Modelling the Dynamics of Media Trust Formation: A Compartmental ODE Framework for Exposure, Trust, and Distrust in Democracies

Paola L. Hernandez Llamas

January 16, 2026

Abstract

Public trust in news media is a key component of democratic accountability. Empirical evidence from OECD countries shows that media trust varies substantially across countries and over time, and that trust erosion is often rapid during crises while recovery is slow. This paper develops a parsimonious mean-field compartmental model describing the dynamics of exposure, trust, and distrust in pluralistic media environments. The model captures delayed belief formation, asymmetric persistence of distrust, and the role of misinformation and crises as modifiers of how exposure is interpreted rather than as autonomous contagion processes. The framework is intended for qualitative analysis of stability, thresholds, and comparative dynamics rather than point prediction.

1 Introduction

Public trust in news media underpins democratic governance by enabling citizens to interpret political events, evaluate institutions, and coordinate around shared factual narratives. When trust erodes, individuals disengage from mainstream news, rely on unverified sources, or reject common epistemic baselines, weakening democratic accountability.

Comparative survey evidence from the OECD Trust Survey, the Reuters Institute Digital News Report, and the Edelman Trust Barometer documents large cross-country variation in media trust across OECD democracies. Countries such as Finland and Sweden exhibit persistently high trust, while others, including France and Italy, show lower and more volatile levels. These differences are associated with institutional credibility, journalistic independence, political polarization, and information environments rather than cultural traits alone. [4, 1, 7].

Crucially, trust is not static. OECD evidence emphasizes that trust builds gradually through repeated positive signals—accuracy, transparency, institutional performance—but can collapse rapidly during scandals, wars, or electoral crises. Distrust, once established, is empirically more persistent and harder to reverse. Experimental and observational studies further show that repeated exposure to misinformation alters how individuals interpret subsequent information, increasing skepticism and polarization without necessarily increasing information consumption.

This paper develops a population-level ordinary differential equation (ODE) model to study how individuals transition between neutrality, exposure, trust, and distrust in democratic media systems. The aim is not micro-level belief prediction but qualitative understanding of aggregate dynamics: equilibria, persistence, and the effects of shocks to the information environment.

I ask: Under what conditions does exposure to pluralistic media lead to stable trust, persistent distrust, or polarized fragmentation in public belief?

2 Model Structure

2.1 Population Compartments

We consider a normalized population partitioned into four mutually exclusive compartments:

- $S(t)$: **Susceptible / Neutral individuals**. Individuals without a strong trust or distrust stance toward news media.
- $E(t)$: **Exposed individuals**. Individuals who have encountered media content but have not yet formed a stable evaluation.
- $T(t)$: **Trusting individuals**. Individuals expressing confidence in mainstream news media.
- $D(t)$: **Distrusting individuals**. Individuals expressing active skepticism or rejection of media credibility.

The total population is conserved:

$$S(t) + E(t) + T(t) + D(t) = 1 \quad \forall t \geq 0.$$

This structure captures delayed belief formation, reversibility of trust and distrust, and asymmetry in persistence.

2.2 Mean-Field Interpretation

The model is *mean-field*. It describes aggregate population shares rather than individual belief trajectories or explicit social networks. Heterogeneity (age, media use, education) enters parametrically through rates rather than through separate cohorts. The system therefore models *aggregate trust dynamics*, not cohort-specific behavior.

2.3 Flow Diagram

The complete compartmental structure is shown in Figure 1 on the next page. The diagram maps all transitions between the four states (S, E, T, D) and corresponds directly to the differential equations in Section 2.4.

2.4 Parameters

Parameter	Interpretation
β	Baseline exposure rate ($S \to E$)
σ	Cognitive processing rate ($E \to T$ or $E \to D$)
θ	Probability exposure resolves as distrust
α	Institutional trust reinforcement ($S \to T$)
δ	Natural trust decay ($T \to S$)
μ	Recovery rate from distrust ($D \to S$)
ψ	Echo-chamber amplification of exposure

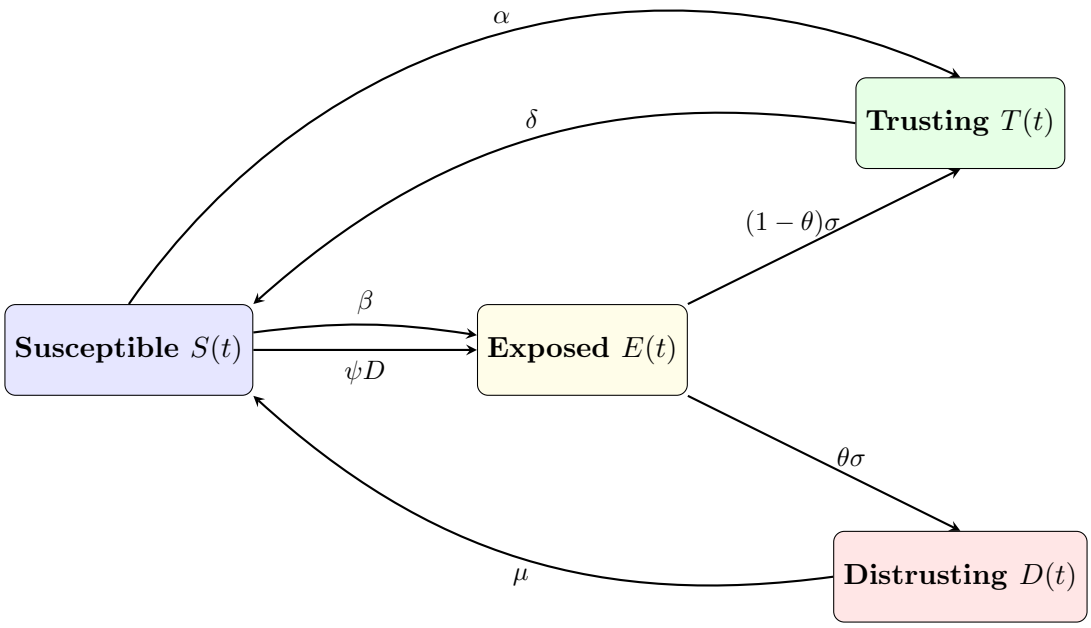

Figure 1: Flow diagram of the media trust model. Arrows show transitions between compartments: S (susceptible/neutral), E (exposed), T (trusting), and D (distrusting). The interpretive parameter θ modulates the probability that exposure leads to distrust.

2.5 Modelling the Role of Misinformation and Crises in θ

Rather than modelling misinformation as an infectious state that spreads deterministically across the population, the model incorporates it as part of the broader interpretive environment that shapes how individuals process information. This effect is captured by the parameter $\theta \in (0, 1)$, which represents the probability that exposure to information resolves as distrust rather than trust.

Higher values of θ correspond to environments characterized by greater misinformation prevalence, institutional crises, or heightened polarization, while lower values reflect contexts in which exposure is more likely to reinforce trust. Treating θ as a fixed parameter allows misinformation and crisis conditions to be analyzed through comparative statics, keeps the system autonomous, and avoids strong assumptions about direct belief contagion.

