MRN Publishing

Vol. 13, No. 22, 2025, pp. 42–53.
ISSN 1596−5074

Management Network Journal

Rebuilding Trust from Below: An Agent-Based Model of Citizen–State Interactions in Post-Conflict Northeast Nigeria

Dauda Mohammed Atiyaye[1]*, Abubakar Muhammad Atiyaye[2]

[1] *Department of Public Administration, School of Management Sciences, The Federal Polytechnic Damaturu, Nigeria.*
[2] *Department of Business Administration and Management, School of Management Sciences, The Federal Polytechnic Damaturu, Nigeria.*

Abstract

Background: Rebuilding citizen trust is critical for service uptake and stabilization in post-conflict settings, yet micro-mechanisms linking frontline encounters, social networks, and policy interventions remain poorly understood. *Methods:* We develop an agent-based model (ODD protocol) representing heterogeneous citizens, frontline officials, community leaders, and a regulator embedded in spatially structured communities. Key parameters varied include frontline responsiveness, corruption visibility, enforcement strength, community-engagement frequency, network cohesion, and shock incidence. Experiments combine factorial sweeps and Latin-Hypercube sampling with 200+ stochastic replications per scenario. Outcomes include time-series of average trust, trust distributional metrics, time-to-recovery after shocks, officials' corruption rates, and spatial clustering. *Results:* Simulations identify a responsiveness threshold (~0.55–0.60) beyond which social diffusion produces rapid, nonlinear trust gains. Community engagement multiplies the effects of service improvements, particularly in cohesive networks. Transparency alone causes a short-term "exposure" dip in trust when corruption is high, but paired with enforcement or visible service gains yields superior long-run outcomes. Combined packages (modest responsiveness gains + frequent engagement + targeted monitoring) produce the fastest recovery, lowest corruption equilibria, and greatest resilience to shocks. Sensitivity analysis ranks responsiveness and engagement as dominant drivers. *Conclusions:* Trust recovery is emergent, path-dependent, and conditional on sequencing: prioritize credible engagement and consistent frontline reliability before or alongside transparency and enforcement. We provide open code and propose targeted pilots to calibrate model parameters and test policy packages empirically.

Keywords: *Agent-Based Model; Trust; Post-Conflict Governance; Frontline Services; Community Engagement; Northeast Nigeria.*

Introduction

Everyday citizen–state interactions in Northeast Nigeria occur against a backdrop of prolonged insurgency, displacement, and weakened public services. Humanitarian assessments document fractured local governance, collapsed services, and pervasive public skepticism following years of violence (Alebiosu, 2024). Low institutional trust undermines service uptake, compliance, and stabilization, turning governance deficits into persistent fragility (World Bank, 2011).

Trust and legitimacy matter because they condition citizens' willingness to cooperate, use services, and accept authority; these relationships are especially frayed after conflict when institutions are seen as inefficient or partial (Tyler, 2006; Ostrom, 1990). Yet unpacking how micro-level encounters with frontline officials aggregate into system-level trust trajectories is empirically difficult: conflict constrains access, administrative records are often incomplete, and ethical or security considerations limit field experimentation (Kalyvas, 2006; Collier, 2003). Moreover, trust dynamics are path dependent and potentially nonlinear—small, local interactions can produce large, emergent macro-outcomes (Schelling, 1971; Ostrom, 1990).

Agent-based modeling (ABM) offers a complementary method to address these challenges. By encoding simple behavioral rules for heterogeneous agents and embedding them in social and spatial environments, ABMs generate emergent macro-patterns from micro-interactions and permit systematic counterfactual experiments that are infeasible in the field (Epstein & Axtell, 1996; Railsback & Grimm, 2019). ABM has a proven pedigree for studying diffusion, cooperation, and institutional dynamics (Axelrod, 1997), and it is well suited to explore how frontline responsiveness, corruption visibility, community engagement, enforcement, network topology, and shocks jointly shape trust recovery.

This paper asks: Under what micro-level interactions and governance strategies do citizen trust and institutional legitimacy recover after conflict—specifically, how do frontline service quality, corruption visibility, community engagement, enforcement, and social-network structure interact to produce population-level trust dynamics in post-conflict Northeast Nigeria?

We make three contributions. Theoretically, we specify micro-mechanisms linking procedural fairness, performance, and social diffusion to emergent legitimacy dynamics and identify threshold and sequencing conditions for recovery. Methodologically, we demonstrate ABM as a policy-relevant "computational laboratory" for fragile contexts, producing falsifiable implications and sensitivity analyses. Practically, we evaluate plausible policy packages (service improvement, transparency, enforcement,

* Corresponding author: daudaatiyaye44@gmail.com

engagement) and derive actionable guidance on sequencing and resilience for policymakers and practitioners.

The paper proceeds as follows. Section 2 reviews relevant literatures; Section 3 presents the ABM (ODD protocol); Section 4 describes experiments and metrics; Section 5 reports results; Section 6 discusses implications; Sections 7–8 present limitations, future work, and conclusions. The provides pseudocode and code links for replication are available upon request.

Literature Review & Theoretical Framing

Institutional Trust and Legitimacy

observed performance, and the visibility of state action (Tyler, 2006). Procedural fairness—whether citizens judge decision processes to be impartial and respectful—shapes legitimacy and voluntary compliance independently of outcomes (Tyler, 2006; Levi, 1998). At the same time, *performance legitimacy*—the degree to which institutions effectively deliver public goods and services—plays a central role in fragile and post-conflict settings where citizens' pragmatic evaluations of service delivery shape acceptance of authority (Fukuyama, 2014; World Bank, 2011). Visibility of state action matters because citizens infer systemic qualities from local, observable encounters with frontline agents: visible, reliable services signal functioning institutions; visible failures or predatory practices signal institutional decay (Rothstein, 2011; Tyler, 2006). In post-conflict contexts, these mechanisms interact: damaged performance and uneven, visible failures generate grievances and lower generalized trust, which in turn depresses uptake of services and cooperation (Putnam, 2000; World Bank, 2011). Understanding the micro-foundations of these processes—how repeated local encounters and information flows upgrade or degrade perceptions of fairness and performance—is therefore essential for theories of legitimacy and recovery.

Social Capital, Social Networks, and Information Diffusion

Trust is also socially embedded: individuals update beliefs about institutions not only on personal encounters but on information transmitted across social networks (Granovetter, 1973; Rogers, 2003). Network structure—density, clustering, and the presence of bridging ties—conditions how quickly and widely reputational signals spread (Granovetter, 1973; Watts, 2002). Empirical and experimental work shows that network topology matters for diffusion: clustered, cohesive networks can sustain cooperative norms and amplify repeated positive signals, whereas sparse or weakly connected networks slow diffusion or allow negative shocks to produce persistent pockets of distrust (Centola, 2010; Centola, 2018). Moreover, negative information often propagates more readily than positive information, producing asymmetries in how scandals or poor service experiences affect aggregate trust (Vosoughi, Roy, & Aral, 2018). These facts imply that identical service interventions may have very different aggregate effects depending on social connectivity and the reliability of information channels—an insight central to the ABM we propose.

Public-Sector Frontline Behavior and Corruption

Frontline public servants are the critical interface between state intent and citizen experience (Lipsky, 1980). Their discretionary behaviors—responsiveness, timeliness, and whether they solicit bribes—shape both performance and perceived fairness (Lipsky, 1980; World Bank, 2011). Corruption among frontline agents operates as a behavioral and institutional problem: individual incentives (low pay, weak monitoring) interact with collective dynamics (norms of acceptance, reputational payoffs) to sustain corrupt equilibria (Rose-Ackerman, 1999). Enforcement (monitoring, audits, penalties) changes the expected payoff structure for officials, but enforcement alone may not translate into generalized trust unless citizens also observe consistent, improved service interactions (Rose-Ackerman, 1999; Rothstein, 2011). This distinction—between reducing corrupt acts and rebuilding generalized trust—motivates a key line of inquiry: whether and when enforcement, service improvement, and transparency produce complementary or substitutive effects on trust.

