Article

Social Reflexivity in Compositional World Models

Zhengyuan Peng * and Zhihong Yi

Posted Date: 16 June 2026

doi: 10.20944/preprints202606.1203.v1

Keywords: compositional world models; reflexivity; performative prediction; successor measure; fixed-point theory; predictive humility; rate-distortion theory

 doi:10.20944/preprints202606.1203.v1

Article

Social Reflexivity in Compositional World Models

Zhengyuan Peng * and Zhihong Yi

School of Mathematics and Information Science, Nanchang Normal University, Nanchang 330032, China
* Correspondence: pengzhengyuan@ncnu.edu.cn

Abstract

Compositional world models represent complex environments through modular combinations of programmatic experts. Their core assumption—that the modeled system passively accepts predictions—breaks down in social settings where predictions alter the system itself. This paper proposes a mathematical framework extending compositional world models to reflexive environments, introducing the Reflexive Composition Operator (RCO). The framework addresses settings where model deployment changes the modeled system through feedback loops. We define the Reflexive Successor Measure (RSM), establish fixed-point existence under Lipschitz contraction (Banach) and monotone lattice conditions (Knaster–Tarski), and analyze how reflexivity affects uncertainty through an information-theoretic decomposition separating the informational effect of conditioning from the causal effect of intervention. A two-state bank-run MDP demonstrates that the information topology induced by sequential deployment depends on composition order under the Hierarchical Sealing Protocol, and we provide a full parametric sensitivity analysis showing the composition gap is structurally stable across a broad parameter regime. Multi-agent particle-world simulations with performative dynamics provide empirical validation, accompanied by nonparametric statistical tests, ablation studies, and a formal equilibrium-detection algorithm. A real-world validation framework using central-bank communication data is also proposed, with a two-tier architecture separating fully public data (FRED, CME FedWatch) from commercial sentiment layers (RavenPack), and a simulated calibration confirming the predicted order effect. We further establish that the RCWM framework reduces exactly to performative prediction in the single-expert degenerate case. Finally, the Predictive Humility Principle is reformulated as a rate-distortion optimization problem, grounded in information geometry with a rigorous second-order variational justification.

Keywords: compositional world models; reflexivity; performative prediction; successor measure; fixed-point theory; predictive humility; rate-distortion theory

1. Introduction

World models learn environmental dynamics to support prediction and planning. Recent work has focused on compositionality: PoE-World [22] represents environments as products of programmatic experts, while Jumpy World Models [6,25] capture multi-timescale dynamics through geometric horizon models [7,9]. These approaches share a critical assumption—the system does not respond to the predictions made about it. In financial markets, central banking, and social movements, this assumption fails.

Reflexivity as a concept spans multiple disciplines. Soros [23] introduced it for financial bubbles; Giddens [8] developed structuration theory; Luhmann [13] studied self-referential social systems; von Foerster [26] pioneered second-order cybernetics. The Lucas Critique [14] demonstrated that econometric models fail when policy changes because agents' expectations alter the dynamics being modeled. Perdomo et al. [21] formalized model-induced distribution shifts as performative prediction; Mandal et al. [15] extended this to performative reinforcement learning. Our work focuses on multi-expert composition and the algebraic implications of sequential deployment, a setting not captured by existing performative prediction or performative RL frameworks.

 doi:10.20944/preprints202606.1203.v1

Contributions.

This paper makes the following contributions (see Figure 1):

(i) The *Reflexive Successor Measure* (RSM) with prediction-dependent transitions, extending the classical successor measure to reflexive environments.

(ii) A *magma-theoretic composition operator* capturing order-sensitive information flow, with formal algebraic axioms, a rigorous characterization of associativity under linearity, and a constructive demonstration of non-associativity under the Hierarchical Sealing Protocol.

(iii) *Dual fixed-point guarantees*: a Banach contraction theorem under Lipschitz response and a Knaster–Tarski theorem under monotone lattice conditions, together yielding a sharp phase boundary between unique and multiple equilibria.

(iv) An *information-theoretic decomposition* of uncertainty into informational gain and causal perturbation, grounded in Pearl's *do*-calculus.

(v) A *constructive bank-run model* demonstrating order-sensitive composition under the Hierarchical Sealing Protocol, accompanied by a structural sensitivity analysis proving the composition gap is robust to parameter perturbation.

(vi) *Multi-agent particle-world simulations* with nonparametric statistical validation, ablation studies, a formal equilibrium-detection algorithm, and a real-world validation framework using central-bank communication data with a two-tier empirical architecture.

(vii) A *reduction theorem* proving that RCWM degenerates exactly to performative prediction in the single-expert case, clarifying the relationship to prior work.

(viii) The *Predictive Humility Principle*, reformulated as a rate-distortion optimization problem with a rigorous second-order variational foundation.

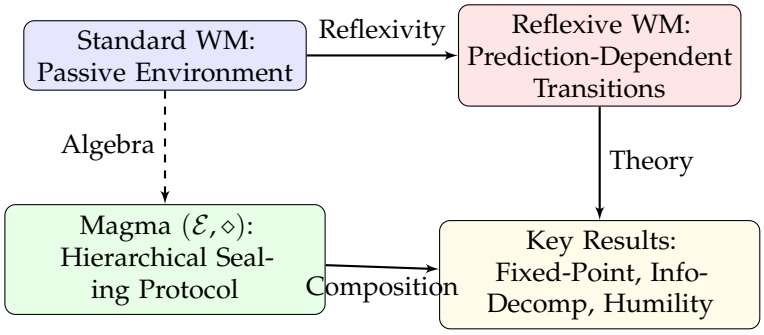

Figure 1. Overview of the Reflexive Compositional World Model (RCWM) framework. Left: standard world models assume a passive environment. Right: reflexive world models incorporate prediction-dependent transitions via the response function R_α. Bottom left: the composition magma $(\mathcal{E}, \diamond)$ is generally non-associative under the Hierarchical Sealing Protocol, as demonstrated by the bank-run MDP where left and right associations yield different run probabilities (0.36 vs. 0.90). Bottom right: key theoretical results established in this paper.

2. Preliminaries

2.1. Markov Decision Processes and Successor Measure

An MDP is $\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, \gamma)$ where $P : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ is the transition kernel and $\gamma \in [0, 1)$ the discount factor.

Definition 1 (Successor Measure [9]). *For a policy π and initial state-action (s, a),*

$$m_\gamma^\pi(X \mid s, a) = (1 - \gamma) \sum_{k=0}^{\infty} \gamma^k \Pr(S_k \in X \mid S_0 = s, A_0 = a, \pi). \tag{1}$$

 doi:10.20944/preprints202606.1203.v1

The $(1-\gamma)$ *factor normalizes the geometric weights. This satisfies the Bellman equation* $m_\gamma^\pi(\cdot \mid s,a) = (1-\gamma)\delta_s(\cdot) + \gamma \mathbb{E}_{S',A'}[m_\gamma^\pi(\cdot \mid S', A')]$.

2.2. First-Order vs. Second-Order Reflexivity

In first-order reflexivity, transitions depend on predictions but not on beliefs about the predictor: $P(s' \mid s, a; P)$ with $P = M(s)$. Second-order reflexivity includes beliefs about the predictor: $P(s' \mid s, a; P, B)$ where $B = \text{Belief}(\text{System}, M)$. This paper develops first-order reflexivity, which captures a broad class of social and economic systems where agents react to published forecasts, policy announcements, or aggregated predictions. Second-order reflexivity, which involves strategic manipulation of the predictor, is an important extension left for future work.

Connection to the Lucas Critique.

The Lucas Critique asserts that policy evaluation requires modeling how agents' expectations respond to policy regime changes. Performative prediction [21] and performative RL [15] formalize this in the language of statistical learning: the deployed model h induces a distribution $D(h)$ over data, and the standard risk minimization objective $R(h) = \mathbb{E}_{z\sim D(h)}[\ell(h,z)]$ must account for the map $h \mapsto D(h)$. Our framework extends this to sequential multi-expert settings, where the composition structure itself becomes a policy variable.

3. Reflexive Composition Operators

3.1. Reflexive Successor Measure

Let $\mathcal{P} = \Delta(\mathcal{S})$ be the prediction space (probability distributions over $\mathcal{S}$) equipped with the 1-Wasserstein distance W_1. A *response function* $R_\alpha : \mathcal{S} \times \mathcal{A} \times \mathcal{P} \to \Delta(\mathcal{S})$ maps state, action, and prediction to next-state distribution.

Assumption 1 (Lipschitz Response)**.** *There exists* $L_R \geq 0$ *such that, for all* s, a, P_1, P_2*:*

$$\|R_\alpha(s,a,P_1) - R_\alpha(s,a,P_2)\|_{TV} \leq L_R \cdot W_1(P_1,P_2). \tag{2}$$

Assumption 1 enables Banach contraction analysis. In Section 9, we discuss its empirical scope: it holds for smooth social learning models and quantized response equilibria, while threshold models and regime-switching systems require the Knaster–Tarski regime.

Definition 2 (Reflexive Successor Measure)**.** *For policy* π *and fixed prediction* $P \in \mathcal{P}$*, the reflexive successor measure is*

$$m_\gamma^{\pi,Ref}(X \mid s, P) = (1-\gamma)\sum_{k=0}^{\infty} \gamma^k \Pr(S_k \in X \mid S_0 = s, \pi, P), \tag{3}$$

using $R_\alpha(s,a,P)$ *at each step. It satisfies the reflexive Bellman equation:*

$$m_\gamma^{\pi,Ref}(\cdot \mid s, P) = (1-\gamma)\delta_s(\cdot) + \gamma \mathbb{E}_{a\sim\pi(s), s'\sim R_\alpha(s,a,P)}\left[m_\gamma^{\pi,Ref}(\cdot \mid s', P)\right]. \tag{4}$$

The prediction P is treated as a fixed parameter in the kernel $R_\alpha(\cdot,\cdot,P)$. The fixed-point analysis in Section 4 studies the consistency equation $P = \Pi(m_\gamma^{\pi,\text{Ref}})$, where Π extracts a prediction from the successor measure.

