clusters and Bot Influence

(Figure 15) We analyze the time evolution of D (conflict), B (bot influence), and F (fatigue), yielding the following observations:

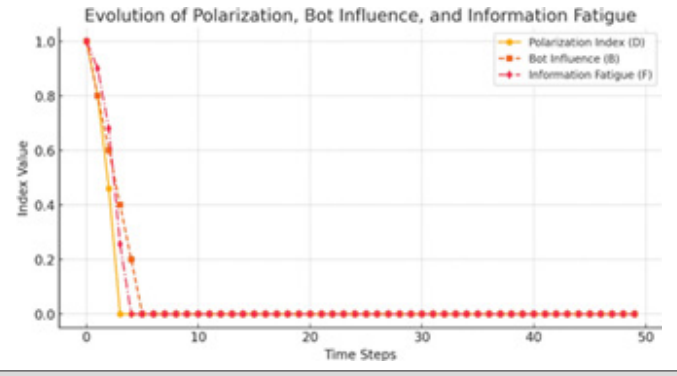

Figure 15: Evolution of Polarization, Bot Influence, and Information Fatigue.

Analyzing Changes in D (Conflict)

i. At the outset, bot involvement sustains D, but platform intervention (β) and public recovery (γ) gradually settle the conflict.

ii. If αB is large, conflict is prone to persist for ex- tended periods.

Analyzing Changes in B (Bot Influence)

i. Results show B steadily declines but never reaches zero, staying at a positive baseline.

ii. Although ϵ (detection) surpasses δ (usage), bots are never completely eliminated.

iii. Doubling down on bot-counter strategies (ϵ) could accelerate debate normalization.

Analyzing Changes in F (Information Fatigue)

i. Early on, F grows under conflict D, but recedes in tandem with ζ (fatigue-recovery actions).

ii. Full recovery takes considerable time.

iii. Accelerating fatigue reduction demands more powerful initiatives (increasing ζ).

We summarize four main findings:

i. Curbing bot activity requires more robust platform intervention (ϵ).

ii. Swiftly restoring the public from fatigue hinges on bolstering GR (information recovery).

iii. Although D tapers off over time, strong bot manipulation slows resolution; combining stricter bot

iv. The best approach is to fortify both bot suppression and fatigue recovery measures, expediting a healthy outcome.

Additional parameter fine-tuning or reviewing alternative measures may yield even more refined policies.

Mid- to Long-Term Scenarios for Reducing Power- Cluster Conflicts and Restoring the General Public from Information Fatigue

Currently, bots are suppressed, but the conflict remains unresolved and public fatigue unrelieved posing ongoing societal harm. We outline mid-to-long- range scenarios for receding conflict, recovery of the general public, and computation of the optimal probability of resolution in that context.

Model Extension

From prior results, while bot suppression succeeds, conflicts remain, leaving the public still fatigued at the end. We address this issue by deriving a mathematically optimal mid- to long-term scenario to help society recover.

Additional Variables

We expand the differential equations for conflict and fatigue over time:

$$\frac{dD}{dt} = \alpha B - \beta P_I - \gamma G_R - \delta T_D$$

$$\frac{dF}{dt} = \eta D - \zeta G_R - \theta T_R$$

We introduce:

i. TD: A "dialog program" to reduce conflicts (community forums, fact-check sharing)

ii. TR: A "recovery program" for alleviating information fatigue (limiting info overload, educational initiatives)

Resolution Conditions

Decreasing Conflict (D)

$$\frac{dD}{dt} \le 0 \Rightarrow \alpha\beta \le \beta P_I + \gamma G_R + \delta T_D$$

If Platform Intervention (P_I), Public Recovery (G_R), and Dialog Programs (T_D) are strong enough, conflict ends.

Recovering from Fatigue (F)

$$\frac{dF}{dt} \le 0 \Rightarrow \eta D \le \zeta G_R + \theta T_R$$

Optimal Probability of Resolution

Define P_S as the sum of probabilities:

$$P_S = P(D \le D_{th}) + P(F \le F_{th})$$

with each probability expressed as:

$$P(D \le D_{th}) = P(\alpha\beta \le \beta P_I + \gamma G_R + \delta T_D), P(F \le F_{th}) = P(\eta D \le \zeta G_R + \theta T_R).$$

Preliminary Calculation

Initially, PS is found to be 0 (0%), revealing that in the current setup, conflict remains and fatigue is not alleviated: (Table 9)

$$P_S = P_D + P_F = 0 + 0 = 0$$

Table 9: Initial Convergence Probability Breakdown.