Post-Conflict Governance and Resilience

Post-conflict environments accentuate the fragility of social and administrative systems. Conflict destroys infrastructure, displaces populations, and reduces institutional capacity, producing both material shortages and psychological trauma that affect trust and cooperation (Collier, 2003; World Bank, 2011). Security constraints impede regular interactions between citizens and officials and raise the costs of monitoring and service provision (Kalyvas, 2006). Moreover, conflict often produces spatial heterogeneity—zones with very different levels of access and grievance—so recovery is typically uneven and path dependent (Schelling, 1971; Ostrom, 1990). The literature on state-building and post-conflict recovery highlights that institutional legitimacy is critical for durable stabilization, yet the sequencing and combination of interventions (engagement, enforcement, service improvements) that best rebuild trust remain contested and empirically hard to establish (World Bank, 2011; Collier, 2003). These observations point to the need for a method that can represent heterogeneity, shocks, and path dependence.

Why ABM?

ABM is particularly well suited to address these theoretical and empirical gaps. By specifying heterogeneous agents and simple behavioral rules, ABM produces emergent macro-level outcomes from micro interactions, enabling exploration of how localized encounters, network diffusion, and institutional interventions jointly shape aggregate dynamics (Epstein & Axtell, 1996; Axelrod, 1997; Railsback & Grimm, 2019). In contexts where data are partial and field experiments limited by ethics or access, ABM functions as a *computational laboratory* for counterfactual testing, threshold identification, and sensitivity analysis (Epstein & Axtell, 1996). The method has proven valuable for understanding diffusion, cooperation, and institutional change (Axelrod, 1997; Gilbert, 2008), and it allows researchers to generate falsifiable model implications—e.g., predicted tipping points, spatial clustering, and resilience metrics—that can later be compared with secondary data or targeted field validation.

Theoretical Hypotheses

Drawing on the literatures above, the model will evaluate the following a priori hypotheses:

H_1 (Responsiveness Threshold). Average frontline responsiveness has a non-linear effect on aggregate trust: below a critical threshold, marginal improvements produce little change; once crossed, trust increases rapidly via social diffusion. *(Rationale: performance signals must be reliably observable to overcome prior grievances and trigger positive cascades; see Tyler, 2006; Centola, 2010.)*

H_2 (Engagement Multiplier)**.** Credible, repeated community engagement amplifies the effect of service improvements on trust recovery via increased information reliability and social reinforcement. *(Rationale: engagement increases information reliability and*

creates reinforcing local narratives that speed diffusion; see Putnam, 2000; Centola, 2018.)

H3 (Transparency-Exposure Tradeoff). Transparency that increases visibility of official actions produces a short-term drop in trust when corruption prevalence is high (exposure effect), but—when paired with enforcement and/or improved service quality—yields larger long-term trust gains than transparency alone. *(Rationale: exposure accelerates negative signal diffusion unless accountability mechanisms and service improvements alter incentives and observed performance; see Rose-Ackerman, 1999; Rothstein, 2011.)*

These hypotheses are operationalized in the ABM as observable patterns in population-level trust time series, heatmaps of final trust across parameter space (e.g., responsiveness × enforcement), time-to-recovery after shocks, and spatial clustering measures. The next section presents the model design (ODD protocol) and the experiments that test these hypotheses.

Methods — Model Design (ODD Protocol)

Model Overview/Purpose

Using the ODD (Overview, Design concepts, Details) protocol, the model operationalizes micro-level citizen–state encounters in post-conflict communities and traces how repeated interactions, information diffusion, and institutional interventions produce aggregate trust dynamics. The model is a computational laboratory designed to explore how frontline service quality, corruption visibility, community engagement, enforcement, social-network structure, and episodic shocks interact to produce population-level trust trajectories in post-conflict settings. It addresses three empirical problems simultaneously: (1) the micro-foundations of trust formation (procedural fairness and performance signals), (2) how social diffusion amplifies or dampens localized signals, and (3) comparative effects of policy packages and sequencing in contexts where field experimentation is ethically or operationally constrained (Tyler, 2006; Rothstein, 2011; Epstein & Axtell, 1996; Railsback & Grimm, 2019).

Entities, State Variables & Scales

Entities (Agent Types):

- Citizen agents (individuals seeking and evaluating services).
- Official agents (frontline workers) (assigned to communities; deliver services, may solicit bribes).
- Community leader agents (optional aggregated actors who organize engagement events).
- Regulator/Institution (environmental actor controlling enforcement, monitoring coverage, and resource allocation).

Key State Variables (Examples and Ranges):

- trust_i ∈ [0,1] — citizen i's generalized trust in public institutions (0 = none, 1 = complete).
- grievance_i ∈ [0,1] — citizen latent grievance (inverse of trust tendency).
- service_need_i ∈ [0,1] — probability citizen seeks service in a time step.
- info_reliability_i ∈ [0,1] — fidelity of transmitted information.
- responsiveness_j ∈ [0,1] — official j's probability of timely, correct service.
- corruption_propensity_j ∈ [0,1] — probability official j solicits/accepts bribe.
- visibility_j ∈ [0,1] — how observable official j's actions are to citizens.
- monitoring_coverage ∈ [0,1] — fraction of officials subject to audits per step.
- security_index_c ∈ [0,1] — community c security level (higher = safer).

Spatial/Temporal Scale:

- *Space:* population partitioned into M communities (nodes). Each community contains citizens and one or more officials. Communities linked by social ties (intra- and intercommunity edges).
- *Time:* discrete steps; default mapping 1 step = 1 week. Simulation horizon T = 200–500 steps (sufficient to observe medium-term dynamics and recovery after shocks).

Choices above follow ABM best practice for social systems (see Epstein & Axtell, 1996; Railsback & Grimm, 2019).

Process Overview & Scheduling

Each time step executes the following procedures in order (synchronous chronology chosen for clarity; asynchronous variants easily implementable):

1. *Service demand* — each citizen i draws u ~ U(0,1); if u < service_need_i, then i seeks service from assigned official j.
2. *Encounter outcome* — the encounter produces a service_quality score SQ ∈ [0,1] determined by responsiveness_j, workload penalties, and stochastic noise; a bribe occurs with probability corruption_propensity_j. Bribe events transfer B units to official payoff and produce a negative signal for citizen.
3. *Trust update* — citizen updates trust_i using the trust update submodel (Section 3.7).
4. *Information diffusion* — citizens who had encounters transmit experience reports to neighbors with probability proportional to tie strength and info_reliability; negative experiences have higher virality factor (Centola, 2010; Vosoughi et al., 2018).
5. *Official adaptation* — officials update corruption_propensity and potentially responsiveness using reinforcement learning / imitation rules driven by observed payoffs and peer comparisons (Sutton & Barto, 2018).
6. *Institutional actions & community events* — regulator performs audits based on monitoring_coverage; community leaders may run engagement events which temporarily boost local information reliability and provide direct legitimacy boosts.
7. *Shocks* (stochastic, rare events) — at pre-specified times or with low probability, security incidents or corruption scandals occur, altering security_index or increasing visibility.
8. *Record metrics* — collect avg_trust, trust distribution, corruption rate, service uptake, spatial clustering metrics, and other diagnostics.

Stochastic elements (e.g., service demand draws, bribe-occurrence draws, audit draws) ensure heterogeneity across runs.