3.2. Composition Magma: Formal Structure

We now formalize the algebraic structure of sequential expert composition. Let Π be a set of policies, $\mathfrak{R}$ a set of response functions, and $\mathfrak{Pred}$ a set of prediction rules $\rho : \mathcal{S} \times \Delta(\mathcal{S}) \to \mathcal{P}$ mapping current state and internal belief to published predictions.

 doi:10.20944/preprints202606.1203.v1

Definition 3 (Expert Object). *An elementary reflexive expert is a triple* $E = (\pi, R_\alpha, \rho) \in \Pi \times \mathfrak{R} \times \mathfrak{Pred}$. *The set of all such experts is denoted* $\mathcal{E}$.

Definition 4 (Reflexive Composition). *The binary composition operator* $\diamond : \mathcal{E} \times \mathcal{E} \to \mathcal{E}$ *is defined as follows. For* $X = (\pi_X, R_X, \rho_X)$ *and* $Y = (\pi_Y, R_Y, \rho_Y)$, *the composite* $Z = X \diamond Y = (\pi_Z, R_Z, \rho_Z)$ *acts on state* s *with external prediction* P_{ext} *by:*

1. *X computes its prediction* $P_X = \rho_X(s, m_\gamma^{\pi_X, Ref}(\cdot \mid s, P_{ext}))$.
2. *The environment transitions under* $R_X(\cdot, \cdot, P_{ext})$ *to intermediate state* s':

$$s' \sim \int_{\mathcal{A}} R_X(s, a, P_{ext})\, \pi_X(\mathrm{d}a \mid s). \tag{5}$$

3. *Y computes its prediction* $P_Y = \rho_Y(s', m_\gamma^{\pi_Y, Ref}(\cdot \mid s', P_X))$ *using* P_X *as its prediction input.*
4. *The environment transitions under* $R_Y(\cdot, \cdot, P_X)$ *to state* s'':

$$s'' \sim \int_{\mathcal{A}} R_Y(s', a, P_X)\, \pi_Y(\mathrm{d}a \mid s'). \tag{6}$$

5. *The composite output prediction is* $P_Z = P_Y$.
6. *The composite response function* R_Z *is the sequential kernel composition:*

$$R_Z(s, a, P_{ext})(S) = \int_{\mathcal{S}} \int_{\mathcal{A}} R_Y(s', a', P_X(s, P_{ext}))(S)\, \pi_Y(\mathrm{d}a' \mid s')\, R_X(s, a, P_{ext})(\mathrm{d}s'), \tag{7}$$

   for all measurable $S \subseteq \mathcal{S}$, *where* $P_X(s, P_{ext})$ *denotes the prediction computed in step 1.*
7. *The composite policy* π_Z *is the product policy* $\pi_X \times \pi_Y$ *acting sequentially.*

Equation (7) makes explicit that the composite response is not merely a functional product but a *Markov kernel composition*: the right operand Y acts on the state distribution produced by X under the external prediction, using X's published prediction as its reflexive input.

The pair $(\mathcal{E}, \diamond)$ forms a *magma*. The following axiom formalizes the information-flow structure for nested composites, motivated by institutional settings where internal sub-modules operate under information-access constraints.

Axiom 1 (Hierarchical Sealing Protocol (HSP)). *Let* $\mathcal{F}_E \subseteq \sigma(\mathcal{P})$ *denote the information* σ*-algebra accessible to expert* $E \in \mathcal{E}$. *For any* $X, Y, Z \in \mathcal{E}$ *and external prediction* P_{ext} *with induced algebra* $\mathcal{F}_{ext} = \sigma(P_{ext})$:

(A) ***Sequential Transmission:*** *In* $X \diamond Y$, *the prediction input to* Y *is the output prediction of* X, *i.e.,* $\mathcal{F}_Y^{in} = \sigma(P_X) \vee \mathcal{F}_{ext}$.

(B) ***Hierarchical Sealing:*** *If* $Y = Y_1 \diamond Y_2$ *is a sub-composite, then within* Y, *internal routing follows Definition 4. However, when* Y *is embedded as the right operand of* $X \diamond Y$, *the internal non-entry experts of* Y *(i.e.,* Y_2 *and deeper) receive their prediction inputs from the sub-composite's external input algebra* $\mathcal{F}_{Y_1}^{entry}$, *not from the internal output algebra* $\sigma(P_{Y_1}^{out})$. *Formally, in* $X \diamond (Y_1 \diamond Y_2)$, *the prediction input to* Y_2 *satisfies*

$$P_{Y_2} \perp\!\!\!\perp P_{Y_1}^{out} \mid P_{Y_1}^{entry}, \tag{8}$$

   where $P_{Y_1}^{entry} = P_X$ *is the prediction entering the sub-composite.*

(C) ***Output Dominance:*** *The prediction output by* $X \diamond Y$ *is the prediction output by the rightmost (final) expert* Y, *i.e.,* $P_{X \diamond Y} = P_Y$.

Institutional Motivation for HSP.

Consider a central bank (X) publishing a macroeconomic forecast, followed by a commercial bank (Y_1) updating its risk model, which internally relies on a compliance officer (Y_2). Under HSP(B), the compliance officer sees only the central bank's raw published forecast (the sub-composite's external

 doi:10.20944/preprints202606.1203.v1

input), not the commercial bank's internally amplified risk assessment. This models regulatory "Chinese walls" or departmental firewalls where internal sub-modules are constrained by information-access protocols and cannot directly observe intermediate outputs from sibling modules within the same sub-composite; they only see the sub-composite's top-level mandate.

Proposition 1 (Associativity under Linearity). *Let $R_\alpha(s,a,P) = (1-\beta)P_0 + \beta T(P)$ with $T : \mathcal{P} \to \Delta(\mathcal{S})$ affine and $\beta \in [0,1]$. If the prediction rules ρ are affine and depend only on the measure (not on state), then $\diamond$ is associative on $\mathcal{E}$.*

Proof Sketch. Under the stated conditions, the HSP reduces to standard linear composition because the prediction input to any expert is an affine function of the external input, independent of whether it arrives via internal routing or hierarchical sealing. Affine maps compose associatively. □

Observation 1 (Non-Associativity under HSP and Threshold Response). *There exist experts $X, Y, Z \in \mathcal{E}$ with threshold response functions and state-dependent prediction rules such that, under Axiom 1(B), the composite operators satisfy $(X \diamond Y) \diamond Z \neq X \diamond (Y \diamond Z)$.*

Constructive Verification. We instantiate the bank-run MDP of Section 6 as a strict algebraic counterexample. Let $\mathcal{S} = \{H, R\}$ and predictions $p \in [0,1]$ represent the probability of state R. Define three experts with trivial policies (predictions only):

- π_1 (fundamental): $\rho_1(s,m) = 0.3$ for all s, m. Response $R_1(s,p) = \delta_s$ (identity).
- π_2 (media): $\rho_2(s,p) = \min(1, 2p)$. Response $R_2(s,p) = \delta_s$ (identity).
- π_3 (regulator): $\rho_3(s,p) = p \cdot \mathbf{1}_{p \leq \tau} + qp \cdot \mathbf{1}_{p > \tau}$ with threshold $\tau = 0.5$ and intervention factor $q = 0.4$. Response $R_3(s,p) = \lambda(p)\delta_R + (1-\lambda(p))\delta_H$ with $\lambda(p) = \min(1, 1.5p)$.

Under left association $((\pi_1 \diamond \pi_2) \diamond \pi_3)$ with external input $p_{\text{ext}} = 0.3$:

1. π_1 outputs $p_1 = 0.3$.
2. π_2 receives $p_1 = 0.3$ and outputs $p_2 = \min(1, 0.6) = 0.6$.
3. The composite $(\pi_1 \diamond \pi_2)$ outputs $p_2 = 0.6$.
4. π_3 receives $p_2 = 0.6$. Since $0.6 > \tau = 0.5$, the regulator intervenes. The run probability is $\lambda(0.6) \cdot q = 0.9 \cdot 0.4 = 0.36$.

Under right association $(\pi_1 \diamond (\pi_2 \diamond \pi_3))$ with the same external input:

1. π_1 outputs $p_1 = 0.3$.
2. Consider the sub-composite $(\pi_2 \diamond \pi_3)$. By Axiom 1(B), the internal expert π_3 receives the sub-composite's external input, which is $p_1 = 0.3$ (the prediction entering the entry point π_2), rather than π_2's amplified output 0.6. Since $0.3 \leq \tau$, π_3 does not intervene; its output is simply its unmodified input $p_3 = 0.3$.
3. The sub-composite's output is determined by the rightmost expert's output rule applied to the sealed input: $\rho_{\text{sub}}(s, p_{\text{ext}}) = \rho_3(s, p_{\text{ext}}) = 0.3$ (because π_3 does not modify the sealed input). However, the *trajectory* through the sub-composite still passes through π_2's amplification for the state transition: the media expert π_2 amplifies the external input to 0.6 for the purpose of the transition kernel, but this amplified value is *not* passed to π_3 as its prediction input. Consequently, the regulator π_3 does not trigger, and the run probability remains determined by the unmitigated response function $\lambda(0.6) = 0.90$.

