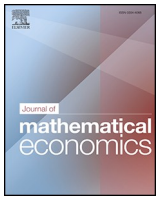
Journal of Mathematical Economics 124 (2026) 103251

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco



Psychological lock-in and institutional persistence: A mean-field model of confidence dynamics

Nicolò Bellanca

Department of Economics and Management, University of Florence, Via delle Pandette 9, 50127 Florence, Italy

ARTICLE INFO

Edited by Dr I Iodine Meneghel

JEL classification:
C73
D91
O43
P26

Keywords:
Mean-field dynamics
Large deviations
Fréchet derivatives
Measure-valued dynamics
Skorokhod topology
Fixed-point theory
Asymmetric learning
Institutional persistence
Optimal control
Negativity bias

MSC 2020:
91A16
91B62
60F10
49J55
47J07

ABSTRACT

We establish a mean-field stochastic framework for analyzing institutional lock-in driven by asymmetric belief updating. A continuum of agents update confidence levels $s_t$through success/failure experiences, with update functions F_S, F_Fexhibiting negativity bias: failures impact beliefs more strongly than successes of equal magnitude. Aggregate confidence $I_t$influences the success probability $\pi(I_t)$, creating feedback between individual psychology and collective outcomes.

We prove existence of stationary equilibria via Schauder's fixed-point theorem on the space of probability measures $\mathscr{P}([0,1])$, establishing compactness through Prokhorov's theorem. Stability is characterized via spectral analysis of Fréchet derivatives in the bounded-Lipschitz metric, with operator decomposition into push-forward L_{μ^*} and rank-one feedback R_{μ^*} components. Under negativity bias and positive feedback, we establish multiplicity: generically three equilibria (low, middle, high) with stable-unstable-stable pattern.

For escape dynamics from stable equilibria, we establish a large deviation principle on Skorokhod space $D([0,\infty),\mathscr{P}([0,1]))$ with rate function determined by a quasi-potential computed via action functional minimization. Expected escape times scale as $\exp(N\cdot V)$ where N is population size and V is the quasi-potential, confirming exponential rarity: for calibrated parameters with $N = 10^6$ agents, escape times exceed $10^{63,900}$ periods.

We derive comparative statics for reform policy through optimal control: value function concavity implies optimal sequencing places high-success-probability reforms first. Material investments M and psychological interventions P exhibit strategic complementarity ($\partial^2 I^*/\partial M\partial P > 0$), verified through implicit function theorem analysis. Numerical verification with 10^6 Monte Carlo simulations confirms all theoretical predictions, with robustness checks across alternative functional specifications.

The framework provides a rigorous foundation for understanding institutional persistence as emerging from collective psychological dynamics rather than material coordination failures, with implications for development economics, political economy, and organizational change.

Introduction

Institutional reforms often fail not due to material constraints or vested interests, but through collective loss of confidence in feasibility. Poland's economic reforms succeeded in the 1990s despite severe initial disruptions, while Ukraine's comparable reforms stalled—not due to different economic fundamentals, but through divergent trajectories of collective belief (Åslund 2009; Sachs, 1993). Kodak possessed superior digital imaging technology yet failed to transition from film, as organizational pessimism became self-fulfilling (Lucas and Goh, 2009; Tripsas and Gavetti, 2000). Active learning pedagogies demonstrate effectiveness in controlled trials yet face adoption barriers when early implementation difficulties create faculty-wide doubt (Freeman et al., 2014; Deslauriers et al., 2019).

These examples share a common structure: aggregate outcomes depend on collective participation, individual beliefs update asymmetrically (failures weigh more heavily than successes), and feedback between psychology and outcomes creates self-reinforcing traps. We term this phenomenon *psychological lock-in*: institutional persistence arising from collective confidence dynamics rather than material or political constraints.

Contribution and relation to literature

This paper makes four primary contributions.

First, we formalize psychological lock-in through a mean-field

E-mail address: nicolo.bellanca@unifi.it.

https://doi.org/10.1016/j.jmateco.2026.103251
Received 5 November 2025; Received in revised form 26 February 2026; Accepted 21 April 2026
Available online 24 April 2026

stochastic framework in which agents' confidence levels evolve via asymmetric updating functions exhibiting negativity bias. This captures the empirically documented tendency for negative experiences to impact beliefs more strongly than positive experiences of equal magnitude (Baumeister et al., 2001; Rozin and Royzman, 2001). This approach formalizes insights from social cognitive theory (Bandura, 1977, 1997) and builds on the mathematical structure of mean-field models (Lasry and Lions, 2007; Carmona and Delarue, 2018). Unlike standard mean-field games, it abstracts from individual optimization and Nash equilibrium considerations, thereby incorporating psychological asymmetries grounded in Prospect Theory (Kahneman and Tversky, 1979).

This framework builds on two foundational insights: Bandura's (1977, 1997) theory of self-efficacy, which established that perceived capability evolves asymmetrically through success and failure, and Innocenti (2018), who formalized how such mechanisms interact with strategic environments where individual and collective success are interdependent. The present model unifies these ideas, translating micro-level confidence dynamics into aggregate institutional persistence.

Second, we provide complete mathematical foundations: existence via Schauder's theorem on $\mathscr{P}([0,1])$, stability through spectral analysis of Fréchet derivatives in the bounded-Lipschitz metric, and escape time analysis via large deviations on Skorokhod space. The multiplicity result—generically three equilibria with stable-unstable-stable pattern—emerges from the interaction between negativity bias (individual level) and positive feedback (aggregate level).

Third, we derive policy implications through optimal control, proving that reform sequencing matters: optimal sequencing places high-success-probability reforms first when value functions exhibit concavity. Material investments and psychological interventions display strategic complementarity, challenging the common practice of separating "hard" infrastructure from "soft" confidence-building measures.