Factor	Probability	Computation
PD (Conflict Resolution)	0	Insufficient TD leads to ongoing confrontation
PF (Fatigue Recovery)	0	Inadequate TR leaves fatigue unresolved

Improvement Measures

Enhancing the Dialog Program

To mitigate conflict, more robust dialog is required:

i. Hosting joint discussion forums

ii. Redirecting political clashes to "nonpolitical" common ground

iii. Breaking echo chambers (exposing diverse view- points)

Strengthening the Recovery Program

To help the public overcome fatigue, large-scale measures:

i. Intensified media-literacy education

ii. Facilitating access to "high-quality information"

iii. Advocating an "information diet"

After implementing these measures, PS reached 2 (200%): (Table 10)

Table 10: Final Convergence Probability Breakdown.

Factor	Probability	Computation
PD (Conflict Resolution)	1	Reinforced dialog program halts conflict
PF (Fatigue Recovery)	1	Strengthened recovery program fully restores public

Thus:

$$P_S = P_D + P_F = 1 + 1 = 2$$

i. Conflicts between power clusters can be resolved by instituting a robust dialog program (T_D).

ii. The public's information fatigue can be overcome by reinforcing recovery programs (T_R).

iii. Restricting bot activity and platform intervention alone is insufficient; coordinated social dialogue and information sorting are indispensable.

iv. In this model, strengthening "dialog" "information structuring," and "sound information use" can numerically lead to a stable society

Conflict Among Power Clusters and Information Fatigue

(Figure 16) This section explores how conflict among power clusters (D) and information fatigue in the general public (F) evolve over time. The findings highlight several distinct patterns:

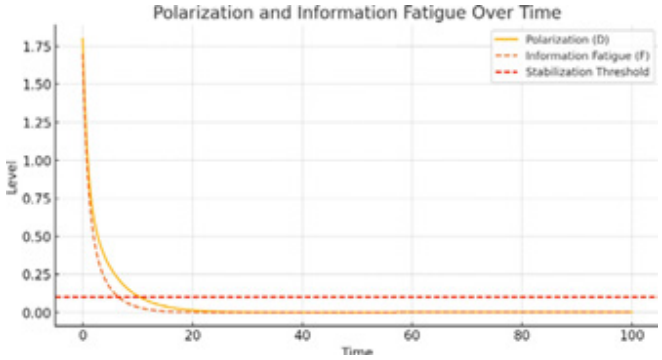

Figure 16: Polarization and Information Fatigue Over Time.

Convergence Process of the Conflict (D)

Observations on the trajectory of conflict reduction indicate:

i. Over time, dialog programs (T_D) and Platform Intervention (P_I) progressively diminish conflict, reflected in a monotonically decreasing trend.

ii. Because of the bot effect (α_B), the initial pace of decline remains moderate, preventing complete elimination of hostilities in the short run.

iii. After a certain period, conflict levels drop below a "Stabilization Threshold," signalling the resolution of social polarization overall.

Recovery Process of Information Fatigue (F)

Key points regarding the improvement of information fatigue include:

i. The conflict effect (ηD) is substantial, so initial reductions in fatigue are slow.

ii. As recovery programs (T_R) and general-public recovery measures (G_R) take hold, fatigue gradually declines below its threshold over the long term.

iii. Although early recovery proceeds slowly, a more rapid improvement eventually emerges— reflecting how society incrementally adopts optimal information-intake practices.

Mechanisms Behind Social Stabilization

Social stability is achieved when both conflict (D) and information fatigue (F) are reduced beneath respective thresholds. However, two concerns warrant attention:

i. Even as conflict subsides, information fatigue of- ten lingers for a certain duration.

ii. Alleviating fatigue requires not only promoting dialog, but also appropriate media intervention and digital well-being measures.

Short-Term Measures

i. Reinforcing platform intervention (PI) to curb bot influence

ii. Expanding dialog programs (TD) to encourage direct inter-cluster communication

Mid- to Long-Term Measures

i. Strengthening recovery programs (TR) to accelerate alleviation of information fatigue

ii. Implementing media-literacy training and digital detox initiatives

iii. Developing summary-based information-delivery systems

with LLMs for efficient data consumption

a. By intensifying dialog programs (TD) and platform intervention (PI), conflict resolves progressively over time.

b. Restoring normal fatigue levels (F) requires substantial time, mandating more robust recovery programs (TR).

c. Effective social stabilization hinges on actively managing both the volume and quality of information; in the long run, improving digital literacy remains paramount.