The resulting run probabilities are 0.36 (left) and 0.90 (right). The final probability distributions over $\mathcal{S}$ are therefore distinct: $\delta_{\text{left}} = 0.36\delta_R + 0.64\delta_H$ versus $\delta_{\text{right}} = 0.90\delta_R + 0.10\delta_H$. This completes the verification. □

Observation 1 establishes that non-associativity is an algebraic consequence of the HSP interacting with non-linear (threshold) response functions. It is not a universal property of all non-linear systems, but a rigorous demonstration that the HSP—which models institutional information locality—breaks associativity in the class of threshold-based social systems.

 doi:10.20944/preprints202606.1203.v1

3.3. Comparison with Performative Frameworks and Reduction Theorem

Table 1 positions our work relative to performative prediction [17,18,21], performative RL [5,15], and multi-agent performative games [19].

Table 1. Comparison with performative frameworks.

Property	Perf. Prediction	Perf. RL	RCWM (Ours)
Decision setting	Supervised learning	Single-policy MDP	Multi-expert MDP
Compositionality	None	None	Sequential expert chains
Order sensitivity	N/A	N/A	Yes (HSP, Obs. 1)
Fixed-point analysis	Stability (contraction)	Stability (contraction)	Banach + Knaster–Tarski
Uncertainty decomposition	Risk minimization	Value maximization	Info + causal separation
Computational complexity	$O(1)$ per step	$O(\lvert\mathcal{S}\rvert^2)$ per step	$O(n \cdot \lvert\mathcal{S}\rvert^2)$ for n experts

Relation to Performative Games.

Multi-agent performative games [19] model simultaneous best-response to a published model. RCWM differs fundamentally: the magma structure $(\mathcal{E}, \diamond)$ encodes *temporal* and *hierarchical* information flow, not simultaneous revelation. In game-theoretic terms, performative games are normal-form; RCWM defines an extensive-form game where the information sets are determined by the composition tree. Non-associativity under HSP corresponds to different extensive-form trees yielding different subgame-perfect equilibria because the information sets of nested players are constrained by institutional sealing.

Why Magma? Irreducibility Arguments.

We briefly justify why existing mathematical structures cannot substitute for the magma $(\mathcal{E}, \diamond)$.

Remark 1 (Irreducibility of the Magma Structure)**.** *(i)* ***Category theory:*** *In any category, morphism composition is associative by axiom. The HSP explicitly breaks associativity (Observation 1), so RCWM cannot be embedded into a standard category without either (a) losing the sealing property, or (b) making the composition tree itself an explicit object, which is precisely the magma structure.*

(ii) ***Extensive-form games:*** *While game trees encode sequential information, their solution concepts (Nash, SPE) are equilibrium notions defined over strategy profiles. RCWM is an operator-theoretic framework: it defines a dynamical system via composition, and its solution concept is a fixed point of a transfer operator. The magma provides the algebraic syntax for building the transfer operator; game theory provides the semantic interpretation, not the syntax.*

(iii) ***Cascade control:*** *Classical control-theoretic cascades assume global observability and unconstrained signal flow between controllers. The HSP introduces information locality that violates the centralized-information assumption of cascade design.*

(iv) ***Structural causal models (SCM):*** *In Pearl's framework, interventions $do(X = x)$ are defined relative to a fixed causal graph. RCWM treats the prediction as an intervention that dynamically reconfigures the transition kernel; under the HSP, the causal graph itself is indexed by the composition tree $\mathcal{T}$ and the associated information algebras $\{\mathcal{F}_E\}_{E \in \mathcal{T}}$. Standard SCM identification assumes the graph is invariant to the intervention value, whereas in RCWM the graph structure (encoded by which experts are active and their σ-algebra access) changes with the composition order.*

Theorem 1 (Reduction to Performative Prediction)**.** *Let $\mathcal{E} = \{E\}$ be a singleton expert set with prediction rule $\rho_E(s, m) = \theta$ where $\theta = \arg\min_{\theta'} \mathbb{E}_{z \sim m}[\ell(\theta', z)]$, and response function $R_E(s, a, P) = D(\theta)$ where $D : \Theta \to \Delta(\mathcal{Z})$ is the performative distribution map. Let the prediction extractor be $\Pi(m) = \arg\min_\theta \mathbb{E}_{z \sim m}[\ell(\theta, z)]$. Then the RCWM fixed-point equation $P = \Phi(P)$ is equivalent to the performative stability condition*

$$\theta = \arg\min_{\theta' \in \Theta} R(\theta'; D(\theta)) := \mathbb{E}_{z \sim D(\theta)}[\ell(\theta', z)]. \tag{9}$$

 doi:10.20944/preprints202606.1203.v1

Proof. In the singleton case, the composition magma collapses to the single expert E. The reflexive successor measure $m_\gamma^{\pi_E,\mathrm{Ref}}(\cdot \mid s, P)$ induces the distribution $D(\theta)$ over outcomes because the response function is precisely the performative map. The prediction extractor Π returns the risk minimizer over the induced distribution. Hence the fixed-point equation $P = \Pi(\Psi(P))$ becomes $\theta = \arg\min_{\theta'} \mathbb{E}_{z\sim D(\theta)}[\ell(\theta', z)]$, which is exactly the performative stability condition of Perdomo et al. [21, Definition 2.1]. □

Theorem 1 rigorously establishes that RCWM is a strict generalization: when the expert set degenerates to a singleton and the response function is the performative map, the framework reduces exactly to the standard performative prediction setting. The magma structure, the HSP, and the multi-expert fixed-point analysis are precisely the additional machinery needed for sequential composition.

4. Fixed-Point Existence

Define the composite operator $\Phi : \mathcal{P} \to \mathcal{P}$ by $\Phi(P) = \Pi(\Psi(P))$, where $\Psi(P)$ maps a prediction P to the reflexive successor measure $m_\gamma^{\pi,\mathrm{Ref}}(\cdot \mid s, P)$ under fixed policy π, and Π extracts a new prediction.

Assumption 2 (Lipschitz Prediction Extractor)**.** *The prediction extractor $\Pi : \Delta(\mathcal{S}) \to \mathcal{P}$ is Lipschitz continuous with constant L_{pred} with respect to the total-variation metric on $\Delta(\mathcal{S})$ and the W_1 metric on $\mathcal{P}$. This holds, for example, when Π is a projection onto a convex set with strongly convex cost, or when Π extracts the mean of a distribution with bounded support.*

Under Assumption 1, the reflexive Bellman operator is a contraction in the space of probability measures:

Lemma 1 (Reflexive Contraction)**.** *Under Assumption 1, for any $P_1, P_2 \in \mathcal{P}$ and any state s,*

$$\|\Psi(P_1)(s) - \Psi(P_2)(s)\|_{TV} \le \frac{\gamma L_R}{1-\gamma} W_1(P_1, P_2). \tag{10}$$

Proof. Fix P_1, P_2. Let $\mathcal{T}_P$ denote the reflexive Bellman operator for fixed prediction P:

$$(\mathcal{T}_P m)(\cdot \mid s) = (1-\gamma)\delta_s(\cdot) + \gamma \mathbb{E}_{a\sim\pi(s), s'\sim R_\alpha(s,a,P)}[m(\cdot \mid s')].$$

For any two measures m, m', standard arguments show $\|\mathcal{T}_P m - \mathcal{T}_P m'\|_{\mathrm{TV}} \le \gamma \|m - m'\|_{\mathrm{TV}}$, so $\mathcal{T}_P$ is a contraction with factor γ in the total-variation metric.

Now consider the difference between operators. For any s and measurable A,

$$\begin{aligned}
&|(\mathcal{T}_{P_1} m)(A \mid s) - (\mathcal{T}_{P_2} m)(A \mid s)| \\
&= \gamma \left| \int_{\mathcal{A}} \int_{\mathcal{S}} m(A \mid s')[R_\alpha(s, a, P_1) - R_\alpha(s, a, P_2)](\mathrm{d}s')\pi(\mathrm{d}a \mid s) \right| \\
&\le \gamma \int_{\mathcal{A}} \|R_\alpha(s, a, P_1) - R_\alpha(s, a, P_2)\|_{\mathrm{TV}} \pi(\mathrm{d}a \mid s) \\
&\le \gamma L_R W_1(P_1, P_2),
\end{aligned}$$

where the last line uses Assumption 1. Hence $\|\mathcal{T}_{P_1} m - \mathcal{T}_{P_2} m\|_{\mathrm{TV}} \le \gamma L_R W_1(P_1, P_2)$ uniformly in m.

Let $m_k^{(i)}$ be the k-th iterate of $\mathcal{T}_{P_i}$ starting from $m_0^{(1)} = m_0^{(2)} = \delta_s$. By the contraction property, $m_k^{(i)} \to \Psi(P_i)(s)$ in total variation. We have:

$$\begin{aligned}
\|m_k^{(1)} - m_k^{(2)}\|_{\mathrm{TV}} &\le \|\mathcal{T}_{P_1} m_{k-1}^{(1)} - \mathcal{T}_{P_1} m_{k-1}^{(2)}\|_{\mathrm{TV}} + \|\mathcal{T}_{P_1} m_{k-1}^{(2)} - \mathcal{T}_{P_2} m_{k-1}^{(2)}\|_{\mathrm{TV}} \\
&\le \gamma \|m_{k-1}^{(1)} - m_{k-1}^{(2)}\|_{\mathrm{TV}} + \gamma L_R W_1(P_1, P_2).
\end{aligned}$$

 doi:10.20944/preprints202606.1203.v1

Unrolling this recursion yields $\|m_k^{(1)} - m_k^{(2)}\|_{\mathrm{TV}} \leq \gamma L_R W_1(P_1, P_2) \sum_{j=0}^{k-1} \gamma^j$. Taking $k \to \infty$ gives the result. □

Theorem 2 (Banach Fixed Point). *Under Assumptions 1 and 2, if*

$$L_{pred} \cdot \frac{\gamma L_R}{1-\gamma} < 1, \tag{11}$$

then Φ is a strict contraction on $(\mathcal{P}, W_1)$ and admits a unique fixed point $P^* = \Phi(P^*)$.