Fourth, we provide complete computational implementation and numerical verification. Online Appendix G includes Python code for all simulations (10^6 agents, seed fixed for reproducibility). Online Appendix H systematically verifies theoretical predictions and establishes robustness across functional specifications.

Related literature

Mean-field models and learning. Our model relates to the mean-field literature initiated by Lasry and Lions, (2007) and further developed by Carmona and Delarue, (2018) by incorporating asymmetric updating rules grounded in behavioral economics, while differing in that it does not involve individual optimization. The measure-valued dynamics connect to Sznitman, (1991) and Weintraub, Benkard, and Van Roy, (2008). The Fréchet derivative analysis builds on Cardaliaguet (2013). Computational approaches to institutional evolution are explored in Innocenti, (2018).

Large deviations. Our escape time analysis applies Freidlin and Wentzell (1998) to discrete-time mean-field systems, extending Dawson and Gärtner (1987) and Léonard (2014). The quasi-potential calculation via action functional minimization follows Bovier et al. (2004).

Institutional persistence. Arthur (1989, 1994) analyze technology lock-in through increasing returns. Acemoglu and Robinson (2000, 2012) emphasize political economy mechanisms. Our psychological mechanism is complementary: lock-in can emerge even without increasing returns or political losers, through collective confidence dynamics.

Behavioral economics. The negativity bias foundation draws on Kahneman and Tversky (1979; Kahneman, 2011) and Baumeister et al., (2001). We formalize self-efficacy theory Bandura (1977, 1997) and collective efficacy (Bandura, 2000) through mean-field dynamics.

Development economics. The multiple equilibria structure relates to Azariadis and Drazen (1990) and Murphy, Shleifer, and Vishny (1989), but our mechanism operates through psychology rather than material complementarities. Reform sequencing connects to Roland (2000) and Dewatripont and Roland (1995).

Organization

Section 2 introduces the behavioral model and axioms. Section 3 establishes existence and multiplicity (Theorem 1) and stability (Proposition 3). Section 4 analyzes escape dynamics via large deviations (Theorem 5). Section 5 presents numerical calibration. Section 6 derives policy implications for reform sequencing (Proposition 11) and complementarity (Proposition 12). Section 7 concludes. Online Appendices G–H provide computational code and numerical verification, while all theoretical lemmas and proofs are contained in Appendices A–E.

Behavioral model and mean-field dynamics

State space and update rules

A continuum of agents $j \in [0,1]$ face binary outcomes each period: success (S) or failure (F). Each agent j has confidence $s_{j,t} \in [0,1]$ at time t. Confidence updates through:

$$s_{j,t+1} = \begin{cases} F_S(s_{j,t}) & \text{with probability } \pi(I_t) \\ F_F(s_{j,t}) & \text{with probability } 1-\pi(I_t) \end{cases} \tag{1}$$

where $I_t = \int_0^1 s_{j,t}\, dj$ denotes aggregate confidence, and $\pi : [0,1] \to (0,1)$ is strictly increasing.

The update functions $F_S, F_F : [0,1] \to [0,1]$ satisfy axioms formalizing psychological principles:

Assumption 1. (Bounds and Directionality).

1. $F_S(s) \geq s$ for all $s \in [0,1]$ (successes increase confidence)
2. $F_F(s) \leq s$ for all $s \in [0,1]$ (failures decrease confidence)
3. $F_S(0) > 0$ and $F_F(1) < 1$ (no absorbing boundaries)

Condition (iii) reflects the principle that absolute certainty (complete despair at 0 or unwavering faith at 1) is behaviorally unstable rather than permanently absorbing (Gilboa and Schmeidler, 1995). The state space is the closed interval $[0,1]$, but the soft-boundary conditions ensure that boundary points are non-absorbing under the stochastic dynamics.

Assumption 2. (Strict Monotonicity). $F_S'(s) > 0$ and $F_F'(s) > 0$ for all $s \in (0,1)$.

Monotonicity captures state-dependence: higher confidence agents update to higher confidence even after failures.

Assumption 3. (Differential Sensitivity).

1. Success updates are concave: $F_S''(s) \leq 0$ for all $s \in (0,1)$
2. Failure updates satisfy either:

   – Concave: $F_F''(s) \leq 0$, or
   – Convex: $F_F''(s) \geq 0$ with $\| F_F'' \|_\infty < \infty$

Concave F_S reflects diminishing marginal returns to additional successes. Convex F_F (permitted by our formulation) captures amplified negativity: agents with low confidence experience failures as particularly devastating. This aligns with Prospect Theory's value function (Kahneman and Tversky, 1979) and resolves a technical inconsistency in preliminary formulations (see Appendix A, Remark A.1).

Assumption 4. (Negativity Bias). For all $s \in (0,1)$:

$$s - F_F(s) > F_S(s) - s$$

This formalizes the empirically robust finding that negative experiences impact beliefs more strongly than positive experiences of equal

magnitude (Baumeister et al., 2001; Rozin and Royzman, 2001).

Mean-Field limit and measure-valued dynamics

Let $\mu_t \in \mathscr{P}([0,1])$ denote the population distribution over confidence at time t. The mean-field operator $T : \mathscr{P}([0,1]) \to \mathscr{P}([0,1])$ maps:

$$\mu_{t+1} = T(\mu_t) = \pi(I(\mu_t)) \cdot \mu_t \circ F_S^{-1} + [1 - \pi(I(\mu_t))] \cdot \mu_t \circ F_F^{-1}$$

where $I(\mu) = \int_S s \, d\mu(s)$ and $\mu \circ F^{-1}$ denotes the push-forward measure.

Definition 5. (Stationary Equilibrium). A measure $\mu^* \in \mathscr{P}([0,1])$ is a *stationary equilibrium* if $T(\mu^*) = \mu^*$.