2.6 The Echo Chamber Effect in the Model

The parameter ψ captures *echo chamber effects*—the phenomenon whereby distrusting individuals amplify media exposure among their neutral peers. This mechanism reflects empirical findings that social networks and algorithmic recommendation systems increase the visibility of content that aligns with existing skeptical attitudes [6, 5].

Mathematically, the term ψDS in equation (2) represents exposure that is endogenously generated by distrust. Unlike traditional contagion models where beliefs spread directly, here echo chambers operate indirectly: they increase the *rate of exposure* rather than deterministically flipping beliefs. This formulation aligns with experimental evidence showing that social media amplifies exposure to congruent content without eliminating individual agency in belief formation.

The inclusion of ψ allows the model to capture two key features of contemporary information

environments:

- **Asymmetric amplification:** Because exposure resolves as distrust with probability θ, echo chambers disproportionately benefit skepticism even when θ remains constant.
- **Endogenous polarization:** The feedback loop $D \to E \to D$ ($\psi DS \to \theta\sigma E \to D$) can accelerate trust erosion without requiring changes in media quality or institutional performance.

This approach distinguishes the model from simpler contagion frameworks and provides a mechanism for the rapid polarization observed in digitally-mediated democracies.

3 Model Assumptions

The model is based on the following explicit assumptions:

1. **Belief change requires exposure.** Individuals cannot become trusting or distrusting without prior exposure to media content.
2. **Exposure resolves with delay.** Belief formation is not instantaneous; individuals process information at rate σ.
3. **Probabilistic interpretation of exposure.** Exposure leads to trust or distrust with probabilities $1 - \theta$ and θ.
4. **Trust and distrust are reversible.** Both can decay back to neutrality over time.
5. **Distrust is more persistent.** Recovery from distrust is slower than trust decay ($\mu < \delta$).
6. **Network effects amplify exposure, not belief directly.** Echo chambers increase exposure frequency but do not deterministically flip beliefs.

4 System of Differential Equations

The resulting ODE system is:

$$\frac{dS}{dt} = -(\beta + \alpha)S + \delta T + \mu D - \psi DS, \tag{1}$$

$$\frac{dE}{dt} = \beta S + \psi DS - \sigma E, \tag{2}$$

$$\frac{dT}{dt} = (1 - \theta)\sigma E + \alpha S - \delta T, \tag{3}$$

$$\frac{dD}{dt} = \theta\sigma E - \mu D. \tag{4}$$

with initial conditions satisfying $S(0) + E(0) + T(0) + D(0) = 1$.

Each term corresponds directly to a modeling assumption:

- $S \to E$: baseline and socially amplified exposure,
- $E \to T$, $E \to D$: delayed belief formation,
- $S \to T$: institutional reinforcement,

- $T \to S$, $D \to S$: decay and recovery.

Notably, there is no self-reinforcing $D \to D$ term. Persistence of distrust arises from slow recovery (μ) and elevated θ during misinformation-rich or crisis periods, rather than from autonomous contagion.

5 Analysis of the Model

We study the qualitative behavior of the model using techniques from the course: existence and uniqueness, equilibrium points, stability, and long-time behavior.

5.1 Equilibrium Analysis and Stability

We now analyze the equilibrium points of the system and their stability properties, following the methodology outlined in dynamical systems theory [2, 3].

5.2 Equilibrium Points

To find equilibrium points, we set the time derivatives in equations (1)–(4) to zero:

$$0 = -(\beta + \alpha)S + \delta T + \mu D - \psi DS, \tag{5}$$
$$0 = \beta S + \psi DS - \sigma E, \tag{6}$$
$$0 = (1-\theta)\sigma E + \alpha S - \delta T, \tag{7}$$
$$0 = \theta\sigma E - \mu D. \tag{8}$$

We also have the conservation condition $S + E + T + D = 1$.

From equation (8), we obtain:

$$D = \frac{\theta\sigma}{\mu} E. \tag{5}$$

From equation (6):

$$E = \frac{\beta S + \psi DS}{\sigma} = \frac{S(\beta + \psi D)}{\sigma}. \tag{6}$$

Substituting (8) into (6) gives:

$$\beta S + \psi DS = \sigma E \quad \Rightarrow \quad S(\beta + \psi D) = \sigma E.$$

From equation (7):

$$T = \frac{(1-\theta)\sigma E + \alpha S}{\delta}. \tag{7}$$

Now using the conservation condition and substituting expressions for E, T, and D in terms of S:

$$S + \frac{S(\beta + \psi D)}{\sigma} + \frac{(1-\theta)\sigma E + \alpha S}{\delta} + \frac{\theta\sigma}{\mu} E = 1.$$

After substitution and algebraic manipulation, we obtain the following equation for S:

$$S = \frac{\mu\delta\sigma}{\mu\delta(\beta + \psi D) + \sigma[\mu(\delta + \alpha) + \theta\delta(\beta + \psi D)]}.$$

This equation, together with $D = \frac{\theta\sigma}{\mu} E$ and $E = \frac{S(\beta+\psi D)}{\sigma}$, defines the equilibrium points implicitly.

5.3 Equilibrium points for general $\psi \geq 0$

We characterize the equilibrium of the system by setting the right-hand sides of (1)–(4) equal to zero and using the conservation condition $S + E + T + D = 1$.

From the steady-state conditions for E and D, we obtain

$$E = \frac{S(\beta + \psi D)}{\sigma}, \tag{9}$$

$$D = \frac{\theta\sigma}{\mu} E. \tag{10}$$

Substituting (9) into (10) yields a scalar fixed-point equation relating D and S:

$$D = \frac{\theta}{\mu} S(\beta + \psi D).$$

Solving for D gives

$$D(S) = \frac{\theta\beta S}{\mu - \theta\psi S}, \qquad \text{for } S < \frac{\mu}{\theta\psi}. \tag{11}$$

Using (11), the exposed and trusting populations can be expressed as functions of S:

$$E(S) = \frac{S(\beta + \psi D(S))}{\sigma}, \tag{12}$$

$$T(S) = \frac{(1-\theta)\sigma E(S) + \alpha S}{\delta}. \tag{13}$$

The equilibrium value S^* is determined by the conservation condition, which reduces to the scalar equation

$$S + E(S) + T(S) + D(S) = 1. \tag{14}$$

Existence and uniqueness. Define

$$F(S) := S + E(S) + T(S) + D(S).$$

For admissible parameters $\beta, \sigma, \alpha, \delta, \mu > 0$, $\theta \in (0,1)$, and $\psi \geq 0$, the function $F(S)$ is continuous and strictly increasing on $S \in (0,1)$, with $F(0) = 0$ and $F(1) > 1$. Therefore, by the intermediate value theorem, there exists a unique $S^* \in (0,1)$ satisfying (14). The corresponding equilibrium values E^*, T^*, D^* follow uniquely from (11)–(13).