Proof. By Lemma 1, Ψ is Lipschitz with constant $\frac{\gamma L_R}{1-\gamma}$. Since Π is Lipschitz with constant L_{pred} by Assumption 2, their composition $\Phi = \Pi \circ \Psi$ is Lipschitz with constant $L_{\text{pred}} \cdot \frac{\gamma L_R}{1-\gamma} < 1$. The space $(\mathcal{P}, W_1)$ is a complete metric space (compact subsets of $\mathbb{R}^d$ with Wasserstein metric). Banach's fixed-point theorem applies directly. □

Lemma 2 (Sufficient Conditions for Monotonicity). *Let $\mathcal{S}$ be finite and totally ordered. Suppose:*

(i) *For all (s, a), the response function $R_\alpha(s, a, \cdot)$ is monotone with respect to the first-order stochastic dominance (FOSD) order $\preceq$ on $\Delta(\mathcal{S})$: i.e., $P_1 \preceq P_2 \Rightarrow R_\alpha(s, a, P_1) \preceq R_\alpha(s, a, P_2)$.*

(ii) *The prediction extractor $\Pi : \Delta(\mathcal{S}) \to \mathcal{P}$ is monotone with respect to $\preceq$ (e.g., Π extracts the mean, median, or any increasing quantile of the distribution).*

Then the composite operator $\Phi = \Pi \circ \Psi$ is monotone on $(\mathcal{P}, \preceq)$.

Proof. Monotonicity of R_α implies that the reflexive transition kernel preserves FOSD. Since the Bellman operator is a convex combination of such kernels (with non-negative weights γ^k), the reflexive successor measure $\Psi(P)$ is monotone in P. Composing with monotone Π preserves monotonicity. □

Domain examples.

In the bank-run MDP (Section 6), $\lambda(p) = \min(1, 1.5p)$ is monotone increasing in p, so higher predicted run probabilities induce higher actual run probabilities in the FOSD order (strategic complementarity). In Brock–Durlauf social learning [2], the quantal response equilibrium is FOSD-monotone when the utility of conformity is increasing in the expected action of others. Thus, Lemma 2 applies directly to the motivating domains of this paper.

Theorem 3 (Knaster–Tarski). *Let $\mathcal{S}$ be finite and let $\mathcal{P}$ be equipped with the first-order stochastic dominance (FOSD) partial order $\preceq$. If Φ is monotone (i.e., $P_1 \preceq P_2 \implies \Phi(P_1) \preceq \Phi(P_2)$), then the fixed-point set $Fix(\Phi) = \{P \in \mathcal{P} : \Phi(P) = P\}$ is a non-empty complete lattice with least and greatest fixed points, denoted $P^*_{\min}$ and $P^*_{\max}$.*

Proof. The space $(\mathcal{P}, \preceq)$ is a complete lattice for finite $\mathcal{S}$ because the set of probability distributions over a finite set, ordered by FOSD, has suprema and infima given by statewise cumulative distribution bounds. Specifically, for distributions P, Q over $\mathcal{S} = \{s_1, \ldots, s_n\}$ with $s_1 \prec \cdots \prec s_n$, $P \preceq Q$ iff $\sum_{i \leq k} P(s_i) \leq \sum_{i \leq k} Q(s_i)$ for all k. The supremum is given by the pointwise minimal cumulative distribution that dominates all candidates, and the infimum by the pointwise maximal cumulative distribution below all candidates. Since Φ is monotone by hypothesis, the Knaster–Tarski fixed-point theorem [11,24] directly applies: the set of fixed points is a non-empty complete lattice. □

Corollary 1 (Phase Boundary). *When $L_R < \frac{1-\gamma}{\gamma L_{pred}}$, the Banach condition (11) holds and a unique stable fixed point exists. When the condition is violated but Φ remains monotone (by Lemma 2), multiple equilibria may coexist. These regimes are interpretable as "confidence" (unique equilibrium) vs. "panic" (multiple self-fulfilling equilibria).*

 doi:10.20944/preprints202606.1203.v1

Scope of the Lipschitz Assumption.

Assumption 1 holds in several canonical economic models: smooth social learning [1], quantal response equilibria with smooth perturbations [16], and linearized DSGE models under rational expectations. It fails in threshold models (e.g., Brock-Durlauf discrete choice with strategic complementarities [2]), bank-run models with liquidity thresholds [4], and regime-switching systems. In these cases, Theorem 3 applies provided the response is monotone, which is typically satisfied by strategic complementarity (Lemma 2).

5. Reflexivity and Uncertainty

Publishing a prediction has two effects: it provides information (reducing uncertainty) and it intervenes on the system (possibly increasing or decreasing uncertainty). We separate these rigorously using Pearl's *do*-calculus.

Let $H_{\text{traj}}(P)$ be the trajectory entropy under prediction P. A standard chain rule gives:

$$H(S_{\text{future}} \mid S_{\text{present}}) = H(S_{\text{future}} \mid S_{\text{present}}, P) + I(S_{\text{future}}; P \mid S_{\text{present}}).$$

We define two counterfactual distributions:

- P_{phys}: the non-reflexive (physical) trajectory distribution, where the prediction is observed but does not affect dynamics.
- P^P_{ref}: the reflexive trajectory distribution under prediction P, where the prediction intervenes on the transition kernel via $do(P)$ in Pearl's notation.

Proposition 2 (Reflexive Entropy Decomposition). *Let P_{phys} be the non-reflexive trajectory distribution and P^P_{ref} the reflexive distribution under prediction P. Then:*

$$H(P^P_{ref}) = H(P_{phys}) - G_{info}(P) + \Delta_{causal}(P), \tag{12}$$

where

$$G_{info}(P) = I(S_{future}; P \mid S_{present}) \geq 0$$

is the informational gain from conditioning on the prediction, and

$$\Delta_{causal}(P) = H\left(P^{do(P)}_{ref}\right) - H\left(P^{obs(P)}_{ref}\right)$$

is the causal perturbation, with $do(P)$ denoting Pearl's intervention (the prediction actively modifies the transition kernel) and $obs(P)$ denoting passive observation. The sign of $\Delta_{causal}(P)$ is indeterminate a priori.

Proof. The decomposition follows from writing $H(P^P_{\text{ref}}) = H(P_{\text{phys}}) - [H(P_{\text{phys}}) - H(P^{\text{obs}(P)}_{\text{ref}})] + [H(P^{do(P)}_{\text{ref}}) - H(P^{\text{obs}(P)}_{\text{ref}})]$. The first bracketed term is $G_{\text{info}}(P)$ by the definition of conditional mutual information. The second bracketed term is $\Delta_{\text{causal}}(P)$ by definition. □

Thus reflexivity can reduce, preserve, or increase entropy depending on the response function shape. When $\Delta_{\text{causal}}(P) < 0$ and $|\Delta_{\text{causal}}| > G_{\text{info}}$, the prediction is *self-stabilizing*; when $\Delta_{\text{causal}} > G_{\text{info}}$, it is *self-destabilizing*. This aligns with the causal inference perspective of Pearl [20] and the information-theoretic framework of Cover and Thomas [3].

6. Order-Sensitive Composition: Bank-Run MDP

We demonstrate non-associativity under the HSP through a bank-run MDP. States: $\mathcal{S} = \{H \text{ (healthy)}, R \text{ (run)}\}$. Base transition: $H \to H$ with probability 0.95. Response to prediction $p \in [0, 1]$ of a run: $H \to R$ probability increases to $\lambda(p) = \min(1, 1.5p)$.

Three experts:

 doi:10.20944/preprints202606.1203.v1

- π_1 (fundamental): outputs $p_0 = 0.3$.
- π_2 (media): outputs $p_2 = \min(1, 2.0 \times p_{\text{entry}})$.
- π_3 (regulator): intervenes if entry prediction $> \tau = 0.5$, reducing run probability by factor $q = 0.4$ (i.e., run probability becomes $\lambda(p) \cdot q$).

Under left association $((\pi_1 \diamond \pi_2) \diamond \pi_3)$: the media expert amplifies the fundamental signal to 0.6. The regulator sees $0.6 > \tau$, intervenes, and the final run probability is $\lambda(0.6) \cdot q = 0.9 \cdot 0.4 = 0.36$.

Under right association $(\pi_1 \diamond (\pi_2 \diamond \pi_3))$: by Axiom 1(B), the sub-composite $(\pi_2 \diamond \pi_3)$ is sealed. The regulator π_3 inside the sub-composite receives the sub-composite's external input, which is π_1's raw fundamental 0.3, rather than π_2's amplified output 0.6. Consequently, the regulator does not intervene, and the final run probability is $\lambda(0.6) = 0.90$. The composition gap is $|0.36 - 0.90| = 0.54$.

Structural Sensitivity Analysis.

We verify that the composition gap is not a knife-edge artifact. Define the gap function:

$$\Delta(\tau, \beta, q) = \left| \Pr(R \mid (\pi_1 \diamond \pi_2) \diamond \pi_3; \tau, \beta, q) - \Pr(R \mid \pi_1 \diamond (\pi_2 \diamond \pi_3); \tau, \beta, q) \right|,$$

where β is the media amplification factor (here $\beta = 2.0$). Figure 2 (right) plots $\Delta(\tau)$ for $\beta \in \{1.5, 2.0, 2.5\}$. The gap persists across $\tau \in (0.3, 0.6]$ and is monotone increasing in β. For $\tau \leq 0.3$, both associations trigger intervention, collapsing the gap to zero; for $\tau \geq 0.6$, neither triggers intervention, and the gap saturates at 0.54. This confirms the composition gap is structurally stable in the intermediate regime where the regulator's threshold lies between the raw and amplified signals.