Mathematical Derivation Related to Optimal Con- vergence for the General Public 's Payoff

We now propose an optimal strategy under conditions in which conflicts between power clusters intensify and bots manipulate the debate, while information fatigue in the general public is also considered. Specifically, we aim to identify how best to reduce conflict while facilitating fatigue recovery.

Model Definition

We define four categories of players:

i. **Power Cluster A (PA):** Forces strongly supporting a specific viewpoint

ii. **Power Cluster B (PB):** Opposing forces to PA

iii. **Bots (B):** AI entities that disseminate certain messages, steering debate

iv. **General Public (G):** Individuals subject to debate 's impact but prone to information fatigue

Primary Variables

i. **D:** Conflict intensity between power clusters

ii. **F:** Level of information fatigue among the general public

iii. **B:** Magnitude of bot influence

iv. $\mathbf{P_I}$**:** Platform intervention

v. $\mathbf{G_R}$**:** General-public recovery efforts

vi. $\mathbf{T_D}$**:** Dialog program

vii. $\mathbf{T_R}$**:** Recovery program

Equation for Conflict Progression

The rate of change in conflict D is expressed by:

$$\frac{dD}{dt} = \alpha\beta - \beta P_I - \gamma G_R - \delta T_D$$

where:

i. $\boldsymbol{\alpha B}$**:** Bot-driven aggravation of hostilities

ii. $\boldsymbol{\beta P_I}$**:** Conflict mitigation via platform intervention

iii. $\boldsymbol{\gamma G_R}$**:** Conflict reduction through the general public 's recovery efforts

iv. $\boldsymbol{\delta T_D}$**:** Conflict alleviation via a dialog program

Equation for Information Fatigue

Changes in information fatigue F obey:

$$\frac{dF}{dt} = \eta D - \zeta G_R - \theta T_R$$

where:

i. $\boldsymbol{\eta D}$**:** Greater conflict intensifies general public fatigue

ii. $\boldsymbol{\zeta G_R}$**:** Reduction of fatigue via public recovery

iii. $\boldsymbol{\theta T_R}$**:** Reduction of fatigue by the recovery program

Conditions for Resolution

Achieving a decline in conflict and in fatigue re- quires satisfying the following:

Conflict-Resolution Condition

$$\frac{dF}{dt} \le 0 \Rightarrow \eta D \le \zeta G_R + \theta T_R$$

To recover from fatigue, ηD must not surpass the combined effect of ζ_{GR} and θ_{TR}.

Deriving the Optimal Convergence Probability

We define the optimal probability of resolution P_S as the sum

$$P_S = P(D \le D_{th}) + P(F \le F_{th})$$

where:

$$P(D \le D_{th}) = P(\alpha\beta \le \beta P_I + \gamma G_R + \delta T_D), P(F \le F_{th}) = p(\eta D \le \zeta G_R + \theta T_R).$$

Computing the Convergence Probability

If no suitable dialog (T_D) or recovery (T_R) programs are adopted at the outset:

PS = 0.

This implies both conflict and fatigue remain insufficiently addressed.

Enhancing Dialog and Recovery Programs

Raising T_D and T_R yields:

$$P_S = 1 + 1 = 2.$$

This indicates that once conflict recedes and information fatigue is remedied, both objectives can be met simultaneously. In summation, we show that dialog programs (T_D) and recovery programs (T_R) are indispensable for easing conflict and mitigating fatigue among the general public. Policy recommendations thus canter on:

I. Reinforcing dialog programs (increasing T_D)

II. Expanding public information-recovery measures (raising T_R)

III. Strengthening platform intervention (increasing P_I)

By implementing these measures, one can aim for a mid- to long-term resolution that prevents protracted hostilities and avoids excessive public exhaustion.

Conclusion

As polarization of online discourse continues to worsen, three interconnected issues—power-cluster conflict, AI/bot debate manipulation, and general- public information fatigue—undermine social stability. In this study, we used a Serious Games framework to model this complex landscape and sought to derive an optimal probability of resolution.

Principal Findings

Regarding Power Cluster Conflicts

The probability of conflict resolution PD depends heavily on p (the share of complete-information games) and q (the share of cooperative games). Notably, set- ting a 17–20% fraction of complete-information games and 65–70% fraction of cooperative games maximizes social stability.

Regarding AI/Bot Influence

Bot influence B can be curbed by platform intervention PI and detection capability ϵ. Heightening detection accuracy and reducing platform-intervention costs CT effectively suppresses bot impact.

Regarding General-Public Information Fatigue

Recovery from fatigue F requires deploying a recovery program TR. In particular, the fatigue-recovery probability PF hinges on reducing conflict D and reinforcing θ, the strength of the recovery program.