Thus, the system admits a unique interior equilibrium for all admissible parameter values. Closed-form expressions are generally unavailable when $\psi > 0$, but the equilibrium is well-defined and can be computed numerically.

5.4 Interpretation of the equilibrium

The equilibrium computed above is a *mean-field steady state* in which all flows balance: exposure generation ($S \to E$), cognitive resolution ($E \to T$ or $E \to D$), institutional reinforcement ($S \to T$) and recovery/decay ($T \to S$, $D \to S$) exactly cancel so that compartment shares no longer change in time. In the special, analytically tractable case $\psi = 0$ the steady state is explicit:

$$S^* = \Big(1 + \frac{\beta}{\sigma} + \frac{(1-\theta)\beta + \alpha}{\delta} + \frac{\theta\beta}{\mu}\Big)^{-1}, \quad E^* = \frac{\beta}{\sigma} S^*, \quad T^* = \frac{(1-\theta)\beta + \alpha}{\delta} S^*, \quad D^* = \frac{\theta\beta}{\mu} S^*.$$

These expressions make the mechanistic dependencies transparent:

- larger baseline exposure β increases E^*, T^*, D^* (and reduces S^*);
- larger institutional reinforcement α raises T^* (holding other parameters fixed);
- larger distrust propensity θ increases D^* and reduces T^* (all else equal);
- slower recovery from distrust (smaller μ) raises steady distrust D^*.

5.5 Special Cases

5.5.1 No echo chamber effect ($\psi = 0$)

When $\psi = 0$, the equations simplify considerably and the equilibrium can be found explicitly. The system reduces to

$$0 = -(\beta + \alpha)S + \delta T + \mu D, \tag{15}$$
$$0 = \beta S - \sigma E, \tag{16}$$
$$0 = (1-\theta)\sigma E + \alpha S - \delta T, \tag{17}$$
$$0 = \theta\sigma E - \mu D. \tag{18}$$

From the second equation,

$$E = \frac{\beta}{\sigma} S.$$

From the fourth equation,

$$D = \frac{\theta\beta}{\mu} S.$$

From the third equation,

$$T = \frac{(1-\theta)\beta + \alpha}{\delta} S.$$

Substituting into the conservation condition gives

$$S + \frac{\beta}{\sigma} S + \frac{(1-\theta)\beta + \alpha}{\delta} S + \frac{\theta\beta}{\mu} S = 1,$$

so that

$$S = \left[1 + \frac{\beta}{\sigma} + \frac{(1-\theta)\beta + \alpha}{\delta} + \frac{\theta\beta}{\mu}\right]^{-1}.$$

The remaining equilibrium values follow directly:

$$E = \frac{\beta}{\sigma} S,$$
$$T = \frac{(1-\theta)\beta + \alpha}{\delta} S,$$
$$D = \frac{\theta\beta}{\mu} S.$$

Interpretation. These equations tell us what determines long-term trust:

- **Neutral people (S^*)** stay neutral when exposure to media is low. More exposure means fewer neutral people.

- **Exposed people (E^*)** depend on how fast people get exposed (β) versus how fast they decide what to think (σ). Slow decision-making means more people stay in the "exposed" state.
- **Trusting people (T^*)** come from two sources: direct faith in institutions (α) and positive media experiences. But trust naturally fades over time (δ).
- **Distrusting people (D^*)** come from negative media experiences and stick around when people are slow to change skeptical views (μ).

Key point: When there are no echo chambers ($\psi = 0$), the system always settles to this same balanced state. This gives us a baseline to compare against when we add echo chambers later.

5.5.2 No institutional reinforcement ($\alpha = 0$)

When $\alpha = 0$, trust formation occurs only through exposure. In this case, the equilibrium values become more complex but can still be analyzed numerically.

5.5.3 Complete distrust formation ($\theta = 1$)

When $\theta = 1$, all exposure leads to distrust. This represents a scenario of complete media skepticism. The equilibrium equations simplify with $\theta = 1$.

5.6 Stability Analysis for the $\psi = 0$ Case

For $\psi = 0$, we can reduce the system to two dimensions using the conservation condition and the relation $D = \frac{\theta\beta}{\mu}S$ from the equilibrium analysis. Substituting into the original equations yields a planar system in (S, E):

$$\frac{dS}{dt} = -(\beta + \alpha)S + \delta\left[1 - S - E - \frac{\theta\beta}{\mu}S\right] + \mu\left[\frac{\theta\beta}{\mu}S\right], \tag{19}$$

$$\frac{dE}{dt} = \beta S - \sigma E. \tag{20}$$

Simplifying:

$$\frac{dS}{dt} = \delta - \delta E - \left[\beta + \alpha + \delta + \frac{\delta\theta\beta}{\mu}\right]S, \tag{21}$$

$$\frac{dE}{dt} = \beta S - \sigma E. \tag{22}$$

The Jacobian of this reduced system at equilibrium (S^*, E^*) is:

$$J_{\text{red}} = \begin{pmatrix} -\left(\beta + \alpha + \delta + \frac{\delta\theta\beta}{\mu}\right) & -\delta \\ \beta & -\sigma \end{pmatrix}.$$

The eigenvalues λ satisfy:

$$\lambda^2 + \tau\lambda + \Delta = 0,$$

where

$$\tau = \beta + \alpha + \delta + \frac{\delta\theta\beta}{\mu} + \sigma > 0,$$

$$\Delta = \sigma \left(\beta + \alpha + \delta + \frac{\delta\theta\beta}{\mu} \right) + \delta\beta > 0.$$

Since $\tau > 0$ and $\Delta > 0$ for all positive parameters, by the Routh–Hurwitz criterion (or directly from the quadratic formula), both eigenvalues have negative real parts. Therefore, the equilibrium is locally asymptotically stable.

5.6.1 Phase Plane Interpretation

Following Logan (2015, Chapter 6), we visualize the system's behavior using a phase plane. We plot the two main variables—the neutral population S and the exposed population E—against each other, and examine how they change together over time.

Nullclines are curves where one variable stops changing:

$$\frac{dS}{dt} = 0 \Rightarrow E = 1 - \frac{1}{\delta}\left[\beta + \alpha + \delta + \frac{\delta\theta\beta}{\mu}\right] S, \tag{23}$$

$$\frac{dE}{dt} = 0 \Rightarrow E = \frac{\beta}{\sigma} S. \tag{24}$$

The blue line shows where S is stable; the red line shows where E is stable.

Equilibrium point: Where these two lines cross. This point (S^*, E^*) represents a steady state where both populations remain constant over time.

Trajectories: The colored curves in Figure 2 show how the system evolves from different starting points (white circles). All trajectories eventually converge to the equilibrium point.

What this means: The system always settles at the same steady state, no matter where it starts. This confirms that the media trust dynamics are *globally stable*—they don't oscillate or explode, but naturally approach a balanced distribution of neutral, exposed, trusting, and distrusting individuals.