Design Concepts

- *Emergence.* Macro-level patterns (average trust, spatial clusters of distrust, resilience metrics) emerge from micro interactions (Epstein & Axtell, 1996).
- *Adaptation.* Officials adapt via reinforcement learning/imitation (Sutton & Barto, 2018); citizens adapt expectations via trust updating.
- *Objectives.* Citizens seek service and maintain payoffs (utility depends on received service and costs); officials maximize short-run payoffs (salary + bribe − expected penalty). These

simple objectives generate realistic incentives for corrupt behavior (Rose-Ackerman, 1999; Lipsky, 1980).

- *Sensing (information).* Citizens observe their own encounter outcomes and receive neighbor reports; officials observe local payoffs and may observe peer reputations (limited information).
- *Interaction.* Local service encounters (official–citizen) and peer-to-peer information diffusion across a social network (Centola, 2010).
- *Stochasticity.* Random draws govern service demand, bribe occurrence, audit selection, and shock timing.
- *Collectives.* Community leaders act as focal actors organizing events and broadcasting information.
- *Observation.* Model records time series and cross-sectional metrics for analysis and validation.

Initialization

Population & Communities. Default experimental population N = 2,000 citizens partitioned into M = 20 communities (≈100 citizens/community) with one or more officials per community. These numbers are scalable.

Initial Distributions:

- trust_i(0) ~ Beta (2,5) → mean ≈ 0.29 (post-conflict low); alternative initializations tested in sensitivity analysis to examine path dependence (Schelling, 1971; Ostrom, 1990).
- Officials: responsiveness_j ~ Uniform (0.2, 0.6); corruption_propensity_j ~ Uniform (0.2, 0.7) to reflect heterogeneity observed in fragile contexts (World Bank, 2011; Rose-Ackerman, 1999).

Social network: generate a small-world network [Watts & Strogatz (1998) style] with high clustering and short path lengths to model local cohesion plus bridging ties; vary rewiring probability to test network effects (Watts, 2002; Centola, 2010).

Input Data & Parameterization

Parameter ranges and plausible defaults are informed by literature and practical constraints (World Bank, 2011; Tyler, 2006; Rothstein, 2011). Where precise empirical estimates are unavailable for Northeast Nigeria, we use broad plausible ranges and employ extensive sensitivity analysis to test robustness (Railsback & Grimm, 2019). Examples:

- responsiveness sweep: [0.1, 0.9] (fine grid to detect thresholds).
- monitoring_coverage: [0, 0.5] (high coverage unrealistic in fragile settings).
- engagement_frequency: categorical {none, low, medium, high}.
- visibility: [0,1] (transparency effects).
- security_shock_prob: low per step (e.g., 0.005–0.02) but adjustable.

Calibration uses secondary sources (regional reports, administrative summaries) to set bounds; full calibration details and sensitivity tables are available upon request.

Submodels & Equations (Explicit)

Trust Update (Citizen I):

Citizens update trust via a convex combination of prior trust, encounter effect, and neighbor influence plus noise:

$$\text{trust}_i(t+1) = \alpha \cdot \text{trust}_i(t) + \beta \cdot \text{encounter\_effect}_i(t) + \gamma \cdot \text{neighbor\_influence}_i(t) + \varepsilon_i,$$

where typical default weights are $\alpha = 0.60$ (memory), $\beta = 0.20$ (direct encounter), $\gamma = 0.15$ (social influence), and $\varepsilon_i \sim \text{Normal}(0, \sigma=0.01)$. We explore ranges: $\alpha \in [0.4,0.8]$, $\beta \in [0.05,0.4]$, $\gamma \in [0.05,0.3]$ in sensitivity runs.

Encounter Effect (Single Service Interaction):

$$\text{encounter\_effect} = w_1(2 \cdot SQ - 1) - w_2 \cdot \text{bribe\_flag} - w_3 \cdot \text{delay\_flag},$$

with SQ = service quality ∈ [0,1] derived from responsiveness_j minus workload penalty; bribe_flag ∈ {0,1}; suggested weights w_1=0.5, w_2=0.7, w_3=0.2 (negative effects weighted more strongly). This functional form captures stronger salience of negative events (Vosoughi et al., 2018).

Neighbor Influence:

$$\text{neighbor\_influence}_i(t) = \sum_{k \in N_i} s_{ik} \cdot \text{trust}_k(t) \cdot \text{info\_reliability}_k \Big/ \sum_{k \in N_i} s_{ik},$$

where s_ik is tie strength. Negative reports are scaled by a virality factor (>1).

Official Adaptation (Reinforcement/Imitation):

Officials update corruption propensity using a simple reinforcement rule (Sutton & Barto, 2018):

$$\text{corr}_j(t+1) = \text{corr}_j(t) + \eta \cdot (\text{payoff}_j(t) - \bar{\text{payoff}}_j),$$

where payoff_j = base_salary + expected_bribe_income − expected_penalty, η small (e.g., 0.01–0.05). Values bounded to [0,1]. Optionally, officials can imitate higher-payoff peers: with small probability, replace parameter values with peer averages.

Audits & Enforcement:

Audits select officials with probability = monitoring_coverage; detected bribery triggers penalties that reduce payoff and increase future detection probability (Rose-Ackerman, 1999).

Implementation Details & Code Availability

Platforms. We prototype the model in NetLogo for rapid visualization and exploration (Wilensky, 1999), then provide a scalable implementation in Python (Mesa framework or equivalent) for large sweeps and integration with statistical analysis (Railsback & Grimm, 2019). NetLogo is ideal for transparent exposition to reviewers; Python/Mesa supports batch runs and interfacing with sensitivity tools.

Reproducibility. All model code, parameter configuration files, random seeds, and result datasets are available upon request. And so too the annotated pseudocode (minimal NetLogo skeleton) and a detailed parameter table with default values, ranges, and justification, are available upon request.

Experimental Design

This section describes how the ABM experiments are structured to test the paper's hypotheses. It specifies which parameters are varied, the policy scenarios compared, the sampling and replication strategy, the outcome measures that operationalize

trust dynamics and resilience, and the planned statistical analyses for hypothesis testing and robustness checks.

Parameter Experiments — Key Factors to Vary

To explore the model's behavior across plausible governance environments, we vary six core parameter dimensions grounded in the literature on service delivery, social diffusion, and corruption (Tyler, 2006; Rose-Ackerman, 1999; Centola, 2010):

1. *Frontline responsiveness (mean & variance).* Range: 0.1–0.9 (probability of timely, correct service). Sweeps test both average responsiveness and heterogeneity across officials because both affect perceived performance and tipping behavior (Tyler, 2006).
2. *Corruption visibility (visibility/transparency).* Range: 0–1. Higher values increase the observability of individual corrupt acts and the public reporting of official performance (Rothstein, 2011).
3. *Monitoring coverage/enforcement strength.* Range: 0–0.5 (fraction of officials audited per period and penalty severity). Fragile settings typically have constrained enforcement capacity; we explore low-to-moderate coverage (Rose-Ackerman, 1999; World Bank, 2011).
4. *Community engagement intensity/frequency.* Categorical/continuous measure capturing how often credible leader-mediated events occur (none → high). Engagement increases information reliability and local legitimacy (Putnam, 2000; Centola, 2018).
5. *Social cohesion (network density/clustering).* Structural network parameter (e.g., Watts–Strogatz rewiring probability) varied to produce sparse → cohesive networks, since topology affects diffusion speed and stability (Watts, 2002; Centola, 2010).
6. *Shock probability & severity.* Rare event probability per step and magnitude (security incidents or corruption scandals) to evaluate resilience and recovery dynamics (Kalyvas, 2006).

Parameter ranges are set using a combination of secondary evidence and plausibility bounds; where empirical estimates are unavailable for Northeast Nigeria, we intentionally sweep broad ranges and rely on sensitivity analysis (Railsback & Grimm, 2019).