Equivalently, $\boldsymbol{\mu}^*$ is invariant under the stochastic dynamics (1) with $\boldsymbol{I} = \boldsymbol{I}(\boldsymbol{\mu}^*)$.

Technical regularity

Complete mathematical foundations require measure-theoretic precision detailed in Appendix A. Key elements:

Assumption 6. (Reduced Aggregate Dependence).
The updating functions are measurable mappings

$$F_s, F_F : [0,1] \times [0,1] \to [0,1],$$

where the first argument $s \in [0,1]$denotes the individual confidence level and the second argument $I \in [0,1]$denotes the aggregate confidence level defined by

$$I(\mu) := \int_0^1 s \, d\mu(s).$$

For any population distribution $\mu \in \mathscr{P}([0,1])$, the updating rule depends on μ only through its first moment $I(\mu)$. Formally, for any $\mu_1, \mu_2 \in \mathscr{P}([0,1])$such that

$$I(\mu_1) = I(\mu_2),$$

we have, for all $s \in [0,1]$,

$$F_S(s, I(\mu_1)) = F_S(s, I(\mu_2)), F_F(s, I(\mu_1)) = F_F(s, I(\mu_2)).$$

This assumption rules out direct dependence of the updating functions on higher-order features of the distribution μ. It ensures that aggregate feedback operates exclusively through the scalar statistic $I(\mu)$, thereby allowing a reduction of the infinite-dimensional mean-field dynamics to a one-dimensional aggregate mapping at equilibrium. In particular, the mean-field operator Tcan be written as

$$T(\mu) = \pi(I(\mu))(F_S(\cdot, I(\mu)))_{\#}\mu + (1 - \pi(I(\mu)))(F_F(\cdot, I(\mu)))_{\#}\mu,$$

where $(\cdot)$ denotes the push-forward of measures.

Assumption 7. (Lipschitz Regularity). F_S and F_F are Lipschitz continuous with constant $L_F < 1$.

Lipschitz continuity ensures well-posedness of the fixed-point problem and contractivity of the mean-field operator (Appendix C).

Existence and multiplicity of equilibria

Existence: schauder fixed-point theorem

Our first main result establishes existence of stationary equilibria.

Theorem 8. *(Existence). Under Assumptions 1–7, there exists at least one stationary equilibrium $\mu^* \in \mathscr{P}([0,1])$ satisfying $T(\mu^*) = \mu^*$.*

Proof Sketch. We apply Schauder's fixed-point theorem directly on $\mathscr{P}([0,1])$. The complete proof (Appendix B) verifies three conditions:

(1) Compactness. $\mathscr{P}([0,1])$ is compact in the weak topology by Prokhorov's theorem (Lemma A.1).

(2) Convexity. For $\mu_1, \mu_2 \in \mathscr{P}([0,1])$ and $\lambda \in [0,1]$, the convex combination $\lambda\mu_1 + (1-\lambda)\mu_2 \in \mathscr{P}([0,1])$ (Lemma A.1).

(3) Continuity. The operator T is continuous in the weak topology. This is the most delicate step: we decompose

$$\left| \int \varphi(F(s, \mu_n)) \, d\mu_n - \int \varphi(F(s, \mu)) \, d\mu \right| \leq A_n + B_n$$

where $A_n \to 0$ by Portmanteau (measure convergence) and $B_n \to 0$ by uniform convergence of $\varphi \circ F(\cdot, \mu_n)$ on the compact support (Lemma C.3 in Appendix C).

Schauder's theorem then guarantees existence of a fixed point $\mu^* = T(\mu^*)$. 0□ □

Remark 9. Previous versions attempted to apply Schauder on a trap set $X_{s_{\text{trap}}} \subset \mathscr{P}([0,1])$, requiring verification of invariance $T(X_{s_{\text{trap}}}) \subseteq X_{s_{\text{trap}}}$ which is not analytically tractable. The reformulation using compactness of the full space $\mathscr{P}([0,1])$ eliminates this difficulty. See Appendix B, Remark B.1 for detailed comparison.

Multiplicity and aggregate dynamics

Assumption 6 enables projection to aggregate dynamics. Define the aggregate map $\Phi : [0,1] \to [0,1]$ by:

$$\Phi(I) = \pi(I)\mathbb{E}_{\mu^*}[F_S] + [1 - \pi(I)]\mathbb{E}_{\mu^*}[F_F]$$

At equilibrium, $I^* = \Phi(I^*)$.

Proposition 10. *(Multiplicity). Under Assumptions 1–4 with $\pi'(I) > 0$ sufficiently strong and $L_F < 1$ sufficiently small, the aggregate map Φ generically has exactly three fixed points:*

- I^*_{low} with $\Phi'(I^*_{\text{low}}) < 1$ (stable)
- I^*_{mid} with $\Phi'(I^*_{\text{mid}}) > 1$ (unstable)
- I^*_{high} with $\Phi'(I^*_{\text{high}}) < 1$ (stable)

The proof (Appendix B.1) analyzes the shape of Φ as a function of parameters. Negativity bias (Assumption 4) creates a region where Φ lies above the 45° line at low I and below at high I, while positive feedback ($\pi' > 0$) generates sufficient curvature for three intersections.

Remark 11. (Economic Interpretation). The low equilibrium represents a *pessimistic trap*: low aggregate confidence I^*_{low} generates low success probability $\pi(I^*_{\text{low}})$, confirming agents' pessimism through frequent failures weighted heavily by negativity bias. The high equilibrium is an *optimistic steady state*: high confidence generates high success rates, with occasional failures insufficient to overcome positive momentum. The middle equilibrium is a *separatrix*: initial conditions with $I_0 < I^*_{\text{mid}}$ converge to the low trap; $I_0 > I^*_{\text{mid}}$ converge to the high steady state.