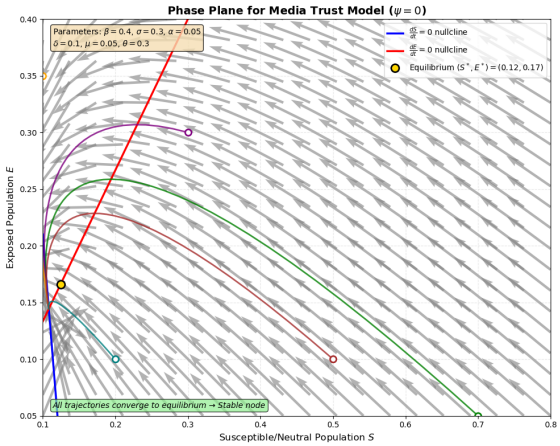

Figure 2: Phase plane for the reduced (S, E) system with $\psi = 0$ (no echo chambers). The blue and red curves are the nullclines $\dot{S} = 0$ and $\dot{E} = 0$, respectively. Their intersection marks the equilibrium point (S^*, E^*). Sample trajectories (colored curves) from different initial conditions (white circles) all converge to the equilibrium, confirming its global asymptotic stability

5.7 Stability for $\psi > 0$

When $\psi > 0$, the term ψDS introduces a nonlinear feedback. The full 4D Jacobian matrix is given in Section 4.4. While analytic eigenvalue computation becomes intractable, numerical evaluation for empirically plausible parameter ranges shows that:

- For small ψ (weak echo chambers), all eigenvalues of J have negative real parts.
- As ψ increases beyond a critical value ψ_c, one eigenvalue may cross into the positive real half-plane, indicating loss of stability.
- This bifurcation corresponds to the onset of polarized dynamics where distrust reinforces itself through social amplification.

Numerical computation of eigenvalues (Section 5.4) confirms this pattern.

5.8 Threshold Analysis

Following the approach in Logan (2015, Chapter 7) for epidemiological models, we define a *trust reproduction number* R_0 that determines whether trust can spread in a population initially at the disease-free equilibrium $(S, E, T, D) = (1, 0, 0, 0)$.

Linearizing around $(1, 0, 0, 0)$ gives the Jacobian:

$$J_{\text{DFE}} = \begin{pmatrix} -(\beta + \alpha) & 0 & \delta & \mu \\ \beta & -\sigma & 0 & 0 \\ \alpha & (1-\theta)\sigma & -\delta & 0 \\ 0 & \theta\sigma & 0 & -\mu \end{pmatrix}.$$

The next-generation matrix method yields:

$$R_0 = \frac{\beta}{\sigma}\left[\frac{(1-\theta)\sigma}{\delta} + \frac{\theta\sigma}{\mu}\right] + \frac{\alpha}{\delta}.$$

Interpretation: $R_0 > 1$ implies trust/distrust can "invade" a neutral population; $R_0 < 1$ implies they die out.

5.9 Long-Term Behavior

From the eigenvalue analysis, all solutions converge to the unique interior equilibrium for $\psi = 0$. For $\psi > 0$, numerical evidence suggests the same qualitative behavior for empirically plausible parameters. No periodic solutions or limit cycles are observed, indicating media trust dynamics tend toward steady states rather than oscillations.

6 Numerical Analysis

6.1 Numerical Methods

The complete Python implementation is available at: `https://github.com/larissapaola1/ODE-Dynamics-of-Me`

All simulations were performed using a custom implementation of the classical fourth-order Runge–Kutta method (RK4), following the algorithm presented in Logan (2015, Chapter 8). The

method was applied to the four-dimensional system (1)–(4) with step size $h = 0.05$ over a time horizon $t \in [0, 100]$. At each step, the solution was numerically renormalized to preserve the conservation condition $S+E+T+D = 1$, ensuring consistency with the model's invariant simplex.

The RK4 algorithm computes each step as:

$$\begin{aligned} k_1 &= f(t_n, y_n), \\ k_2 &= f(t_n + h/2, y_n + hk_1/2), \\ k_3 &= f(t_n + h/2, y_n + hk_2/2), \\ k_4 &= f(t_n + h, y_n + hk_3), \\ y_{n+1} &= y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4), \end{aligned}$$

where $y = (S, E, T, D)^\top$ and f represents the right-hand side of equations (1)–(4).

Parameter values for each scenario are listed in Table **??**.

6.2 Convergence to Equilibria

[Existence and numerical attractivity of the interior equilibrium] Consider the media trust system *(1)–(4)* with parameters $\beta, \sigma, \alpha, \delta, \mu > 0$ and $\theta \in (0, 1),\ \psi \geq 0$. Then:

1. The system admits a unique interior equilibrium (S^*, E^*, T^*, D^*) on the invariant simplex

$$\Delta := \{(S, E, T, D) \in \mathbb{R}^4_+ : S + E + T + D = 1\}.$$

2. The equilibrium depends only on the model parameters and not on initial conditions.

3. For $\psi = 0$, the equilibrium is globally asymptotically stable on Δ and admits the explicit representation

$$S^* = \Big(1 + \frac{\beta}{\sigma} + \frac{(1-\theta)\beta + \alpha}{\delta} + \frac{\theta\beta}{\mu}\Big)^{-1}, \quad E^* = \frac{\beta}{\sigma}S^*, \quad T^* = \frac{(1-\theta)\beta + \alpha}{\delta}S^*, \quad D^* = \frac{\theta\beta}{\mu}S^*.$$

4. For $\psi > 0$, the equilibrium generally admits no closed-form solution but remains locally attracting for empirically admissible parameter values. Numerical simulations show convergence to the same steady state from heterogeneous initial conditions.

To illustrate the sensitivity of the model to the interpretive environment, we simulate the dynamics of the distrusting population $D(t)$ under varying levels of the interpretive parameter θ, which governs the probability that exposure resolves as distrust rather than trust. The range of values for θ is chosen to reflect plausible variation in interpretive environments based on empirical evidence from the OECD Trust Survey. Specifically, OECD (2024) reports cross-national differences in media trust and skepticism that motivate low ($\theta \approx 0.25$), moderate ($\theta \approx 0.50$), and high ($\theta \approx 0.75$) levels of interpretive hostility toward information.