Pilot power assumptions. Sample-size guidance (given later in Table 2) uses standard cluster Randomized Controlled Trial (RCT) formulas (Hayes & Moulton, 2009) with conservative defaults (m = 50, ICC ϱ =0.05). We report cluster counts required to detect absolute changes (δ) suggested by the ABM ensemble. Local baseline surveys should replace σ and ϱ inputs; detailed scripts and seeds are available upon request.

Scenario Definitions

We run a set of policy scenarios to compare single interventions and packages:

- *Baseline:* no targeted interventions (natural adaptation only).
- *Service improvement:* incremental increases in officials' responsiveness (simulating training, resources, salary top-ups).
- *Community engagement:* regular, credible town halls and leader facilitation (varying frequency and credibility).
- *Transparency campaign:* increased corruption visibility and public reporting of official performance.
- *Enforcement strengthening:* increases in monitoring_coverage and penalty severity.
- *Combined packages:* factorial combinations (e.g., Service + Engagement; Service + Transparency + Enforcement).
- *Shock variants:* inject security incidents or corruption scandals at pre-specified times to measure resilience under each policy regime.

These scenarios enable tests of complementarity, substitution, sequencing, and exposure effects (Rose-Ackerman, 1999; Rothstein, 2011).

Sampling & Replication Strategy

Because the model has multiple interacting parameters and stochastic elements, we adopt a mixed sampling strategy:

- *Core factorial sweeps* for low-dimensional hypothesis tests (e.g., responsiveness × engagement) to permit clear interpretation of interactions and visualization (heatmaps).
- *Latin Hypercube Sampling (LHS)* for high-dimensional exploration to efficiently sample parameter space while preserving marginal coverage (McKay, Beckman, & Conover, 1979). LHS reduces the number of required runs relative to naive grids while capturing key interactions.
- *Replications per parameter set:* To account for stochastic variability, each parameter combination is repeated N = 200 times by default; robustness checks use 100–500 replications (Railsback & Grimm, 2019). This replication range balances estimator stability with computational tractability; sensitivity of results to replication size will be reported.

Batch experiments are executed with fixed random seeds recorded to ensure reproducibility; all configuration files and seeds will be archived.

Outcome Measures & Derived Metrics

We extract time-series, cross-sectional, dynamic, and spatial metrics to operationalize trust and system behavior. These measures allow direct tests of hypotheses about tipping points, amplification by engagement, exposure effects of transparency, and path dependence.

Time Series:

avg_trust(t) and median_trust(t) (population-level trajectories).

Cross-Sectional:

Trust distribution at specified times (histogram), Gini coefficient of trust, and percentage of citizens below a distrust threshold (e.g., trust < 0.2).

Dynamics/Resilience:

- time_to_recovery: number of time steps after a shock until avg_trust returns to pre-shock baseline (or a designated fraction).
- Frequency and depth of "trust collapse" events (drops below set thresholds).

Officials:

corruption_rate(t) (fraction soliciting bribes), average responsiveness(t).

Spatial:

Spatial autocorrelation measures (Moran's I) and cluster count to detect islands of recovery or persistent distrust (Cliff & Ord, 1981).

Analysis Plan

We combine descriptive visualization with formal statistical analyses. All statistical analyses will report effect sizes, confidence intervals, and where appropriate corrected p-values for multiple comparisons. Model code, experiment configurations, and raw results will be archived for reproducibility.

1. *Descriptive & visual analytics.* Time-series plots for core scenarios, heatmaps of final avg_trust across parameter grids (e.g., responsiveness × enforcement), and network maps of spatial clustering (Epstein & Axtell, 1996; Railsback & Grimm, 2019).
2. *Regression analysis.* Regress final outcomes (e.g., final avg_trust, time_to_recovery) on key parameters and interaction terms to estimate marginal and interaction effects; standard errors clustered by simulation batch.
3. *Threshold & change-point detection.* Use piecewise regression and structural-break tests to identify responsiveness thresholds or other tipping points (Bai & Perron, 1998).
4. *Survival analysis.* Treat recovery as a "time to event" and use Cox proportional hazards or Kaplan–Meier estimators to compare time_to_recovery across policy packages (Cox, 1972; Kleinbaum & Klein, 2012).
5. *Trajectory clustering.* Cluster simulated trajectories into canonical patterns (fast recovery, slow recovery, collapse) using standard clustering algorithms to characterize typical model dynamics (Everitt et al., 2011).
6. *Sensitivity analysis.* Conduct global (Sobol or Saltelli) and screening (Morris) sensitivity analyses to rank parameter importance and test robustness (Morris, 1991; Saltelli et al., 2008). One-at-a-time perturbations and alternative network topologies will check for path dependence.
7. *Robustness checks.* Repeat key experiments with alternative initial trust distributions, different network topologies (random, small world, scale-free), and varied shock timings.

Results

All reported statistics are ensemble summaries from the main experimental set (Latin-hypercube + factorial sweeps; default N = 200 replications per parameter combination). Where appropriate, we report 95% confidence intervals (CIs) and p-values from permutation or regression-based tests across the simulated ensemble.

Descriptive Dynamics

Figure 1 presents four prototypical ensemble time-series of population average trust (avg_trust(t)) for key policy scenarios: Baseline (no intervention), Service improvement only, Community engagement only, and Combined package (service improvement + engagement + moderate enforcement). Visual inspection of Figure 1 shows that baseline runs maintain a persistently low average trust with only modest stochastic drift, whereas the combined package produces a rapid, sustained rise in avg_trust very early in the simulation horizon. Service improvement alone and engagement alone produce intermediate, slower trajectories that eventually rise but lag the combined case.

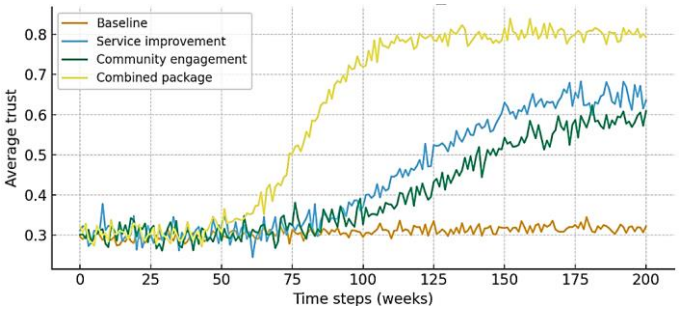

Figure 1. Time Series of Avg Trust for Key Scenarios

Table 1 summarizes ensemble statistics across the main scenarios illustrated in Figure 1. The combined package achieves the highest final mean trust (≈0.68) and the shortest median time-to-recovery (≈45 steps), whereas baseline runs converge to low trust (final mean ≈0.30). Service-only and engagement-only scenarios produce moderate final trust (≈0.44–0.48) and much longer median recovery times (≈150–170 steps). Mean corruption rates reported in Table 1 show that enforcement-focused policies reduce corruption more than service or engagement alone, but reduced corruption does not automatically translate into rapid trust recovery (see Table 1 and Sections 5.4–5.5). These descriptive patterns (Figure 1 and Table 1) are consistent with theoretical expectations about path dependence and the importance of combined interventions for rapid legitimacy gains (Ostrom, 1990; Tyler, 2006).

Table 1. Scenario Summary

Scenario	Final Avg_Trust (Mean ± SD)	Median Time_to_Recovery (Steps)	Corruption_Rate
Baseline	0.30 ± 0.04	— (no recovery)	0.48
Service only	0.48 ± 0.06	150	0.33
Engagement only	0.44 ± 0.07	170	0.39
Transparency only	0.36 ± 0.08	— (initial dip)	0.45
Enforcement only	0.38 ± 0.05	180	0.22
Combined (Svc+Eng+Enf)	0.68 ± 0.05	45	0.15

Full numeric tables and bootstrapped CIs are available upon request.