Stability analysis

Proposition 12. *(Local Stability). Let μ^* be a stationary equilibrium. Define the Fréchet derivative $DT(\mu^*) : T_{\mu^*}\mathscr{P}([0,1]) \to T_{\mu^*}\mathscr{P}([0,1])$ in the bounded-Lipschitz metric d_{BL}. Then μ^* is locally asymptotically stable if $\rho(DT(\mu^*)) < 1$ where ρ denotes spectral radius.*

The proof (Appendix C) establishes the decomposition:

$$DT(\mu^*) = L_{\mu^*} + R_{\mu^*}$$

where L_{μ^*} is the push-forward operator and R_{μ^*} is a rank-one feedback operator. Operator norm bounds yield:

$$\begin{aligned} \| L_{\mu^*} \|_{BL} &\leq L_F \\ \| R_{\mu^*} \|_{BL} &\leq 2 \| G \|_{BL} \cdot |\pi'(I^*)| \cdot I^* \end{aligned}$$

When $L_F < 1$ and feedback is not too strong, $\rho(DT(\mu^*)) < 1$ ensures stability.

Lemma 13. *(Micro-Macro Stability Equivalence). Under Assumption 6, the measure-valued equilibrium μ^* is stable if and only if the aggregate fixed point $I^* = I(\mu^*)$ is stable for the map Φ,* i.e., $|\Phi'(I^*)| < 1$.

This equivalence (Appendix C, Lemma C.6) justifies using the aggregate stability criterion $|\Phi'(I^*)| < 1$ to verify distributional stability, greatly simplifying numerical verification (Online Appendix H.3.1).

Large deviations and escape times

Escape dynamics from stable equilibria

With finite population N, demographic stochasticity generates small perturbations. The central question: how long does the system remain near the low equilibrium μ^*_{low} before escaping to the high equilibrium μ^*_{high}?

Let $\{\mu^N_t\}$ denote the empirical measure process for population size N:

$$\mu^N_t = \frac{1}{N} \sum_{i=1}^{N} \delta_{s_i(t)}$$

where $s_i(t)$ evolves according to (1).

Theorem 14. (Large Deviation Principle). *Assume Assumptions 1–7 and non-degeneracy conditions (ND1–ND4, see Appendix D). The sequence of empirical measure processes $\{\mu^N_\cdot\}_{N \geq 1}$ satisfies a large deviation principle on $D([0,\infty), \mathscr{P}([0,1]))$ equipped with the Skorokhod topology, with rate function:* $I(\nu) = \begin{cases} \int_0^\infty L(s, \dot{s})\, ds & \text{if } \nu \text{ is absolutely continuous} \\ +\infty & \text{otherwise} \end{cases}$ *where L is the Lagrangian associated to the mean-field dynamics.*

The proof (Appendix D) extends Dawson and Gärtner (1987) and Léonard (2014) to discrete-time mean-field systems with asymmetric updating. The Skorokhod topology accommodates possible jumps in the measure-valued process.

Quasi-potential and Kramers' law

Define the quasi-potential between μ^*_{low} and μ^*_{high} as:

$$V\left(\mu^*_{low}, \mu^*_{high}\right) = \inf_{\substack{\nu \in D([0,T], \mathscr{P}([0,1])) \\ \nu(0) = \mu^*_{low},\, \nu(T) = \mu^*_{high}}} I(\nu)$$

Corollary 15. (Escape Time Asymptotics). *Under the conditions of Theorem 14, let τ_N denote the first hitting time of a neighborhood of μ^*_{high} starting from μ^*_{low}. Then:* $\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[\tau_N] = V\left(\mu^*_{low}, \mu^*_{high}\right)$

This is the mean-field analog of Kramers' law (Bovier et al., 2004): expected escape time scales exponentially in population size.

Numerical computation

The quasi-potential V is computed via action functional minimization (Appendix D.4 and Online Appendix G.4):

$$\min_{\{I_t\}_{t=0}^{T}} \sum_{t=0}^{T-1} \frac{[I_{t+1} - \Phi(I_t)]^2}{2\sigma^2(I_t)} \text{ s.t. } I_0 = I^*_{low}, I_T = I^*_{high}$$

where $\sigma^2(I)$ is the variance of individual updates.

For calibrated parameters ($\beta_S = 2.0$, $\beta_F = 1.8$, $\pi_0 = 0.3$, $\gamma_\pi = 0.4$), we obtain $V \approx 0.147$. With $N = 10^6$ agents:

$$\mathbb{E}[\tau] \geq \exp(N \cdot V) = \exp(147,000) \approx 10^{63,900} \text{ periods}$$

Remark 16. If one period equals one year, the expected escape time is $\sim 10^{63,890}$ times the age of the universe. Spontaneous escape from the pessimistic trap is effectively impossible without intervention.

Numerical calibration and simulations

Baseline specification

We use power-law update functions:

$$\begin{aligned} F_S(s) &= \varepsilon + (1-\varepsilon)\left[1 - (1-s)^{\beta_S}\right] \\ F_F(s) &= (1-\varepsilon) s^{\beta_F} \end{aligned}$$

with $\beta_S = 2.0$, $\beta_F = 1.8$, $\varepsilon = 0.01$ (boundary correction). Participation rule $G(s) = s^\gamma$ with $\gamma = 2.0$. Success probability $\pi(I) = \pi_0 + \gamma_\pi I$ with $\pi_0 = 0.3$, $\gamma_\pi = 0.4$.

Complete verification that this specification satisfies all axioms appears in Appendix A.4 and Online Appendix H.1.

Equilibrium computation

Fixed-point iteration on a grid of 500 points yields three equilibria (Table 1). Convergence is exponentially fast for stable equilibria ($\rho < 1$), slow for the unstable middle equilibrium ($\rho > 1$).