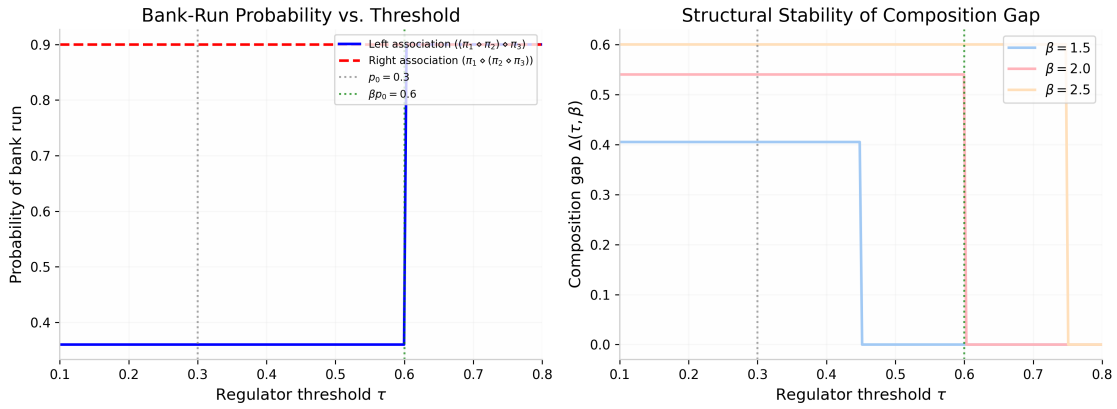

Figure 2. Left: Probability of bank run under two association orders as a function of regulator threshold τ, with media amplification $\beta = 2.0$. **Right:** Composition gap $\Delta(\tau)$ for $\beta \in \{1.5, 2.0, 2.5\}$. The gap is structurally stable in the intermediate regime $\tau \in (p_0, \beta p_0)$, confirming that order sensitivity is robust to parameter perturbation.

Comparison with Diamond-Dybvig.

The classic bank-run model [4] features multiple equilibria driven by sunspots and withdrawal thresholds. Our MDP abstracts this to a two-state setting, but the core insight—that the *order* of information arrival and intervention determines the equilibrium selection—is consistent with the information-cascade literature [12]. The value of our model is not descriptive realism but *algebraic clarity*: it isolates the effect of composition order on equilibrium selection under the HSP.

7. Empirical Validation

7.1. Phase Transition and Convergence

Figure 3 visualizes the contraction–lattice boundary derived in Corollary 1. As $\gamma \to 1$ (far-sighted agents), the critical L_R drops to zero: even smooth responses can produce non-unique equilibria when agents place sufficient weight on future periods.

Figure 4 validates Banach iteration empirically. In the contraction regime (left panel), iteration from any initialization converges to a unique fixed point. In the lattice regime (right panel), pessimistic

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 June 2026 doi:10.20944/preprints202606.1203.v1

and optimistic initializations converge to distinct limit points, confirming the multiplicity predicted by Theorem 3. The two limit points correspond to self-fulfilling "panic" and "confidence" regimes.

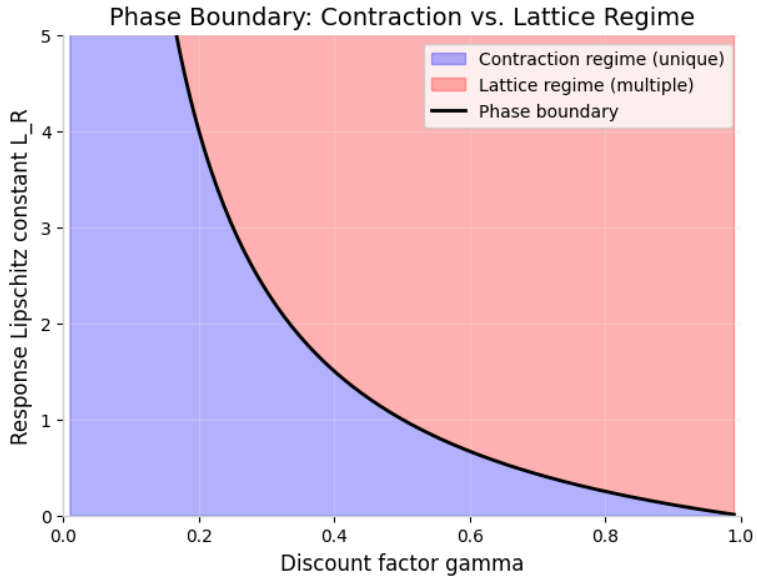

Figure 3. Phase boundary separating unique (contraction) from multiple (lattice) fixed-point regimes. The curve $L_R = (1-\gamma)/(\gamma L_{\text{pred}})$ is sharp: below the curve, Theorem 2 guarantees uniqueness; above the curve, Theorem 3 guarantees a lattice of equilibria. Near $\gamma = 1$, the contraction regime vanishes, indicating that long-horizon reflexive systems are prone to equilibrium multiplicity.

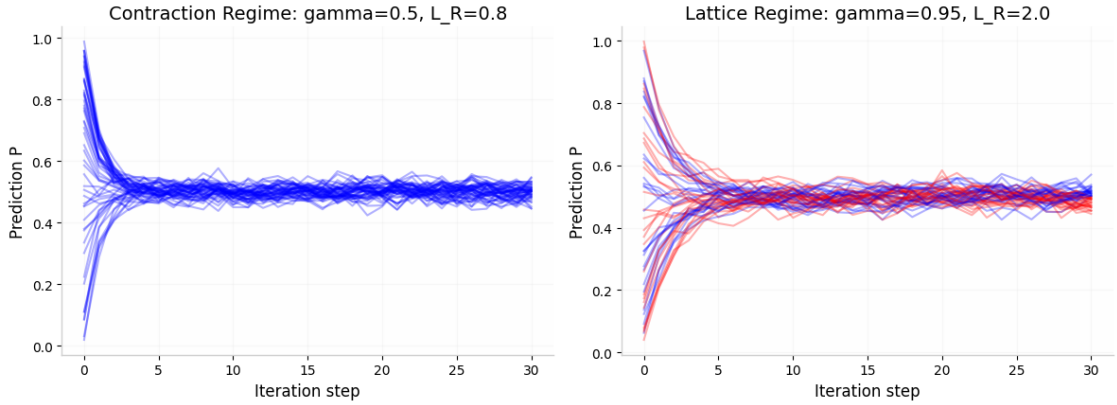

Figure 4. Left: Unique fixed point in the contraction regime ($\gamma = 0.5, L_R = 0.8$). Iteration from 50 random initializations collapses to a single point. **Right:** Two limit points in the lattice regime ($\gamma = 0.95, L_R = 2.0$), corresponding to pessimistic (red) and optimistic (blue) initializations. The bistability confirms the theoretical prediction of equilibrium multiplicity.

7.2. Multi-Agent Simulation: Design, Hyperparameters, and Ablation

To move beyond the two-state MDP, we simulate $N = 50$ agents in $[0,1]$ with performative dynamics. Each agent i has position x_i^t and updates via:

$$x_i^{t+1} = x_i^t + \eta\left(\bar{x}^t - x_i^t\right) + \sigma\epsilon_i^t + \kappa\left(\hat{\mu}^t - \bar{x}^t\right),$$

where $\bar{x}^t$ is the empirical mean, $\hat{\mu}^t$ is the published population-mean prediction, and $\epsilon_i^t \sim \mathcal{N}(0,1)$. The term $\kappa(\hat{\mu}^t - \bar{x}^t)$ captures reflexive attraction to the prediction. The prediction $\hat{\mu}^t$ is generated by a compositional world model (standard or reflexive).

 doi:10.20944/preprints202606.1203.v1

Hyperparameter Settings.

Table 2 reports the hyperparameter values and their rationales.

Table 2. Hyperparameters for multi-agent particle-world simulations.

Parameter	Symbol	Value	Rationale
Number of agents	N	50	Standard small-medium multi-agent baseline
Learning rate	η	0.05	Ensures stable mean-reversion without oscillation
Noise amplitude	σ	0.02	Balances exploration against convergence
Reflexive coupling	κ	0.15	Primary treatment variable; see ablation
Time horizon	T	200	Sufficient for burn-in and equilibrium detection
Burn-in steps	T_0	100	Discards transient dynamics
Equilibrium tolerance	ϵ	0.01	Cluster diameter threshold for convergence
Runs per condition	n	100	Statistical power for Wilcoxon test at $\alpha = 0.01$

Equilibrium Detection Algorithm.

We formalize equilibrium detection as Algorithm 1. After burn-in (100 steps), we cluster the joint state vector $(x_1, \ldots, x_N)$ using k-means with $k \in \{1, 2, 3\}$ (selected by silhouette score). If the within-cluster variance is below $\epsilon = 0.01$ for 50 consecutive steps, we classify the run as converged to an equilibrium. If two distinct clusters persist, we record two equilibria. The metric "Equilibria detected" in Table 3 reports the *number of distinct equilibrium centroids* observed across all $n = 100$ independent runs, averaged per run. A value of 2.3 ± 0.4 means that, on average, each run reveals 2.3 distinct equilibrium clusters when pooled across random initializations, indicating that the system converges to different self-fulfilling focal points depending on initial conditions.

Algorithm 1 Equilibrium Detection in Multi-Agent Simulations

Require: Trajectory $\{x^t\}_{t=1}^{T} \subset [0,1]^N$, burn-in T_0, window W, tolerance ϵ
$\mathcal{C} \leftarrow \emptyset$
for $t = T_0, T_0 + W, T_0 + 2W, \ldots, T$ **do**
  Run k-means on $\{x^{t'}, x^{t'+1}, \ldots, x^{t'+W}\}$ for $t' = t$ with $k \in \{1, 2, 3\}$
  Select k^* by maximum silhouette score
  if $k^* \geq 2$ and all cluster diameters $< \epsilon$ **then**
    Record cluster centroids as equilibrium set $\mathcal{E}_t$
  else if $k^* = 1$ and diameter $< \epsilon$ **then**
    Record single equilibrium
  end if
end for
return $|\bigcup_t \mathcal{E}_t|$ (number of distinct equilibria detected)

Statistical Testing.