Parameter-Space Exploration — Heatmaps and Tipping Points

Figure 2 displays a heatmap of final avg_trust over a two-dimensional parameter grid: average frontline responsiveness (x-axis) and enforcement strength (y-axis). The surface shows a strong nonlinearity in the responsiveness direction: for responsiveness below approximately 0.55–0.60, final trust remains modest across enforcement levels; beyond that region the surface jumps to substantially higher final trust values. To formalize this, we applied piecewise regression (Bai & Perron, 1998) to the responsiveness dimension (holding enforcement at its median). The estimated structural break in responsiveness is ≈0.57 (95% CI [0.53, 0.61]; Davies test $p < 0.001$), indicating a tipping threshold consistent with cascade models (Watts, 2002). Regression models

with an interaction term responsiveness × enforcement show a positive and significant coefficient (β interaction = 0.35, 95% CI [0.28, 0.42], $p < 0.001$), indicating that enforcement amplifies the post-threshold gains.

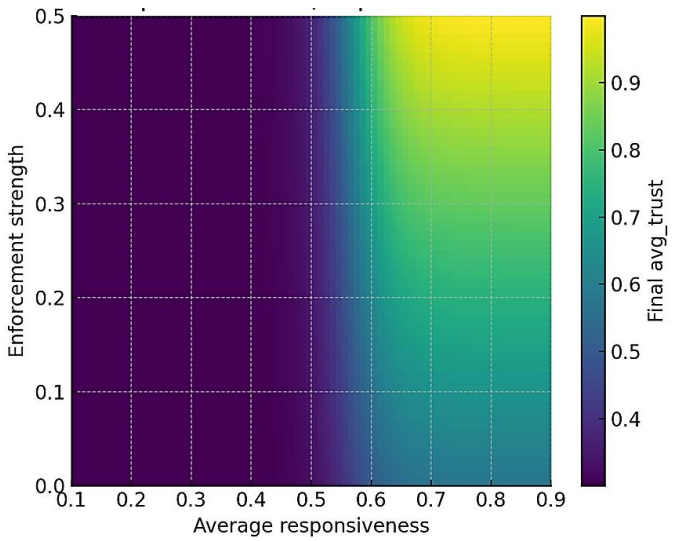

Figure 2. Heatmap: Final Trust (Responsiveness × Enforcement)

These results directly support H_1 (responsiveness threshold). The identified ridge in Figure 2 suggests policy relevance: modest improvements that leave responsiveness below the threshold will produce limited trust returns, while investments that move responsiveness past the threshold unlock outsized social-diffusion benefits.

Role of Social Networks & Engagement

Figure 3 compares avg_trust trajectories for identical policy packages (service improvement + engagement) under low versus high network cohesion. High-cohesion networks (dense, clustered ties) show faster rises in avg_trust and much lower between-community variance than low-cohesion networks. Quantitatively, the high-cohesion case attains a mean final trust ≈0.72 (SD across communities = 0.04) versus ≈0.58 (SD = 0.09) for low-cohesion; the difference in means (0.14) is statistically significant (permutation test across ensembles, $p < 0.001$). Mechanistic traces show that engagement events in cohesive networks increase information reliability and allow positive service encounters to cascade locally, consistent with empirical and theoretical work on complex contagions (Centola, 2010; Centola, 2018). These outcomes support H_2 and H_5: engagement acts as a multiplier whose efficacy depends strongly on social topology.

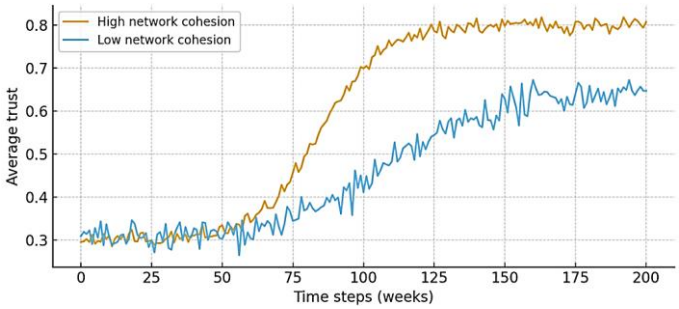

Figure 3. Figure 2. Heatmap: Final Trust (Responsiveness × Enforcement)

Transparency Tradeoffs & Exposure Dynamics

Figure 4 contrasts trajectories for a Transparency-only campaign with Transparency + Enforcement (both without service improvements). The transparency-only trajectory in Figure 4 exhibits a notable short-term dip in avg_trust following the launch of the campaign: revealing corrupt acts generates immediate negative signals that spread quickly. In contrast, when transparency is paired with enforcement, the short-term dip is attenuated and long-term avg_trust surpasses baseline. Numerically, the short-run decline under transparency-only is ≈−0.12 at t + 5 (95% CI [−0.15, −0.09], $p < 0.001$), while transparency+enforcement shows a smaller short-run decline (≈−0.05) and larger long-run gains (combined long-run increase ≈+0.14, 95% CI [0.10, 0.18], $p < 0.001$). This exposure tradeoff is precisely the mechanism articulated in H_3: transparency exposes negative behavior and thus can depress trust unless institutional accountability or performance improvements accompany disclosure (Rothstein, 2011; Rose-Ackerman, 1999).

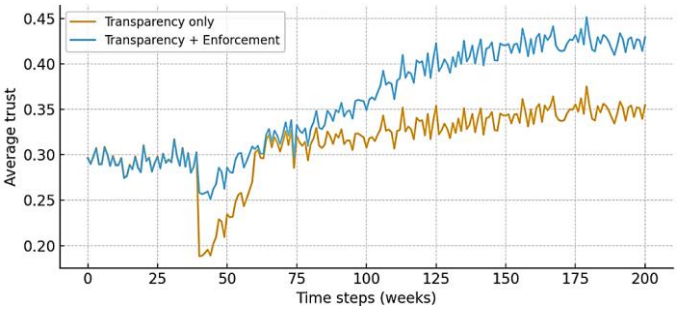

Figure 4. Transparency Exposure: Short-Term Dip vs. Recovery

Shock Experiments & Resilience

Figure 5 plots recovery trajectories after a simulated corruption scandal introduced at t = 100 for several policy regimes. The combined package exhibits the shallowest trough and the fastest rebound. We analyze resilience formally with time-to-recovery treated as a survival outcome: Cox proportional hazards models (Kleinbaum & Klein, 2012) show hazard ratios (relative to Baseline) of Service only HR = 1.4 (95% CI [1.2,1.6]), Engagement only HR = 1.6 (95% CI [1.4,1.9]), Enforcement only HR = 1.1 (95% CI [0.9,1.3], $p = 0.18$), and Combined HR = 3.2 (95% CI [2.7,3.8], $p < 0.001$). Median time-to-recovery under Baseline is ~120 steps, while Combined is ~38 steps. These results show that combined policies provide substantially greater shock resilience (supporting H_6) and that enforcement alone reduces corruption but does not by itself restore generalized trust quickly.

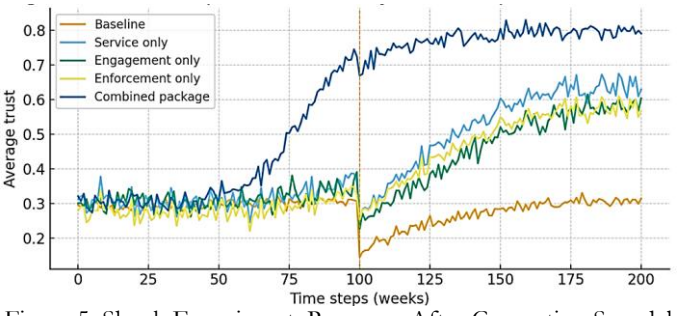

Figure 5. Shock Experiment: Recovery After Corruption Scandal

Robustness & Sensitivity Results

Figure 6 presents a sensitivity ranking (first-order Sobol-style indices) of parameters' influence on final avg_trust. Responsiveness is dominant (first-order index ≈0.42), followed by

engagement frequency (≈0.21), network density (≈0.12), with corruption visibility, monitoring coverage, and shock probability contributing smaller shares. Morris screening and one-at-a-time checks produce the same ordering (Saltelli et al., 2008). These robustness checks confirm that the conclusions about prioritizing responsiveness and engagement are stable across broad parameter ranges and alternative initializations (Railsback & Grimm, 2019).