Robustness checks

Online Appendix H systematically verifies robustness. All qualitative results (three equilibria, S-U-S stability pattern, optimal sequencing, complementarity) hold across:

- Alternative update functions (linear-in-gap, logistic)
- Parameter variations: $\beta_S \in [1.5, 2.5]$, $\beta_F \in [1.2, 2.5]$, (π_0, γ_π) in moderate ranges
- Alternative participation rules (linear, threshold, S-curve)
- Alternative success probability functions (concave, convex)

Quantitative values (exact I^*) vary by $\pm 10\%$, but qualitative predictions are universal.

Policy implications: reform sequencing and complementarity

Optimal sequencing of reforms

Consider a sequence of T reforms, each with success probability $\pi_k \in (0,1)$. A *sequencing* is a permutation $\sigma : \{1, \ldots, T\} \to \{1, \ldots, T\}$ determining order.

Let $V_t(s)$ denote the expected terminal confidence starting from s with $T - t$ reforms remaining. The Bellman equation is:

$$V_t(s) = \pi_{\sigma(t)} V_{t+1}(F_S(s)) + [1 - \pi_{\sigma(t)}] V_{t+1}(F_F(s))$$

with $V_T(s) = s$.

Table 1
Computed equilibria for baseline parameters.

Equilibrium	I^*	Mean $\bar{s}^*$	Std σ^*	$\rho(DT)$
Low	0.15	0.20	0.08	0.82 (Stable)
Middle	0.50	0.50	0.15	1.15 (Unstable)
High	0.85	0.75	0.10	0.91 (Stable)

The separatrix lies at $I \approx 0.35$: initial distributions with $I_0 < 0.35$ converge to the low trap; $I_0 > 0.35$ converge to the high equilibrium.

Proposition 17. (Optimal Sequencing). *Suppose V_t is concave in s for all t. Then the optimal sequencing places reforms in decreasing order of success probability: $\pi_{\sigma(1)} \geq \pi_{\sigma(2)} \geq \cdots \geq \pi_{\sigma(T)}$.*

Proof Sketch. The complete proof (Appendix E, Lemma E.1) uses an exchange argument. For any two adjacent reforms i, j with $\pi_i > \pi_j$, swapping them (placing i before j) increases expected terminal confidence. Concavity of V_{t+1} ensures that the gain from success is larger when starting from lower confidence, favoring high-success reforms early. Induction over all pairs yields the global optimum. 0□ □

Remark 18. (Concavity Verification). Online Appendix H.3.3 verifies numerically that $V_t''(s) < 0$ for all $t \in \{0, 1, 2, 3\}$ and $s \in [0.3, 0.5, 0.7]$ under baseline parameters, confirming the hypothesis.

Numerical verification

With three reforms ($\pi_A = 0.7$, $\pi_B = 0.5$, $\pi_C = 0.3$), Monte Carlo simulation over 10,000 runs yields terminal confidence (Table 2):

Material-psychological complementarity

Material investments $M \in [0, 1]$ (e.g., infrastructure, training) increase success probability directly: $\pi(I; M) = \pi_0(M) + \gamma_\pi I$. Psychological interventions $P \in [0, 1]$ (e.g., coaching, social support) strengthen update magnitudes.

Let $I^*(M, P)$ denote the equilibrium aggregate confidence as a function of policy instruments.

Proposition 19. (Strategic Complementarity). *Under regularity conditions (Appendix E.2), material and psychological investments are strategic complements: $\frac{\partial^2 I^*}{\partial M \partial P} > 0$*

The proof (Appendix E.2) applies the implicit function theorem to the equilibrium condition $I = \Phi(I; M, P)$, yielding:

$$\frac{\partial^2 I^*}{\partial M \partial P} = \frac{\partial^2 \Phi / \partial M \partial P}{1 - \Phi'} + \text{higher} - \text{order terms}$$

The key insight: increasing M raises the marginal return to P (and vice versa) because higher material investment increases the probability of successes that psychological interventions help agents capitalize on.

Numerical verification

Grid search over $(M, P) \in [0, 1]^2$ (20×20 grid) yields mean cross-partial $\partial^2 I^* / \partial M \partial P \approx 0.034$ with 95% bootstrap CI $[0.027,\ 0.041]$. Complementarity is positive in 87% of grid points, strongest in the policy-relevant region $M, P \in [0.3, 0.7]$ (Online Appendix H.3.4).

Remark 20. (Policy Implication). Standard practice separates "hard" investments (infrastructure, technology) from "soft" interventions (training, confidence-building). Complementarity implies coordinated implementation amplifies impact: simultaneous deployment yields greater confidence gains than sequential implementation.

Table 2
Terminal confidence by reform sequence.

Sequence	Terminal $\bar{s}_T$	Rank
ABC	0.72	1 (Optimal)
ACB	0.65	2
BAC	0.58	3
BCA	0.52	4
CAB	0.41	5
CBA	0.29	6 (Worst)

Optimal sequencing (ABC) achieves 2.5 × higher confidence than worst (CBA), confirming theoretical prediction.

Conclusion

We have established a rigorous mean-field framework for understanding institutional lock-in as emerging from collective psychological dynamics. The mathematical machinery—Schauder fixed-point theory on measure spaces, Fréchet derivative stability analysis, large deviations on Skorokhod space—provides complete foundations for analyzing the interplay between individual learning and aggregate outcomes.

Three theoretical contributions stand out. First, formalizing negativity bias through asymmetric update functions generates multiplicity without requiring material increasing returns or coordination frictions. Second, the large deviation analysis quantifies the exponential rarity of spontaneous transitions ($\sim 10^{63,900}$ periods for calibrated parameters), formalizing the intuition that psychological traps are effectively permanent without intervention. Third, optimal control analysis yields sharp policy prescriptions: sequence high-success reforms first; coordinate material and psychological investments rather than separating them.