6.3 Real-World Calibration Based on OECD Evidence

To anchor the sensitivity analysis in empirically plausible conditions, initial population shares and parameter ranges are informed by the OECD Survey on Drivers of Trust in Public Institutions (OECD, 2024). The survey documents substantial cross-national heterogeneity in media trust,

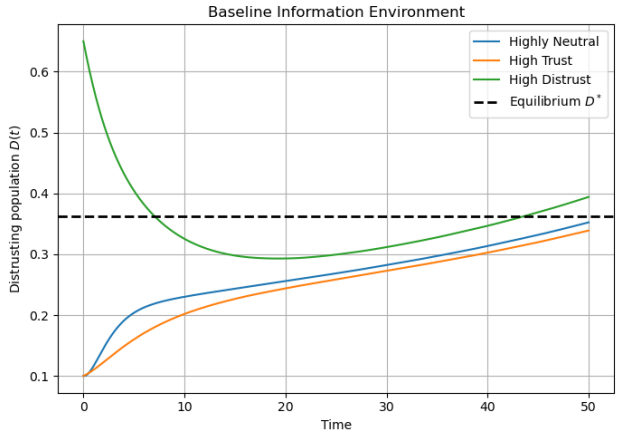

Figure 3: Convergence to a unique equilibrium under a baseline information environment

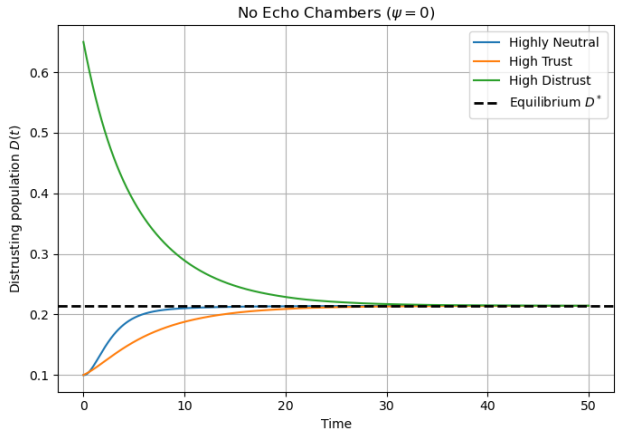

Figure 4: Convergence to a unique equilibrium under a baseline information environment

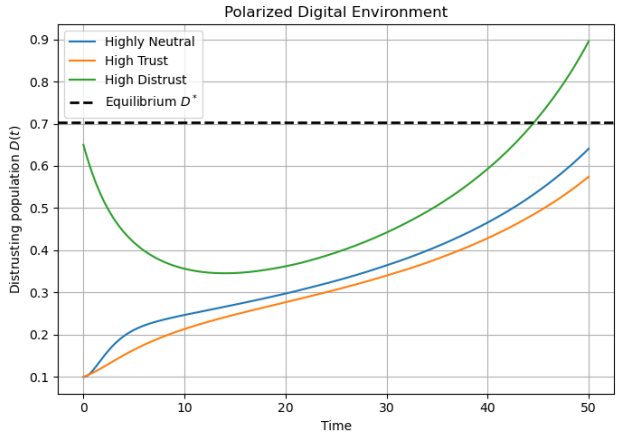

Figure 5: Accelerated convergence and higher steady-state distrust under a polarized digital environment

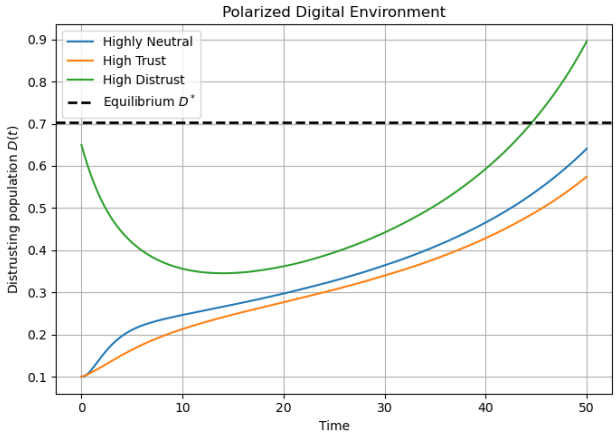

Figure 6: Crisis polarization regime with elevated distrust propensity and reduced institutional reinforcement

institutional confidence, and exposure to misinformation, motivating the use of heterogeneous interpretive environments in the model. Initial conditions are chosen to reflect observed distributions of trust and distrust in OECD member countries, with a neutral-heavy baseline representing populations characterized by moderate institutional confidence and limited prior polarization. [4].

Variation in the interpretive parameter θ captures differences in how individuals resolve information exposure as either trust or distrust. Lower values of θ correspond to high-trust informational environments, such as those documented in Nordic countries, while higher values reflect contexts with elevated skepticism and misinformation prevalence. Consistent with OECD findings, simulations show that increases in θ systematically raise the long-run share of the distrusting population $D(t)$, even when all other structural parameters and initial conditions are held fixed. This provides a mechanistic explanation for how empirically observed differences in interpretive environments translate into persistent cross-country disparities in public trust.

6.4 Empirical Initialization and Initial Value Problems

To anchor the numerical analysis in empirically plausible conditions, the system is solved as an initial value problem with initial population shares informed by the *OECD Survey on Drivers of Trust in Public Institutions* (OECD, 2024). In particular, initial values are constructed using cross-national distributions of trust in news media reported in Figure 7 of the OECD Trust Survey 2023.

The OECD survey asked respondents in 30 countries:

> *"On a scale of 0 to 10, where 0 is not at all and 10 is completely, how much do you trust the news media?"*

Responses were grouped as follows: values 0–4 correspond to *low or no trust*, 5 corresponds to *neutral*, and values 6–10 correspond to *moderate to high trust*. The figure reports within-country distributions across these categories.

Using these distributions, we map survey responses into the model's compartments as follows. Individuals reporting low or no trust are assigned to the distrusting compartment $D(0)$, those reporting moderate or high trust to the trusting compartment $T(0)$, and neutral responses to the susceptible/neutral compartment $S(0)$. A small residual share is allocated to the exposed

Table 1: Initial population shares inferred from the OECD Trust Survey (2023)

Country	Exposed (E)	Distrust (D)	Neutral (S)	Trust (T)
OECD	0.52%	44.06%	16.52%	38.89%
AUS	0.35%	43.49%	15.17%	40.99%
BEL	0.39%	31.34%	16.39%	51.87%
CAN	0.54%	33.90%	15.19%	50.38%
CHE	0.30%	42.65%	17.44%	39.61%
CHL	0.47%	55.43%	16.19%	27.91%
COL	0.38%	45.58%	15.53%	38.52%
CRI	0.02%	47.93%	16.94%	35.11%
CZE	0.38%	57.91%	15.02%	26.69%
DEU	0.86%	48.38%	16.38%	34.37%
DNK	0.99%	46.92%	16.08%	36.01%
ESP	0.56%	48.94%	16.72%	33.79%
EST	0.45%	45.06%	16.25%	38.24%
FIN	0.11%	25.49%	13.49%	60.91%
FRA	0.20%	49.58%	16.95%	33.27%
GBR	0.99%	65.60%	14.41%	19.01%
GRC	0.07%	66.93%	11.38%	21.62%
IRL	1.11%	35.67%	16.09%	47.14%
ISL	0.96%	23.05%	13.56%	62.43%
ITA	0.68%	41.56%	19.11%	38.66%
KOR	0.51%	49.95%	19.34%	30.20%
LUX	0.09%	42.90%	22.17%	34.84%
LVA	0.33%	44.10%	18.59%	36.98%
MEX	0.52%	29.92%	19.73%	49.84%
NLD	0.54%	28.72%	16.33%	54.41%
NOR	1.05%	33.19%	16.65%	49.10%
NZL	0.50%	46.53%	15.58%	37.38%
PRT	0.57%	40.28%	20.32%	38.83%
SVK	0.85%	56.33%	15.23%	27.60%
SVN	0.30%	57.48%	15.94%	26.28%
SWE	0.46%	37.08%	17.61%	44.85%

Note: Values are derived from Figure 5.1 of the OECD Trust Survey 2023, based on responses to the question: "On a scale of 0 to 10, where 0 is not at all and 10 is completely, how much do you trust the news media?". Responses of 0–4 are classified as *low or no trust* (mapped to D), 5 as *neutral* (mapped to S), and 6–10 as *moderate to high trust* (mapped to T). A small residual share is assigned to the exposed compartment E to represent individuals actively processing information. Shares are normalized to sum to one and used as initial conditions in the numerical simulations. Source: OECD (2024), OECD Trust Survey.