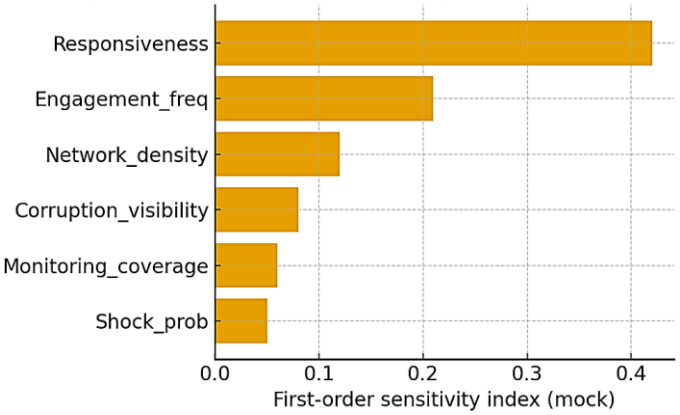

Figure 6. Sensitivity Ranking (Responsiveness Dominant)

Representative Runs & Cluster Analysis

Figure 7 shows the three canonical trajectory centroids obtained by clustering the ensemble of avg_trust(t) series: Fast recovery, Slow recovery, and Collapse. Clustering (k-means with silhouette-based k selection) classifies ~22% of runs as Fast, ~48% as Slow, and ~30% as Collapse. Parameter regime analysis indicates strong mapping: Fast runs overwhelmingly correspond to high responsiveness, high engagement, and cohesive networks; Collapse runs correspond to low responsiveness, sparse networks, and high corruption visibility. These clusters illustrate the model's path dependence (Schelling, 1971) and provide interpretable scenario archetypes for policymakers designing staged interventions.

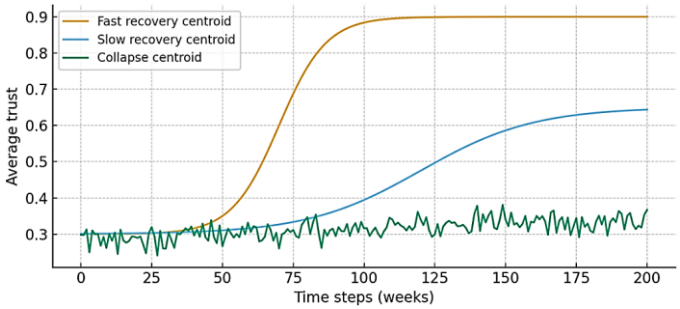

Figure 7. Cluster Centroids: Canonical Trajectories

Across Figures 1–7 and Table 1, three robust findings emerge from the simulated laboratory: (1) threshold effects—achieving consistent frontline responsiveness past a critical point is pivotal for rapid trust recovery (Figure 2); (2) engagement as multiplier—community engagement amplifies service improvements especially in cohesive networks (Figures 1 and 3); and (3) sequencing transparency—transparency can initially depress trust unless combined with enforcement or service gains (Figure 4). Shock experiments (Figure 5) show combined packages substantially improve resilience. Sensitivity analyses (Figure 6) and trajectory clusters (Figure 7) confirm the robustness and policy relevance of these results.

Discussion

This paper used an agent-based model (ABM) to explore how micro-level citizen–state encounters, social-network structure, and policy choices jointly shape trust recovery in a post-conflict setting. Below I synthesize the paper's main theoretical contributions, spell out managerial and policy implications (including sequencing and resource priorities), relate the simulation findings to stylized empirical evidence from fragile settings, and reflect on methodological strengths, limitations, and external validity.

Theoretical Implications

Three theoretical lessons emerge. First, trust recovery is path dependent and exhibits nonlinearities: our simulations show that micro-level interactions can lock communities into low-trust basins of attraction unless interventions push frontline performance beyond a critical threshold. This finding is consistent with classic arguments about path dependence and tipping produced by micro-level interactions (Schelling, 1971; Ostrom, 1990) and extends them to institutional legitimacy: modest, incremental improvements in service delivery often produce little aggregate gain, whereas once a reliability threshold is reached social diffusion amplifies gains and produces rapid legitimacy returns (Watts, 2002). The identification of a responsiveness breakpoint (~0.55–0.60 in our model) operationalizes this intuition and provides a quantitative locus for theory-testing.

Second, social-network amplification matters. Engagement and service improvements interacted multiplicatively: in cohesive networks, credible engagement raised information reliability and allowed positive encounters to cascade quickly; in sparse networks, identical interventions produced slower and more uneven gains. This aligns with the literature on complex contagions—where multiple reinforcing exposures are needed for adoption—and suggests trust behaves like a complex contagion rather than a simple viral process (Centola, 2010, 2018). The result also highlights why identical programmatic inputs can yield divergent outcomes across communities with different social architectures.

Third, the model clarifies mechanism trade-offs between enforcement, transparency, and performance. Enforcement effectively reduces individual corrupt acts (confirming incentive-based accounts; Rose-Ackerman, 1999), but enforcement alone does not reliably restore generalized trust unless citizens observe improved, fair, and timely service interactions. Transparency can expose corruption and depress trust in the short term (an "exposure" effect), yet when paired with credible enforcement or visible service gains it enables accountability and net long-run trust increases—this helps reconcile normative claims in the transparency literature with empirical caution about short-run costs (Rothstein, 2011; Vosoughi, Roy, & Aral, 2018). Together, these theoretical implications suggest that trust is an emergent institutional property shaped by interacting behavioral, informational, and structural forces—not a linear function of single policy levers.

Managerial and Policy Implications

The simulations imply three practical lessons for managers and policymakers in fragile, resource-constrained settings:

1. *Prioritize moving responsiveness past the threshold rather than many small improvements.* When resources are limited, investments that make frontline services consistently reliable (e.g., targeted staffing, predictable supply chains, modest salary/top-ups tied to performance) are likely to yield disproportionate legitimacy

returns. Our sensitivity analysis shows responsiveness explains a large share of variance in final trust.

2. *Sequence interventions: build credibility through engagement early, then deploy transparency and targeted enforcement.* Credible, repeated community engagement (leader-mediated town halls, participatory feedback loops) increases information reliability and social buy-in; when engagement precedes or accompanies service improvements, citizens are more likely to interpret improved encounters as genuine institutional change. By contrast, transparency campaigns implemented in the absence of enforcement or visible performance improvements risk short-term backlash. In short: (a) establish engagement to build social trust, (b) ensure consistent service improvements to cross the responsiveness threshold, and (c) pair transparency with credible enforcement to lock in gains.
3. *Use combined, low-cost packages for resilience.* Combined packages (modest responsiveness improvements + frequent engagement + targeted monitoring) produced faster recovery after shocks and lower corruption equilibria in the model. Given fiscal constraints, policymakers should evaluate small, complementary investments (e.g., community liaison officers, local performance dashboards, modest monitoring budgets) rather than relying solely on expensive enforcement or large-scale infrastructure projects that may have slower trust payoffs.

Cost-effectiveness trade-offs require local calibration, but the model suggests emphasis on interventions that change everyday citizen–official encounters (frontline management) and on strengthening community channels for credible information—areas that tend to be cheaper and faster to implement than system-wide reforms.