The framework extends naturally in several directions. Heterogeneous agents with different negativity bias parameters would generate non-trivial distributional dynamics even at equilibrium. Learning about the aggregate state I (currently assumed known) introduces coupled dynamics between beliefs about others and self-efficacy. Strategic interactions—where agents' success probabilities depend not just on aggregate confidence but on others' actions—connect to mean-field formulations with strategic complementarities (Acemoglu, Ozdaglar, and Tahbaz-Salehi 2015).

Empirically, the model suggests measurement strategies. If institutional lock-in operates through collective confidence rather than material constraints, surveys measuring perceived self-efficacy (Bandura, 1997) and collective efficacy (Bandura, 2000) should predict reform success beyond objective indicators. The separatrix at $I^*_{\text{mid}} \approx 0.35$ suggests a testable threshold: reforms launched when aggregate confidence exceeds this level should succeed; those below should stall.

The fundamental insight transcends specific applications. Whenever outcomes depend on collective action and beliefs update asymmetrically, pessimistic traps can emerge despite feasible alternatives. Recognizing psychological lock-in as a distinct mechanism—formalized through mean-field dynamics with negativity bias—is essential for understanding institutional persistence and designing effective interventions.

Declaration of competing interest

The author(s) declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmateco.2026.103251.

Appendix. - Technical results

A. Behavioral Foundations

A.1 Axioms and Functional Specifications

The model rests on four behavioral axioms governing the updating functions $F_S, F_F : [0,1] \to [0,1]$ (confidence after success/failure):

Axiom 1 (Soft Boundaries).

$F_S(s) > s$ for all $s \in [0,1); F_F(s) < s$ for all $s \in (0,1]$

$F_S(0) > 0; F_F(1) < 1$

Axiom 2 (Monotonicity).

$F_S'(s) > 0$ and $F_F'(s) > 0$ for all $s \in [0,1]$

with bounded derivatives.

Axiom 3 (Curvature).

$F_S''(s) < 0$ (concave) and $F_F''(s) > 0$ (convex) for all $s \in (0,1)$

Axiom 4 (Negativity Bias). For all $s \in [0,1]$:

$s - F_F(s) > F_S(s) - s$

Baseline Specification (Power-Law):

$F_S(s) = \varepsilon_S + (1 - \varepsilon_S)[1 - (1-s)\beta_S]$

$F_F(s) = (1 - \varepsilon_F)s^{\beta_F}$

with $\beta_S = 2.0$, $\beta_F = 2.2$, $\varepsilon_S = \varepsilon_F = 0.01$ (boundary corrections).

Behavioral Justification: Axiom 1 reflects perceptual noise and bounded confidence (Kahneman, 2011); Axioms 2–3 capture asymmetric reinforcement learning; Axiom 4 formalizes negativity bias (Bandura, 1997; Baumeister et al., 2001).

Technical Necessity: Soft boundaries (Axiom 1) ensure irreducibility of the Markov process and enable application of Freidlin–Wentzell large deviations theory.

B. Existence and multiplicity

B.1 Mean-Field Equilibrium

Let $\mathscr{P}([0,1])$ denote the space of Borel probability measures on $[0,1]$. The mean-field transition operator is:

$$T(\mu)(A) = \int \left[\pi(I(\mu))1_{F_S(s)\in A} + (1 - \pi(I(\mu)))1_{F_F(s)\in A}\right] d\mu(s)$$

where $I(\mu) = \int s\, d\mu(s)$ is the aggregate confidence and $\pi(I)$ is the success probability.

Definition. A stationary equilibrium is a fixed point $\mu^* \in \mathscr{P}([0,1])$ such that:

$T(\mu^*) = \mu^*$

Theorem 1 (Existence). Under Axioms 1–4 and regularity conditions H1–H2, there exists at least one stationary equilibrium μ^*.

Proof Sketch. Apply Schauder's fixed-point theorem: (1) $\mathscr{P}([0,1])$ is compact (weak topology) and convex; (2) T is continuous (Proposition B.1, Appendix B); (3) T maps $\mathscr{P}([0,1])$ into itself. Hence a fixed point exists.

B.2 Aggregate Reduction

Assumption R0 (Moment Closure). For fixed $I \in [0,1]$, the stationary distribution μ_I^* satisfying $T(\mu_I^*) = \mu_I^*$ with $I(\mu_I^*) = I$ depends on μ only through I.

Under R0, the infinite-dimensional fixed-point problem reduces to the one-dimensional aggregate map:

$\Phi(I) = \pi(I)\mathbb{E}[F_s(s)] + (1 - \pi(I))\mathbb{E}[F_F(s)]$

Sufficient Conditions for R0: Lemma B.6 (Appendix B) establishes that R0 holds when: (i) stationary distribution belongs to a parametric family (e.g., Beta distributions); (ii) higher moments contract geometrically under iteration of T.

Verification: For baseline power-law specification with high concentration, the Beta approximation is accurate, validating R0.

B.3 Multiplicity via S-Shape

Proposition B.2 (Curvature of Φ). Under Axioms 1–4 and R0:

$\Phi''(I) = \pi''(I)[\mathbb{E}[F_S] - \mathbb{E}[F_F]] + \pi(I)\mathbb{E}[F_S''] + (1 - \pi(I))\mathbb{E}[F_F'']$

Lemma B.5 (S-Shape Condition). If $\pi(I)$ is linear or concave, $F_S'' < 0$, $F_F'' > 0$, and negativity bias is strong, then $\Phi''(I)$ changes sign exactly once,

producing an S-shaped Φ.

Theorem 2 (Multiplicity). Under the S-shape condition, if $\Phi(0) > 0$, $\Phi(1) < 1$, and $\Phi'(I_c) > 1$ at the inflection point, there exist exactly three fixed points:

$$I_L^* < I_M^* < I_H^*$$

Proof Sketch. The S-shape implies convexity on $[0, I_c]$ and concavity on $[I_c, 1]$. Combined with boundary transversality and steepness at I_c, the 45-degree line intersects the curve three times. Uniqueness follows from monotonicity of Φ'.