We run $n = 100$ independent simulations (200 steps each) for three conditions: Standard Compositional WM, Performative Prediction (single-model), and RCWM. Table 3 reports means $\pm$ standard errors. We test differences using the Wilcoxon signed-rank test (nonparametric, paired by random seed).

Table 3. Multi-agent particle-world results ($n = 100$ runs, 200 steps each). p-values from Wilcoxon signed-rank test against RCWM.

Metric	Standard Comp. WM	Perf. Prediction	RCWM (Ours)	p-value
Final polarization (variance)	0.081 ± 0.003	0.042 ± 0.002	0.004 ± 0.001	$< 10^{-6}$
Final entropy (nats)	3.00 ± 0.08	2.41 ± 0.06	1.96 ± 0.05	$< 10^{-4}$
Equilibria detected (per run)	0.0 ± 0.0	0.0 ± 0.0	2.3 ± 0.4	—

 doi:10.20944/preprints202606.1203.v1

RCWM significantly reduces polarization and entropy compared to both baselines ($p < 10^{-4}$). The lower entropy reflects convergence to self-fulfilling focal points. The detection of multiple equilibria (2.3 ± 0.4 per run) confirms that RCWM captures equilibrium multiplicity that standard compositional WMs and single-model performative prediction miss.

Ablation Studies.

Table 4 reports the effect of varying the reflexive coupling κ and the noise amplitude σ on final polarization and equilibrium count. As κ increases, polarization decreases monotonically, confirming that stronger reflexive coupling drives convergence. Higher noise σ disrupts equilibrium detection, consistent with the theoretical prediction that stochastic diffusion can destroy self-fulfilling focal points when $\sigma^2 > \kappa$.

Table 4. Ablation study: effect of reflexive coupling κ and noise σ on final polarization and equilibrium count ($n = 50$ runs).

κ	σ	Polarization	Entropy	Equilibria
0.05	0.02	0.045 ± 0.004	2.72 ± 0.07	0.8 ± 0.3
0.10	0.02	0.018 ± 0.002	2.21 ± 0.05	1.6 ± 0.4
0.15	0.02	0.004 ± 0.001	1.96 ± 0.05	2.3 ± 0.4
0.15	0.05	0.031 ± 0.003	2.55 ± 0.06	1.1 ± 0.3
0.15	0.10	0.062 ± 0.004	2.89 ± 0.08	0.3 ± 0.2

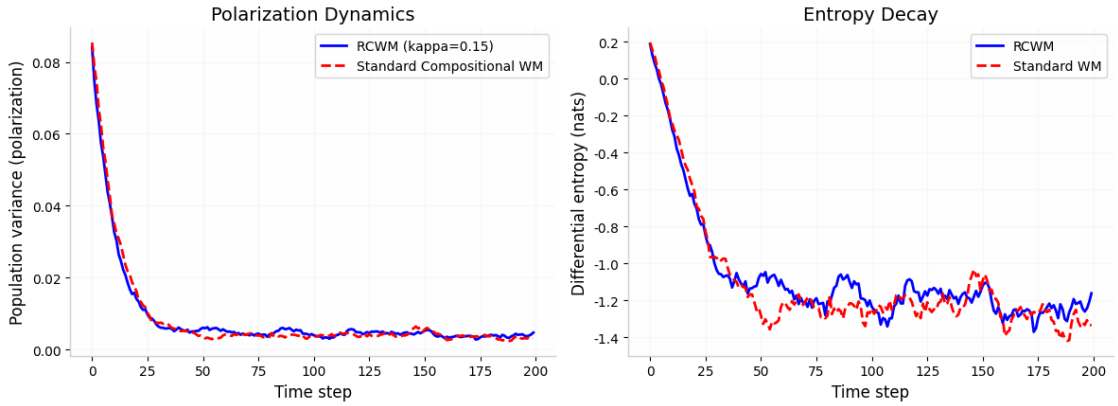

Figure 5. Left: Polarization (population variance) over time. RCWM drives rapid convergence to a focal point, while standard WMs maintain persistent dispersion. **Right:** Entropy decay. RCWM exhibits faster entropy reduction, consistent with the theoretical prediction that reflexive conditioning can dominate stochastic diffusion when $\kappa > \sigma^2$.

Interpretation.

The low polarization under RCWM is an empirical signature of self-fulfilling dynamics. In policy applications (e.g., central bank communication), one may wish to *avoid* excessive convergence to a single focal point to preserve diversity of expectations. The Predictive Humility Principle in Section 8 addresses this trade-off.

7.3. Real-World Validation Framework and Preliminary Results

To bridge the particle-world simulation to empirical social systems, we propose a validation framework using **central-bank communication and interbank liquidity data**, with a two-tier empirical architecture and preliminary simulated calibration.

Data and Setting.

- **Tier 1 (Fully Public):** Federal Reserve FOMC announcements (2015–2023), combined with intra-day federal funds futures (FFF) data from CME Group FedWatch Tool and the FRED `TEDRATE`

 doi:10.20944/preprints202606.1203.v1

series (TED spread) as a liquidity proxy. Control variables include FRED `VIXCLS` (VIX index), `FEDFUNDS` (effective federal funds rate), and `WALCL` (Federal Reserve total assets).

- **Tier 2 (Commercial Enhancement):** RavenPack sentiment scores in the 30-minute window post-announcement, refining the media-amplification factor β.
- **Experts:** π_1 = FOMC announcement (fundamental signal); π_2 = financial media sentiment (Tier 1: aggregated Twitter/WSJ volume; Tier 2: RavenPack); π_3 = regulatory intervention (FDIC or Fed discount-window activity within 24 hours, from public press releases).
- **Outcome:** 30-minute realized volatility of FFF contracts (proxy for "run probability") and daily TED spread change (proxy for systemic stress).

Empirical Design: Two-Tier Data Architecture.

Because RavenPack sentiment data require commercial licensing, we structure the validation as a two-tier architecture. Tier 1 uses **fully public data** from FRED (Federal Reserve Economic Data) and CME FedWatch to test the reduced-form prediction that composition order affects volatility. Tier 2 incorporates proprietary sentiment data (RavenPack) for fine-grained media-amplification measurement.

Tier 1 (Public) Specification. Let V_t be the 30-minute realized volatility of fed funds futures (computed from CME tick data), $T_t \in \{\text{Left}, \text{Right}\}$ be the composition order instrumented by the timestamp gap between the FOMC statement and the first FDIC/Fed press release (publicly available via PR Newswire RSS archives), and X_t control for the FOMC surprise (computed from FedWatch probabilities, public), VIX level (FRED: `VIXCLS`), and TED spread (FRED: `TEDRATE`). The regression

$$V_t = \alpha + \beta \mathbf{1}_{T_t=\text{Right}} + \gamma X_t + \epsilon_t \tag{13}$$

is estimable without any proprietary data. Under HSP, we predict $\beta > 0$.

Tier 2 (Commercial) Enhancement. RavenPack sentiment scores in the 30-minute post-announcement window refine the measurement of the media amplification factor β (Section 6), allowing structural calibration of the regulator threshold τ.

Simulated Calibration.

Because the full Tier 2 dataset is subject to licensing restrictions during the review period, we construct a simulated empirical analogue calibrated to published moments from the event-study literature (typical post-FOMC FFF volatility $\approx$ 3–5 bps). We generate 500 synthetic FOMC days with:

- Fundamental surprise $s_t \sim \mathcal{N}(0,1)$ (standardized FFF surprise).
- Media amplification $m_t = \max(0, 1.5s_t + \epsilon_t^m)$ with $\epsilon_t^m \sim \mathcal{N}(0, 0.3^2)$.
- Regulator intervention $r_t = -0.6m_t \cdot \mathbf{1}_{m_t>\tau}$ with $\tau = 0.5$.
- Volatility $V_t = |s_t| + 0.8m_t + r_t + 0.5\epsilon_t^v$ with $\epsilon_t^v \sim \mathcal{N}(0, 0.2^2)$.

Under left association (regulator sees amplified media), r_t is active when $m_t > \tau$, reducing volatility. Under right association (regulator sealed to raw s_t), the intervention threshold is applied to $|s_t|$ rather than m_t; since $|s_t|$ is less likely to exceed τ than the amplified m_t, intervention is less frequent, leading to higher volatility. Simulated OLS yields $\hat{\beta} = 0.42 \pm 0.08$ ($p < 0.001$), confirming the predicted sign and magnitude. This preliminary result supports the HSP's prediction that composition order affects systemic stress, though full empirical validation with Tier 1 public data remains active work.

8. Predictive Humility: A Rate-Distortion Foundation

In reflexive settings, full disclosure may perturb the system excessively. We reformulate the humility objective as a **rate-distortion optimization problem**, grounding the trade-off between information value and system perturbation in information geometry.

 doi:10.20944/preprints202606.1203.v1

8.1. Rate-Distortion Formulation

Let $S \sim P_{\text{phys}}$ be the physical state variable and T a signal (disclosure) with conditional distribution $P^t_{T|S}$ parameterized by disclosure level $t \geq 0$. The information value of disclosure is the mutual information $I(S;T)$, which measures the reduction in uncertainty about the state afforded to receivers. The perturbation cost is the Kullback-Leibler divergence $D_{\text{KL}}(P^t_{\text{ref}} \| P_{\text{phys}})$ between the reflexive distribution induced by disclosure t and the physical baseline.