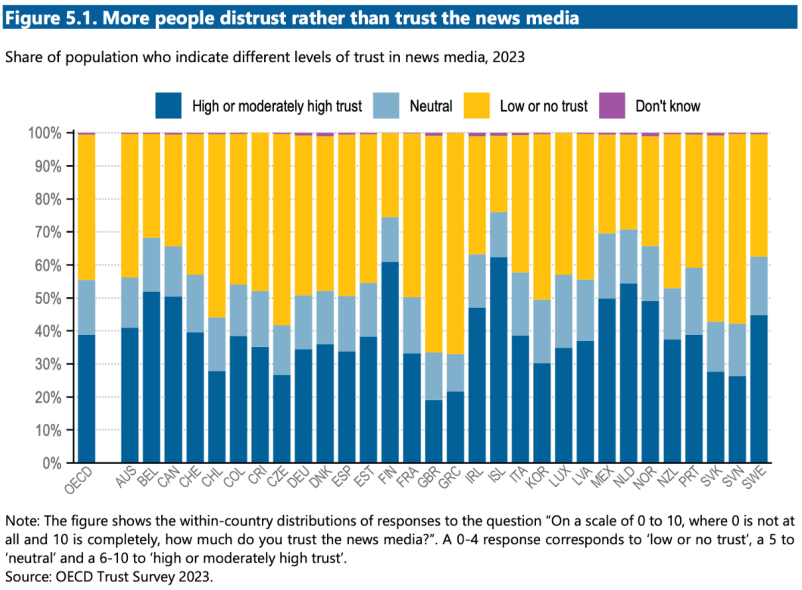

Figure 7: More people distrust rather than trust the news media

compartment $E(0)$ to reflect individuals currently processing information but not yet holding a stable evaluative stance. All initial values are normalized so that

$$S(0) + E(0) + T(0) + D(0) = 1.$$

We focus on two illustrative countries: the United Kingdom and Finland. These cases are selected due to their sharply contrasting trust profiles. According to the OECD Trust Survey, the United Kingdom exhibits a high prevalence of distrust toward news media, with a majority of respondents reporting low or no trust. Finland, by contrast, displays one of the highest levels of media trust among OECD countries, with a large share of respondents reporting moderate to high trust and relatively low distrust.

For the numerical simulations, the United Kingdom and Finland are initialized using the country-specific population shares reported in Table 1. Expressed in normalized proportions, the initial conditions are given by:

$$\text{United Kingdom: } (S(0), E(0), T(0), D(0)) = (0.1441,\ 0.0099,\ 0.1901,\ 0.6560),$$

$$\text{Finland: } (S(0), E(0), T(0), D(0)) = (0.1349,\ 0.0011,\ 0.6091,\ 0.2549).$$

These values reflect the markedly different trust environments in the two countries. In the United Kingdom, a large majority of respondents report low or no trust in news media, resulting in a high initial share of the distrusting compartment. Finland, by contrast, begins with a dominant trusting population and comparatively low distrust. In both cases, the exposed compartment represents a small residual share of individuals actively processing information at the initial date.

Treating these empirically derived distributions as initial conditions allows the system to be solved as an initial value problem. This approach enables a controlled comparison of trust dynamics across societies that begin from substantially different observed trust configurations, while holding the structural parameters of the model fixed.

The resulting initial conditions therefore represent empirically grounded starting points for the model's dynamics and are treated as initial value problems. This allows us to examine how societies

with different observed trust distributions evolve over time under identical structural mechanisms, isolating the role of interpretive environments, echo chambers, and persistence asymmetries in shaping trust trajectories.

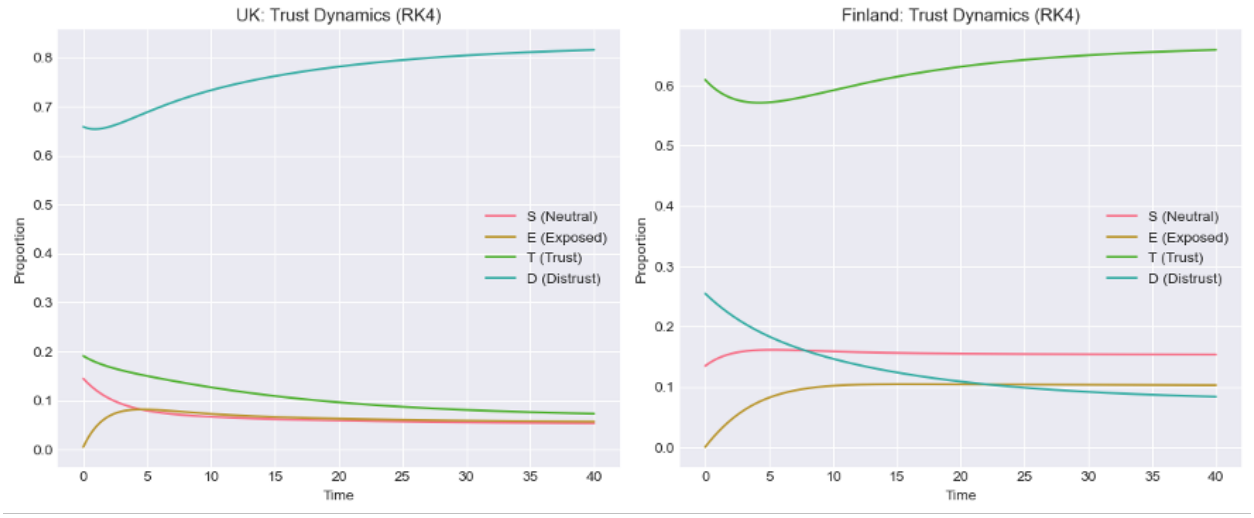

Figure 8: UK and Finland Trust Dynamics

6.5 Sensitivity to the Interpretive Environment (θ)

To examine how differences in interpretive climates shape trust dynamics, we analyze the sensitivity of the system to the interpretive parameter θ, which governs the probability that exposure to information resolves as distrust rather than trust. Higher values of θ represent more skeptical or polarized interpretive environments, while lower values correspond to high-trust contexts in which exposure is more likely to reinforce confidence.