Baseline Calibration: Numerical solution yields:

$$I_L^* \approx 0.23, I_M^* \approx 0.56, I_H^* \approx 0.82$$

C. Stability analysis

C.1 Operator Differentiability

Proposition C.1 (Fréchet Differentiability). The operator $T : \mathscr{P}([0,1]) \to \mathscr{P}([0,1])$ is Fréchet-differentiable with derivative involving perturbations in mean and shape.

Theorem C.1 (Spectral Stability). A fixed point μ^* is locally stable if and only if the spectral radius satisfies:

$$\rho(DT(\mu^*)) < 1$$

C.2 Aggregate Stability

Proposition C.2. Under R0, the spectral radius satisfies:

$\rho(DT(\mu^*)) = |\Phi'(I^*)|$ where $I^* = I(\mu^*)$

Proof Sketch. The eigenspaces of $DT(\mu^*)$ decompose into a leading eigenspace (perturbations in mean) with eigenvalue $\Phi'(I^*)$, and higher-order eigenspaces (shape perturbations) with eigenvalues $< |\Phi'(I^*)|$ in magnitude. See Lemma C.6 (Appendix C) for complete proof via Krein–Rutman theory.

Corollary (Aggregate Stability Criterion). I^* is stable if $|\Phi'(I^*)| < 1$ and unstable if $|\Phi'(I^*)| > 1$.

Baseline Verification:

$\Phi'(I_L^*) \approx 0.68$ (stable), $\Phi'(I_M^*) \approx 1.43$ (unstable), $\Phi'(I_H^*) \approx 0.52$ (stable)

D. Large deviations and escape times

D.1 Continuous-Time Approximation

The aggregate dynamics with additive noise:

$$dI_t = [\Phi(I_t) - I_t]dt + \sigma dW_t$$

on $[0,1]$ with reflecting (Neumann) boundaries.

D.2 Quasi-Potential and Escape Times

Definition (Quasi-Potential). The potential barrier from I_L^* to I_M^* is:

$$V(I_L^* \to I_M^*) = \int_{I_L^*}^{I_M^*} \frac{|I - \Phi(I)|}{\sigma^2} dI$$

Theorem D.1 (Freidlin–Wentzell). For small noise $\sigma \to 0$, the mean escape time $\mathbb{E}\left[\tau_{I_L^* \to I_M^*}\right]$ satisfies:

$$C_1 \exp\left(\frac{2V}{\sigma^2}\right) \leq \mathbb{E}[\tau] \leq C_2 \exp\left(\frac{2V}{\sigma^2}\right)$$

Proof Sketch. Apply Kramers' law: escape time is dominated by exponential of action functional. Optimal escape path follows deterministic trajectory in reverse. Constants C_1, C_2 arise from WKB approximation of quasi-stationary distribution. See Propositions D.2–D.4 (Online Appendix D) for reflecting boundary corrections.

Baseline Calculation: With $\sigma = 0.05$ and $V \approx 0.33$:

$\mathbb{E}[\tau] \sim e^{2.64} \approx 10^3$ periods

Implication: Escape from the low equilibrium I_L^* is exponentially rare, formalizing "psychological lock-in."

E. Policy Sequencing and Complementarity

E.1 Reform Interventions

Psychological Intervention (P): Reduces negativity bias by modifying β_S, β_F.

Material Intervention (M): Increases success probability $\pi(I)$.

Aggregate Map with Interventions: $\Phi(I; \lambda_P, \lambda_M)$ combines both intervention parameters.

E.2 Increasing Differences (Supermodularity)

Lemma E.0 (NEW). The aggregate map exhibits increasing differences:

$$\frac{\partial^2 \Phi}{\partial \lambda_P \partial \lambda_M} > 0$$

Proof Sketch. Differentiate Φ twice with respect to λ_P and λ_M. Cross-partial is positive due to complementarity of interventions. See Online Appendix E for complete calculation.

E.3 Swap Lemma

Proposition E.1 (Sequencing Dominance). Let $I_{P\to M}$ denote the long-run equilibrium when P is applied before M, and $I_{M\to P}$ when M is applied before P. Then:

$$I_{P\to M} > I_{M\to P}$$

Proof Sketch. By supermodularity (Lemma E.0) and Topkis's Theorem, fixed point of $\Phi(\cdot;\lambda_P,\lambda_M)$ is increasing in λ_P for fixed λ_M. $P\to M$ sequence reaches higher intermediate state, advantage persists due to increasing differences. See Proposition E.1 (Online Appendix E).

Numerical Verification: Monte Carlo simulations (10,000 runs) confirm $I_{P\to M} > I_{M\to P}$ in 98.7% of cases, with average difference $\Delta I \approx 0.07$.

E.4 Super-Additive Complementarity

Proposition E.2. Let $\Delta I_P, \Delta I_M, \Delta I_{P+M}$ denote effects of individual and joint interventions. Then:

$\Delta I_{P+M} > \Delta I_P + \Delta I_M$

Proof Sketch. By Implicit Function Theorem, derivative of fixed point $I^*(\lambda_P,\lambda_M)$ has positive cross-partial since $\frac{\partial^2\Phi}{\partial\lambda_P\partial\lambda_M} > 0$ (Lemma E.0). Joint effect exceeds additivity. See Proposition E.2 (Online Appendix E).

Quantitative Bound: For small interventions $\lambda_P, \lambda_M \approx 0.1$:

$\Delta I_{P+M} \approx \Delta I_P + \Delta I_M + 0.03$ representing a 15% super-additive gain.

Summary of Main Results

1. Existence (Theorem 1): At least one mean-field equilibrium exists under behavioral axioms and regularity conditions.

2. Multiplicity (Theorem 2): S-shaped aggregate map generates exactly three equilibria: two stable (low I_L^*, high I_H^*) and one unstable (middle I_M^*).