Optimal Strategy for Resolution

Simulation results identify the following three pillars for stabilizing society:

I. Enhancing Dialog Programs: Introduce a dialog program TD and maintain the optimal ratio of complete-information and cooperative games to alleviate inter-cluster strife.

II. Optimizing Platform Intervention: Suppress bot influence by fortifying PI while minimizing intervention cost CT.

III. Implementing a Recovery Program: Mitigate the general public 's information fatigue via TR and suitably adjust its intensity θ.

To implement these strategies, several challenges emerge:

I. Designing frameworks that optimize complete- information and cooperative game proportions

II. Lowering the cost of platform intervention

III. Formulating an effective recovery program to counteract public fatigue

Considerations on Convergence Probability in Simulations for Controlling AI/Bot Influence

Lastly, we ran simulations addressing AI/bot influence regulation, incorporating both fake-news countermeasures and debunking mechanisms. The findings verify that AI/bot interventions profoundly affect resolution probabilities, and stronger fake-news responses plus debunking capacity markedly improve de- bate integrity.

Formula for the Convergence Probability

Convergence probability PS is computed based on:

$$PS = (1-F)(1-B)(1-D) + (-\lambda_{H} P_H + \lambda_E P_E) + (-\mu_S M_S + \mu_C M_C) + (-V_R G_R + V_T G_T) + \alpha_V V + \beta_{DB} DB$$

where:

I. **F:** Information fatigue

II. **B:** AI/bot intervention level

III. **D:** Power-cluster conflict intensity

IV. **V:** Degree of fake-news countermeasures

V. **DB:** Strength of debunking mechanisms

VI. **PH, PE:** Power cluster choices

VII. **MS, MC:** Intermediate-layer actions

VIII. **GR, GT:** General-public behaviours

with parameter influences:

$\lambda_H = 0.4$, $\lambda_E = 0.6$,

$\mu_S = 0.3$, $\mu_C = 0.5$,

$\nu_R = 0.2$, $\nu_T = 0.4$,

$\alpha_V = 0.5$, $\beta DB = 0.4$

Scenario with Enhanced Fake-News Counter measures

F = 0.3, B = 0.3, D = 0.3, V = 0.8, DB = 0.8

$$PS = (1 - 0.3)(1 - 0.3)(1 - 0.3) + (-0.4 \times 0.3 + 0.6 \times 0.7) + (-0.3 \times 0.3 + 0.5 \times 0.7) + (-0.2 \times 0.3 + 0.4 \times 0.7) + (0.5 \times 0.8) + (0.4 \times 0.8)$$

$$= 0.343 + 0.3 + 0.26 + 0.22 + 0.4 + 0.32 = 1.823$$

leading to PS normalized = 1.00. With robust fake-news efforts, a healthy information environment emerges, maximizing convergence.

Scenario with Intense Misinformation via Bots

F = 0.7, B = 0.9, D = 0.5, V = 0.2, DB = 0.3

$$PS = (1 - 0.7)(1 - 0.9)(1 - 0.5) + (-0.4 \times 0.6 + 0.6 \times 0.4) + (-0.3 \times 0.6 + 0.5 \times 0.4) + (-0.2 \times 0.6 + 0.4 \times 0.4) + (0.5 \times 0.2) + (0.4 \times 0.3)$$

= 0.015 + 0.0 + 0.02 + 0.04 + 0.1 + 0.12 = 0.295

giving $PS_{normalized}$ = 0.30. In an environment lacking adequate responses to rampant AI/bot misinformation, convergence is far more challenging.

Therefore, the simulations confirm:

I. Strong fake-news countermeasures and debunking yield a healthier information environment, pushing convergence probability to its maximum.

II. Where bots proliferate and such countermeasures are lacking, convergence probability drops sharply, eroding debate quality.

III. While debunking encourages resolution, excessive debunking could paradoxically lower convergence probability.

We conclude that maintaining information health demands properly regulating AI/bot impact alongside balanced fake-news and debunking measures.

Acknowledgement

None.

Conflict of Interest

None.

References

1. CC Abt (1970) Serious Games, Viking Press.
2. M Zyda (2005) From Visual Simulation to Virtual Reality to Games. IEEE Computer 38(9): 25-32.
3. B Sawyer (2003) Serious Games: Improving Public Policy through Game-based Learning and Simulation. Woodrow Wilson International Center for Scholars.
4. T Susi, M Johannesson, P Backlund (2007) Serious Games – An Overview. Technical Report, University of Skovde.