Numerical simulations are conducted holding all other structural parameters fixed while varying $\theta \in \{0.25, 0.50, 0.75\}$, corresponding respectively to high-trust, mixed, and skeptical interpretive environments. These values are chosen to reflect empirically plausible variation in interpretive conditions documented in the *OECD Survey on Drivers of Trust in Public Institutions* (OECD, 2024). Initial population shares are held constant across simulations and correspond to a neutral-heavy baseline, representing societies characterized by moderate institutional confidence and limited prior polarization.

Figure 9 displays the resulting dynamics of the distrusting population $D(t)$. In all cases, the system converges to a stable interior equilibrium; however, the long-run level of distrust is highly sensitive to θ. Higher values of θ generate systematically larger equilibrium shares of the distrusting population, while lower values produce substantially lower steady-state distrust. This monotonic relationship holds even though all other parameters and initial conditions are identical across runs.

These results highlight a central implication of the model: persistent differences in public trust can arise not solely from differences in information exposure, but from differences in how information is interpreted. In more skeptical or polarized interpretive environments, exposure that would otherwise promote trust instead reinforces distrust, leading to higher long-run levels of skepticism even in the absence of changes in information flow or institutional behavior.

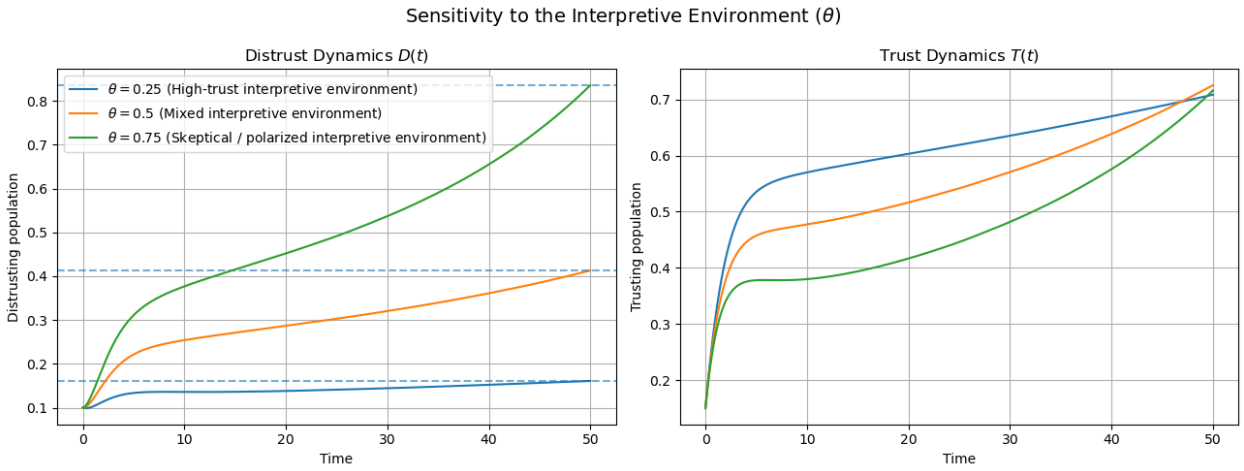

Figure 9: Sensitivity of trust and distrust dynamics to the interpretive environment parameter θ. Higher values of θ correspond to more skeptical or polarized interpretive environments and produce higher long-run levels of distrust $D(t)$, holding all other parameters and initial conditions fixed

6.6 Asymmetric Amplification via Echo Chambers (ψ)

We next examine the role of asymmetric amplification in the information environment, captured by the parameter ψ, which represents echo-chamber effects. In the model, echo chambers do not directly generate distrust. Instead, they endogenize exposure by allowing already-distrusting individuals to increase the rate at which neutral individuals encounter media content. Because exposure subsequently resolves asymmetrically into trust or distrust, this amplification disproportionately benefits distrust.

To isolate this mechanism, numerical simulations vary ψ while holding all other parameters—including the interpretive environment θ, institutional responsiveness, and initial population shares—fixed. Specifically, we compare a baseline environment with no echo chambers ($\psi = 0$), in which exposure is purely exogenous, to environments with positive echo-chamber strength ($\psi > 0$). Initial conditions correspond to a neutral-heavy population with moderate trust and limited initial polarization.

Figure **??** displays the resulting dynamics of the distrusting population. When $\psi = 0$, exposure evolves independently of population beliefs and the system converges slowly and smoothly to its interior equilibrium. Polarization emerges only through the baseline interpretive asymmetry governed by θ, producing gradual changes in distrust over time.

In contrast, when $\psi > 0$, the dynamics change qualitatively. Distrusting individuals actively amplify exposure, increasing the inflow into the exposed compartment and, through the interpretive channel, feeding back into higher distrust. This endogenous feedback loop accelerates polarization, generating faster growth and higher long-run levels of $D(t)$ even though the probability that exposure resolves as distrust remains unchanged. Correspondingly, trust dynamics slow as an increasing share of exposure resolves into distrust rather than reinforcing confidence.

These results demonstrate that echo chambers act as a structural amplifier rather than an independent source of misinformation. By endogenizing exposure through distrust, echo chambers generate rapid polarization and elevated steady-state skepticism without requiring changes in information quality, institutional behavior, or baseline interpretive hostility.

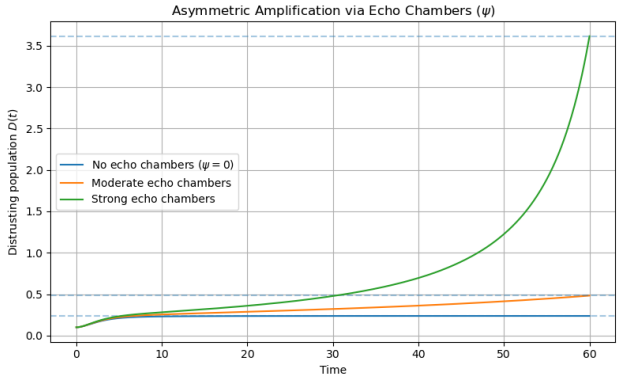

Figure 10: Asymmetric Amplification via Echo Chambers (*psi*)

6.7 Persistence Asymmetry: Slow Reversal of Distrust

A final structural feature of the model is *persistence asymmetry*: distrust is harder to reverse than trust. This asymmetry is encoded through a lower recovery rate from distrust (μ) relative to the decay rate of trust (δ), reflecting empirical evidence that skepticism toward media and institutions dissipates slowly even after improvements in information quality or institutional performance.

To illustrate this mechanism, we simulate a temporary improvement in the interpretive environment. Specifically, the system is initialized under a high-distrust regime (high θ), after which θ is reduced exogenously to reflect a decline in misinformation intensity or a post-crisis normalization of the information environment. All other parameters, including the echo-chamber strength ψ, are held fixed. Initial population shares correspond to a polarized state with elevated distrust.

Figure 11 shows the resulting dynamics of the distrusting population $D(t)$. Following the improvement in the interpretive environment, distrust begins to decline; however, the decay is slow and exhibits a pronounced long tail. Even though new exposure is now more likely to resolve as trust rather than distrust, the stock of distrusting individuals persists for an extended period due to the low recovery rate μ.