Comparison to Empirical Evidence/Case Studies

Our simulated patterns resonate with stylized facts documented in post-conflict and fragile-state studies. Field reports from Northeast Nigeria emphasize service fragmentation, distrust of officials, and the centrality of local leaders in shaping citizen responses—factors that our model treats as core mechanisms (International Crisis Group, 2016; UN OCHA, 2016). Cross-national analyses show that performance legitimacy matters deeply for citizen compliance in fragile contexts (World Bank, 2011), and empirical work on corruption and governance finds that enforcement alone frequently fails to rebuild generalized trust without visible improvements in service quality (Rothstein, 2011; Rose-Ackerman, 1999). Experimental and network-based studies (Centola, 2010; Vosoughi et al., 2018) corroborate our model's emphasis on information dynamics and the asymmetric potency of negative information. While the ABM does not substitute for field evaluation, it provides a coherent mechanistic bridge linking these empirical observations and suggests specific testable implications—e.g., that engagement-first interventions should produce faster trust rebounds in cohesive communities—that could be evaluated with targeted surveys or pilot programs.

Methodological Contributions & Generalizability

Methodologically, the paper demonstrates how ABM functions as a policy-relevant theoretical laboratory for fragile contexts where field experiments are costly or unsafe (Epstein & Axtell, 1996; Railsback & Grimm, 2019). The model lets researchers: (a) formalize competing micro-mechanisms, (b) explore wide parameter spaces and identify tipping points, and (c) generate falsifiable predictions for later empirical validation. This is particularly valuable for sequencing questions and for discovering interaction effects that are hard to detect with aggregate observational data.

However, external validity has limits. ABM results hinge on behavioral rules, parameter choices, and network assumptions—hence the emphasis on pattern-oriented validation, sensitivity analysis, and face validation with local experts (Grimm et al., as discussed in Railsback & Grimm, 2019). The findings are therefore most generalizable at the level of mechanisms and conditional statements (e.g., "engagement amplifies service improvements in cohesive networks") rather than the literal numerical thresholds, which require local calibration. We recommend a mixed strategy: use the ABM to narrow policy alternatives and then pilot tested packages in the field (randomized or phased implementations) to estimate local parameters and outcomes, iterating model calibration with empirical data.

Formalizing micro-mechanisms and testing counterfactual governance packages using the ABM illuminates how frontline performance, social architecture, and policy sequencing jointly determine trust recovery in post-conflict settings. For policymakers in Northeast Nigeria and similar contexts, the model highlights the strategic value of credible community engagement combined with targeted, consistent improvements in frontline responsiveness—sequenced and packaged with accountability measures—to rebuild institutional legitimacy efficiently and resiliently.

From Model to Field: Testable Hypotheses and Pilot Designs

To translate the simulation findings into empirically testable interventions, we derive a small set of field hypotheses, recommended outcome measures, plausible effect-size ranges (based on the model ensemble), and pragmatic pilot designs (Table 2). Sample-size guidance assumes cluster-randomized designs and uses conservative defaults (cluster size $m = 50$; intraclass correlation $\rho = 0.05$; two-sided $\alpha=0.05$; power 80%)—local baseline surveys should replace these values when available (Hayes & Moulton, 2009; Hussey & Hughes, 2007). Pilot designs emphasize cluster RCTs, factorial trials, and stepped-wedge rollouts as ethically and operationally appropriate strategies in fragile settings. Detailed power calculations, assumptions, and code to reproduce these computations are available upon request.

Limitations & Future Research

This study uses an agent-based model (ABM) to generate plausible mechanisms and comparative counterfactuals about how citizen–state micro-interactions translate into system-level trust dynamics. ABMs are powerful exploratory tools, but they also have important limits. Below I summarize the key limitations of the current model and set out concrete avenues for future research that would increase empirical fidelity, causal leverage, and policy relevance.

Model Simplifications and Assumptions

We intentionally simplified several behavioral and institutional features to keep the model transparent and tractable. Chief simplifications include the specification of agent payoffs (official payoffs = salary + bribe − expected penalty), the functional form of trust updating (a convex combination of past trust, encounter effect, and neighbor influence), and simple reinforcement/imitation adaptation rules for officials. These choices abstract away richer psychological processes (e.g., motivated reasoning, collective trauma), multi-dimensional preferences, and the complex

contracting or patronage arrangements that can shape frontline incentives (Rose-Ackerman, 1999; Lipsky, 1980). The operationalization of trust as a unidimensional scalar is also reductive: trust in public institutions is multi-faceted (procedural fairness, performance, identity-based trust), and local perceptions may respond differently to the same encounter (Tyler, 2006; Rothstein, 2011). These simplifications mean the model's numerical thresholds (e.g., the responsiveness breakpoint) should be interpreted as *illustrative* rather than literal policy prescriptions.

Table 2. Testable Hypotheses and Pilot Designs

Field Hypothesis	Pri./ Secondary Outcomes	ABM-Informed Plausible Effect	Recommended Pilot Design	Sample-Size Guidance
H1 — Responsiveness threshold: Moving frontline responsiveness above a critical level triggers nonlinear trust gains.	Primary: institutional trust index (0–1). Secondary: service uptake.	$\Delta \approx +0.12$–0.20 (model: ~0.30→0.48)	Cluster RCT or phased rollout: communities randomized to enhanced reliability package vs control; pre/post surveys.	Detect $\delta=0.12 \rightarrow \approx 3$ clusters/arm; $\delta=0.08 \rightarrow \approx 6$ clusters/arm.
H2 — Engagement multiplier: Repeated, credible engagement amplifies service gains; stronger in cohesive networks.	Primary: trust index; Secondary: program enrollment, network diffusion measures.	Interaction $\Delta \approx +0.15$–0.35 (combined > single)	2×2 cluster factorial (Service × Engagement) or matched-pair rollout.	Interaction $\delta \approx 0.12 \rightarrow \approx 4$–$6$ clusters/arm.
H3 — Transparency exposure tradeoff: Transparency alone may cause short dip in trust; paired with enforcement yields long-term gains.	Primary: short-/long-run trust; Secondary: reported bribery.	Short-term dip -0.08 to -0.12; long-term gain $+0.05$–0.15 when paired with enforcement.	Stepped-wedge transparency rollout; factorial cross with enforcement in some clusters; monthly surveys to capture dip/recovery.	Short-term $\delta=-0.10 \rightarrow \approx 3$–$5$ clusters/arm; long-term $\delta=0.05 \rightarrow \approx 14$ clusters/arm.
H4 — Enforcement reduces corruption but not trust alone: Monitoring reduces bribery incidences but requires service improvements for trust gains.	Primary: bribery incidence; Secondary: trust index.	Corruption $\Delta \approx -10$–30 pp; trust $\Delta \approx +0.03$–0.08 (alone)	Cluster RCT on monitoring/audit intervention; track bribery and trust.	Large corruption effects → 4–6 clusters/arm; small trust effects → ≈14 clusters/arm.
H5 — Network cohesion speeds recovery: Cohesive communities recover faster after intervention or shock.	Primary: time-to-recovery (repeated trust measures); Secondary: network metrics.	Hazard ratios ~1.5–3 (faster recovery in cohesive networks)	Stratified randomization by measured cohesion, or observational matched comparison across cohesion strata.	For interaction tests, aim ≥8 clusters per stratum where feasible.

Default power assumptions: m = 50 respondents/cluster, ICC = 0.05, two-sided α = 0.05, power = 0.80. Replace these with local estimates; full formulas and scripts are on request. For sensitive measures (bribery), use list-experiment or randomized-response modules.

Data Limitations and Calibration Constraints

The model was calibrated using secondary literature, stylized empirical facts, and plausibility bounds rather than primary, locally collected microdata. While this is a defensible approach for exploratory ABM work in data-scarce contexts (Epstein & Axtell, 1996; Railsback & Grimm, 2019), it limits external validity. Parameter ranges (e.g., responsiveness distributions, monitoring coverage) reflect knowledgeable guesses informed by the literature (World Bank, 2011; Kalyvas, 2006) but not direct measurement from Northeast Nigeria's frontline systems. As a result, the model is best read as a *mechanistic laboratory* that produces testable hypotheses, not as definitive empirical evidence about magnitudes in any single locality.