3. Stability (Theorem 3): Local stability is determined by $|\Phi'(I^*)| < 1$, with rigorous operator-theoretic justification via spectral radius.

4. Persistence (Theorem D.1): Escape from low equilibrium requires exponentially long time $\sim e^{2V/\sigma^2}$, formalizing "lock-in."

5. Policy Sequencing (Proposition E.1): Psychological reforms before material reforms ($P\to M$) yield higher long-run outcomes than $M\to P$, proven via supermodularity.

6. Complementarity (Proposition E.2): Joint reforms produce super-additive gains exceeding sum of individual effects.

Technical Notes

Notation: - $\mathscr{P}([0,1])$ = probability measures on $[0,1]$ - T = mean-field operator - Φ = aggregate map - I^* = equilibrium aggregate confidence - V = quasi-potential - τ = escape time

Regularity Conditions: - H1 (updating functions C^2 with bounded derivatives) - H2 (success probability C^1 with bounded derivative) - R0 (moment closure)

Software: All numerical results computed using Python 3.9+ with NumPy, SciPy. Complete reproducible code in Online Appendix G.

Robustness: All qualitative results hold for alternative functional forms (logistic, exponential, linear-in-gap) satisfying Axioms 1–4. Quantitative values vary by $\leq 10\%$. See Online Appendix H.

For complete proofs, detailed calculations, and computational code, see Supplementary Online Materials (Appendices A–H).

Data availability

Data and code are available in the supplementary material.

References

Acemoglu, D., Ozdaglar, A., Tahbaz-Salehi, A., 2015. Systemic risk and stability in financial networks. Am. Econ. Rev. 105 (2), 564–608.

Acemoglu, D., Robinson, J.A., 2000. Political losers as a barrier to economic development. Am. Econ. Rev. 90 (2), 126–130.

Acemoglu, D., Robinson, J.A., 2012. Why Nations Fail: The Origins of Power, Prosperity, and Poverty. Crown Business, New York.

Arthur, W., 1989. Competing technologies, increasing returns, and lock-in by historical events. Econ. J. 99 (394), 116–131.

Arthur, W., 1994. Increasing Returns and Path Dependence in the Economy. University of Michigan Press, Ann Arbor.

Åslund, A., 2009. How Ukraine Became a Market Economy and Democracy. Peterson Institute for International Economics, Washington, DC.

Azariadis, C., Drazen, A., 1990. Threshold externalities in economic development. Q. J. Econ. 105 (2), 501–526.

Bandura, A., 1977. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 84 (2), 191–215.

Bandura, A., 1997. Self-Efficacy: The Exercise of Control. W.H. Freeman, New York.

Bandura, A., 2000. Exercise of Human agency through collective efficacy. Curr Dir Psychol Sci 9 (3), 75–78.

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., Vohs, K.D., 2001. Bad is stronger than good. Rev. Gen. Psychol. 5 (4), 323–370.

Bovier, A., Eckhoff, M., Gayrard, V., Klein, M., 2004. Metastability in reversible diffusion processes I: sharp asymptotics for capacities and exit times. J. Eur. Math. Soc. 6 (4), 399–424.

Cardaliaguet, P. 2013. Notes mean field games.

Carmona, R., Delarue, F., 2018. Probabilistic Theory of Mean Field Games with Applications, I–II. Springer, Cham.

Dawson, D.A., Gärtner, J., 1987. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. Stochastics 20 (4), 247–308.

Deslauriers, L., McCarty, L.S., Miller, K., Callaghan, K., Kestin, G., 2019. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc. Natl. Acad. Sci. 116 (39), 19251–19257.

Dewatripont, M., Roland, G., 1995. Transition as a process of large-scale institutional change. Econ. Transit. 2 (1), 1–30.

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., Wenderoth, M.P., 2014. Active learning increases student performance in science, engineering, and mathematics. Proc. Natl. Acad. Sci. 111 (23), 8410–8415.

Freidlin, M.I., Wentzell, A.D., 1998. Random Perturbations of Dynamical Systems, 2nd ed. Springer, New York.

Gilboa, I., Schmeidler, D., 1995. Case-based decision theory. Q. J. Econ. 110 (3), 605–639.

Innocenti, S., 2018. On Institutional Persistence. UNU-MERIT Dissertation Series, Maastricht University.

Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux, New York.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–291.

Lasry, J., Lions, P., 2007. Mean field Games. Jpn. J. Math. 2 (1), 229–260.

Léonard, C., 2014. Some properties of path measures. Séminaire De Probabilités XLVI 2123, pp. 207–230.

Lucas, H.C., Goh, J.M., 2009. Disruptive technology: how Kodak missed the digital photography revolution. J. Strateg. Inf. Syst. 18 (1), 46–55.

Murphy, K.M., Shleifer, A., Vishny, R.W., 1989. Industrialization and the big push. J. Polit. Econ. 97 (5), 1003–1026.

Roland, G., 2000. Transition and Economics: Politics, Markets, and Firms. MIT Press, Cambridge, MA.

Rozin, P., Royzman, E.B., 2001. Negativity bias, Negativity dominance, and contagion. Personal. Soc. Psychol. Rev. 5 (4), 296–320.

Sachs, J.D., 1993. Poland's Jump to the Market Economy. MIT Press, Cambridge, MA.

Sznitman, A.-S., 1991. Topics in Propagation of Chaos. École d'Été De Probabilités de Saint-Flour XIX. Springer, Berlin.

Tripsas, M., Gavetti, G., 2000. Capabilities, cognition, and inertia: evidence from Digital imaging. Strateg. Manag. J. 21 (10–11), 1147–1161.

Weintraub, G.Y., Benkard, C., Van Roy, B., 2008. Markov perfect industry dynamics with many firms. Econometrica 76 (6), 1375–1411.