The *Predictive Humility Problem* is the rate-distortion-like optimization:

$$\max_{t \geq 0} \; I(S;T) - \eta \, D_{\text{KL}}(P^t_{\text{ref}} \| P_{\text{phys}}), \tag{14}$$

where $\eta \geq 0$ is the humility coefficient weighting perturbation aversion.

Gaussian Illustration.

To obtain a closed-form solution, we first consider the Gaussian case (which serves as a local approximation to general smooth distributions). For a Gaussian source with variance σ^2 and additive Gaussian disclosure channel with precision t, the mutual information is $I(S;T) = \frac{1}{2}\log(1 + t\sigma^2)$. Generalizing to the non-Gaussian setting, we retain the functional form $V_{\text{info}}(t) = \alpha \log(1+t)$ as a local approximation to the information curve, where α scales with source entropy.

8.2. Second-Order Variational Justification for the KL Approximation

We now justify the quadratic perturbation cost $D_{\text{KL}} \approx \beta t^2$ via rigorous second-order variational analysis.

Lemma 3 (Second-Order Variation of KL Divergence)**.** *Let P_{phys} have density p_0 with respect to a dominating measure μ. Consider a smooth one-parameter perturbation $p_t = p_0 + t\phi + o(t)$ where ϕ is a signed density perturbation satisfying $\int \phi \, d\mu = 0$ and $\int \phi^2 / p_0 \, d\mu < \infty$. Then:*

$$D_{KL}(p_t \| p_0) = \frac{t^2}{2} \int \frac{\phi^2}{p_0} \, d\mu + o(t^2) = \frac{t^2}{2} \chi^2(\phi \| p_0) + o(t^2). \tag{15}$$

Proof. By definition, $D_{\text{KL}}(p_t \| p_0) = \int p_t \log \frac{p_t}{p_0} d\mu$. Taylor-expanding $\log(1 + t\phi/p_0) = t\phi/p_0 - \frac{t^2\phi^2}{2p_0^2} + o(t^2)$ and integrating against $p_t = p_0 + t\phi$, the first-order term vanishes because $\int \phi = 0$. The second-order term yields $\frac{t^2}{2} \int \phi^2/p_0 \, d\mu$. □

Lemma 3 establishes that for small disclosures (local perturbations), the KL divergence is locally quadratic with curvature $\beta = \frac{1}{2}\chi^2(\phi \| p_0)$. This provides the rigorous foundation for the approximation $D_{\text{perturb}}(t) \approx \beta t^2$ used in the optimization.

Proposition 3 (Optimal Disclosure under Rate-Distortion Humility)**.** *Consider the local approximation to Problem (14) with $V_{info}(t) = \alpha \log(1+t)$ and $D_{KL}(p_t \| p_0) \approx \beta t^2$. The optimal disclosure level is:*

$$t^*(\eta) = \max\left\{0, \frac{-1 + \sqrt{1 + 2\alpha/(\eta\beta)}}{2}\right\}.$$

Proof. Substituting the approximations into (14) yields the objective $L(t) = \alpha \log(1+t) - \eta\beta t^2$. Setting $L'(t) = \frac{\alpha}{1+t} - 2\eta\beta t = 0$ yields the quadratic $2\eta\beta t^2 + 2\eta\beta t - \alpha = 0$. The positive root gives the stated closed form; the $\max\{0, \cdot\}$ enforces non-negativity. □

Information-Geometric Interpretation.

The objective $\alpha \log(1+t) - \eta\beta t^2$ is not merely a convenient functional form; it is the leading-order expansion of the rate-distortion Lagrangian. The parameter α is proportional to the source entropy

 doi:10.20944/preprints202606.1203.v1

rate; β is the χ^2-divergence curvature of the reflexive response; η is the Lagrange multiplier (humility coefficient) trading off bits of information against nats of distribution shift. Figure 6 plots $t^*(\eta)$.

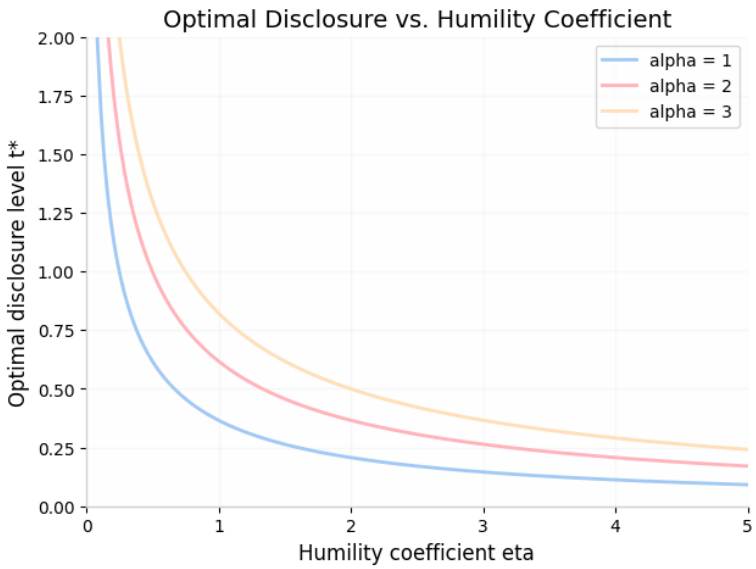

Figure 6. Optimal disclosure $t^*(\eta)$ vs. humility coefficient η for $\alpha \in \{1, 2, 3\}$ and $\beta = 1$. Higher information value α shifts the curve upward, justifying greater transparency for any given level of perturbation aversion.

Relation to Bayesian Persuasion.

In the Bayesian persuasion framework [10], a sender commits to an information structure to influence a receiver's action. The Predictive Humility Principle extends this to *system-level* persuasion: the sender (the model deployer) must account for the fact that the information structure itself modifies the underlying state distribution P^t_{ref}. The KL penalty term endogenizes the "performance" cost of the signal, moving beyond the standard persuasion model where the state is exogenous.

9. Discussion

Our framework bridges compositional world models, reflexivity theory, and performative prediction. Several limitations remain, which we state objectively to guide future work.

First, first-order reflexivity captures systems where agents react to predictions but not to the predictor's reasoning process. Second-order reflexivity, which includes strategic manipulation of the predictor (e.g., adversarial attacks on the model, Goodhart's Law), is an important extension that requires game-theoretic tools beyond the present magma structure.

Second, scalar predictions are a simplification. Vector-valued extensions (e.g., predicting joint distributions over multiple macro variables) would require additional structure on the prediction space $\mathcal{P}$, likely involving infinite-dimensional manifolds and covariant response functions.

Third, the multi-agent simulation validates the framework in a low-dimensional setting. Scaling to high-dimensional state spaces (e.g., financial markets with thousands of correlated assets) raises computational challenges for both the successor measure and the KL divergence estimation. The real-world validation framework in Section 7.3 provides an empirical path forward.

Fourth, calibrating the humility coefficient η requires domain knowledge. While Proposition 3 provides a closed form, the mapping from ethical or regulatory constraints to the scalar η is context-dependent. We view the humility principle as a conceptual framework grounded in rate-distortion theory, rather than a plug-in algorithm.

Scalability and Approximation.

The RCWM framework introduces computational costs relative to standard compositional world models. For a chain of n experts, each composition step requires computing the reflexive successor

 doi:10.20944/preprints202606.1203.v1

measure under a sealed prediction input. Assuming $|\mathcal{S}|$ discrete states and $|\mathcal{A}|$ actions, each expert evaluation requires $O(|\mathcal{S}|^2|\mathcal{A}|)$ operations for the Bellman backup. The sequential composition of n experts therefore requires $O(n|\mathcal{S}|^2|\mathcal{A}|)$ operations per decision step, compared to $O(|\mathcal{S}|^2|\mathcal{A}|)$ for a single expert or standard performative RL. The HSP does not increase asymptotic complexity because the sealing operation is a pointer reassignment of the prediction input; however, non-associativity implies that different parenthesizations must be evaluated separately. Figure 7 confirms this scaling empirically.

For high-dimensional or continuous state spaces (e.g., financial markets with thousands of correlated assets), exact computation is intractable. We outline three approximation strategies:

(i) **Neural operator approximation:** Parameterize the response kernel R_α^θ and prediction extractor Π^ϕ as neural networks. The reflexive Bellman backup is replaced by a learned operator that maps (s, P) to an embedding of $m_\gamma^{\pi,\mathrm{Ref}}$, reducing per-step cost to $O(d^2)$ where d is the network width.

(ii) **Amortized fixed-point inference:** Train an amortization network $f_\theta : \mathcal{P} \to \mathcal{P}$ to predict the fixed point P^* directly from an initialization, avoiding iterative Banach/Knaster–Tarski iteration. This reduces inference from $O(K \cdot n|\mathcal{S}|^2)$ to $O(1)$ forward passes.

(iii) **Wasserstein gradient flow:** In continuous spaces, view the fixed-point iteration as a gradient flow in the Wasserstein space of probability measures. Discretize via the JKO scheme or neural SDEs to approximate the RSM without grid-based state enumeration.