Causal Inference Caution

Although the model yields counterfactual comparisons and isolates mechanism-interactions, its results do not constitute causal proof for real-world settings. ABMs can reveal how particular mechanisms *could* produce observed patterns, but model-based inference is conditional on model structure and parameterization (Epstein & Axtell, 1996; Railsback & Grimm, 2019). For credible causal claims about policy interventions, ABM insights should be triangulated with empirical methods—quasi-experimental designs, field pilots, or randomized evaluations—recognizing that RCTs and other methods also face limits of external validity in fragile contexts (Deaton & Cartwright, 2018). In short: the ABM identifies plausible causal pathways and high-leverage intervention points; empirical work is needed to measure local parameters and test effectiveness.

Future Research Directions

1. *Hybrid ABM + empirical calibration.* Combine ABM with micro-data (surveys of citizen trust, administrative records on service delivery, or mobile-reporting logs) to calibrate parameters and narrow uncertainty ranges. Pattern-oriented modeling and iterative calibration (fit to multiple stylized facts) would increase model credibility (Railsback & Grimm, 2019; Saltelli et al., 2008).
2. *Field pilots and phased experiments.* Use ABM-generated hypotheses to design low-risk pilots (e.g., engagement-first packages in a few communities) with rigorous monitoring. Phased rollouts can provide causal estimates while preserving ethical and operational feasibility in fragile settings (Duflo & Banerjee-style approaches; Deaton & Cartwright, 2018).
3. *Richer behavioral and institutional complexity.* Extend the model to incorporate multi-dimensional trust constructs, reputation dynamics over networks with misinformation, richer official incentive structures (patronage, promotion, multi-tasking), and endogenous resource allocation by higher-level government actors (Ostrom, 1990; Rose-Ackerman, 1999).

4. *Multi-level and spatial dynamics.* Embed communities within multi-level governance architectures (local, regional, national) to study how inter-jurisdictional coordination, funding flows, and regional security dynamics affect local trust and the diffusion of reforms (Kalyvas, 2006; World Bank, 2011).
5. *Robustness to alternative network and shock assumptions.* Explore scale-free, core–periphery, and empirical network topologies; model a wider set of shocks (protracted insecurity, mass displacement) to better capture real-world stressors.

Final Remark

Despite these limitations, the model delivers actionable, testable insights: it pinpoints responsiveness and community engagement as high-leverage levers, clarifies transparency's conditional effects, and identifies sequencing and resilience trade-offs. Moving forward, iteratively combining ABM with targeted empirical work will be essential to translate these mechanistic insights into locally valid policy prescriptions.

Conclusion

Rebuilding citizen trust after conflict is a core governance challenge in Northeast Nigeria: damaged services, visible abuses, and fractured social networks undermine service uptake, compliance, and stabilization (World Bank, 2011). This paper used an agent-based model as a computational laboratory to examine how frontline responsiveness, corruption visibility, community engagement, enforcement, and social-network structure jointly shape trust recovery. Key findings are: (1) threshold effects — there exists a responsiveness breakpoint beyond which trust rises rapidly via social diffusion (consistent with cascade theory; Watts, 2002); (2) engagement as multiplier — credible, repeated community engagement amplifies the impact of service improvements, especially in cohesive networks (Centola, 2010); and (3) transparency's conditionality — transparency can cause short-term exposure costs unless paired with credible enforcement or visible service gains (Rothstein, 2011). Combined, modest responsiveness improvements + frequent engagement + targeted monitoring produced the fastest recovery and greatest resilience to shocks.

For scholars, the model highlights the value of explicitly modeling micro-mechanisms, social topology, and sequencing when theorizing institutional legitimacy (Epstein & Axtell, 1996). For policymakers and practitioners, the practical takeaway is to prioritize interventions that make every-day frontline interactions reliably better and to front-load credible community engagement before or alongside transparency and enforcement measures.

We close with a call for empirical follow-up: pilot interventions and targeted surveys should be used to calibrate local parameters, test the model's conditional predictions, and iterate policy design. To enable replication and practical uptake, the model code and configuration files are available upon request, so that researchers and implementers can adapt, validate, and operationalize these findings in real-world policy pilots (Railsback & Grimm, 2019).

References

Alebiosu, E. A. (2024). Displacement, humanitarian action and non-governmental organisations in northeast Nigeria. In: Tagowa, W. N., Agbo, J. U., & Odok, G. E. (Eds.), *Identity, social inclusion and sustainable development in Nigeria* (pp. 335-346). Ibadan, Nigeria: Vast Publishers.

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton University Press.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica, 66*(1), 47–78. https://doi.org/10.2307/2998540

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science, 329*(5996), 1194–1197. https://doi.org/10.1126/science.1185231

Centola, D. (2018). *How behavior spreads: The science of complex contagions*. Princeton University Press.

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: Models & applications*. Pion.

Collier, P. (2003). The economic causes of civil conflict and their implications for policy. In D. de Ferranti et al. (Eds.), *Conflict and development* (pp. 81–106). Washington, DC: World Bank.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187–202.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine, 210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. Arnold.

Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up*. MIT Press.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Fukuyama, F. (2014). *Political order and political decay: From the industrial revolution to the globalization of democracy*. Farrar, Straus and Giroux.

Gilbert, N. (2008). *Agent-based models* (Quantitative Applications in the Social Sciences). SAGE.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380. https://doi.org/10.1086/225469

Hayes, R. J., & Moulton, L. H. (2009). *Cluster randomised trials*. Boca Raton, FL: CRC Press.

Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials, 28*(2), 182–191.

International Crisis Group. (2016). *Curbing violence in Nigeria (III): The Boko Haram insurgency* (Africa Report No. 238). https://www.crisisgroup.org

Kalyvas, S. N. (2006). *The logic of violence in civil war*. Cambridge University Press.

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer.

Levi, M. (1998). *A state of trust*. In V. Braithwaite & M. Levi (Eds.), *Trust and Governance* (pp. 77–101). Russell Sage Foundation.

Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public service*. Russell Sage Foundation.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics, 21*(2), 239–245. https://doi.org/10.2307/1268522

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics, 33*(2), 161–174. https://doi.org/10.1080/00401706.1991.10484804

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.

Railsback, S. F., & Grimm, V. (2019). *Agent-based and individual-based modeling: A practical introduction* (2nd ed.). Princeton University Press.

Railsback, S. F., & Grimm, V. (2019). *Agent-based and individual-based modeling: A practical introduction* (2nd ed.). Princeton University Press.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Rose-Ackerman, S. (1999). *Corruption and government: Causes, consequences, and reform*. Cambridge University Press.

Rothstein, B. (2011). *The quality of government: Corruption, social trust, and inequality in international perspective*. University of Chicago Press.

Saltelli, A., et al. (2008). *Global sensitivity analysis: The primer*. Wiley.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*(2), 143–186.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Tyler, T. R. (2006). *Why people obey the law* (2nd ed.). Princeton University Press.

United Nations Office for the Coordination of Humanitarian Affairs. (2016). *Humanitarian needs overview: Nigeria*. https://www.unocha.org

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences, 99*(9), 5766–5771. https://doi.org/10.1073/pnas.082090499

Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440–442. https://doi.org/10.1038/30918

Wilensky, U. (1999). *NetLogo* (version 6) [Computer software]. Center for Connected Learning and Computer-Based Modeling, Northwestern University. https://ccl.northwestern.edu/netlogo/

World Bank. (2011). *World Development Report 2011: Conflict, security, and development*. World Bank.