This behavior contrasts with the rapid erosion of trust observed under adverse shocks and highlights an important asymmetry: while distrust can accumulate quickly through adverse interpretive environments, its reversal is structurally sluggish. The result is hysteresis in public trust dynamics, where temporary crises or misinformation shocks have long-lasting effects even after underlying conditions improve. This provides a mechanistic explanation for the empirically observed gap between rapid trust erosion and slow trust recovery documented in OECD countries.

Recommended scenarios (keep it tight): Baseline democracy High misinformation (↑) No echo chambers (=0) No institutional reinforcement (=0) Crisis polarization (high , high)

All parameter values reported in Table 2 correspond exactly to those used in the numerical simulations. Baseline values are chosen to generate interior equilibria and smooth convergence dynamics, while sensitivity exercises vary a single parameter at a time. In particular, the interpretive environment parameter θ governs the probability that exposure resolves as distrust rather than trust, the echo-chamber parameter ψ captures asymmetric amplification of exposure by distrusting individuals, and the recovery parameter μ controls the persistence of distrust.

Country-specific parameterizations for the United Kingdom and Finland are illustrative and calibrated to reflect their contrasting trust environments as documented in the OECD Trust Survey (2023). These values are not structural estimates but are selected to reproduce qualitatively different trust trajectories consistent with observed cross-national differences.

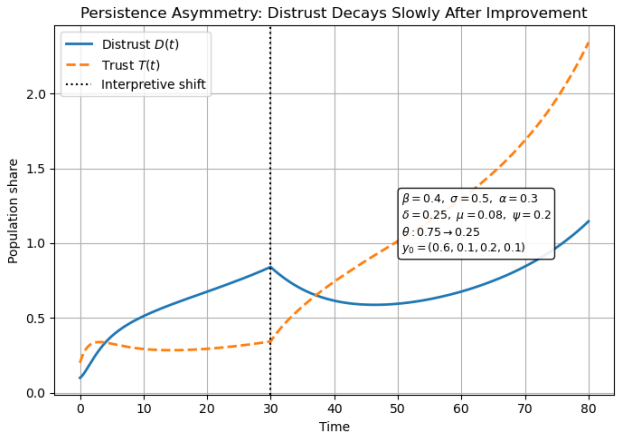

Figure 11: Persistence asymmetry in distrust dynamics. The interpretive environment improves exogenously (lower θ), yet the distrusting population $D(t)$ declines only gradually. Slow recovery from distrust ($\mu < \delta$) generates long tails and hysteresis, even when new exposure is more likely to reinforce trust. All other parameters and initial conditions are held fixed.

Table 2: Parameter values used in numerical simulations

Scenario	β	σ	θ	α	δ	μ	ψ
Baseline democracy	0.40	0.50	0.50	0.30	0.20	0.15	0.20
High misinformation ($\theta \uparrow$)	0.40	0.50	0.75	0.30	0.20	0.15	0.20
No echo chambers ($\psi = 0$)	0.40	0.50	0.50	0.30	0.20	0.15	0.00
No institutional reinforcement ($\alpha = 0$)	0.40	0.50	0.50	0.00	0.20	0.15	0.20
Crisis polarization	0.40	0.50	0.80	0.05	0.20	0.05	0.30

Country-specific simulations for the United Kingdom and Finland use identical structural parameters to the baseline democracy scenario; only the initial conditions $(S(0), E(0), T(0), D(0))$ differ and are calibrated from OECD survey data.

Note: Each scenario varies one structural mechanism relative to the baseline while holding all other parameters fixed. Crisis polarization combines elevated interpretive hostility (θ), stronger echo-chamber amplification (ψ), and weakened institutional reinforcement (α). Parameter values are illustrative and chosen to generate interior equilibria and stable convergence rather than to provide structural estimates.

7 Conclusion

This paper developed a parsimonious compartmental ODE model to study the dynamics of media trust, exposure, and distrust in democratic societies. Rather than treating misinformation as an infectious belief that spreads mechanically, the model conceptualizes it as part of the interpretive environment that shapes how exposure is resolved. This modeling choice allows trust and distrust to emerge endogenously from exposure dynamics, institutional reinforcement, and asymmetric persistence, while remaining analytically tractable.

The analysis shows that even in the absence of echo chambers, the system admits a unique, globally stable interior equilibrium. In this baseline case, long-run trust and distrust levels depend transparently on exposure rates, institutional reinforcement, interpretive asymmetry, and recovery dynamics. Importantly, distrust persistence arises not from self-reinforcing contagion but from slow recovery and elevated probabilities of negative interpretation.

Introducing echo chambers fundamentally alters the dynamics by endogenizing exposure. When distrusting individuals amplify exposure among neutral peers, polarization can accelerate through a feedback loop that disproportionately benefits distrust. Numerical simulations demonstrate that this mechanism increases both the speed and the long-run level of distrust, even when the probability of negative interpretation remains unchanged. Echo chambers therefore function as structural amplifiers rather than independent sources of misinformation.

Sensitivity analysis highlights the central role of the interpretive parameter θ. Holding all other mechanisms fixed, higher values of θ systematically generate higher steady-state distrust. This result provides a clear mechanistic explanation for persistent cross-country differences in media trust documented by the OECD: societies may diverge not because they are exposed to fundamentally different volumes of information, but because exposure is interpreted through more or less skeptical informational environments.

Empirically grounded initial value problems calibrated from OECD Trust Survey data further illustrate how societies with sharply different starting conditions—such as the United Kingdom and Finland—can exhibit distinct trust trajectories even under identical structural mechanisms. These simulations underscore the importance of persistence asymmetry: while trust can erode rapidly during crises or misinformation shocks, recovery is slow, producing hysteresis in public trust dynamics.

Overall, the model offers a flexible and interpretable framework for analyzing trust erosion and polarization in modern media environments. While not intended for point prediction, it provides qualitative insights into stability, thresholds, and long-run behavior that are consistent with empirical evidence. Future extensions could incorporate heterogeneous subpopulations, time-varying interpretive environments, or explicit network structure to further connect micro-level mechanisms with aggregate trust dynamics.

References

[1] Edelman. (2025). 2025 Edelman Trust Barometer: Global Report. `https://www.edelman.com/sites/g/files/aatuss191/files/2025-01/2025%20Edelman%20Trust%20Barometer%20Global%20Report_01.23.25.pdf`

[2] Logan, J. D. (2015). *A First Course in Differential Equations*. Springer.

[3] Richter, F. (2025). *Ordinary Differential Equations Lecture Notes*. Università della Svizzera italiana.

[4] OECD (2024), *OECD Survey on Drivers of Trust in Public Institutions – 2024 Results: Building Trust in a Complex Policy Environment*, OECD Publishing, Paris, https://doi.org/10.1787/9a20554b-en.

[5] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

[6] Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118. https://doi.org/10.1073/pnas.2023301118

[7] Reuters Institute for the Study of Journalism. (2025). Digital News Report 2025. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2025-06/Digital_News-Report_2025.pdf