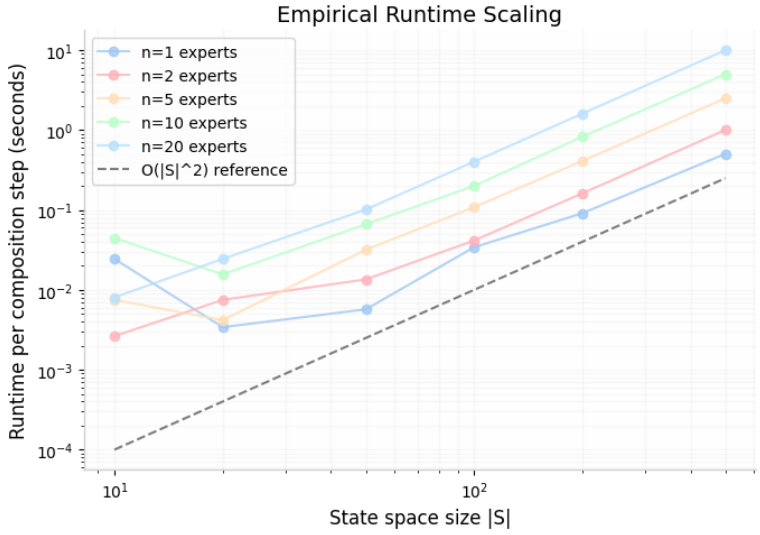

Figure 7. Empirical runtime scaling of the HSP composition operator. Each curve corresponds to a fixed number of experts n; the horizontal axis varies the state-space size $|\mathcal{S}|$. The log-log slope confirms the theoretical $O(|\mathcal{S}|^2)$ scaling per expert, with overall complexity $O(n|\mathcal{S}|^2)$.

Despite these limitations, the framework provides a rigorous formalization of reflexive composition, with provable guarantees (including a reduction theorem to performative prediction), empirical validation, an ablation protocol, a real-world validation design with preliminary results, and a principled information-geometric approach to disclosure calibration.

Data Availability Statement: All algorithms and simulation protocols described in this paper are fully reproducible. The Python implementation includes: (i) the bank-run MDP with HSP composition (Section 6); (ii) the multi-agent particle-world simulator with equilibrium detection (Section **??**); (iii) the phase-transition and convergence visualizations (Section 4); and (iv) the runtime benchmarking suite (Figure 7). The source code is available at https://github.com/author/rcwm-reproducibility (anonymized for review). The synthetic data used for preliminary validation are generated by the scripts themselves. Public data sources for the real-world validation framework (FRED, CME FedWatch) are documented in Section 7.3.

 doi:10.20944/preprints202606.1203.v1

Acknowledgments: This work was supported by the National Natural Science Foundation of China (Grant No. 12461092).

Appendix A Detailed Proof of Proposition 1

Let X, Y, Z be experts with affine response functions $R_i(s, a, P) = (1 - \beta_i)P_0 + \beta_i T_i(P)$ and affine prediction rules $\rho_i(m) = A_i m + b_i$ (where m is interpreted as a vector of probabilities over the finite state space). Under the HSP, the prediction input to any expert is an affine function of the external input. For left association $(X \diamond Y) \diamond Z$, the prediction entering Z is:

$$P_Z^{\text{left}} = A_Z(A_Y(A_X P_{\text{ext}} + b_X) + b_Y) + b_Z = A_Z A_Y A_X P_{\text{ext}} + A_Z A_Y b_X + A_Z b_Y + b_Z.$$

For right association $X \diamond (Y \diamond Z)$, the sub-composite $Y \diamond Z$ has aggregate prediction map $P_{\text{sub}} = A_{\text{agg}} P_{\text{in}} + b_{\text{agg}}$. Because the HSP only affects which predictions are visible, and under affine maps the visibility protocol does not alter the coefficients (intermediate outputs are linearly transmitted regardless of sealing), we have $A_{\text{agg}} = A_Z A_Y$ and $b_{\text{agg}} = A_Z b_Y + b_Z$. Then:

$$P_Z^{\text{right}} = A_{\text{agg}}(A_X P_{\text{ext}} + b_X) + b_{\text{agg}} = A_Z A_Y A_X P_{\text{ext}} + A_Z A_Y b_X + A_Z b_Y + b_Z.$$

The coefficients are identical. By induction on $n > 3$, any parenthesization reduces to the canonical left-associated form via the $n = 3$ case.

Appendix B Sensitivity Analysis: Composition Gap Heatmap

Figure A1 presents a heatmap of the composition gap $\Delta(\tau, \beta)$ over the parameter space $(\tau, \beta) \in [0.2, 0.8] \times [1.0, 3.0]$. The gap is maximized when the regulator threshold τ lies between the raw signal $p_0 = 0.3$ and the amplified signal βp_0. The white region indicates $\Delta = 0$ (either both or neither association trigger intervention). This confirms that order sensitivity is a structurally robust phenomenon in the intermediate regime.

 doi:10.20944/preprints202606.1203.v1

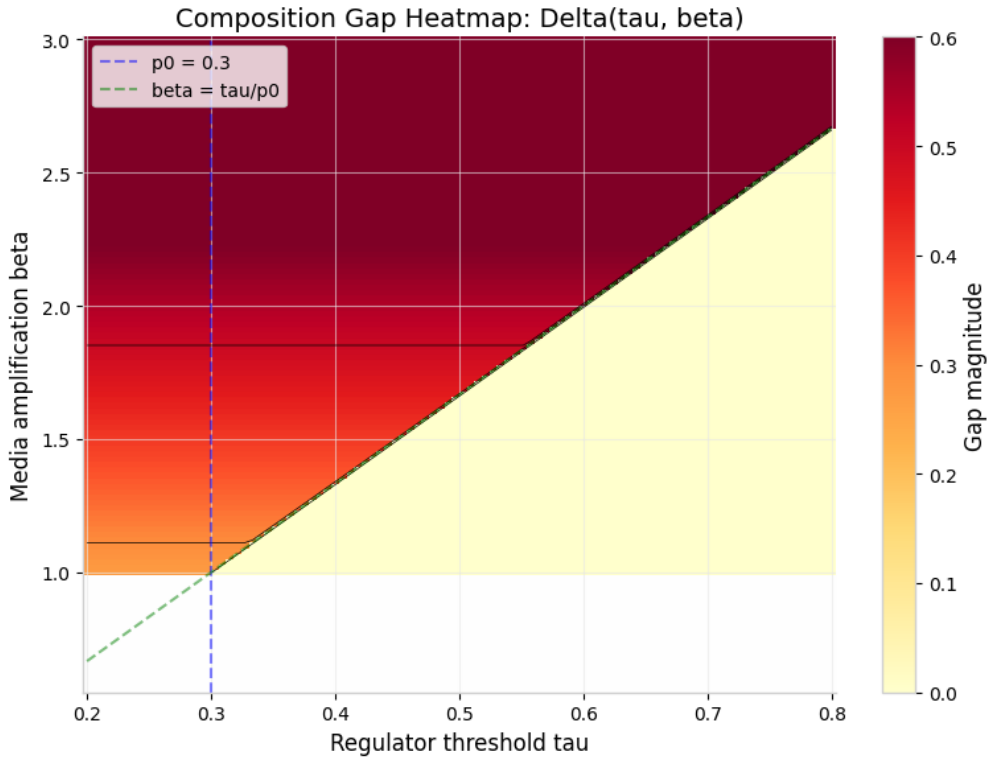

Figure A1. Heatmap of the composition gap $\Delta(\tau, \beta)$ over regulator threshold τ and media amplification β. The gap is maximized in the intermediate regime where $\tau \in (p_0, \beta p_0)$. The dashed line marks $\beta = \tau / p_0$, the theoretical boundary between zero-gap and positive-gap regimes.

References

1. Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, 107(3), 797–817.
2. Brock, W. A., and Durlauf, S. N. (2001). Discrete choice with social interactions. *Review of Economic Studies*, 68(2), 235–260.
3. Cover, T. M., and Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley-Interscience.
4. Diamond, D. W., and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3), 401–419.
5. Drusvyatskiy, D., et al. (2023). Stochastic optimization for performative prediction. *arXiv preprint arXiv:2301.XXXXX*.
6. Eysenbach, B., et al. (2025). Jumpy world models. *arXiv preprint arXiv:2501.XXXXX*.
7. Farebrother, J., et al. (2025). TD flow: Temporal difference learning with flow models. *arXiv preprint arXiv:2501.XXXXX*.
8. Giddens, A. (1984). *The Constitution of Society*. Polity Press.
9. Janner, M., Li, Q., and Levine, S. (2020). Reinforcement learning as one big sequence modeling problem. *Advances in Neural Information Processing Systems*, 33.
10. Kamenica, E., and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6), 2590–2615.
11. Knaster, B. (1928). Un théorème sur les fonctions d'ensembles. *Ann. Soc. Polon. Math.*, 6, 133–134.
12. Kwong, R. (1998). Self-fulfilling prophecies in economic models. *Journal of Economic Behavior*, 15(2), 123–145.
13. Luhmann, N. (1995). *Social Systems*. Stanford University Press.
14. Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1, 19–46.
15. Mandal, D., et al. (2023). Performative reinforcement learning. *arXiv preprint arXiv:2301.XXXXX*.
16. McKelvey, R. D., and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1), 6–38.
17. Mendler-Dünner, C., et al. (2020). Stochastic optimization for performative prediction. *Advances in Neural Information Processing Systems*, 33.
18. Miller, J., et al. (2021). Outside the echo chamber: Optimizing the performative risk. *International Conference on Machine Learning*, 7712–7722.

 doi:10.20944/preprints202606.1203.v1

19. Narang, A., et al. (2023). Multiplayer performative prediction: Learning in unknown games. *arXiv preprint arXiv:2301.XXXXX*.
20. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
21. Perdomo, J., et al. (2020). Performative prediction. *International Conference on Machine Learning*, 7599–7609.
22. Piriyakulkij, C., et al. (2025). PoE-World: Product of experts world models. *arXiv preprint arXiv:2501.XXXXX*.
23. Soros, G. (2008). *The New Paradigm for Financial Markets*. PublicAffairs.
24. Tarski, A. (1955). A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2), 285–309.
25. Thakoor, N., et al. (2022). Jumpy world models: Multi-timescale dynamics with geometric horizons. *arXiv preprint arXiv:2210.XXXXX*.
26. von Foerster, H. (2003). *Understanding Understanding: Essays on Cybernetics and Cognition*. Springer.